ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

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

A. Jentzen and D. Salimova and T. Welti

Research Report No. 2018-34
September 2018

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

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

Arnulf Jentzen[1], Diyora Salimova[2], and Timo Welti[3]

[1]Seminar for Applied Mathematics, Department of Mathematics, ETH Zurich, Switzerland, e-mail: arnulf.jentzen@sam.math.ethz.ch

[2]Seminar for Applied Mathematics, Department of Mathematics, ETH Zurich, Switzerland, e-mail: diyora.salimova@sam.math.ethz.ch

[3]Seminar for Applied Mathematics, Department of Mathematics, ETH Zurich, Switzerland, e-mail: timo.welti@sam.math.ethz.ch

September 24, 2018

Abstract

In recent years deep artificial neural networks (DNNs) have very successfully been employed in numerical simulations for a multitude of computational problems including, for example, object and face recognition, natural language processing, fraud detection, computational advertisement, and numerical approximations of partial differential equations (PDEs). Such numerical simulations indicate that DNNs seem to admit the fundamental flexibility to overcome the curse of dimensionality in the sense that the number of real parameters used to describe the DNN grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the dimension $d \in \mathbb{N}$ of the function which the DNN aims to approximate in such computational problems. There is also a large number of rigorous mathematical approximation results for artificial neural networks in the scientific literature but there are only a few special situations where results in the literature can rigorously explain the success of DNNs when approximating high-dimensional functions. The key contribution of this article is to reveal that DNNs do overcome the curse of dimensionality in the numerical approximation of Kolmogorov PDEs with constant diffusion and nonlinear drift coefficients. We prove that the number of parameters used to describe the employed DNN grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the PDE dimension $d \in \mathbb{N}$. A crucial ingredient

in our proof is the fact that the artificial neural network used to approximate the solution of the PDE is indeed a deep artificial neural network with a large number of hidden layers.

Contents

1 Introduction **2**

2 Existence of a realization on a suitable artificial probability space **6**
2.1 Markov-type estimates 6
2.2 Existence of a realization with the desired approximation properties . 7

3 The Feynman-Kac formula revisited **8**

4 Stochastic differential equations (SDEs) **9**
4.1 A priori bounds for SDEs 9
4.2 A priori bounds for Brownian motions 10
4.3 Strong perturbations of SDEs 10
4.4 Weak perturbations of SDEs 13

5 Deep artificial neural network (DNN) calculus **20**
5.1 Sums of DNNs with the same architecture 22
5.2 Compositions of DNNs involving artificial identities 24
5.3 Representations of the d-dimensional identities 29

6 DNN approximations for partial differential equations (PDEs) **30**
6.1 DNN approximations with general activation functions 31
6.2 Rectified DNN approximations 39
6.3 Rectified DNN approximations on the d-dimensional unit cube 41

1 Introduction

In recent years deep artificial neural networks (DNNs) have very successfully been employed in numerical simulations for a multitude of computational problems including, for example, object and face recognition (cf., e.g., [37, 41, 60, 62, 64] and the references mentioned therein), natural language processing (cf., e.g., [15, 25, 31, 36, 39, 65] and the references mentioned therein), fraud detection (cf., e.g., [12, 56] and the references mentioned therein), computational advertisement (cf., e.g., [63, 68] and the references mentioned therein), and numerical approximations of partial differential equations (PDEs) (cf., e.g., [4, 5, 6, 17, 19, 20, 23, 26, 28, 30, 40, 46, 48, 55, 61]). Such numerical simulations indicate that DNNs seem to admit the fundamental flexibility to overcome the curse of dimensionality in the sense that the number of real parameters used to describe the DNN grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the dimension $d \in \mathbb{N}$ of the function which the DNN aims to approximate in such computational problems. There is also a large number of rigorous mathematical approximation results for artificial neural networks in the scientific literature (see, for instance, [1, 2, 3, 7, 8, 9, 10, 11, 13, 14, 16, 18, 20, 21, 22, 24, 26, 29, 32, 33, 34, 35, 42, 43, 44, 45, 47, 49, 50, 51, 52, 53, 54, 57, 58, 59, 61, 66, 67] and the references mentioned therein) but there are only a few special situations where results in the literature can rigorously explain the success of DNNs when approximating high-dimensional functions.

The key contribution of this article is to reveal that DNNs do overcome the curse of dimensionality in the numerical approximation of Kolmogorov PDEs with constant diffusion and nonlinear drift coefficients. More specifically, the main result of this article, Theorem 6.3 in Subsection 6.2 below, proves that the number of parameters used to describe the employed DNN grows at most polynomially in both the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ and the PDE dimension $d \in \mathbb{N}$ and, thereby, we establish that DNN approximations do indeed overcome the curse of dimensionality in the numerical approximation of such PDEs. To illustrate the statement of Theorem 6.3 below in more details, we now present the following special case of Theorem 6.3.

Theorem 1.1. *Let $A_d = (a_{d,i,j})_{(i,j)\in\{1,\dots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be symmetric positive semidefinite matrices, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $f_{0,d}\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $f_{1,d}\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be functions, let $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be the functions which satisfy for all $d \in \mathbb{N}$, $x = (x_1,\dots,x_d) \in \mathbb{R}^d$ that $\mathbf{A}_d(x) = (\max\{x_1,0\},\dots,\max\{x_d,0\})$, let*

$$\mathcal{N} = \cup_{L\in\{2,3,4,\dots\}} \cup_{(l_0,l_1,\dots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n})\right), \tag{1}$$

let $\mathcal{P}\colon \mathcal{N} \to \mathbb{N}$ and $\mathcal{R}\colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k,\mathbb{R}^l)$ be the functions which satisfy for all $L \in \{2,3,4,\dots\}$, $l_0,l_1,\dots,l_L \in \mathbb{N}$, $\Phi = ((W_1,B_1),\dots,(W_L,B_L)) \in (\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}}\times \mathbb{R}^{l_n}))$, $x_0 \in \mathbb{R}^{l_0}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, n \in \mathbb{N}\cap[1,L)\colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ that $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1}+1)$, $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0},\mathbb{R}^{l_L})$, and

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{2}$$

let $T,\kappa \in (0,\infty)$, $(\phi^{m,d}_{\varepsilon})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x,y \in \mathbb{R}^d$ that $\mathcal{R}(\phi^{0,d}_{\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{R}(\phi^{1,d}_{\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R}^d)$, $|f_{0,d}(x)| + \sum_{i,j=1}^{d}|a_{d,i,j}| \leq \kappa d^{\kappa}(1+\|x\|^{\kappa}_{\mathbb{R}^d})$, $\|f_{1,d}(x) - f_{1,d}(y)\|_{\mathbb{R}^d} \leq \kappa\|x-y\|_{\mathbb{R}^d}$, $\|(\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \leq \kappa(d^{\kappa}+\|x\|_{\mathbb{R}^d})$, $\sum_{m=0}^{1}\mathcal{P}(\phi^{m,d}_{\varepsilon}) \leq \kappa d^{\kappa}\varepsilon^{-\kappa}$, $|(\mathcal{R}\phi^{0,d}_{\varepsilon})(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(y)| \leq \kappa d^{\kappa}(1+\|x\|^{\kappa}_{\mathbb{R}^d}+\|y\|^{\kappa}_{\mathbb{R}^d})\|x-y\|_{\mathbb{R}^d}$, and

$$|f_{0,d}(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(x)| + \|f_{1,d}(x) - (\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \leq \varepsilon\kappa d^{\kappa}(1+\|x\|^{\kappa}_{\mathbb{R}^d}), \tag{3}$$

and for every $d \in \mathbb{N}$ let $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ be an at most polynomially growing viscosity solution of

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\tfrac{\partial}{\partial x}u_d)(t,x)\, f_{1,d}(x) + \sum_{i,j=1}^{d} a_{d,i,j}\,(\tfrac{\partial^2}{\partial x_i \partial x_j}u_d)(t,x) \tag{4}$$

with $u_d(0,x) = f_{0,d}(x)$ for $(t,x) \in (0,T)\times\mathbb{R}^d$. Then for every $p \in (0,\infty)$ there exist $(\psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$, $c \in \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq c\, d^c \varepsilon^{-c}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, and

$$\left[\int_{[0,1]^d} |u_d(T,x) - (\mathcal{R}\psi_{d,\varepsilon})(x)|^p\, dx\right]^{1/p} \leq \varepsilon. \tag{5}$$

Theorem 1.1 is an in immediate consequence from Corollary 6.4 in Subsection 6.3 below. Corollary 6.4, in turn, is a special case of Theorem 6.3. Next we add some comments regarding the mathematical objects appearing in Theorem 1.1. Theorem 1.1 is an approximation result for rectified DNNs and for every $d \in \mathbb{N}$ the function $\mathbf{A}_d \colon \mathbb{R}^d \to \mathbb{R}^d$ in Theorem 1.1 above describes the d-dimensional rectifier function. The set $\mathcal{N}$ in (1) in Theorem 1.1 above is a set of tuples of real numbers which, in turn, represents the set of all artificial neural networks. For every artificial neural network $\Phi \in \mathcal{N}$ in Theorem 1.1 above we have that $\mathcal{R}(\Phi) \in \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ represents the function associated to the artificial neural network Φ (cf. (2) in Theorem 1.1). The function $\mathcal{R} \colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ from the set $\mathcal{N}$ of all artificial neural networks to the union $\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ of continuous functions thus describes the realizations associated to the artificial neural networks. Moreover, for every artificial neural network $\Phi \in \mathcal{N}$ in Theorem 1.1 above we have that $\mathcal{P}(\Phi) \in \mathbb{N}$ represents the number of real parameters which are used to describe the artificial neural network Φ. In particular, for every artificial neural network $\Phi \in \mathcal{N}$ in Theorem 1.1 we can think of $\mathcal{P}(\Phi) \in \mathbb{N}$ as a quantity related to the amount of memory storage which is needed to store the artificial neural network. The real number $\kappa > 0$ in Theorem 1.1 is an arbitrary constant used to formulate the hypotheses in Theorem 1.1 (cf. (3) in Theorem 1.1 above) and the real number $T > 0$ in Theorem 1.1 describes the time horizon under consideration. Our key hypothesis in Theorem 1.1 is the assumption that both the possibly nonlinear initial value functions $f_{0,d} \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and the possibly nonlinear drift coefficient functions $f_{1,d} \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, of the PDEs in (4) can be approximated without the curse of dimensionality by means of DNNs (see (3) above for details). Simple examples for the functions $f_{0,d} \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $f_{1,d} \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, which fulfill the hypotheses of Theorem 1.1 above are, for instance, provided by the choice that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that $f_{0,d}(x) = \max\{x_1, x_2, \dots, x_d\}$ and $f_{1,d}(x) = x\,[1 + \|x\|^2_{\mathbb{R}^d}]^{-1}$. A natural example for the matrices $A_d = (a_{d,i,j})_{(i,j)\in\{1,\dots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, fulfilling the hypotheses in Theorem 1.1 above is, for instance, provided by the choice that for all $d \in \mathbb{N}$ it holds that $A_d \in \mathbb{R}^{d\times d}$ is the d-dimensional identity matrix in which case the second order term in (4) reduces to the d-dimensional Laplace operator. Roughly speaking, Theorem 1.1 above proves that if both the initial value functions and the drift coefficient functions in the PDEs in (4) can be approximated without the curse of dimensionality by means of DNNs, then the solutions of the PDEs can also be approximated without the curse of dimensionality by means of DNNs (see (5) above for details). In numerical simulations involving DNNs for computational problems from data science (e.g., object and face recognition, natural language processesing, fraud detection, computational advertisement, etc.) it is often not entirely clear how to precisely describe what the involved DNN approximations should achieve and it is thereby often not entirely clear how to precisely specify the approximation error of the employed DNN. The recent articles [17, 28] (cf., e.g., also [4, 5, 6, 19, 20, 23, 26, 30, 40, 46, 48, 55, 61]) suggest to use machine learning algorithms which employ DNNs to approximate solutions and derivatives of solutions, respectively, of PDEs and in the framework of these references it is perfectly clear what the involved DNN approximations should achieve as well as how to specify the approximation error: the DNN should approximate the unique deterministic function which is the solution of the given deterministic PDE (cf., e.g., Han et al. [28, *Neural Network Architecture* on page 5] and Beck et al. [4, Proposition 2.7 and (103)]. The above named references thereby open up the possibility for a complete and rigorous mathematical error analysis for the involved deep learning algorithms and Theorems 1.1 and 6.3, in particular, provide some first contributions to this new research topic. The statements of Theorems 1.1 and 6.3 and their strategies of proof, respectively, are inspired by the article Grohs et al. [26] (cf., e.g., Theorem 1.1 in [26]) in which similar results as Theorems 1.1 and 6.3, respectively, but for Kolmogorov PDEs with affine linear drift and diffusion coefficient functions have been proved. The main difference of the arguments in [26] to this paper is the deepness of the involved artificial neural networks. Roughly speaking, the affine linear structure of the coefficients of the Kolmogorov PDEs in [26] allowed the authors in [26] to essentially employ a flat artificial neural network for approximating the solution flow mapping of such PDEs. In this work the drift coefficient is nonlinear and, in view of this property, we employ in our proofs of Theorem 1.1 and Theorem 6.3, respectively, iterative Euler-type discretizations for the underlying stochastic dynamics associated to the PDEs in (4). The iterative Euler-type discretizations result in multiple compositions which, in turn, result in deep artificial neural networks with a large number of hidden layers. In particular, in our proof of Theorem 1.1 and Theorem 6.3, respectively, the artificial neural networks $\psi_{d,\varepsilon} \in \mathcal{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, approximating the solutions of the PDEs in (4) (see (5) above) are also deep artificial neural networks with a large number of hidden layers even if the artificial neural networks approximating or representing $f_{0,d} \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $f_{1,d} \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, are flat with one hidden layer only. Moreover, our proofs of Theorem 1.1 and Theorem 6.3, respectively, reveal that the number of hidden layers increases to infinity as the prescribed approximation accuracy $\varepsilon > 0$ decreases to zero and the PDE dimension increases to infinity, respectively (cf. (149) and (168) below).

Theorem 1.1 above and Theorem 6.3, respectively, are purely deterministic approximation results for DNNs and solutions of a class of deterministic PDEs. Our proofs of Theorem 1.1 and Theorem 6.3, respectively, are, however, heavily relying on probabilistic arguments on a suitable artificial probability space. Roughly speaking, in our proof of Theorem 6.3 we

(I) design a suitable random DNN on this artificial probability space,

(II) show that this suitable random DNN is in a suitable sense close to the solution of the considered deterministic PDE, and

(III) employ items (I)–(II) above to establish the existence of a realization with the desired approximation properties on the artificial probability space.

The specific realization of this random DNN is then a deterministic DNN approximation of the solution of the considered deterministic PDE with the desired approximation properties. The main work of the paper is the construction and the analysis of this random DNN. For the construction of the random DNN we need suitable general flexibility results for rectified DNNs which, roughly speaking, demonstrate how rectified DNNs can be composed with a moderate growth of the number of involved parameters (see Subsection 5.2 below for details). The construction of the random DNN (cf. (I) above) is essentially performed in Section 5 and Section 6 and the analysis of the random DNN (cf. (II) above) is essentially the subject of Section 3, Section 4, and Subsection 6.1. The argument for the existence of the realization with the desired approximation properties on the artificial probability space (cf. (III) above) is provided in Section 2 and Subsection 6.1.

2 On the existence of a realization with the desired approximation properties on a suitable artificial probability space

In this section we establish in Corollary 2.4 in Subsection 2.2 below on a very abstract level, roughly speaking, the argument that good approximation properties of the random DNN (cf. items (I)–(II) in Section 1 above) imply the existence of a realization with the desired approximation properties on the artificial probability space (cf. item (III) in Section 1 above). The function $u\colon \mathbb{R}^d \to \mathbb{R}$ in Corollary 2.4 will essentially take the role of the solution of the considered deterministic PDE and the random field $X\colon \mathbb{R}^d \times \Omega \to \mathbb{R}$ will essentially take the role of the random DNN. Our proof of Corollary 2.4 is based on an application of Proposition 2.3 in Subsection 2.2 below. Proposition 2.3 is, very loosely speaking, an abstract generalized version of Corollary 2.4. Our proof of Proposition 2.3 is based on an application of the elementary Markov-type estimate in Lemma 2.2 in Subsection 2.1 below. Lemma 2.2, in turn, follows from the Markov inequality in Lemma 2.1 in Subsection 2.1 below. For completeness we also provide the short proof of the Markov inequality in Lemma 2.1. Results related to Lemma 2.2 and Proposition 2.3 can, e.g., be found in Grohs et al. [26, Subsection 3.1]. In particular, Lemma 2.2 is somehow an elementary extension of [26, Proposition 3.3 in Subsection 3.1].

2.1 Markov-type estimates

Lemma 2.1 (Markov inequality)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $\varepsilon \in (0,\infty)$, and let $X\colon \Omega \to [0,\infty]$ be an $\mathcal{F}/\mathcal{B}([0,\infty])$-measurable function. Then*

$$\mu\big(X \geq \varepsilon\big) \leq \frac{\int_\Omega X \, d\mu}{\varepsilon}. \tag{6}$$

Proof of Lemma 2.1. Note that the fact that $X \geq 0$ proves that

$$\mathbb{1}_{\{X\geq\varepsilon\}} = \frac{\varepsilon \cdot \mathbb{1}_{\{X\geq\varepsilon\}}}{\varepsilon} \leq \frac{X \cdot \mathbb{1}_{\{X\geq\varepsilon\}}}{\varepsilon} \leq \frac{X}{\varepsilon}. \tag{7}$$

Integration with respect to μ hence establishes (6). The proof of Lemma 2.1 is thus completed. □

Lemma 2.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $X\colon \Omega \to [-\infty,\infty]$ be a random variable, and let $\varepsilon, q \in (0,\infty)$. Then*

$$\big[\mathbb{P}\big(|X| \geq \varepsilon\big)\big]^{1/q} \leq \frac{\big(\mathbb{E}\big[|X|^q\big]\big)^{1/q}}{\varepsilon}. \tag{8}$$

Proof of Lemma 2.2. Observe that Lemma 2.1 ensures that

$$[\mathbb{P}(|X| \geq \varepsilon)]^{1/q} = [\mathbb{P}(|X|^q \geq \varepsilon^q)]^{1/q} \leq \left[\frac{\mathbb{E}\big[|X|^q\big]}{\varepsilon^q}\right]^{1/q} = \frac{\big(\mathbb{E}\big[|X|^q\big]\big)^{1/q}}{\varepsilon}. \tag{9}$$

The proof of Lemma 2.2 is thus completed. □

2.2 Existence of a realization with the desired approximation properties

Proposition 2.3. *Let $\varepsilon \in (0, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X \colon \Omega \to [-\infty, \infty]$ be a random variable which satisfies that*

$$\inf\nolimits_{q\in(0,\infty)} \big(\mathbb{E}\big[|X|^q\big]\big)^{1/q} < \varepsilon. \tag{10}$$

Then there exists $\omega \in \Omega$ such that $|X(\omega)| < \varepsilon$.

Proof of Proposition 2.3. First, observe that Lemma 2.2 assures that for all $q \in (0, \infty)$ it holds that

$$\big[\mathbb{P}\big(|X| \geq \varepsilon\big)\big]^{1/q} \leq \frac{\big(\mathbb{E}\big[|X|^q\big]\big)^{1/q}}{\varepsilon}. \tag{11}$$

Next note that the hypothesis that $\inf_{q\in(0,\infty)} \big(\mathbb{E}\big[|X|^q\big]\big)^{1/q} < \varepsilon$ demonstrates that there exists $q \in (0, \infty)$ such that

$$\big(\mathbb{E}\big[|X|^q\big]\big)^{1/q} < \varepsilon. \tag{12}$$

Combining this with (11) proves that

$$\big[\mathbb{P}\big(|X| \geq \varepsilon\big)\big]^{1/q} < 1. \tag{13}$$

Hence, we obtain that

$$\mathbb{P}\big(|X| \geq \varepsilon\big) < 1. \tag{14}$$

This shows that

$$\mathbb{P}\big(|X| < \varepsilon\big) = 1 - \mathbb{P}\big(|X| \geq \varepsilon\big) > 0. \tag{15}$$

Therefore, we obtain that

$$\{|X| < \varepsilon\} = \big\{\omega \in \Omega \colon |X(\omega)| < \varepsilon\big\} \neq \emptyset. \tag{16}$$

The proof of Proposition 2.3 is thus completed. □

Corollary 2.4 (Existence of approximating realizations of a random field)**.** *Let $d \in \mathbb{N}$, $p, \varepsilon \in (0, \infty)$, let $u \colon \mathbb{R}^d \to \mathbb{R}$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\nu \colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be a probability measure on $\mathbb{R}^d$, let $X : \mathbb{R}^d \times \Omega \to \mathbb{R}$ be $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R})$-measurable, and assume that*

$$\left[\int_{\mathbb{R}^d} \mathbb{E}\big[|u(x) - X(x)|^p\big]\, \nu(dx)\right]^{1/p} < \varepsilon. \tag{17}$$

Then there exists $\omega \in \Omega$ such that

$$\left[\int_{\mathbb{R}^d} |u(x) - X(x, \omega)|^p\, \nu(dx)\right]^{1/p} < \varepsilon. \tag{18}$$

Proof of Corollary 2.4. Throughout this proof let $Y\colon \Omega \to [-\infty, \infty]$ be the random variable given by

$$Y = \left[\int_{\mathbb{R}^d} |u(x) - X(x)|^p \, \nu(dx) \right]^{1/p}. \tag{19}$$

Observe that Fubini's theorem and (17) ensure that

$$\begin{aligned} \big(\mathbb{E}\big[|Y|^p\big]\big)^{1/p} &= \left(\mathbb{E}\left[\int_{\mathbb{R}^d} |u(x) - X(x)|^p \, \nu(dx) \right] \right)^{1/p} \\ &= \left(\int_{\mathbb{R}^d} \mathbb{E}\big[|u(x) - X(x)|^p\big] \, \nu(dx) \right)^{1/p} < \varepsilon. \end{aligned} \tag{20}$$

Hence, we obtain that

$$\inf_{q \in (0,\infty)} \big(\mathbb{E}\big[|Y|^q\big]\big)^{1/q} \leq \big(\mathbb{E}\big[|Y|^p\big]\big)^{1/p} < \varepsilon. \tag{21}$$

This allows us to apply Proposition 2.3 to obtain that there exists $\omega \in \Omega$ such that

$$|Y(\omega)| < \varepsilon. \tag{22}$$

Combining this with (19) establishes (18). The proof of Corollary 2.4 is thus completed. ◻

3 The Feynman-Kac formula revisited

Theorem 6.3 in Subsection 6.2 below (the main result of this article) and Theorem 1.1 in the introduction, respectively, are, as mentioned above, purely deterministic approximation results for DNNs and a class of deterministic PDEs. In contrast, our proofs of Theorem 6.3 and Theorem 1.1, respectively, are based on a probabilistic argument on a suitable artificial probability space on which we, roughly speaking, design random DNNs. Our construction of the random DNNs is based on suitable Monte Carlo approximations of the solutions of the considered deterministic PDEs. These suitable Monte Carlo approximations, in turn, are based on the link between deterministic Kolmogorov PDEs and solutions of SDEs which is provided by the famous Feynman-Kac formula. In this section we recall in Theorem 3.1 below a special case of this famous formula (cf., e.g., Hairer et al. [27, Subsection 4.4]). Theorem 3.1 below will be used in our proof of Theorem 6.3 below (cf. (145) and (150) in the proof of Proposition 6.1, Proposition 6.1, Corollary 6.2, and Theorem 6.3).

Theorem 3.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $T \in (0, \infty)$, $d \in \mathbb{N}$, $B \in \mathbb{R}^{d \times m}$, let $W\colon [0,T] \times \Omega \to \mathbb{R}^m$ be a standard Brownian motion, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be the d-dimensional Euclidean norm, let $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional Euclidean scalar product, let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a continuous function, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ be a locally Lipschitz continuous function, and assume that*

$$\inf_{p \in (0,\infty)} \sup_{x \in \mathbb{R}^d} \left[\frac{|\varphi(x)|}{(1 + \|x\|^p)} + \frac{\|\mu(x)\|}{(1 + \|x\|)} \right] < \infty. \tag{23}$$

Then

(i) there exist unique stochastic processes $X^x\colon [0,T]\times\Omega \to \mathbb{R}^d$, $x\in\mathbb{R}^d$, *with continuous sample paths which satisfy for all* $x \in \mathbb{R}^d$, $t\in[0,T]$ *that*

$$X_t^x = x + \int_0^t \mu(X_s^x)\,ds + BW_t, \tag{24}$$

(ii) there exists a unique function $u\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ *such that for all* $x\in\mathbb{R}^d$ *it holds that* $u(0,x)=\varphi(x)$, *such that* $\inf_{p\in(0,\infty)}\sup_{(t,x)\in[0,T]\times\mathbb{R}^d}\frac{|u(t,x)|}{1+\|x\|^p}<\infty$, *and such that u is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u)(t,x) = \big\langle(\nabla_x u)(t,x),\mu(x)\big\rangle + \tfrac{1}{2}\operatorname{Trace}\big(BB^*(\operatorname{Hess}_x u)(t,x)\big) \tag{25}$$

for $(t,x)\in(0,T)\times\mathbb{R}^d$, *and*

(iii) it holds for all $t\in[0,T]$, $x\in\mathbb{R}^d$ *that* $\mathbb{E}\big[|\varphi(X_t^x)|\big]<\infty$ *and*

$$u(t,x) = \mathbb{E}\big[\varphi(X_t^x)\big]. \tag{26}$$

4 Stochastic differential equations (SDEs)

In our proofs of Theorem 1.1 above and Theorem 6.3 below (the main result of this article), respectively, we design and analyse (cf. items (I)–(II) in Section 1 above) a suitable random DNN. The construction of this suitable random DNN is based on Euler-Maruyama discretizations of solutions of the SDEs associated to the Kolmogorov PDEs in (4) and for our error analysis of this suitable random DNN we employ appropriate weak error estimates for Euler-Maruyama discretizations of solutions of SDEs. These weak error estimates are established in Lemma 4.5 and Proposition 4.6 in Subsection 4.4 below. Our proofs of Lemma 4.5 and Proposition 4.6, respectively, use suitable strong error estimates for Euler-Maruyama discretizations. These strong error estimates are the subject of Proposition 4.4 in Subsection 4.3 below. Proposition 4.4 follows from an application of the deterministic perturbation-type inequality in Lemma 4.3 in Subsection 4.3 below. Perturbation estimates which are related to Lemma 4.3 and Proposition 4.4 can, e.g., be found in Hutzenthaler et al. [38, Proposition 2.9 and Corollary 2.12]. In particular, our proof of Lemma 4.3 is inspired by the proof of Proposition 2.9 in Hutzenthaler et al. [38]. Furthermore, our proof of Proposition 4.6 employs the elementary a priori estimate in Lemma 4.1 in Subsection 4.1 below. Lemma 4.1, in turn, is a straightforward consequence from Gronwall's integral inequality (see, e.g., Grohs et al. [26, Lemma 2.11]) and its proof is therefore omitted. In our proof of Theorem 6.3 we will also employ the elementary a priori estimate for standard Brownian motions in Lemma 4.2 in Subsection 4.2 below. Lemma 4.2 is a straightforward consequence from Itô's formula and its proof is therefore also omitted.

4.1 A priori bounds for SDEs

Lemma 4.1. *Let* $d,m\in\mathbb{N}$, $\xi\in\mathbb{R}^d$, $p\in[1,\infty)$, $c,C,T\in[0,\infty)$, $B\in\mathbb{R}^{d\times m}$, *let* $\|\cdot\|:\mathbb{R}^d\to[0,\infty)$ *be the d-dimensional Euclidean norm, let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability* *space, let* $W\colon [0,T]\times\Omega\to\mathbb{R}^m$ *be a standard Brownian motion, let* $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ *be a* $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$*-measurable function which satisfies for all* $x\in\mathbb{R}^d$ *that* $\|\mu(x)\|\le C+c\|x\|$*, let* $\chi\colon[0,T]\to[0,T]$ *be a* $\mathcal{B}([0,T])/\mathcal{B}([0,T])$*-measurable function which satisfies for all* $t\in[0,T]$ *that* $\chi(t)\le t$*, and let* $X\colon[0,T]\times\Omega\to\mathbb{R}^d$ *be a stochastic process with continuous sample paths which satisfies for all* $t\in[0,T]$ *that*

$$\mathbb{P}\bigg(X_t=\xi+\int_0^t\mu\big(X_{\chi(s)}\big)\,ds+BW_t\bigg)=1. \tag{27}$$

Then it holds that

$$\sup_{t\in[0,T]}\big(\mathbb{E}[\|X_t\|^p]\big)^{1/p}\le\Big(\|\xi\|+CT+\big(\mathbb{E}\big[\|BW_T\|^p\big]\big)^{1/p}\Big)\,e^{cT}. \tag{28}$$

4.2 A priori bounds for Brownian motions

Lemma 4.2. *Let* $d,m\in\mathbb{N}$*,* $T\in[0,\infty)$*,* $p\in(0,\infty)$*,* $B\in\mathbb{R}^{d\times m}$*, let* $\|\cdot\|:\mathbb{R}^d\to[0,\infty)$ *be the* d*-dimensional Euclidean norm, let* $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space, and let* $W\colon[0,T]\times\Omega\to\mathbb{R}^m$ *be a standard Brownian motion. Then it holds for all* $t\in[0,T]$ *that*

$$\big(\mathbb{E}\big[\|BW_t\|^p\big]\big)^{1/p}\le\sqrt{\max\{1,p-1\}\operatorname{Trace}(B^*B)\,t}. \tag{29}$$

4.3 Strong perturbations of SDEs

Lemma 4.3. *Let* $d\in\mathbb{N}$*,* $L,T\in[0,\infty)$*,* $\delta\in(0,\infty)$*,* $p\in[2,\infty)$*, let* $\|\cdot\|:\mathbb{R}^d\to[0,\infty)$ *be the* d*-dimensional Euclidean norm, let* $\mu\colon\mathbb{R}^d\to\mathbb{R}^d$ *be a function which satisfies for all* $v,w\in\mathbb{R}^d$ *that*

$$\|\mu(v)-\mu(w)\|\le L\|v-w\|, \tag{30}$$

let $X,Y\colon[0,T]\to\mathbb{R}^d$ *be continuous functions, let* $a\colon[0,T]\to\mathbb{R}^d$ *be a* $\mathcal{B}([0,T])/\mathcal{B}(\mathbb{R}^d)$*-measurable function, and assume for all* $t\in[0,T]$ *that* $\int_0^t\|a_s\|\,ds<\infty$ *and*

$$X_t-Y_t=X_0-Y_0+\int_0^t\big[\mu(X_s)-a_s\big]\,ds. \tag{31}$$

Then it holds for all $t\in[0,T]$ *that*

$$\begin{aligned}&\|X_t-Y_t\|^p\\&\le\exp\Big(\Big[L+\tfrac{(1-1/p)}{\delta}\Big]\,p\,t\Big)\bigg(\|X_0-Y_0\|^p+\delta^{(p-1)}\int_0^t\|a_s-\mu(Y_s)\|^p\,ds\bigg).\end{aligned} \tag{32}$$

Proof of Lemma 4.3. Throughout this proof let $\langle\cdot,\cdot\rangle\colon\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ be the d-dimensional Euclidean scalar product and let $\alpha\in(0,\infty)$ be the real number given by

$$\alpha=\Big[L+\tfrac{(1-1/p)}{\delta}\Big]\,p. \tag{33}$$

Note that (31) ensures that the function $([0,T] \ni t \mapsto (X_t - Y_t) \in \mathbb{R}^d)$ is absolutely continuous. The fundamental theorem of calculus and the chain rule hence prove that for all $t \in [0,T]$ it holds that

$$\begin{aligned}\frac{\|X_t - Y_t\|^p}{\exp(\alpha t)} &= \|X_0 - Y_0\|^p + \int_0^t \frac{p\,\|X_s - Y_s\|^{p-2}\langle X_s - Y_s, \mu(X_s) - a_s\rangle}{\exp(\alpha s)}\,ds \\ &\quad - \int_0^t \frac{\alpha\|X_s - Y_s\|^p}{\exp(\alpha s)}\,ds \\ &= \|X_0 - Y_0\|^p + \int_0^t \frac{p\,\|X_s - Y_s\|^{p-2}\langle X_s - Y_s, \mu(X_s) - \mu(Y_s)\rangle}{\exp(\alpha s)}\,ds \\ &\quad + \int_0^t \frac{p\,\|X_s - Y_s\|^{p-2}\langle X_s - Y_s, \mu(Y_s) - a_s\rangle - \alpha\|X_s - Y_s\|^p}{\exp(\alpha s)}\,ds.\end{aligned} \tag{34}$$

Next observe that (30) and the Cauchy-Schwartz inequality ensure that for all $s \in [0,T]$ it holds that

$$\langle X_s - Y_s, \mu(X_s) - \mu(Y_s)\rangle \leq \|X_s - Y_s\|\|\mu(X_s) - \mu(Y_s)\| \leq L\|X_s - Y_s\|^2. \tag{35}$$

This and (34) demonstrate that for all $t \in [0,T]$ it holds that

$$\begin{aligned}&\frac{\|X_t - Y_t\|^p}{\exp(\alpha t)} \leq \|X_0 - Y_0\|^p + \int_0^t \frac{pL\|X_s - Y_s\|^p}{\exp(\alpha s)}\,ds \\ &\quad + \int_0^t \frac{p\,\|X_s - Y_s\|^{p-2}\langle X_s - Y_s, \mu(Y_s) - a_s\rangle - \alpha\|X_s - Y_s\|^p}{\exp(\alpha s)}\,ds \\ &= \|X_0 - Y_0\|^p + \int_0^t \frac{p\,\|X_s - Y_s\|^{p-2}\langle X_s - Y_s, \mu(Y_s) - a_s\rangle - (\alpha - pL)\|X_s - Y_s\|^p}{\exp(\alpha s)}\,ds \\ &= \|X_0 - Y_0\|^p + \int_0^t \frac{p\,\|X_s - Y_s\|^{p-2}\langle X_s - Y_s, \mu(Y_s) - a_s\rangle - \frac{(p-1)}{\delta}\|X_s - Y_s\|^p}{\exp(\alpha s)}\,ds.\end{aligned} \tag{36}$$

Next observe that the Cauchy-Schwartz inequality and Young's inequality prove that for all $s \in [0,T]$ it holds that

$$\begin{aligned}&\|X_s - Y_s\|^{p-2}\langle X_s - Y_s, \mu(Y_s) - a_s\rangle \leq \|X_s - Y_s\|^{p-1}\|\mu(Y_s) - a_s\| \\ &= \delta^{(1-p)/p}\|X_s - Y_s\|^{p-1}\delta^{(p-1)/p}\|\mu(Y_s) - a_s\| \\ &\leq \tfrac{(p-1)}{p}\big[\delta^{(1-p)/p}\|X_s - Y_s\|^{p-1}\big]^{p/(p-1)} + \tfrac{1}{p}\big[\delta^{(p-1)/p}\|\mu(Y_s) - a_s\|\big]^p \\ &= \tfrac{(p-1)}{\delta p}\|X_s - Y_s\|^p + \tfrac{\delta^{(p-1)}}{p}\|\mu(Y_s) - a_s\|^p.\end{aligned} \tag{37}$$

Combining this with (36) assures that for all $t \in [0,T]$ it holds that

$$\begin{aligned}&\frac{\|X_t - Y_t\|^p}{\exp(\alpha t)} \\ &\leq \|X_0 - Y_0\|^p + \int_0^t \frac{\frac{(p-1)}{\delta}\|X_s - Y_s\|^p + \delta^{(p-1)}\|\mu(Y_s) - a_s\|^p - \frac{(p-1)}{\delta}\|X_s - Y_s\|^p}{\exp(\alpha s)}\,ds \\ &= \|X_0 - Y_0\|^p + \int_0^t \frac{\delta^{(p-1)}\|\mu(Y_s) - a_s\|^p}{\exp(\alpha s)}\,ds \\ &\leq \|X_0 - Y_0\|^p + \int_0^t \delta^{(p-1)}\|\mu(Y_s) - a_s\|^p\,ds.\end{aligned} \tag{38}$$

This implies (32). The proof of Lemma 4.3 is thus completed. □

Proposition 4.4 (Perturbation)**.** *Let $d, m \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $L, T \in [0,\infty)$, $\delta \in (0,\infty)$, $p \in [2,\infty)$, $B \in \mathbb{R}^{d\times m}$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $W: [0,T]\times\Omega \to \mathbb{R}^m$ be a standard Brownian motion, let $\mu: \mathbb{R}^d \to \mathbb{R}^d$ be a function which satisfies for all $v, w \in \mathbb{R}^d$ that*

$$\|\mu(v) - \mu(w)\| \leq L\|v - w\|, \tag{39}$$

let $X, Y : [0,T]\times\Omega \to \mathbb{R}^d$ be stochastic processes with continuous sample paths, let $a: [0,T]\times\Omega \to \mathbb{R}^d$ be a $(\mathcal{B}([0,T])\otimes\mathcal{F})/\mathcal{B}(\mathbb{R}^d)$-measurable function, and assume for all $t \in [0,T]$ that $\int_0^t \|a_s\|\, ds < \infty$, $Y_t = y + \int_0^t a_s\, ds + BW_t$, and

$$X_t = x + \int_0^t \mu(X_s)\, ds + BW_t. \tag{40}$$

Then it holds for all $t \in [0,T]$ that

$$\begin{aligned} &\big(\mathbb{E}\big[\|X_t - Y_t\|^p\big]\big)^{1/p} \\ &\leq \exp\!\Big(\Big[L + \tfrac{(1-1/p)}{\delta}\Big]\, t\Big)\left(\|x-y\| + \delta^{(1-1/p)}\left[\int_0^t \mathbb{E}\big[\|a_s - \mu(Y_s)\|^p\big]\, ds\right]^{1/p}\right). \end{aligned} \tag{41}$$

Proof of Proposition 4.4. First, note that for all $t \in [0,T]$ it holds that

$$X_t - Y_t = x - y + \int_0^t \big[\mu(X_s) - a_s\big]\, ds. \tag{42}$$

Lemma 4.3 hence ensures that for all $t \in [0,T]$ that

$$\begin{aligned} &\|X_t - Y_t\|^p \\ &\leq \exp\!\Big(\Big[L + \tfrac{(1-1/p)}{\delta}\Big]\, p\, t\Big)\left(\|x-y\|^p + \delta^{(p-1)}\int_0^t \|a_s - \mu(Y_s)\|^p\, ds\right). \end{aligned} \tag{43}$$

This implies that for all $t \in [0,T]$ it holds that

$$\begin{aligned} &\mathbb{E}\big[\|X_t - Y_t\|^p\big] \\ &\leq \exp\!\Big(\Big[L + \tfrac{(1-1/p)}{\delta}\Big]\, p\, t\Big)\left(\|x-y\|^p + \delta^{(p-1)}\int_0^t \mathbb{E}\big[\|a_s - \mu(Y_s)\|^p\big]\, ds\right). \end{aligned} \tag{44}$$

The fact that $\forall\, b, c \in \mathbb{R}: |b+c|^{1/p} \leq |b|^{1/p} + |c|^{1/p}$ hence demonstrates that for all $t \in [0,T]$ it holds that

$$\begin{aligned} &\big(\mathbb{E}\big[\|X_t - Y_t\|^p\big]\big)^{1/p} \\ &\leq \exp\!\Big(\Big[L + \tfrac{(1-1/p)}{\delta}\Big]\, t\Big)\left(\|x-y\| + \delta^{(1-1/p)}\left[\int_0^t \mathbb{E}\big[\|a_s - \mu(Y_s)\|^p\big]\, ds\right]^{1/p}\right). \end{aligned} \tag{45}$$

The proof of Proposition 4.4 is thus completed. □

4.4 Weak perturbations of SDEs

Lemma 4.5. *Let $d,m \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $h,T,\varepsilon_0,\varepsilon_1,\varsigma_0,\varsigma_1,L_0,L_1,\ell \in [0,\infty)$, $\delta \in (0,\infty)$, $B \in \mathbb{R}^{d\times m}$, $p \in [2,\infty)$, $q \in (1,2]$ satisfy $1/p + 1/q = 1$, let $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $W\colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard Brownian motion, let $\phi_0\colon \mathbb{R}^d \to \mathbb{R}$, $f_1\colon \mathbb{R}^d \to \mathbb{R}^d$, $\phi_2\colon \mathbb{R}^d \to \mathbb{R}^d$, and $\chi\colon [0,T] \to [0,T]$ be functions, let $f_0\colon \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable function, let $\phi_1\colon \mathbb{R}^d \to \mathbb{R}^d$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable function, assume for all $t \in [0,T]$, $x,y \in \mathbb{R}^d$ that*

$$|\phi_0(x) - f_0(x)| \le \varepsilon_0(1 + \|x\|^{\varsigma_0}), \qquad \|\phi_1(x) - f_1(x)\| \le \varepsilon_1(1 + \|x\|^{\varsigma_1}), \tag{46}$$

$$|\phi_0(x) - \phi_0(y)| \le L_0 \left(1 + \int_0^1 \big[r\|x\| + (1-r)\|y\|\big]^\ell \, dr\right) \|x-y\|, \tag{47}$$

$$\|f_1(x) - f_1(y)\| \le L_1\|x-y\|, \qquad \textit{and} \qquad \chi(t) = \max(\{0,h,2h,\dots\} \cap [0,t]), \tag{48}$$

and let $X,Y\colon [0,T]\times\Omega \to \mathbb{R}^d$ be stochastic processes with continuous sample paths which satisfy for all $t \in [0,T]$ that $Y_t = \phi_2(\xi) + \int_0^t \phi_1\big(Y_{\chi(s)}\big)\,ds + BW_t$ and

$$X_t = \xi + \int_0^t f_1(X_s)\,ds + BW_t. \tag{49}$$

Then it holds that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\le \varepsilon_0\,(1 + \mathbb{E}[\|X_T\|^{\varsigma_0}]) \\
&\quad + L_0\, 2^{\max\{\ell-1,0\}} \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big) \Big[1 + \big(\mathbb{E}\big[\|X_T\|^{\ell q}\big]\big)^{1/q} + \big(\mathbb{E}\big[\|Y_T\|^{\ell q}\big]\big)^{1/q}\Big] \\
&\quad \cdot \Bigg[\|\xi - \phi_2(\xi)\| + \varepsilon_1\, \delta^{(1-1/p)}\, T^{1/p}\Big[1 + \sup_{t\in[0,T]} \big(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\big)^{1/p}\Big] \\
&\quad + h\,\delta^{(1-1/p)}\, T^{1/p} L_1 \Big[\sup_{t\in[0,T]} \big(\mathbb{E}\big[\|\phi_1(Y_t)\|^p\big]\big)^{1/p}\Big] + \delta^{(1-1/p)}\, T^{1/p} L_1 \big(\mathbb{E}\big[\|BW_h\|^p\big]\big)^{1/p}\Bigg].
\end{aligned} \tag{50}$$

Proof of Lemma 4.5. First, note that the triangle inequality ensures that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\le \big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(X_T)\big]\big| + \big|\mathbb{E}\big[\phi_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\le \mathbb{E}\big[|f_0(X_T) - \phi_0(X_T)|\big] + \mathbb{E}\big[|\phi_0(X_T) - \phi_0(Y_T)|\big] \\
&\le \varepsilon_0\, \mathbb{E}[1 + \|X_T\|^{\varsigma_0}] + \mathbb{E}\big[|\phi_0(X_T) - \phi_0(Y_T)|\big].
\end{aligned} \tag{51}$$

This implies that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\le \varepsilon_0\, \mathbb{E}[1 + \|X_T\|^{\varsigma_0}] \\
&\quad + L_0\, \mathbb{E}\bigg[\bigg(1 + \int_0^1 \big[r\|X_T\| + (1-r)\|Y_T\|\big]^\ell \, dr\bigg) \|X_T - Y_T\|\bigg] \\
&\le \varepsilon_0\, \mathbb{E}[1 + \|X_T\|^{\varsigma_0}] \\
&\quad + L_0\, \mathbb{E}\bigg[\bigg(1 + 2^{\max\{\ell-1,0\}} \int_0^1 \|rX_T\|^\ell + \|(1-r)Y_T\|^\ell \, dr\bigg) \|X_T - Y_T\|\bigg].
\end{aligned} \tag{52}$$

Therefore, we obtain that

$$
\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\le \varepsilon_0\,\mathbb{E}[1 + \|X_T\|^{\varsigma_0}] \\
&\quad + L_0\,\mathbb{E}\Bigg[\bigg(1 + 2^{\max\{\ell-1,0\}}\left[\int_0^1 r^\ell\,dr\right]\big[\|X_T\|^\ell + \|Y_T\|^\ell\big]\bigg)\,\|X_T - Y_T\|\Bigg] \\
&= \varepsilon_0\,\mathbb{E}[1 + \|X_T\|^{\varsigma_0}] \\
&\quad + L_0\,\mathbb{E}\Bigg[\bigg(1 + \left[\frac{2^{\max\{\ell-1,0\}}}{(\ell+1)}\right]\big[\|X_T\|^\ell + \|Y_T\|^\ell\big]\bigg)\,\|X_T - Y_T\|\Bigg].
\end{aligned} \tag{53}
$$

Hölder's inequality hence demonstrates that

$$
\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\le \varepsilon_0\,(1 + \mathbb{E}[\|X_T\|^{\varsigma_0}]) \\
&\quad + L_0\left(1 + \left[\frac{2^{\max\{\ell-1,0\}}}{(\ell+1)}\right]\Big[\big(\mathbb{E}\big[\|X_T\|^{\ell q}\big]\big)^{1/q} + \big(\mathbb{E}\big[\|Y_T\|^{\ell q}\big]\big)^{1/q}\Big]\right)(\mathbb{E}[\|X_T - Y_T\|^p])^{1/p}.
\end{aligned} \tag{54}
$$

Next observe that Proposition 4.4 (with $d = d$, $m = m$, $x = \xi$, $y = \phi_2(\xi)$, $L = L_1$, $T = T$, $\delta = \delta$, $p = p$, $B = B$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W = W$, $\mu = f_1$, $X = X$, $Y = Y$, $a = ([0,T] \times \Omega \ni (t,\omega) \mapsto \phi_1(Y_{\chi(t)}(\omega)) \in \mathbb{R}^d)$ in the notation of Proposition 4.4) ensures that

$$
\begin{aligned}
&(\mathbb{E}[\|X_T - Y_T\|^p])^{1/p} \\
&\le \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\,T\Big) \\
&\quad \cdot \left(\|\xi - \phi_2(\xi)\| + \delta^{(1-1/p)}\left[\int_0^T \mathbb{E}\big[\|\phi_1(Y_{\chi(s)}) - f_1(Y_s)\|^p\big]\,ds\right]^{1/p}\right) \\
&\le \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\,T\Big) \\
&\quad \cdot \left(\|\xi - \phi_2(\xi)\| + \delta^{(1-1/p)}\left[\int_0^T \mathbb{E}\big[\|\phi_1(Y_{\chi(s)}) - f_1(Y_{\chi(s)})\|^p\big]\,ds\right]^{1/p}\right) \\
&\quad + \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\,T\Big)\left(\delta^{(1-1/p)}\left[\int_0^T \mathbb{E}\big[\|f_1(Y_{\chi(s)}) - f_1(Y_s)\|^p\big]\,ds\right]^{1/p}\right) \\
&\le \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\,T\Big)\left(\|\xi - \phi_2(\xi)\| + \varepsilon_1\,\delta^{(1-1/p)}\left[\int_0^T \mathbb{E}\big[(1 + \|Y_{\chi(s)}\|^{\varsigma_1})^p\big]\,ds\right]^{1/p}\right) \\
&\quad + \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\,T\Big)\,L_1\,\delta^{(1-1/p)}\left[\int_0^T \mathbb{E}\big[\|Y_{\chi(s)} - Y_s\|^p\big]\,ds\right]^{1/p}.
\end{aligned} \tag{55}
$$

This shows that

$$
\begin{aligned}
&\left(\mathbb{E}[\|X_T - Y_T\|^p]\right)^{1/p} \\
&\le \exp\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, \|\xi - \phi_2(\xi)\| \\
&\quad + \exp\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, \varepsilon_1\, \delta^{(1-1/p)} \left[T^{1/p} + \left[\int_0^T \mathbb{E}\big[\|Y_{\chi(s)}\|^{p\varsigma_1}\big]\, ds\right]^{1/p}\right] \\
&\quad + \exp\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, L_1\, \delta^{(1-1/p)} \\
&\quad \cdot \left[\int_0^T \mathbb{E}\Big[\big\| \textstyle\int_{\chi(s)}^s \phi_1(Y_{\chi(u)})\, du + B(W_s - W_{\chi(s)})\big\|^p\Big]\, ds\right]^{1/p}.
\end{aligned}
\tag{56}
$$

Moreover, observe that the triangle inequality assures that

$$
\begin{aligned}
&\left[\int_0^T \mathbb{E}\Big[\big\| \textstyle\int_{\chi(s)}^s \phi_1(Y_{\chi(u)})\, du + B(W_s - W_{\chi(s)})\big\|^p\Big]\, ds\right]^{1/p} \\
&\le \left[\int_0^T \mathbb{E}\Big[\big\| \textstyle\int_{\chi(s)}^s \phi_1(Y_{\chi(u)})\, du\big\|^p\Big]\, ds\right]^{1/p} + \left[\int_0^T \mathbb{E}\Big[\big\|B(W_s - W_{\chi(s)})\big\|^p\Big]\, ds\right]^{1/p} \\
&= \left[\int_0^T \mathbb{E}\Big[|s-\chi(s)|^p \|\phi_1(Y_{\chi(s)})\|^p\Big]\, ds\right]^{1/p} + \left[\int_0^T \mathbb{E}\Big[\big\|B(W_{s-\chi(s)})\big\|^p\Big]\, ds\right]^{1/p} \\
&\le h\left[\int_0^T \mathbb{E}\Big[\|\phi_1(Y_{\chi(s)})\|^p\Big]\, ds\right]^{1/p} + \left[\int_0^T \mathbb{E}\Big[\big\|B(W_{s-\chi(s)})\big\|^p\Big]\, ds\right]^{1/p}.
\end{aligned}
\tag{57}
$$

This and (56) show that

$$
\begin{aligned}
&\left(\mathbb{E}[\|X_T - Y_T\|^p]\right)^{1/p} \\
&\le \exp\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, \|\xi - \phi_2(\xi)\| \\
&\quad + \exp\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, \varepsilon_1\, \delta^{(1-1/p)} \left[T^{1/p} + T^{1/p}\left[\sup_{t\in[0,T]} \big(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\big)^{1/p}\right]\right] \\
&\quad + \exp\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, h\, L_1\, \delta^{(1-1/p)}\, T^{1/p} \left[\sup_{t\in[0,T]} \big(\mathbb{E}\big[\|\phi_1(Y_{\chi(t)})\|^p\big]\big)^{1/p}\right] \\
&\quad + \exp\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, L_1\, \delta^{(1-1/p)} \left[\int_0^T \mathbb{E}\Big[\big\|B(W_{s-\chi(s)})\big\|^p\Big]\, ds\right]^{1/p}.
\end{aligned}
\tag{58}
$$

Therefore, we obtain that

$$
\begin{aligned}
&\left(\mathbb{E}[\|X_T - Y_T\|^p]\right)^{1/p} \\
&\leq \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big)\, \|\xi - \phi_2(\xi)\| \\
&\quad + \varepsilon_1\, \delta^{(1-1/p)}\, T^{1/p} \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big) \left[1 + \sup_{t\in[0,T]} \big(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\big)^{1/p}\right] \\
&\quad + h\, \delta^{(1-1/p)}\, T^{1/p}\, L_1 \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big) \left[\sup_{t\in[0,T]} \big(\mathbb{E}\big[\|\phi_1(Y_t)\|^p\big]\big)^{1/p}\right] \\
&\quad + \delta^{(1-1/p)}\, T^{1/p}\, L_1 \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big) \big(\mathbb{E}\big[\|BW_h\|^p\big]\big)^{1/p}.
\end{aligned}
\tag{59}
$$

Combining this with (54) demonstrates that

$$
\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq \varepsilon_0\, (1 + \mathbb{E}[\|X_T\|^{\varsigma_0}]) \\
&\quad + L_0 \left(1 + \left[\frac{2^{\max\{\ell-1,0\}}}{(\ell+1)}\right] \left[\big(\mathbb{E}\big[\|X_T\|^{\ell q}\big]\big)^{1/q} + \big(\mathbb{E}\big[\|Y_T\|^{\ell q}\big]\big)^{1/q}\right]\right) \\
&\quad \cdot \exp\!\Big(\Big[L_1 + \tfrac{(1-1/p)}{\delta}\Big]\, T\Big) \left[\|\xi - \phi_2(\xi)\| + \varepsilon_1\, \delta^{(1-1/p)}\, T^{1/p} \left[1 + \sup_{t\in[0,T]} \big(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\big)^{1/p}\right]\right. \\
&\quad \left. + h\, \delta^{(1-1/p)}\, T^{1/p} L_1 \left[\sup_{t\in[0,T]} \big(\mathbb{E}\big[\|\phi_1(Y_t)\|^p\big]\big)^{1/p}\right] + \delta^{(1-1/p)}\, T^{1/p} L_1 \big(\mathbb{E}\big[\|BW_h\|^p\big]\big)^{1/p}\right].
\end{aligned}
\tag{60}
$$

The proof of Lemma 4.5 is thus completed. □

Proposition 4.6. *Let $d, m \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $T \in (0,\infty)$, $c, C, \varepsilon_0, \varepsilon_1, \varepsilon_2, \varsigma_0, \varsigma_1, \varsigma_2, L_0, L_1, \ell \in [0,\infty)$, $h \in [0,T]$, $B \in \mathbb{R}^{d\times m}$, $p \in [2,\infty)$, $q \in (1,2]$ satisfy $1/p + 1/q = 1$, let $\|\cdot\| \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $W \colon [0,T]\times\Omega \to \mathbb{R}^m$ be a standard Brownian motion, let $\phi_0 \colon \mathbb{R}^d \to \mathbb{R}$, $f_1 \colon \mathbb{R}^d \to \mathbb{R}^d$, $\phi_2 \colon \mathbb{R}^d \to \mathbb{R}^d$, and $\chi \colon [0,T] \to [0,T]$ be functions, let $f_0 \colon \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable function, let $\phi_1 \colon \mathbb{R}^d \to \mathbb{R}^d$ be a $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable function, assume that $\|\xi - \phi_2(\xi)\| \leq \varepsilon_2(1 + \|\xi\|^{\varsigma_2})$, assume for all $t \in [0,T]$, $x, y \in \mathbb{R}^d$ that*

$$
|\phi_0(x) - f_0(x)| \leq \varepsilon_0(1 + \|x\|^{\varsigma_0}), \qquad \|\phi_1(x) - f_1(x)\| \leq \varepsilon_1(1 + \|x\|^{\varsigma_1}), \tag{61}
$$

$$
|\phi_0(x) - \phi_0(y)| \leq L_0 \left(1 + \int_0^1 \big[r\|x\| + (1-r)\|y\|\big]^\ell\, dr\right) \|x - y\|, \tag{62}
$$

$$
\|f_1(x) - f_1(y)\| \leq L_1\|x - y\|, \qquad \chi(t) = \max(\{0, h, 2h, \dots\} \cap [0,t]), \tag{63}
$$

and $\|\phi_1(x)\| \leq C + c\|x\|$, let $\varpi_r \in \mathbb{R}$, $r \in (0,\infty)$, satisfy for all $r \in (0,\infty)$ that $\varpi_r = \big(\mathbb{E}\big[\|BW_T\|^r\big]\big)^{1/r}$, and let $X, Y \colon [0,T]\times\Omega \to \mathbb{R}^d$ be stochastic processes with continuous sample paths which satisfy for all $t \in [0,T]$ that $Y_t = \phi_2(\xi) + \int_0^t \phi_1\big(Y_{\chi(s)}\big)\, ds + BW_t$ and

$$
X_t = \xi + \int_0^t f_1(X_s)\, ds + BW_t. \tag{64}
$$

Then it holds that

$$\begin{aligned}&\big|\mathbb{E}\big[f_0(X_T)\big]-\mathbb{E}\big[\phi_0(Y_T)\big]\big|\le\big[\varepsilon_0+\varepsilon_1+\varepsilon_2+(h/T)^{1/2}\big]\\&\cdot e^{(\ell+3+2L_1+[\ell\max\{L_1,c\}+c\max\{\varsigma_1,1\}+L_1\max\{\varsigma_0,1\}+2]T)}\big[\|\xi\|+\max\{1,\varepsilon_2\}(1+\|\xi\|^{\varsigma_2})\\&+\max\{1,C,\|f_1(0)\|\}\max\{1,T\}+\varpi_{\max\{\varsigma_0,\varsigma_1 p,p,\ell q\}}\big]^{\max\{1,\varsigma_0,\varsigma_1\}+\ell}\max\{1,L_0\}.\end{aligned}\tag{65}$$

Proof of Proposition 4.6. First, observe that Lemma 4.5 shows that

$$\begin{aligned}&\big|\mathbb{E}\big[f_0(X_T)\big]-\mathbb{E}\big[\phi_0(Y_T)\big]\big|\\&\le\varepsilon_0\,(1+\mathbb{E}[\|X_T\|^{\varsigma_0}])\\&+L_0\,2^{\max\{\ell-1,0\}}\,e^{[L_1+(1-1/p)]T}\Big[1+\big(\mathbb{E}\big[\|X_T\|^{\ell q}\big]\big)^{1/q}+\big(\mathbb{E}\big[\|Y_T\|^{\ell q}\big]\big)^{1/q}\Big]\\&\cdot\left[\|\xi-\phi_2(\xi)\|+\varepsilon_1 T^{1/p}\left[1+\sup_{t\in[0,T]}\big(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\big)^{1/p}\right]\right.\\&\left.+hT^{1/p}L_1\left[\sup_{t\in[0,T]}\big(\mathbb{E}\big[\|\phi_1(Y_t)\|^p\big]\big)^{1/p}\right]+T^{1/p}L_1\big(\mathbb{E}\big[\|BW_h\|^p\big]\big)^{1/p}\right].\end{aligned}\tag{66}$$

Hence, we obtain that

$$\begin{aligned}&\big|\mathbb{E}\big[f_0(X_T)\big]-\mathbb{E}\big[\phi_0(Y_T)\big]\big|\\&\le\varepsilon_0\,(1+\mathbb{E}[\|X_T\|^{\varsigma_0}])\\&+L_0\,2^{\max\{\ell-1,0\}}\,e^{[L_1+(1-1/p)]T}\Big[1+\big(\mathbb{E}\big[\|X_T\|^{\ell q}\big]\big)^{1/q}+\big(\mathbb{E}\big[\|Y_T\|^{\ell q}\big]\big)^{1/q}\Big]\\&\cdot\left[\|\xi-\phi_2(\xi)\|+\varepsilon_1 T^{1/p}\left[1+\sup_{t\in[0,T]}\big(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\big)^{1/p}\right]\right.\\&+h\,T^{1/p}L_1\varepsilon_1\left[\sup_{t\in[0,T]}\big(\mathbb{E}\big[(1+\|Y_t\|^{\varsigma_1})^p\big]\big)^{1/p}\right]\\&\left.+h\,T^{1/p}L_1\left[\sup_{t\in[0,T]}\big(\mathbb{E}\big[\|f_1(Y_t)\|^p\big]\big)^{1/p}\right]+L_1\varpi_p h^{1/2}T^{1/p-1/2}\right].\end{aligned}\tag{67}$$

In addition, note that for all $x\in\mathbb{R}^d$ it holds that

$$\|f_1(x)\|\le\|f_1(x)-f_1(0)\|+\|f_1(0)\|\le\|f_1(0)\|+L_1\|x\|.\tag{68}$$

This and (67) ensure that

$$
\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq \varepsilon_0 \left(1 + \mathbb{E}\big[\|X_T\|^{\varsigma_0}\big]\right) \\
&+ L_0\, 2^{\max\{\ell-1,0\}}\, e^{[L_1+1-1/p]T} \left[1 + \left(\mathbb{E}\big[\|X_T\|^{\ell q}\big]\right)^{1/q} + \left(\mathbb{E}\big[\|Y_T\|^{\ell q}\big]\right)^{1/q}\right] \\
&\cdot \Bigg[\|\xi - \phi_2(\xi)\| + \varepsilon_1 T^{1/p}\left[1 + \sup_{t\in[0,T]} \left(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\right)^{1/p}\right] \\
&+ hT^{1/p} L_1 \varepsilon_1 \left[1 + \sup_{t\in[0,T]} \left(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\right)^{1/p}\right] \\
&+ hT^{1/p} L_1 \left[\|f_1(0)\| + L_1 \left[\sup_{t\in[0,T]} \left(\mathbb{E}\big[\|Y_t\|^{p}\big]\right)^{1/p}\right]\right] + L_1 \varpi_p h^{1/2} T^{1/p-1/2}\Bigg] \\
&= \varepsilon_0 \left(1 + \mathbb{E}\big[\|X_T\|^{\varsigma_0}\big]\right) \\
&+ L_0\, 2^{\max\{\ell-1,0\}}\, e^{[L_1+1-1/p]T} \left[1 + \left(\mathbb{E}\big[\|X_T\|^{\ell q}\big]\right)^{1/q} + \left(\mathbb{E}\big[\|Y_T\|^{\ell q}\big]\right)^{1/q}\right] \\
&\cdot \Bigg[\|\xi - \phi_2(\xi)\| + \varepsilon_1 T^{1/p}\left[1 + hL_1\right]\left[1 + \sup_{t\in[0,T]} \left(\mathbb{E}\big[\|Y_t\|^{p\varsigma_1}\big]\right)^{1/p}\right] \\
&+ hT^{1/p} L_1 \left[\|f_1(0)\| + L_1 \left[\sup_{t\in[0,T]} \left(\mathbb{E}\big[\|Y_t\|^{p}\big]\right)^{1/p}\right]\right] + L_1 \varpi_p h^{1/2} T^{1/p-1/2}\Bigg].
\end{aligned}
\tag{69}
$$

Next observe that Lemma 4.1 and (68) demonstrate that for all $r \in [1, \infty)$, $t \in [0, T]$ it holds that

$$
\begin{aligned}
\sup_{t\in[0,T]} \left(\mathbb{E}\big[\|Y_t\|^{r}\big]\right)^{1/r} &\leq \left(\|\phi_2(\xi)\| + CT + \left(\mathbb{E}\big[\|BW_T\|^{r}\big]\right)^{1/r}\right) e^{cT} \\
&= \left(\|\phi_2(\xi)\| + CT + \varpi_r\right) e^{cT}
\end{aligned}
\tag{70}
$$

and

$$
\begin{aligned}
\sup_{t\in[0,T]} \left(\mathbb{E}\big[\|X_t\|^{r}\big]\right)^{1/r} &\leq \left(\|\xi\| + \|f_1(0)\|T + \left(\mathbb{E}\big[\|BW_T\|^{r}\big]\right)^{1/r}\right) e^{L_1 T} \\
&= \left(\|\xi\| + \|f_1(0)\|T + \varpi_r\right) e^{L_1 T}.
\end{aligned}
\tag{71}
$$

Combining this with (69) shows that

$$
\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq \varepsilon_0 \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\varsigma_0,1\}}\big]^{\varsigma_0} e^{\varsigma_0 L_1 T}\right) + L_0\, 2^{\max\{\ell-1,0\}}\, e^{[L_1+1-1/p]T} \\
&\cdot \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\ell q,1\}}\big]^{\ell} e^{\ell L_1 T} + \big[\|\phi_2(\xi)\| + CT + \varpi_{\max\{\ell q,1\}}\big]^{\ell} e^{\ell cT}\right) \\
&\cdot \Bigg[\|\xi - \phi_2(\xi)\| + \varepsilon_1 T^{1/p}[1 + hL_1]\left(1 + \big[\|\phi_2(\xi)\| + CT + \varpi_{\max\{\varsigma_1 p,1\}}\big]^{\varsigma_1} e^{\varsigma_1 cT}\right) \\
&+ hT^{1/p} L_1 \left(\|f_1(0)\| + L_1 \big[\|\phi_2(\xi)\| + CT + \varpi_p\big] e^{cT}\right) + L_1 \varpi_p h^{1/2} T^{1/p-1/2}\Bigg].
\end{aligned}
\tag{72}
$$

Hence, we obtain that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq \varepsilon_0 \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\varsigma_0,1\}}\big]^{\varsigma_0} e^{\varsigma_0 L_1 T}\right) \\
&+ L_0\, 2^{\max\{\ell-1,0\}}\, e^{[\max\{\ell L_1, \ell c\} + c\max\{\varsigma_1,1\} + L_1 + 1 - 1/p]T} \max\{1, T^{1/p}\} \\
&\cdot \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\ell q,1\}}\big]^{\ell} + \big[\|\phi_2(\xi)\| + CT + \varpi_{\max\{\ell q,1\}}\big]^{\ell}\right) \\
&\cdot \Bigg[\, \|\xi - \phi_2(\xi)\| + \varepsilon_1[1 + hL_1]\Big(1 + \big[\|\phi_2(\xi)\| + CT + \varpi_{\max\{\varsigma_1 p,1\}}\big]^{\varsigma_1}\Big) \\
&+ hL_1\Big(\|f_1(0)\| + L_1\big[\|\phi_2(\xi)\| + CT + \varpi_p\big]\Big) + (h/T)^{1/2} L_1 \varpi_p \Bigg].
\end{aligned} \tag{73}$$

This implies that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq \varepsilon_0 \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\varsigma_0,1\}}\big]^{\varsigma_0} e^{\varsigma_0 L_1 T}\right) \\
&+ L_0\, 2^{\max\{\ell-1,0\}}\, e^{[\max\{\ell L_1, \ell c\} + c\max\{\varsigma_1,1\} + L_1 + 1]T} \\
&\cdot \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\ell q,1\}}\big]^{\ell} + \big[\|\xi\| + \varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{\ell q,1\}}\big]^{\ell}\right) \\
&\cdot \Bigg[\varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + \varepsilon_1[1 + TL_1]\Big(1 + \big[\|\xi\| + \varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{\varsigma_1 p,1\}}\big]^{\varsigma_1}\Big) \\
&+ (h/T)^{1/2} TL_1\Big(\|f_1(0)\| + L_1\big[\|\phi_2(\xi)\| + CT + \varpi_p\big]\Big) + (h/T)^{1/2} L_1 \varpi_p \Bigg].
\end{aligned} \tag{74}$$

Therefore, we obtain that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq \varepsilon_0 \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\varsigma_0,1\}}\big]^{\varsigma_0} e^{\varsigma_0 L_1 T}\right) \\
&+ L_0\, 2^{\max\{\ell-1,0\}}\, e^{[\max\{\ell L_1, \ell c\} + c\max\{\varsigma_1,1\} + L_1 + 1]T} \\
&\cdot \left(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\ell q,1\}}\big]^{\ell} + \big[\|\xi\| + \varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{\ell q,1\}}\big]^{\ell}\right) \\
&\cdot \Bigg[[\varepsilon_1 + \varepsilon_2] \max\{1, T\}[1 + L_1] \\
&\cdot \left(1 + \big[\|\xi\| + \max\{1, \varepsilon_2\}(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{\varsigma_1 p,1\}}\big]^{\max\{1,\varsigma_1\}}\right) \\
&+ (h/T)^{1/2} \max\{1, T\}[1 + L_1]\Big(\|f_1(0)\| + L_1\big[\|\xi\| + \varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_p\big]\Big)\Bigg].
\end{aligned} \tag{75}$$

This and the fact that $\forall\, x \in [0, \infty)\colon\ \max\{x, 1\} \leq x + 1 \leq e^x$ demonstrate that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq \varepsilon_0 \Big(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\varsigma_0,1\}}\big]^{\varsigma_0} e^{\varsigma_0 L_1 T}\Big) \\
&+ L_0\, 2^{\max\{\ell-1,0\}}\, e^{(L_1+[\max\{\ell L_1,\ell c\}+c\max\{\varsigma_1,1\}+L_1+2]T)} \big[\varepsilon_1 + \varepsilon_2 + (h/T)^{1/2}\big] \\
&\cdot \Big(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\ell q,1\}}\big]^{\ell} + \big[\|\xi\| + \varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{\ell q,1\}}\big]^{\ell}\Big) \\
&\cdot \Big[\max\{1, \|f_1(0)\|\} \\
&+ \max\{1, L_1\}\big[\|\xi\| + \max\{1, \varepsilon_2\}(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{p,\varsigma_1 p\}}\big]^{\max\{1,\varsigma_1\}}\Big].
\end{aligned} \tag{76}$$

Hence, we obtain that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq 2^{\max\{\ell-1,0\}}\, e^{(L_1+[\max\{\ell L_1,\ell c\}+c\max\{\varsigma_1,1\}+\max\{\varsigma_0,1\}L_1+2]T)} \big[\varepsilon_0 + \varepsilon_1 + \varepsilon_2 + (h/T)^{1/2}\big] \\
&\cdot \Big(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\ell q,1\}}\big]^{\ell} + \big[\|\xi\| + \varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{\ell q,1\}}\big]^{\ell}\Big) \\
&\cdot \max\{1, L_0\}\Big[\max\{1, \|f_1(0)\|\} + \max\{1, L_1\}\big[\|\xi\| + \max\{1, \varepsilon_2\}(1 + \|\xi\|^{\varsigma_2}) \\
&+ \max\{C, \|f_1(0)\|\}T + \varpi_{\max\{p,\varsigma_1 p,\varsigma_0\}}\big]^{\max\{1,\varsigma_0,\varsigma_1\}}\Big].
\end{aligned} \tag{77}$$

This and the fact that $\forall\, x \in [0, \infty)\colon\ \max\{x, 1\} \leq x + 1 \leq e^x$ show that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq 2^{\max\{\ell,1\}}\, e^{(2L_1+[\ell\max\{L_1,c\}+c\max\{\varsigma_1,1\}+L_1\max\{\varsigma_0,1\}+2]T)} \big[\varepsilon_0 + \varepsilon_1 + \varepsilon_2 + (h/T)^{1/2}\big] \\
&\cdot \Big(1 + \big[\|\xi\| + \|f_1(0)\|T + \varpi_{\max\{\ell q,1\}}\big]^{\ell} + \big[\|\xi\| + \varepsilon_2(1 + \|\xi\|^{\varsigma_2}) + CT + \varpi_{\max\{\ell q,1\}}\big]^{\ell}\Big) \\
&\cdot \max\{1, L_0\}\big[\|\xi\| + \max\{1, \varepsilon_2\}(1 + \|\xi\|^{\varsigma_2}) \\
&+ \max\{1, C, \|f_1(0)\|\}\max\{1, T\} + \varpi_{\max\{\varsigma_0,\varsigma_1 p,p\}}\big]^{\max\{1,\varsigma_0,\varsigma_1\}}.
\end{aligned} \tag{78}$$

Therefore, we obtain that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_0(X_T)\big] - \mathbb{E}\big[\phi_0(Y_T)\big]\big| \\
&\leq e^{(\ell+3+2L_1+[\ell\max\{L_1,c\}+c\max\{\varsigma_1,1\}+L_1\max\{\varsigma_0,1\}+2]T)} \\
&\cdot \big[\varepsilon_0 + \varepsilon_1 + \varepsilon_2 + (h/T)^{1/2}\big] \max\{1, L_0\}\big[\|\xi\| + \max\{1, \varepsilon_2\}(1 + \|\xi\|^{\varsigma_2}) \\
&+ \max\{1, C, \|f_1(0)\|\}\max\{1, T\} + \varpi_{\max\{\varsigma_0,\varsigma_1 p,p,\ell q\}}\big]^{\max\{1,\varsigma_0,\varsigma_1\}+\ell}.
\end{aligned} \tag{79}$$

The proof of Proposition 4.6 is thus completed. □

5 Deep artificial neural network (DNN) calculus

In Section 6 below we establish the existence of a DNN approximating the solution of the PDE without the curse of dimensionality. To demonstrate the existence of such a DNN, we need a few properties about representation flexibilities of DNNs, which we establish in this section. In particular, we state in the elementary and essentially well-known result in Lemma 5.1 in Subsection 5.1 below that every linear combination of realizations of DNNs with the same architecture is again a realization of a suitable DNN. Results similar to Lemma 5.1 can, e.g., be found in Yarotsky [66].

Moreover, in Proposition 5.2 in Subsection 5.2 below we demonstrate under suitable hypotheses that the composition of the realizations of two DNNs is again a realization of a suitable DNN and the number of parameters of this suitable DNN grows at most additively in the number of parameters of the composed DNNs. For the construction of this suitable DNN in Proposition 5.2 we plug an artificial identity in between the two DNNs and for this we employ in Proposition 5.2 the hypothesis that the identity can within the class of considered fully-connected neural networks (see (93)–(94) in Proposition 5.2 below) be described by a suitable flat artificial neural network. In Proposition 5.2 the tuples ϕ_1 and ϕ_2 represent the DNNs which we intend to compose (where the realization of ϕ_1 is a function from $\mathbb{R}^{d_2}$ to $\mathbb{R}^{d_3}$ and where the realization of ϕ_2 is a function from $\mathbb{R}^{d_1}$ to $\mathbb{R}^{d_2}$), the tuple $\mathbb{I}$ represents the artificial neural network which describes the identity on $\mathbb{R}^{d_2}$, and the tuple ψ represents the DNN whose realization coincides with the composition of the realizations of ϕ_1 and ϕ_2 (the realization of ψ is thus a function from $\mathbb{R}^{d_1}$ to $\mathbb{R}^{d_3}$). The hypothesis of the existence of the artificial neural network $\mathbb{I}$ can, roughly speaking, be viewed as a hypothesis on the activation function $\mathbf{a}\colon \mathbb{R} \to \mathbb{R}$ used in Proposition 5.2. Proposition 5.2, loosely speaking, then asserts that the number of parameters of ψ can up to a constant be bounded by the sum of the number of parameters of ϕ_1 and of the number of parameters of ϕ_2. A straightforward DNN construction of the composition of ϕ_1 and ϕ_2 (without artificially plugging the identity on $\mathbb{R}^{d_2}$ in between ϕ_1 and ϕ_2) would possibly result in a DNN whose number of parameters is essentially equal to the product of the number of parameters of ϕ_1 and of the number of parameters of ϕ_2. Such a construction, in turn, would in our proof of the main result of this article (Theorem 6.3 below) not allow us to conclude that DNNs do indeed overcome the curse of dimensionality in the numerical approximation of the considered PDEs (see (177) in the proof of Proposition 6.1 for details). Moreover, in Proposition 5.3 in Subsection 5.2 below we establish under similar hypotheses as in Proposition 5.2 a result similar to Proposition 5.2 which is tailor-made to the DNNs which we design in the proof of our main result in Theorem 6.3 below. In particular, (109) in Proposition 5.3 is tailor-made to construct a DNN which is based on an Euler discretization of a (stochastic) differential equation. We refer to (147) and (175) in the proof of Proposition 6.1 below for further details.

To apply Proposition 5.2 and Proposition 5.3, respectively, we need to verify that the class of considered ANNs does indeed enjoy the property to be able to represent the identity on $\mathbb{R}^{d_2}$. Fortunately, ANNs with the rectifier function as the activation function do indeed admit this property. This fact is verified in the elementary result in Lemma 5.4 in Subsection 5.3 below. In particular, Lemma 5.4 shows for every $d \in \mathbb{N}$ that the d-dimensional identity can be explicitly represented by a suitable flat rectified ANN (with one hidden layer with $2d$ neurons and the rectifier function as the activation function in front of the $2d$-dimensional hidden layer).

5.1 Sums of DNNs with the same architecture

Lemma 5.1. *Let* $\mathbf{A}_n\colon \mathbb{R}^n \to \mathbb{R}^n$*,* $n \in \mathbb{N}$*, and* $\mathbf{a}\colon \mathbb{R} \to \mathbb{R}$ *be continuous functions which satisfy for all* $n \in \mathbb{N}$*,* $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ *that* $\mathbf{A}_n(x) = (\mathbf{a}(x_1), \ldots, \mathbf{a}(x_n))$*, let*

$$\mathcal{N} = \cup_{L\in\{2,3,4,\ldots\}} \cup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n})\right), \tag{80}$$

let $\mathcal{P}\colon \mathcal{N} \to \mathbb{N}$ *and* $\mathcal{R}\colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ *be the functions which satisfy for all* $L \in \{2, 3, 4, \ldots\}$*,* $l_0, l_1, \ldots, l_L \in \mathbb{N}$*,* $\Phi = ((W_1, B_1), \ldots, (W_L, B_L)) \in (\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}}\times \mathbb{R}^{l_n}))$*,* $x_0 \in \mathbb{R}^{l_0}$*,* $\ldots$*,* $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ *with* $\forall\, n \in \mathbb{N} \cap [1, L)\colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ *that* $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1} + 1)$*,* $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$*, and*

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{81}$$

let $\mathbb{L} \in \{2, 3, 4, \ldots\}$*,* M*,* $\mathfrak{L}_0, \mathfrak{L}_1, \ldots, \mathfrak{L}_{\mathbb{L}} \in \mathbb{N}$*,* $h_1, h_2, \ldots, h_M \in \mathbb{R}$*, and let* $(\phi_m)_{m\in\{1,2,\ldots,M\}} \subseteq (\times_{n=1}^{\mathbb{L}}(\mathbb{R}^{\mathfrak{L}_n\times\mathfrak{L}_{n-1}} \times \mathbb{R}^{\mathfrak{L}_n}))$*. Then there exists* $\psi \in \mathcal{N}$ *such that for all* $x \in \mathbb{R}^{\mathfrak{L}_0}$ *it holds that* $\mathcal{R}(\psi) \in C(\mathbb{R}^{\mathfrak{L}_0}, \mathbb{R}^{\mathfrak{L}_{\mathbb{L}}})$*,* $\mathcal{P}(\psi) \le M^2\mathcal{P}(\phi_1)$*, and*

$$(\mathcal{R}\psi)(x) = \sum_{m=1}^{M} h_m(\mathcal{R}\phi_m)(x). \tag{82}$$

Proof of Lemma 5.1. Throughout this proof let $((W_{m,1}, B_{m,1}), \ldots, (W_{m,\mathbb{L}}, B_{m,\mathbb{L}})) \in (\times_{n=1}^{\mathbb{L}}(\mathbb{R}^{\mathfrak{L}_n\times\mathfrak{L}_{n-1}} \times \mathbb{R}^{\mathfrak{L}_n}))$, $m \in \{1, 2, \ldots, M\}$, satisfy for all $i \in \{1, 2, \ldots M\}$ that $\phi_i = ((W_{i,1}, B_{i,1}), \ldots, (W_{i,\mathbb{L}}, B_{i,\mathbb{L}}))$, let $(l_0, l_1, \ldots, l_{\mathbb{L}}) \in \mathbb{N}^{\mathbb{L}+1}$ satisfy for all $i \in \{1, 2, \ldots, \mathbb{L}-1\}$ that $l_0 = \mathfrak{L}_0$, $l_i = M\mathfrak{L}_i$, and $l_{\mathbb{L}} = \mathfrak{L}_{\mathbb{L}}$, let $((W_1, B_1), \ldots, (W_{\mathbb{L}}, B_{\mathbb{L}})) \in (\times_{n=1}^{\mathbb{L}}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n}))$ satisfy that

$$W_1 = \begin{pmatrix} W_{1,1} \\ W_{2,1} \\ \vdots \\ W_{M,1} \end{pmatrix} \in \mathbb{R}^{(M\mathfrak{L}_1)\times\mathfrak{L}_0} = \mathbb{R}^{l_1\times l_0}, \qquad B_1 = \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ \vdots \\ B_{M,1} \end{pmatrix} \in \mathbb{R}^{(M\mathfrak{L}_1)} = \mathbb{R}^{l_1}, \tag{83}$$

$$W_{\mathbb{L}} = \begin{pmatrix} h_1 W_{1,\mathbb{L}} & h_2 W_{2,\mathbb{L}} & \cdots & h_M W_{M,\mathbb{L}} \end{pmatrix} \in \mathbb{R}^{\mathfrak{L}_{\mathbb{L}}\times(M\mathfrak{L}_{\mathbb{L}-1})} = \mathbb{R}^{l_{\mathbb{L}}\times l_{\mathbb{L}-1}}, \tag{84}$$

$$\text{and} \qquad B_{\mathbb{L}} = \sum_{m=1}^{M} h_m B_{m,\mathbb{L}} \in \mathbb{R}^{\mathfrak{L}_{\mathbb{L}}} = \mathbb{R}^{l_{\mathbb{L}}}, \tag{85}$$

assume for all $i \in \{2, 3, 4, \ldots\} \cap [0, \mathbb{L} - 1]$ that

$$W_i = \begin{pmatrix} W_{1,i} & 0 & \cdots & 0 \\ 0 & W_{2,i} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & W_{M,i} \end{pmatrix} \in \mathbb{R}^{(M\mathfrak{L}_i)\times(M\mathfrak{L}_{i-1})} = \mathbb{R}^{l_i\times l_{i-1}} \tag{86}$$

$$\text{and} \qquad B_i = \begin{pmatrix} B_{1,i} \\ B_{2,i} \\ \vdots \\ B_{M,i} \end{pmatrix} \in \mathbb{R}^{(M\mathfrak{L}_i)} = \mathbb{R}^{l_i}, \tag{87}$$

and let $\psi = ((W_1, B_1), \dots, (W_{\mathbb{L}}, B_{\mathbb{L}})) \in \mathcal{N}$. Note that for all $x \in \mathbb{R}^{l_0}$ it holds that

$$W_1 x + B_1 = \begin{pmatrix} W_{1,1}x + B_{1,1} \\ W_{2,1}x + B_{2,1} \\ \vdots \\ W_{M,1}x + B_{M,1} \end{pmatrix}. \tag{88}$$

Moreover, observe that for all $i \in \mathbb{N} \cap [0, \mathbb{L}-2]$, $x_1, x_2, \dots, x_M \in \mathbb{R}^{l_i}$ it holds that

$$\begin{aligned} W_{i+1}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} + B_{i+1} &= \begin{pmatrix} W_{1,i+1} & 0 & \cdots & 0 \\ 0 & W_{2,i+1} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & W_{M,i+1} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} + \begin{pmatrix} B_{1,i+1} \\ B_{2,i+1} \\ \vdots \\ B_{M,i+1} \end{pmatrix} \\ &= \begin{pmatrix} W_{1,i+1}x_1 + B_{1,i+1} \\ W_{2,i+1}x_2 + B_{2,i+1} \\ \vdots \\ W_{M,i+1}x_M + B_{M,i+1} \end{pmatrix}. \end{aligned} \tag{89}$$

Next note that for all $x_1, x_2, \dots, x_M \in \mathbb{R}^{l_{\mathbb{L}-1}}$ it holds that

$$\begin{aligned} W_{\mathbb{L}}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} + B_{\mathbb{L}} &= \begin{pmatrix} h_1 W_{1,\mathbb{L}} & h_2 W_{2,\mathbb{L}} & \cdots & h_M W_{M,\mathbb{L}} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} + \sum_{m=1}^{M} h_m B_{m,\mathbb{L}} \\ &= \left[\sum_{m=1}^{M} h_m W_{m,\mathbb{L}} x_m\right] + \left[\sum_{m=1}^{M} h_m B_{m,\mathbb{L}}\right] = \sum_{m=1}^{M} h_m \big(W_{m,\mathbb{L}} x_m + B_{m,\mathbb{L}}\big). \end{aligned} \tag{90}$$

This, (88), and (89) ensure that for all $x \in \mathbb{R}^{\mathfrak{L}_0}$ it holds that $\mathcal{R}(\psi) \in C(\mathbb{R}^{\mathfrak{L}_0}, \mathbb{R}^{\mathfrak{L}_{\mathbb{L}}})$ and

$$(\mathcal{R}\psi)(x) = \sum_{m=1}^{M} h_m (\mathcal{R}\phi_m)(x). \tag{91}$$

Moreover, observe that the assumption that for all $i \in \{1, 2, \dots, \mathbb{L}-1\}$ it holds that $l_0 = \mathfrak{L}_0$, $l_i = M\mathfrak{L}_i$, and $l_{\mathbb{L}} = \mathfrak{L}_{\mathbb{L}}$ assures that

$$\begin{aligned} \mathcal{P}(\psi) &= \sum_{n=1}^{\mathbb{L}} l_n(l_{n-1}+1) = l_1(l_0+1) + l_{\mathbb{L}}(l_{\mathbb{L}-1}+1) + \sum_{n=2}^{\mathbb{L}-1} l_n(l_{n-1}+1) \\ &= M\mathfrak{L}_1(\mathfrak{L}_0+1) + \mathfrak{L}_{\mathbb{L}}(M\mathfrak{L}_{\mathbb{L}-1}+1) + \sum_{n=2}^{\mathbb{L}-1} M\mathfrak{L}_n(M\mathfrak{L}_{n-1}+1) \\ &\le M^2\left[\sum_{n=1}^{\mathbb{L}} \mathfrak{L}_n(\mathfrak{L}_{n-1}+1)\right] = M^2\mathcal{P}(\phi_1). \end{aligned} \tag{92}$$

Combining this with (91) establishes (82). The proof of Lemma 5.1 is thus completed. □

5.2 Compositions of DNNs involving artificial identities

Proposition 5.2 (Composition of neural networks)**.** *Let $d_1, d_2, d_3 \in \mathbb{N}$, let $\mathbf{A}_n \colon \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$, and $\mathbf{a} \colon \mathbb{R} \to \mathbb{R}$ be continuous functions which satisfy for all $n \in \mathbb{N}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ that $\mathbf{A}_n(x) = (\mathbf{a}(x_1), \dots, \mathbf{a}(x_n))$, let*

$$\mathcal{N} = \cup_{L \in \{2,3,4,\dots\}} \cup_{(l_0,l_1,\dots,l_L) \in \mathbb{N}^{L+1}} \left(\times_{n=1}^{L} (\mathbb{R}^{l_n \times l_{n-1}} \times \mathbb{R}^{l_n})\right), \tag{93}$$

let $\mathcal{P} \colon \mathcal{N} \to \mathbb{N}$, $\mathcal{L} \colon \mathcal{N} \to \cup_{L \in \{2,3,4,\dots\}} \mathbb{N}^{L+1}$, and $\mathcal{R} \colon \mathcal{N} \to \cup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ be the functions which satisfy for all $L \in \{2, 3, 4, \dots\}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_L, B_L)) \in (\times_{n=1}^{L} (\mathbb{R}^{l_n \times l_{n-1}} \times \mathbb{R}^{l_n}))$, $x_0 \in \mathbb{R}^{l_0}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, n \in \mathbb{N} \cap [1, L) \colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ that $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1}+1)$, $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, $\mathcal{L}(\Phi) = (l_0, l_1, \dots, l_L)$, and

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{94}$$

and let $\phi_1, \phi_2, \mathbb{I} \in \mathcal{N}$, $L_1, L_2 \in \{2, 3, 4 \dots\}$, $\mathfrak{i}, l_{1,0}, l_{1,1}, \dots, l_{1,L_1}, l_{2,0}, l_{2,1}, \dots, l_{2,L_2} \in \mathbb{N}$ satisfy for all $x \in \mathbb{R}^{d_2}$, $i \in \{1, 2\}$ that $\mathcal{R}(\phi_1) \in C(\mathbb{R}^{d_2}, \mathbb{R}^{d_3})$, $\mathcal{R}(\phi_2) \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$, $\mathcal{R}(\mathbb{I}) \in C(\mathbb{R}^{d_2}, \mathbb{R}^{d_2})$, $\mathcal{L}(\phi_i) = (l_{i,0}, l_{i,1}, \dots, l_{i,L_i}) \in \mathbb{N}^{L_i+1}$, $\mathcal{L}(\mathbb{I}) = (d_2, \mathfrak{i}, d_2) \in \mathbb{N}^3$, and $(\mathcal{R}\,\mathbb{I})(x) = x$. Then there exists $\psi \in \mathcal{N}$ such that for all $x \in \mathbb{R}^{d_1}$ it holds that $\mathcal{R}(\psi) \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_3})$, $\mathcal{L}(\psi) = (l_{2,0}, l_{2,1}, \dots, l_{2,L_2-1}, \mathfrak{i}, l_{1,1}, l_{1,2}, \dots, l_{1,L_1}) \in \mathbb{N}^{L_1+L_2+1}$, $\mathcal{P}(\psi) \le \max\{1, 2^{-1}(d_2)^{-2}\, \mathcal{P}(\mathbb{I})\}(\mathcal{P}(\phi_1) + \mathcal{P}(\phi_2))$, and

$$(\mathcal{R}\psi)(x) = (\mathcal{R}\phi_1)\big((\mathcal{R}\phi_2)(x)\big) = \big((\mathcal{R}\phi_1) \circ (\mathcal{R}\phi_2)\big)(x). \tag{95}$$

Proof of Proposition 5.2. Throughout this proof let $(W_{3,1}, B_{3,1}) \in \mathbb{R}^{\mathfrak{i} \times d_2} \times \mathbb{R}^{\mathfrak{i}}$, $(W_{3,2}, B_{3,2}) \in \mathbb{R}^{d_2 \times \mathfrak{i}} \times \mathbb{R}^{d_2}$, and $((W_{j,1}, B_{j,1}), \dots, (W_{j,L_j}, B_{j,L_j})) \in (\times_{n=1}^{L_j} (\mathbb{R}^{l_{j,n} \times l_{j,n-1}} \times \mathbb{R}^{l_{j,n}}))$, $j \in \{1, 2\}$, satisfy for all $j \in \{1, 2\}$ that $\mathbb{I} = ((W_{3,1}, B_{3,1}), (W_{3,2}, B_{3,2}))$ and $\phi_j = ((W_{j,1}, B_{j,1}), \dots, (W_{j,L_j}, B_{j,L_j}))$, let $L_4 = L_1 + L_2$, let $l_{4,0}, l_{4,1}, \dots, l_{4,L_4} \in \mathbb{N}$ satisfy for all $i \in \{0, 1, \dots, L_2 - 1\}$, $j \in \{1, 2, \dots, L_1\}$ that

$$l_{4,i} = l_{2,i}, \qquad l_{4,L_2} = \mathfrak{i}, \qquad \text{and} \qquad l_{4,L_2+j} = l_{1,j}, \tag{96}$$

let $((W_{4,1}, B_{4,1}), \dots, (W_{4,L_4}, B_{4,L_4})) \in (\times_{n=1}^{L_4} (\mathbb{R}^{l_{4,n} \times l_{4,n-1}} \times \mathbb{R}^{l_{4,n}}))$ satisfy for all $i \in \{1, 2, \dots, L_2 - 1\}$, $j \in \{2, 3, \dots, L_1\}$ that

$$(W_{4,i}, B_{4,i}) = (W_{2,i}, B_{2,i}), \tag{97}$$

$$(W_{4,L_2}, B_{4,L_2}) = (W_{3,1} W_{2,L_2}, W_{3,1} B_{2,L_2} + B_{3,1}), \tag{98}$$

$$(W_{4,L_2+1}, B_{4,L_2+1}) = (W_{1,1} W_{3,2}, W_{1,1} B_{3,2} + B_{1,1}), \tag{99}$$

and $(W_{4,L_2+j}, B_{4,L_2+j}) = (W_{1,j}, B_{1,j})$, and let $\psi = ((W_{4,1}, B_{4,1}), \dots, (W_{4,L_4}, B_{4,L_4})) \in \mathcal{N}$. Observe that for all $x \in \mathbb{R}^{l_{4,L_2-1}} = \mathbb{R}^{l_{2,L_2-1}}$, $y \in \mathbb{R}^{l_{4,L_2}} = \mathbb{R}^{\mathfrak{i}}$ it holds that

$$W_{4,L_2} x + B_{4,L_2} = W_{3,1} W_{2,L_2} x + W_{3,1} B_{2,L_2} + B_{3,1} = W_{3,1}(W_{2,L_2} x + B_{2,L_2}) + B_{3,1} \tag{100}$$

and

$$W_{4,L_2+1} y + B_{4,L_2+1} = W_{1,1} W_{3,2} y + W_{1,1} B_{3,2} + B_{1,1} = W_{1,1}(W_{3,2} y + B_{3,2}) + B_{1,1}. \tag{101}$$

This ensures that for all $x \in \mathbb{R}^{d_1}$ it holds that $\mathcal{R}(\psi) \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_3})$ and

$$(\mathcal{R}\psi)(x) = (\mathcal{R}\phi_1)\Big((\mathcal{R}\,\mathbb{I})\big((\mathcal{R}\phi_2)(x)\big)\Big) = (\mathcal{R}\phi_1)\big((\mathcal{R}\phi_2)(x)\big). \tag{102}$$

Moreover, note that

$$\begin{aligned}\mathcal{P}(\psi) &= \sum_{n=1}^{L_4} l_{4,n}(l_{4,n-1}+1) = \sum_{n=1}^{L_1+L_2} l_{4,n}(l_{4,n-1}+1)\\ &= \left[\sum_{n=1}^{L_2-1} l_{4,n}(l_{4,n-1}+1)\right] + \left[\sum_{n=L_2+2}^{L_1+L_2} l_{4,n}(l_{4,n-1}+1)\right]\\ &\quad + l_{4,L_2}(l_{4,L_2-1}+1) + l_{4,L_2+1}(l_{4,L_2}+1)\\ &= \left[\sum_{n=1}^{L_2-1} l_{2,n}(l_{2,n-1}+1)\right] + \left[\sum_{n=2}^{L_1} l_{4,L_2+n}(l_{4,L_2+n-1}+1)\right]\\ &\quad + \mathfrak{i}\,(l_{2,L_2-1}+1) + l_{1,1}(\mathfrak{i}+1).\end{aligned} \tag{103}$$

Hence, we obtain that

$$\begin{aligned}\mathcal{P}(\psi) &= \left[\sum_{n=1}^{L_2-1} l_{2,n}(l_{2,n-1}+1)\right] + \left[\sum_{n=2}^{L_1} l_{1,n}(l_{1,n-1}+1)\right]\\ &\quad + \mathfrak{i}\,(l_{2,L_2-1}+1) + l_{1,1}(\mathfrak{i}+1)\\ &= \left[\sum_{n=1}^{L_2-1} l_{2,n}(l_{2,n-1}+1)\right] + \left[\sum_{n=2}^{L_1} l_{1,n}(l_{1,n-1}+1)\right]\\ &\quad + \frac{\mathfrak{i}}{d_2}\cdot l_{2,L_2}(l_{2,L_2-1}+1) + l_{1,1}\left(l_{1,0}\cdot\frac{\mathfrak{i}}{d_2}+1\right)\\ &\le \max\{1,\mathfrak{i}/d_2\}\left(\left[\sum_{n=1}^{L_2} l_{2,n}(l_{2,n-1}+1)\right] + \left[\sum_{n=1}^{L_1} l_{1,n}(l_{1,n-1}+1)\right]\right)\\ &= \max\{1,\mathfrak{i}/d_2\}\big(\mathcal{P}(\phi_1)+\mathcal{P}(\phi_2)\big).\end{aligned} \tag{104}$$

Next observe that

$$\mathcal{P}(\mathbb{I}) = \mathfrak{i}\,(d_2+1) + d_2(\mathfrak{i}+1) = 2\,\mathfrak{i}\,d_2 + \mathfrak{i} + d_2 > 2\,\mathfrak{i}\,d_2. \tag{105}$$

This and (104) ensure that

$$\mathcal{P}(\psi) \le \max\{1, 2^{-1}(d_2)^{-2}\,\mathcal{P}(\mathbb{I})\}\big(\mathcal{P}(\phi_1)+\mathcal{P}(\phi_2)\big). \tag{106}$$

Combining this with (102) establishes (95). The proof of Proposition 5.2 is thus completed. □

Proposition 5.3. *Let $d \in \mathbb{N}$, let $\mathbf{A}_n\colon \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$, and $\mathbf{a}\colon \mathbb{R} \to \mathbb{R}$ be continuous functions which satisfy for all $n \in \mathbb{N}$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ that $\mathbf{A}_n(x) = (\mathbf{a}(x_1), \ldots, \mathbf{a}(x_n))$, let*

$$\mathcal{N} = \cup_{L\in\{2,3,4,\ldots\}} \cup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n})\right), \tag{107}$$

let $\mathcal{P}\colon \mathcal{N} \to \mathbb{N}$, $\mathcal{L}\colon \mathcal{N} \to \cup_{L\in\{2,3,4,\dots\}} \mathbb{N}^{L+1}$, *and* $\mathcal{R}\colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ *be the functions which satisfy for all* $L \in \{2, 3, 4, \dots\}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_L, B_L)) \in (\times_{n=1}^L (\mathbb{R}^{l_n \times l_{n-1}} \times \mathbb{R}^{l_n}))$, $x_0 \in \mathbb{R}^{l_0}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ *with* $\forall\, n \in \mathbb{N} \cap [1, L)\colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ *that* $\mathcal{P}(\Phi) = \sum_{n=1}^L l_n(l_{n-1}+1)$, $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, $\mathcal{L}(\Phi) = (l_0, l_1, \dots, l_L)$, *and*

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{108}$$

and let $\phi_1, \phi_2, \mathbb{I} \in \mathcal{N}$, $L_1, L_2 \in \{2, 3, 4 \dots\}$, $\mathfrak{i}, l_{1,0}, l_{1,1}, \dots, l_{1,L_1}, l_{2,0}, l_{2,1}, \dots, l_{2,L_2} \in \mathbb{N}$ *satisfy for all* $x \in \mathbb{R}^d$, $i \in \{1, 2\}$ *that* $\mathcal{R}(\phi_i) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\mathcal{L}(\phi_i) = (l_{i,0}, l_{i,1}, \dots, l_{i,L_i}) \in \mathbb{N}^{L_i+1}$, $\mathcal{L}(\mathbb{I}) = (d, \mathfrak{i}, d) \in \mathbb{N}^3$, $(\mathcal{R}\,\mathbb{I})(x) = x$, *and* $l_{1,L_1-1} \le l_{2,L_2-1} + \mathfrak{i}$. *Then there exists* $\psi \in \mathcal{N}$ *such that for all* $x \in \mathbb{R}^d$ *it holds that* $\mathcal{R}(\psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\mathcal{L}(\psi) = (l_{1,0}, l_{1,1}, \dots, l_{1,L_1-1}, l_{2,1} + \mathfrak{i}, l_{2,2} + \mathfrak{i}, \dots, l_{2,L_2-1} + \mathfrak{i}, l_{2,L_2}) \in \mathbb{N}^{L_1+L_2}$, $\mathcal{P}(\psi) \le \mathcal{P}(\phi_1) + (\mathcal{P}(\phi_2) + \mathcal{P}(\mathbb{I}))^3$, *and*

$$(\mathcal{R}\psi)(x) = (\mathcal{R}\phi_1)(x) + (\mathcal{R}\phi_2)\big((\mathcal{R}\phi_1)(x)\big). \tag{109}$$

Proof of Proposition 5.3. Throughout this proof let $(W_{3,1}, B_{3,1}) \in (\mathbb{R}^{\mathfrak{i}\times d} \times \mathbb{R}^{\mathfrak{i}})$, $(W_{3,2}, B_{3,2}) \in (\mathbb{R}^{d\times\mathfrak{i}} \times \mathbb{R}^d)$, and $((W_{j,1}, B_{j,1}), \dots, (W_{j,L_j}, B_{j,L_j})) \in (\times_{n=1}^{L_j} (\mathbb{R}^{l_{j,n} \times l_{j,n-1}} \times \mathbb{R}^{l_{j,n}}))$, $j \in \{1, 2\}$, satisfy for all $j \in \{1, 2\}$ that $\mathbb{I} = ((W_{3,1}, B_{3,1}), (W_{3,2}, B_{3,2}))$ and $\phi_j = ((W_{j,1}, B_{j,1}), \dots, (W_{j,L_j}, B_{j,L_j}))$, let $L_4 = L_1 + L_2 - 1$, let $l_{4,0}, l_{4,1}, \dots, l_{4,L_4} \in \mathbb{N}$ satisfy for all $i \in \{0, 1, \dots, L_1 - 1\}$, $j \in \{0, 1, \dots, L_2 - 2\}$ that

$$l_{4,i} = l_{1,i}, \qquad l_{4,L_1+j} = l_{2,j+1} + \mathfrak{i}, \qquad \text{and} \qquad l_{4,L_4} = l_{2,L_2} = d, \tag{110}$$

let $((W_{4,1}, B_{4,1}), \dots, (W_{4,L_4}, B_{4,L_4})) \in (\times_{n=1}^{L_4} (\mathbb{R}^{l_{4,n} \times l_{4,n-1}} \times \mathbb{R}^{l_{4,n}}))$, assume for all $i \in \{1, 2, \dots, L_1 - 1\}$ that

$$(W_{4,i}, B_{4,i}) = (W_{1,i}, B_{1,i}) \in (\mathbb{R}^{l_{1,i} \times l_{1,i-1}} \times \mathbb{R}^{l_{1,i}}) = (\mathbb{R}^{l_{4,i} \times l_{4,i-1}} \times \mathbb{R}^{l_{4,i}}), \tag{111}$$

$$W_{4,L_1} = \begin{pmatrix} W_{2,1} W_{1,L_1} \\ W_{3,1} W_{1,L_1} \end{pmatrix} \in \mathbb{R}^{(l_{2,1}+\mathfrak{i}) \times l_{1,L_1-1}} = \mathbb{R}^{l_{4,L_1} \times l_{4,L_1-1}}, \tag{112}$$

$$B_{4,L_1} = \begin{pmatrix} W_{2,1} B_{1,L_1} + B_{2,1} \\ W_{3,1} B_{1,L_1} + B_{3,1} \end{pmatrix} \in \mathbb{R}^{(l_{2,1}+\mathfrak{i})} = \mathbb{R}^{l_{4,L_1}}, \tag{113}$$

$$W_{4,L_4} = \begin{pmatrix} W_{2,L_2} & W_{3,2} \end{pmatrix} \in \mathbb{R}^{l_{2,L_2} \times (l_{2,L_2-1}+\mathfrak{i})} = \mathbb{R}^{l_{4,L_4} \times l_{4,L_4-1}}, \tag{114}$$

$$\text{and} \qquad B_{4,L_4} = B_{2,L_2} + B_{3,2} \in \mathbb{R}^{l_{2,L_2}} = \mathbb{R}^{l_{4,L_4}}, \tag{115}$$

assume for all $j \in \mathbb{N} \cap [0, L_2 - 2]$ that

$$W_{4,L_1+j} = \begin{pmatrix} W_{2,j+1} & 0 \\ 0 & W_{3,1} W_{3,2} \end{pmatrix} \in \mathbb{R}^{(l_{2,j+1}+\mathfrak{i}) \times (l_{2,j}+\mathfrak{i})} = \mathbb{R}^{l_{4,L_1+j} \times l_{4,L_1+j-1}} \tag{116}$$

$$\text{and} \qquad B_{4,L_1+j} = \begin{pmatrix} B_{2,j+1} \\ W_{3,1} B_{3,2} + B_{3,1} \end{pmatrix} \in \mathbb{R}^{(l_{2,j+1}+\mathfrak{i})} = \mathbb{R}^{l_{4,L_1+j}}, \tag{117}$$

and let $\psi = ((W_{4,1}, B_{4,1}), \dots, (W_{4,L_4}, B_{4,L_4})) \in \mathcal{N}$. Observe that for all $x \in \mathbb{R}^{l_{4,L_1-1}}$ it holds that

$$\begin{aligned} W_{4,L_1}x + B_{4,L_1} &= \begin{pmatrix} W_{2,1}W_{1,L_1} \\ W_{3,1}W_{1,L_1} \end{pmatrix} x + \begin{pmatrix} W_{2,1}B_{1,L_1} + B_{2,1} \\ W_{3,1}B_{1,L_1} + B_{3,1} \end{pmatrix} \\ &= \begin{pmatrix} W_{2,1}W_{1,L_1}x + W_{2,1}B_{1,L_1} + B_{2,1} \\ W_{3,1}W_{1,L_1}x + W_{3,1}B_{1,L_1} + B_{3,1} \end{pmatrix} \\ &= \begin{pmatrix} W_{2,1}(W_{1,L_1}x + B_{1,L_1}) + B_{2,1} \\ W_{3,1}(W_{1,L_1}x + B_{1,L_1}) + B_{3,1} \end{pmatrix}. \end{aligned} \tag{118}$$

Moreover, note that for all $i \in \mathbb{N} \cap [0, L_2 - 2]$, $x \in \mathbb{R}^{l_{2,i}}$, $y \in \mathbb{R}^{\mathfrak{i}}$ it holds that

$$\begin{aligned} W_{4,L_1+i}\begin{pmatrix} x \\ y \end{pmatrix} + B_{4,L_1+i} &= \begin{pmatrix} W_{2,i+1} & 0 \\ 0 & W_{3,1}W_{3,2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} B_{2,i+1} \\ W_{3,1}B_{3,2} + B_{3,1} \end{pmatrix} \\ &= \begin{pmatrix} W_{2,i+1}x + B_{2,i+1} \\ W_{3,1}W_{3,2}y + W_{3,1}B_{3,2} + B_{3,1} \end{pmatrix} \\ &= \begin{pmatrix} W_{2,i+1}x + B_{2,i+1} \\ W_{3,1}(W_{3,2}y + B_{3,2}) + B_{3,1} \end{pmatrix}. \end{aligned} \tag{119}$$

Next observe that for all $x \in \mathbb{R}^{l_{2,L_2-1}}$, $y \in \mathbb{R}^{\mathfrak{i}}$ it holds that

$$\begin{aligned} W_{4,L_4}\begin{pmatrix} x \\ y \end{pmatrix} + B_{4,L_4} &= \begin{pmatrix} W_{2,L_2} & W_{3,2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + B_{2,L_2} + B_{3,2} \\ &= (W_{2,L_2}x + B_{2,L_2}) + (W_{3,2}y + B_{3,2}). \end{aligned} \tag{120}$$

Moreover, note that the hypothesis that $\forall\, y \in \mathbb{R}^d \colon (\mathcal{R}\,\mathbb{I})(y) = y$ ensures that for all $x \in \mathbb{R}^d$ it holds that

$$W_{3,2}\mathbf{A}_{\mathfrak{i}}(W_{3,1}x + B_{3,1}) + B_{3,2} = x. \tag{121}$$

Combining this, (111), (118), (119), and (120) proves that for all $x \in \mathbb{R}^d$ it holds that

$$(\mathcal{R}\psi)(x) = (\mathcal{R}\phi_1)(x) + (\mathcal{R}\phi_2)\big((\mathcal{R}\phi_1)(x)\big). \tag{122}$$

Next observe that

$$\begin{aligned} \mathcal{P}(\psi) &= \sum_{n=1}^{L_4} l_{4,n}(l_{4,n-1} + 1) = \sum_{n=1}^{L_1+L_2-1} l_{4,n}(l_{4,n-1} + 1) \\ &= \left[\sum_{n=1}^{L_1-1} l_{4,n}(l_{4,n-1} + 1)\right] + \left[\sum_{n=L_1+1}^{L_1+L_2-2} l_{4,n}(l_{4,n-1} + 1)\right] \\ &\quad + l_{4,L_1}(l_{4,L_1-1} + 1) + l_{4,L_1+L_2-1}(l_{4,L_1+L_2-2} + 1) \\ &= \left[\sum_{n=1}^{L_1-1} l_{1,n}(l_{1,n-1} + 1)\right] + \left[\sum_{n=1}^{L_2-2} l_{4,L_1+n}(l_{4,L_1+n-1} + 1)\right] \\ &\quad + (l_{2,1} + \mathfrak{i})(l_{1,L_1-1} + 1) + l_{2,L_2}(l_{2,L_2-1} + \mathfrak{i} + 1). \end{aligned} \tag{123}$$

The hypothesis that $l_{1,L_1-1} \leq l_{2,L_2-1} + \mathfrak{i}$ therefore assures that

$$\begin{aligned}\mathcal{P}(\psi) &= \left[\sum_{n=1}^{L_1-1} l_{1,n}(l_{1,n-1}+1)\right] + \left[\sum_{n=1}^{L_2-2} (l_{2,n+1}+\mathfrak{i})(l_{2,n}+\mathfrak{i}+1)\right] \\ &\quad + (l_{2,1}+\mathfrak{i})(l_{1,L_1-1}+1) + l_{2,L_2}(l_{2,L_2-1}+\mathfrak{i}+1) \\ &\leq \mathcal{P}(\phi_1) + \left[\sum_{n=1}^{L_2-2} \left[l_{2,n+1}(l_{2,n}+1) + l_{2,n+1}\mathfrak{i} + \mathfrak{i}\,(l_{2,n}+1) + \mathfrak{i}^2\right]\right] \\ &\quad + (l_{2,1}+\mathfrak{i})(l_{2,L_2-1}+\mathfrak{i}+1) + l_{2,L_2}(l_{2,L_2-1}+\mathfrak{i}+1) \\ &\leq \mathcal{P}(\phi_1) + \left[\sum_{n=2}^{L_2-1} l_{2,n}(l_{2,n-1}+1)\right] + \mathfrak{i}\left[\sum_{n=1}^{L_2-2}(l_{2,n+1}+l_{2,n})\right] \\ &\quad + (L_2-2)(\mathfrak{i}+\mathfrak{i}^2) + (l_{2,1}+\mathfrak{i})(l_{2,L_2-1}+\mathfrak{i}+1) \\ &\quad + l_{2,L_2}(l_{2,L_2-1}+1) + l_{2,L_2}\mathfrak{i}.\end{aligned} \tag{124}$$

Hence, we obtain that

$$\begin{aligned}\mathcal{P}(\psi) &\leq \mathcal{P}(\phi_1) + \left[\sum_{n=2}^{L_2} l_{2,n}(l_{2,n-1}+1)\right] + \mathfrak{i}\left[\sum_{n=1}^{L_2-2}(l_{2,n+1}+l_{2,n})\right] \\ &\quad + (L_2-2)(\mathfrak{i}+\mathfrak{i}^2) + \mathfrak{i}\,(l_{2,1}+l_{2,L_2-1}+l_{2,L_2}) \\ &\quad + l_{2,1}(l_{2,L_2-1}+1) + \mathfrak{i}\,(\mathfrak{i}+1) \\ &\leq \mathcal{P}(\phi_1) + \mathcal{P}(\phi_2) + 2\,\mathfrak{i}\left[\sum_{n=1}^{L_2} l_{2,n}\right] + (L_2-1)(\mathfrak{i}+\mathfrak{i}^2) + l_{2,1}(l_{2,L_2-1}+1).\end{aligned} \tag{125}$$

Next observe that

$$\mathcal{P}(\phi_2) = \sum_{n=1}^{L_2} l_{2,n}(l_{2,n-1}+1) \geq 2\left[\sum_{n=1}^{L_2} l_{2,n}\right] \geq 2\,L_2 \tag{126}$$

and

$$\mathcal{P}(\mathbb{I}) = \mathfrak{i}\,(d+1) + d\,(\mathfrak{i}+1) \geq 2\,\mathfrak{i} + 2. \tag{127}$$

This and (125) demonstrate that

$$\begin{aligned}\mathcal{P}(\psi) &\leq \mathcal{P}(\phi_1) + \mathcal{P}(\phi_2) + \mathfrak{i}\,\mathcal{P}(\phi_2) + (\mathfrak{i}+\mathfrak{i}^2)\mathcal{P}(\phi_2) + \mathcal{P}(\phi_2)(\mathcal{P}(\phi_2)+1) \\ &= \mathcal{P}(\phi_1) + \mathcal{P}(\phi_2)(2+2\,\mathfrak{i}+\mathfrak{i}^2) + (\mathcal{P}(\phi_2))^2 \\ &\leq \mathcal{P}(\phi_1) + \mathcal{P}(\phi_2)(\mathcal{P}(\mathbb{I}) + (\mathcal{P}(\mathbb{I}))^2) + (\mathcal{P}(\phi_2))^2 \\ &\leq \mathcal{P}(\phi_1) + 2\,\mathcal{P}(\phi_2)(\mathcal{P}(\mathbb{I}))^2 + (\mathcal{P}(\phi_2))^2 \\ &\leq \mathcal{P}(\phi_1) + (\mathcal{P}(\phi_2) + \mathcal{P}(\mathbb{I}))^3.\end{aligned} \tag{128}$$

Combining this with (122) establishes (109). The proof of Proposition 5.3 is thus completed. □

5.3 Representations of the d-dimensional identities

Lemma 5.4 (Artificial neural networks with rectifier functions)**.** *Let $d \in \mathbb{N}$, let $\mathbf{A}_n : \mathbb{R}^n \to \mathbb{R}^n$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ that $\mathbf{A}_n(x) = (\max\{x_1, 0\}, \dots, \max\{x_n, 0\})$, let*

$$\mathcal{N} = \cup_{L \in \{2,3,4,\dots\}} \cup_{(l_0,l_1,\dots,l_L) \in \mathbb{N}^{L+1}} \left(\times_{n=1}^{L} (\mathbb{R}^{l_n \times l_{n-1}} \times \mathbb{R}^{l_n}) \right), \tag{129}$$

and let $\mathcal{P} : \mathcal{N} \to \mathbb{N}$, $\mathcal{L} : \mathcal{N} \to \cup_{L \in \{2,3,4,\dots\}} \mathbb{N}^{L+1}$, and $\mathcal{R} : \mathcal{N} \to \cup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ be the functions which satisfy for all $L \in \{2, 3, 4, \dots\}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_L, B_L)) \in (\times_{n=1}^{L} (\mathbb{R}^{l_n \times l_{n-1}} \times \mathbb{R}^{l_n}))$, $x_0 \in \mathbb{R}^{l_0}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, n \in \mathbb{N} \cap [1, L) : x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ that $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1} + 1)$, $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, $\mathcal{L}(\Phi) = (l_0, l_1, \dots, l_L)$, and

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L. \tag{130}$$

Then there exists $\psi \in \mathcal{N}$ such that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}(\psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\mathcal{L}(\psi) = (d, 2d, d) \in \mathbb{N}^3$, and

$$(\mathcal{R}\psi)(x) = x. \tag{131}$$

Proof of Lemma 5.4. Throughout this proof let $w_1 \in \mathbb{R}^{2 \times 1}$, $w_2 \in \mathbb{R}^{1 \times 2}$, $(W_1, B_1) \in (\mathbb{R}^{(2d) \times d} \times \mathbb{R}^{2d})$, and $(W_2, B_2) \in (\mathbb{R}^{d \times (2d)} \times \mathbb{R}^d)$ satisfy that

$$w_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \in \mathbb{R}^{2 \times 1}, \quad W_1 = \begin{pmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_1 & 0 & \cdots & 0 \\ 0 & 0 & w_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_1 \end{pmatrix} \in \mathbb{R}^{(2d) \times d}, \quad B_1 = 0 \in \mathbb{R}^{2d}, \tag{132}$$

$$w_2 = \begin{pmatrix} 1 & -1 \end{pmatrix} \in \mathbb{R}^{1 \times 2}, \quad W_2 = \begin{pmatrix} w_2 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_2 \end{pmatrix} \in \mathbb{R}^{d \times (2d)}, \quad B_2 = 0 \in \mathbb{R}^d \tag{133}$$

and let $\psi = ((W_1, B_1), (W_2, B_2)) \in \mathcal{N}$. Observe that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$W_1 x + B_1 = W_1 x = \begin{pmatrix} w_1 x_1 \\ w_1 x_2 \\ \vdots \\ w_1 x_d \end{pmatrix} \in \mathbb{R}^{2d}. \tag{134}$$

This ensures that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\mathbf{A}_{(2d)}(W_1 x + B_1) = \begin{pmatrix} \max\{x_1, 0\} \\ \max\{-x_1, 0\} \\ \max\{x_2, 0\} \\ \max\{-x_2, 0\} \\ \vdots \\ \max\{x_d, 0\} \\ \max\{-x_d, 0\}. \end{pmatrix} \in \mathbb{R}^{2d}. \tag{135}$$

Hence, we obtain that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned} W_2 \mathbf{A}_{(2d)}(W_1 x + B_1) + B_2 &= W_2 \begin{pmatrix} \max\{x_1, 0\} \\ \max\{-x_1, 0\} \\ \max\{x_2, 0\} \\ \max\{-x_2, 0\} \\ \vdots \\ \max\{x_d, 0\} \\ \max\{-x_d, 0\}. \end{pmatrix} \\ &= \begin{pmatrix} \max\{x_1, 0\} - \max\{-x_1, 0\} \\ \max\{x_2, 0\} - \max\{-x_2, 0\} \\ \vdots \\ \max\{x_d, 0\} - \max\{-x_d, 0\} \end{pmatrix} = x \in \mathbb{R}^d. \end{aligned} \tag{136}$$

This demonstrates that for all $x \in \mathbb{R}^d$ it holds that

$$(\mathcal{R}\psi)(x) = x. \tag{137}$$

Combining this with the fact that $\mathcal{L}(\psi) = (d, 2d, d) \in \mathbb{N}^3$ establishes (131). The proof of Lemma 5.4 is thus completed. □

6 DNN approximations for partial differential equations (PDEs)

In this section we establish in our main result in Theorem 6.3 in Subsection 6.2 below that rectified DNNs have the capacity to approximate solutions of second-order Kolmogorov PDEs with nonlinear drift and constant diffusion coefficients without suffering from the curse of dimensionality. Our proof of Theorem 6.3 is based on an application of Corollary 6.2 in Subsection 6.1 below. Corollary 6.2, in turn, follows immediately from Proposition 6.1 in Subsection 6.1 below. Proposition 6.1 is, roughly speaking, a generalized version of Theorem 6.3 which covers a more general type of activation function instead of only the rectifier function as the employed activation function. Proposition 6.1 shows for every $p \in [2, \infty)$ that the $L^p(\nu_d; \mathbb{R})$-distance between the solution of the PDE at the time of maturity and the DNN is smaller or equal than the prescribed approximation accuracy $\varepsilon > 0$, where

$\nu_d\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$, $d \in \mathbb{N}$, is a suitable sequence of probability measures. Corollary 6.2 slightly generalizes this result, in particular, by assuming that $p \in (0,\infty)$ is an arbitrary strictly positive real number instead of assuming that $p \in [2,\infty)$ is greater or equal than 2 (cf. Proposition 6.1). Finally, in Corollary 6.4 in Subsection 6.3 below we specialize Theorem 6.3 in Subsection 6.2 to the case where for every $d \in \mathbb{N}$ we have that the probability measure ν_d is nothing else but the uniform distribution on the d-dimensional unit cube $[0,1]^d$. Theorem 1.1 in the introduction in Section 1 above follows directly from Corollary 6.4 in Subsection 6.3.

6.1 DNN approximations with general activation functions

Proposition 6.1. *Let $T, \kappa \in (0,\infty)$, $\eta \in [1,\infty)$, $p \in [2,\infty)$, let $A_d = (a_{d,i,j})_{(i,j)\in\{1,\dots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be symmetric positive semidefinite matrices, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm and let $\nu_d\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure, let $f_{0,d}\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $f_{1,d}\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be functions, let $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $\mathbf{a}\colon \mathbb{R} \to \mathbb{R}$ be continuous functions which satisfy for all $d \in \mathbb{N}$, $x = (x_1,\dots,x_d) \in \mathbb{R}^d$ that $\mathbf{A}_d(x) = (\mathbf{a}(x_1),\dots,\mathbf{a}(x_d))$, let*

$$\mathcal{N} = \cup_{L\in\{2,3,4,\dots\}} \cup_{(l_0,l_1,\dots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{n=1}^{L} (\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n})\right), \tag{138}$$

let $\mathcal{P}\colon \mathcal{N} \to \mathbb{N}$, $\mathcal{L}\colon \mathcal{N} \to \cup_{L\in\{2,3,4,\dots\}}\mathbb{N}^{L+1}$, and $\mathcal{R}\colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k,\mathbb{R}^l)$ be the functions which satisfy for all $L \in \{2,3,4,\dots\}$, $l_0,l_1,\dots,l_L \in \mathbb{N}$, $\Phi = ((W_1,B_1),\dots,(W_L,B_L)) \in (\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n}))$, $x_0 \in \mathbb{R}^{l_0}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, n \in \mathbb{N} \cap [1,L)\colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ that $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1}+1)$, $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0},\mathbb{R}^{l_L})$, $\mathcal{L}(\Phi) = (l_0,l_1,\dots,l_L)$, and

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{139}$$

let $(\phi^{m,d}_{\varepsilon})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$, $(\phi^{2,d})_{d\in\mathbb{N}} \subseteq \mathcal{N}$, and assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x,y \in \mathbb{R}^d$ that $\mathcal{R}(\phi^{0,d}_{\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{R}(\phi^{1,d}_{\varepsilon}), \mathcal{R}(\phi^{2,d}) \in C(\mathbb{R}^d,\mathbb{R}^d)$, $|f_{0,d}(x)| + \sum_{i,j=1}^{d} |a_{d,i,j}| \le \kappa d^{\kappa}(1+\|x\|^{\kappa}_{\mathbb{R}^d})$, $\|f_{1,d}(x) - f_{1,d}(y)\|_{\mathbb{R}^d} \le \kappa\|x-y\|_{\mathbb{R}^d}$, $\|(\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \le \kappa(d^{\kappa} + \|x\|_{\mathbb{R}^d})$, $\mathcal{P}(\phi^{2,d}) + \sum_{m=0}^{1} \mathcal{P}(\phi^{m,d}_{\varepsilon}) \le \kappa d^{\kappa}\varepsilon^{-\kappa}$, $|(\mathcal{R}\phi^{0,d}_{\varepsilon})(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(y)| \le \kappa d^{\kappa}(1+\|x\|^{\kappa}_{\mathbb{R}^d}+\|y\|^{\kappa}_{\mathbb{R}^d})\|x-y\|_{\mathbb{R}^d}$, $\mathcal{L}(\phi^{2,d}) \in \mathbb{N}^3$, $(\mathcal{R}\phi^{2,d})(x) = x$, $\int_{\mathbb{R}^d} \|z\|^{p(2\kappa+1)}_{\mathbb{R}^d}\, \nu_d(dz) \le \eta d^{\eta}$, and

$$|f_{0,d}(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(x)| + \|f_{1,d}(x) - (\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \le \varepsilon\kappa d^{\kappa}(1 + \|x\|^{\kappa}_{\mathbb{R}^d}). \tag{140}$$

Then

(i) there exist unique at most polynomially growing functions $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, such that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $u_d(0,x) = f_{0,d}(x)$ and such that for all $d \in \mathbb{N}$ it holds that u_d is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\tfrac{\partial}{\partial x}u_d)(t,x)\, f_{1,d}(x) + \sum_{i,j=1}^{d} a_{d,i,j}\, (\tfrac{\partial^2}{\partial x_i \partial x_j}u_d)(t,x) \tag{141}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$ and *(ii) there exist* $(\psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$, $c \in \mathbb{R}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon}) \leq c\, d^c \varepsilon^{-c}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, *and*

$$\left[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}\psi_{d,\varepsilon})(x)|^p \, \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{142}$$

Proof of Proposition 6.1. Throughout this proof let $\iota \in \mathbb{R}$ be the real number given by $\iota = \max\{\kappa, 1\}$, let $\mathcal{A}_d \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$ that $\mathcal{A}_d = \sqrt{2A_d}$, let $\Phi^{0,d}_\delta \colon \mathbb{R}^d \to \mathbb{R}$, $\delta \in (0,1]$, $d \in \mathbb{N}$, and $\Phi^{1,d}_\delta \colon \mathbb{R}^d \to \mathbb{R}^d$, $\delta \in (0,1]$, $d \in \mathbb{N}$, be the functions which satisfy for all $m \in \{0,1\}$, $d \in \mathbb{N}$, $\delta \in (0,1]$, $x \in \mathbb{R}^d$ that

$$\Phi^{m,d}_\delta(x) = (\mathcal{R}\phi^{m,d}_\delta)(x), \tag{143}$$

let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $W^{d,m} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $d, m \in \mathbb{N}$, be independent standard Brownian motions, let $\varpi_{d,q} \in \mathbb{R}$, $d \in \mathbb{N}$, $q \in (0,\infty)$, satisfy for all $q \in (0,\infty)$, $d \in \mathbb{N}$ that

$$\varpi_{d,q} = \big(\mathbb{E}\big[\|\mathcal{A}_d W^{d,1}_T\|^q_{\mathbb{R}^d}\big]\big)^{1/q}, \tag{144}$$

let $X^{d,x} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, be stochastic processes with continuous sample paths which satisfy for all $x \in \mathbb{R}^d$, $d \in \mathbb{N}$, $t \in [0,T]$ that

$$X^{d,x}_t = x + \int_0^t f_{1,d}(X^{d,x}_s)\, ds + \mathcal{A}_d W^{d,1}_t \tag{145}$$

(cf. Theorem 3.1), let $\chi_\delta \colon [0,T] \to [0,T]$, $\delta \in (0,1]$, be the functions which satisfy for all $\delta \in (0,1]$, $t \in [0,T]$ that

$$\chi_\delta(t) = \max\big(\{0, \delta^2, 2\delta^2, 3\delta^2, \dots\} \cap [0,t]\big), \tag{146}$$

let $Y^{\delta,d,m,x} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $\delta \in (0,1]$, $d, m \in \mathbb{N}$, $x \in \mathbb{R}^d$, be stochastic processes with continuous sample paths which satisfy for all $x \in \mathbb{R}^d$, $d, m \in \mathbb{N}$, $\delta \in (0,1]$, $t \in [0,T]$ that

$$Y^{\delta,d,m,x}_t = x + \int_0^t \Phi^{1,d}_\delta\big(Y^{\delta,d,m,x}_{\chi_\delta(s)}\big)\, ds + \mathcal{A}_d W^{d,m}_t, \tag{147}$$

let $\mathfrak{M}_{d,\varepsilon} \in \mathbb{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, be the natural numbers which satisfy for all $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ that

$$\begin{aligned}&\mathfrak{M}_{d,\varepsilon}\\ &= \min\left(\mathbb{N} \cap \left[\left(\tfrac{2^{(\kappa+4)} p\kappa d^\kappa e^{\kappa^2 T}}{\varepsilon}\right)^2 \left[1 + \big(\kappa d^\kappa T + \sqrt{2(p\iota - 1)\kappa d^\kappa T}\big)^{p\kappa} + \eta d^\eta\right]^{2/p}, \infty\right)\right),\end{aligned} \tag{148}$$

and let $\mathcal{D}_{d,\varepsilon} \in (0,1]$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, be the real numbers which satisfy for all $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ that

$$\begin{aligned}\mathcal{D}_{d,\varepsilon} &= \varepsilon\big[\max\{2\kappa d^\kappa, 1\} + T^{-1/2}\big]^{-1} e^{-(3+3\kappa+[\kappa^2+2\kappa\iota+2]T)} \big|\max\{1, 2\kappa(\kappa+1)d^\kappa\}\big|^{-1} 2^{-(2\iota+1)}\\ &\cdot \left[\big|2 + \max\{1, \kappa d^\kappa, \|f_{1,d}(0)\|_{\mathbb{R}^d}\} \max\{1,T\} + \sqrt{2(2\iota-1)\kappa d^\kappa T}\big|^{p\iota+p\kappa} + \eta d^\eta\right]^{-\frac{1}{p}}.\end{aligned} \tag{149}$$

Observe that the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{R}(\phi_\varepsilon^{0,d}) \in C(\mathbb{R}^d,\mathbb{R})$ and (140) ensure that $f_{0,d} \in C(\mathbb{R}^d,\mathbb{R})$. This, the fact that for all $d \in \mathbb{N}$ it holds that the function $f_{1,d}\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous, and Theorem 3.1 establish item (i). It thus remains to prove item (ii). For this note that the fact that $\forall\, y,z \in \mathbb{R}, \alpha \in [1,\infty)\colon |y+z|^\alpha \le 2^{\alpha-1}(|y|^\alpha + |z|^\alpha)$ and Theorem 3.1 ensure that for all $M,d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \mathbb{E}\left[\left| u_d(T,x) - \frac{1}{M}\left[\sum_{m=1}^M \Phi_\delta^{0,d}(Y_T^{\delta,d,m,x})\right]\right|^p\right] \nu_d(dx) \\
&\le 2^{p-1} \int_{\mathbb{R}^d} \mathbb{E}\Big[\big|u_d(T,x) - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p\Big] \nu_d(dx) \\
&+ 2^{p-1} \int_{\mathbb{R}^d} \mathbb{E}\left[\left|\mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big] - \frac{1}{M}\left[\sum_{m=1}^M \Phi_\delta^{0,d}(Y_T^{\delta,d,m,x})\right]\right|^p\right] \nu_d(dx) \\
&= 2^{p-1} \int_{\mathbb{R}^d} \big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \nu_d(dx) \\
&+ 2^{p-1} \int_{\mathbb{R}^d} \mathbb{E}\left[\left|\mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big] - \frac{1}{M}\left[\sum_{m=1}^M \Phi_\delta^{0,d}(Y_T^{\delta,d,m,x})\right]\right|^p\right] \nu_d(dx).
\end{aligned} \tag{150}$$

The fact that $2\sqrt{p-1} \le p$ and, e.g., Grohs et al. [26, Corollary 2.5] hence prove that for all $M,d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \mathbb{E}\left[\left| u_d(T,x) - \frac{1}{M}\left[\sum_{m=1}^M \Phi_\delta^{0,d}(Y_T^{\delta,d,m,x})\right]\right|^p\right] \nu_d(dx) \\
&\le 2^{p-1} \int_{\mathbb{R}^d} \big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \nu_d(dx) \\
&+ \frac{2^{p-1}(2\sqrt{p-1})^p}{M^{p/2}} \int_{\mathbb{R}^d} \mathbb{E}\Big[\big|\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x}) - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p\Big] \nu_d(dx) \\
&\le 2^{p-1} \int_{\mathbb{R}^d} \big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \nu_d(dx) \\
&+ \frac{2^{p-1}p^p}{M^{p/2}} \int_{\mathbb{R}^d} \mathbb{E}\Big[\big|\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x}) - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p\Big] \nu_d(dx).
\end{aligned} \tag{151}$$

The fact that $\forall\, y,z \in \mathbb{R}, \alpha \in [1,\infty)\colon |y+z|^\alpha \le 2^{\alpha-1}(|y|^\alpha + |z|^\alpha)$ and Jensen's inequality therefore assure that for all $M,d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \mathbb{E}\left[\left| u_d(T,x) - \frac{1}{M}\left[\sum_{m=1}^M \Phi_\delta^{0,d}(Y_T^{\delta,d,m,x})\right]\right|^p\right] \nu_d(dx) \\
&\le 2^{p-1} \int_{\mathbb{R}^d} \big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \nu_d(dx) \\
&+ \frac{2^{2(p-1)}p^p}{M^{p/2}} \int_{\mathbb{R}^d} \mathbb{E}\Big[\big|\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big|^p\Big] + \mathbb{E}\Big[\big|\mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p\Big] \nu_d(dx) \\
&\le 2^{p-1} \int_{\mathbb{R}^d} \big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \nu_d(dx) \\
&+ \frac{2^{2p-1}p^p}{M^{p/2}} \int_{\mathbb{R}^d} \mathbb{E}\Big[\big|\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big|^p\Big] \nu_d(dx).
\end{aligned} \tag{152}$$

Next observe that for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&2\int_0^1 \big[r\|x\|_{\mathbb{R}^d} + (1-r)\|y\|_{\mathbb{R}^d}\big]^\kappa \, dr \geq \int_0^1 \big[r^\kappa \|x\|_{\mathbb{R}^d}^\kappa + (1-r)^\kappa \|y\|_{\mathbb{R}^d}^\kappa\big] \, dr \\
&= \big[\|x\|_{\mathbb{R}^d}^\kappa + \|y\|_{\mathbb{R}^d}^\kappa\big] \int_0^1 r^\kappa \, dr = \frac{\big[\|x\|_{\mathbb{R}^d}^\kappa + \|y\|_{\mathbb{R}^d}^\kappa\big]}{\kappa+1}.
\end{aligned} \tag{153}$$

This and the hypothesis that $\forall\, d \in \mathbb{N}, \delta \in (0,1], x, y \in \mathbb{R}^d \colon |(\mathcal{R}\phi_\delta^{0,d})(x) - (\mathcal{R}\phi_\delta^{0,d})(y)| \leq \kappa d^\kappa (1 + \|x\|_{\mathbb{R}^d}^\kappa + \|y\|_{\mathbb{R}^d}^\kappa)\|x-y\|_{\mathbb{R}^d}$ prove that for all $d \in \mathbb{N}$, $\delta \in (0,1]$, $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&|\Phi_\delta^{0,d}(x) - \Phi_\delta^{0,d}(y)| = |(\mathcal{R}\phi_\delta^{0,d})(x) - (\mathcal{R}\phi_\delta^{0,d})(y)| \\
&\leq \kappa d^\kappa (1 + \|x\|_{\mathbb{R}^d}^\kappa + \|y\|_{\mathbb{R}^d}^\kappa)\|x-y\|_{\mathbb{R}^d} \\
&\leq \kappa d^\kappa \left(1 + 2(\kappa+1)\int_0^1 \big[r\|x\|_{\mathbb{R}^d} + (1-r)\|y\|_{\mathbb{R}^d}\big]^\kappa \, dr\right) \|x-y\|_{\mathbb{R}^d} \\
&\leq 2\kappa(\kappa+1) d^\kappa \left(1 + \int_0^1 \big[r\|x\|_{\mathbb{R}^d} + (1-r)\|y\|_{\mathbb{R}^d}\big]^\kappa \, dr\right) \|x-y\|_{\mathbb{R}^d}.
\end{aligned} \tag{154}$$

Proposition 4.6 (with $d = d$, $m = d$, $\xi = x$, $T = T$, $c = \kappa$, $C = \kappa d^\kappa$, $\varepsilon_0 = \delta\kappa d^\kappa$, $\varepsilon_1 = \delta\kappa d^\kappa$, $\varepsilon_2 = 0$, $\varsigma_0 = \kappa$, $\varsigma_1 = \kappa$, $\varsigma_2 = 0$, $L_0 = 2\kappa(\kappa+1)d^\kappa$, $L_1 = \kappa$, $\ell = \kappa$, $h = \min\{\delta^2, T\}$, $B = \mathcal{A}_d$, $p = 2$, $q = 2$, $\|\cdot\| = \|\cdot\|_{\mathbb{R}^d}$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W = W^{d,1}$, $\phi_0 = \Phi_\delta^{0,d}$, $f_1 = f_{1,d}$, $\phi_2 = \mathrm{id}_{\mathbb{R}^d}$, $\chi = \chi_{\min\{\delta, \sqrt{T}\}}$, $f_0 = f_{0,d}$, $\phi_1 = \Phi_\delta^{1,d}$, $(\varpi_r)_{r \in (0,\infty)} = (\varpi_{d,r})_{r \in (0,\infty)}$, $X = X^{d,x}$, $Y = Y^{\delta,d,1,x}$ for $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ in the notation of Proposition 4.6) hence ensures that for all $d \in \mathbb{N}$, $\delta \in (0,1]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \leq \big[2\delta\kappa d^\kappa + (\min\{\delta^2, T\}/T)^{1/2}\big]^p \\
&\cdot e^{p(\kappa+3+2\kappa+[\kappa\max\{\kappa,\kappa\}+\kappa\max\{\kappa,1\}+\kappa\max\{\kappa,1\}+2]T)} \big[\|x\|_{\mathbb{R}^d} + \max\{1,0\}(1+1) \\
&+ \max\{1, \kappa d^\kappa, \|f_{1,d}(0)\|_{\mathbb{R}^d}\} \max\{1,T\} + \varpi_{d,\max\{\kappa,2\kappa,2,2\kappa\}}\big]^{p\max\{1,\kappa,\kappa\}+p\kappa} \\
&\cdot \big|\max\{1, 2\kappa(\kappa+1)d^\kappa\}\big|^p \\
&\leq \big[2\delta\kappa d^\kappa + (\delta^2/T)^{1/2}\big]^p e^{p(3+3\kappa+[\kappa^2+2\kappa\iota+2]T)} \big|\max\{1, 2\kappa(\kappa+1)d^\kappa\}\big|^p \\
&\cdot \big[\|x\|_{\mathbb{R}^d} + 2 + \max\{1, \kappa d^\kappa, \|f_{1,d}(0)\|_{\mathbb{R}^d}\} \max\{1,T\} + \varpi_{d,\max\{2\kappa,2\}}\big]^{p\iota+p\kappa}.
\end{aligned} \tag{155}$$

Moreover, note that Lemma 4.2 assures that for all $q \in [2,\infty)$, $d \in \mathbb{N}$ it holds that

$$\varpi_{d,q} \leq \sqrt{(q-1)\,\mathrm{Trace}(\mathcal{A}_d^*\mathcal{A}_d)T} = \sqrt{2(q-1)\,\mathrm{Trace}(A_d)T} \leq \sqrt{2(q-1)\kappa d^\kappa T}. \tag{156}$$

This, (155), and the fact that $\forall\, y, z \in \mathbb{R}, \alpha \in [1,\infty) \colon |y+z|^\alpha \leq 2^{\alpha-1}(|y|^\alpha + |z|^\alpha)$

demonstrate that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \, \nu_d(dx) \\
&\leq \delta^p \left[2\kappa d^\kappa + T^{-1/2}\right]^p e^{p(3+3\kappa+[\kappa^2+2\kappa\iota+2]T)} \big|\max\{1, 2\kappa(\kappa+1)d^\kappa\}\big|^p \\
&\cdot \int_{\mathbb{R}^d} \left[\|x\|_{\mathbb{R}^d} + 2 + \max\{1, \kappa d^\kappa, \|f_{1,d}(0)\|_{\mathbb{R}^d}\} \max\{1,T\} + \sqrt{2(2\iota-1)\kappa d^\kappa T}\right]^{p\iota+p\kappa} \nu_d(dx) \\
&\leq \delta^p \left[2\kappa d^\kappa + T^{-1/2}\right]^p e^{p(3+3\kappa+[\kappa^2+2\kappa\iota+2]T)} \big|\max\{1, 2\kappa(\kappa+1)d^\kappa\}\big|^p \\
&\cdot 2^{p\iota+p\kappa-1}\Bigg(\left[2 + \max\{1, \kappa d^\kappa, \|f_{1,d}(0)\|_{\mathbb{R}^d}\} \max\{1,T\} + \sqrt{2(2\iota-1)\kappa d^\kappa T}\right]^{p\iota+p\kappa} \\
&+ \int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^{p\iota+p\kappa} \, \nu_d(dx)\Bigg).
\end{aligned} \tag{157}$$

Next note that the fact that $\iota \leq \kappa + 1$ and Hölder's inequality prove that for all $d \in \mathbb{N}$ it holds that

$$\begin{aligned}
\int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^{p\iota+p\kappa} \, \nu_d(dx) &\leq \left[\int_{\mathbb{R}^d} \|x\|_{\mathbb{R}^d}^{p(2\kappa+1)} \, \nu_d(dx)\right]^{(\iota+\kappa)/(2\kappa+1)} \\
&\leq [\eta d^\eta]^{(\iota+\kappa)/(2\kappa+1)} \leq \eta d^\eta.
\end{aligned} \tag{158}$$

Combining this and (157) ensures that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
&2^{p-1} \int_{\mathbb{R}^d} \big|\mathbb{E}\big[f_{0,d}(X_T^{d,x})\big] - \mathbb{E}\big[\Phi_\delta^{0,d}(Y_T^{\delta,d,1,x})\big]\big|^p \, \nu_d(dx) \\
&\leq \delta^p \left[2\kappa d^\kappa + T^{-1/2}\right]^p e^{p(3+3\kappa+[\kappa^2+2\kappa\iota+2]T)} \big|\max\{1, 2\kappa(\kappa+1)d^\kappa\}\big|^p 2^{p(2\iota+1)-2} \\
&\cdot \left(\left[2 + \max\{1, \kappa d^\kappa, \|f_{1,d}(0)\|_{\mathbb{R}^d}\} \max\{1,T\} + \sqrt{2(2\iota-1)\kappa d^\kappa T}\right]^{p\iota+p\kappa} + \eta d^\eta\right).
\end{aligned} \tag{159}$$

Next observe that for all $d \in \mathbb{N}$, $\delta \in (0,1]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
|\Phi_\delta^{0,d}(x)| &\leq |\Phi_\delta^{0,d}(x) - f_{0,d}(x)| + |f_{0,d}(x)| \\
&\leq \delta\kappa d^\kappa(1 + \|x\|_{\mathbb{R}^d}^\kappa) + \kappa d^\kappa(1 + \|x\|_{\mathbb{R}^d}^\kappa) \leq 2\kappa d^\kappa(1 + \|x\|_{\mathbb{R}^d}^\kappa).
\end{aligned} \tag{160}$$

Moreover, note that Lemma 4.1 shows that for all $q \in [1,\infty)$, $\delta \in (0,1]$, $d, m \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\big(\mathbb{E}\big[\|Y_T^{\delta,d,m,x}\|_{\mathbb{R}^d}^q\big]\big)^{1/q} &\leq \Big(\|x\|_{\mathbb{R}^d} + \kappa d^\kappa T + \big(\mathbb{E}\big[\|\mathcal{A}_d W_T^{d,m}\|_{\mathbb{R}^d}^q\big]\big)^{1/q}\Big) e^{\kappa T} \\
&= \big(\|x\|_{\mathbb{R}^d} + \kappa d^\kappa T + \varpi_{d,q}\big) e^{\kappa T}.
\end{aligned} \tag{161}$$

This and (156) demonstrate that for all $q \in [2,\infty)$, $\delta \in (0,1]$, $d, m \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that

$$\big(\mathbb{E}\big[\|Y_T^{\delta,d,m,x}\|_{\mathbb{R}^d}^q\big]\big)^{1/q} \leq \big(\|x\|_{\mathbb{R}^d} + \kappa d^\kappa T + \sqrt{2(q-1)\kappa d^\kappa T}\big) e^{\kappa T}. \tag{162}$$

Combining this with (160), the fact that $\forall\, y,z\in\mathbb{R},\alpha\in[1,\infty)\colon |y+z|^\alpha\le 2^{\alpha-1}(|y|^\alpha+|z|^\alpha)$, and Hölder's inequality ensures that for all $\delta\in(0,1]$, $d\in\mathbb{N}$ it holds that

$$\begin{aligned}
&\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|\Phi^{0,d}_\delta(Y^{\delta,d,1,x}_T)\big|^p\Big]\,\nu_d(dx)\le\big(2\kappa d^\kappa\big)^p\int_{\mathbb{R}^d}\mathbb{E}\Big[\Big(1+\|Y^{\delta,d,1,x}_T\|^\kappa_{\mathbb{R}^d}\Big)^p\Big]\,\nu_d(dx)\\
&\le\big(2\kappa d^\kappa\big)^p2^{p-1}\int_{\mathbb{R}^d}\mathbb{E}\Big[1+\|Y^{\delta,d,1,x}_T\|^{p\kappa}_{\mathbb{R}^d}\Big]\,\nu_d(dx)\\
&\le\big(4\kappa d^\kappa\big)^p\left(1+\int_{\mathbb{R}^d}\mathbb{E}\Big[\|Y^{\delta,d,1,x}_T\|^{p\kappa}_{\mathbb{R}^d}\Big]\,\nu_d(dx)\right)\\
&\le\big(4\kappa d^\kappa\big)^p\left(1+\int_{\mathbb{R}^d}\mathbb{E}\Big[\|Y^{\delta,d,1,x}_T\|^{p\iota}_{\mathbb{R}^d}\Big]^{\kappa/\iota}\,\nu_d(dx)\right)\\
&\le\big(4\kappa d^\kappa\big)^p\left(1+\int_{\mathbb{R}^d}\Big[\Big(\|x\|_{\mathbb{R}^d}+\kappa d^\kappa T+\sqrt{2(p\iota-1)\kappa d^\kappa T}\Big)\,e^{\kappa T}\Big]^{p\kappa}\,\nu_d(dx)\right).
\end{aligned}\tag{163}$$

The fact that $\forall\, y,z\in\mathbb{R},\alpha\in[0,\infty)\colon |y+z|^\alpha\le 2^\alpha(|y|^\alpha+|z|^\alpha)$ hence proves that for all $\delta\in(0,1]$, $d\in\mathbb{N}$ it holds that

$$\begin{aligned}
&\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|\Phi^{0,d}_\delta(Y^{\delta,d,1,x}_T)\big|^p\Big]\,\nu_d(dx)\\
&\le\big(4\kappa d^\kappa\big)^p\left(1+2^{p\kappa}\,e^{p\kappa^2T}\left[\int_{\mathbb{R}^d}\|x\|^{p\kappa}_{\mathbb{R}^d}\,\nu_d(dx)+\Big(\kappa d^\kappa T+\sqrt{2(p\iota-1)\kappa d^\kappa T}\Big)^{p\kappa}\right]\right)\\
&\le\big(4\kappa d^\kappa\big)^p2^{p\kappa}\,e^{p\kappa^2T}\left[1+\Big(\kappa d^\kappa T+\sqrt{2(p\iota-1)\kappa d^\kappa T}\Big)^{p\kappa}+\int_{\mathbb{R}^d}\|x\|^{p\kappa}_{\mathbb{R}^d}\,\nu_d(dx)\right].
\end{aligned}\tag{164}$$

Next note that Hölder's inequality shows that for all $d\in\mathbb{N}$ it holds that

$$\begin{aligned}
\int_{\mathbb{R}^d}\|x\|^{p\kappa}_{\mathbb{R}^d}\,\nu_d(dx)&\le\left[\int_{\mathbb{R}^d}\|x\|^{p(2\kappa+1)}_{\mathbb{R}^d}\,\nu_d(dx)\right]^{\kappa/(2\kappa+1)}\\
&\le[\eta d^\eta]^{\kappa/(2\kappa+1)}\le\eta d^\eta.
\end{aligned}\tag{165}$$

Combining this and (164) ensures that for all $\delta\in(0,1]$, $d\in\mathbb{N}$ it holds that

$$\begin{aligned}
&2^{2p-1}p^p\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|\Phi^{0,d}_\delta(Y^{\delta,d,1,x}_T)\big|^p\Big]\,\nu_d(dx)\\
&\le\frac{\big(2^{\kappa+4}p\kappa d^\kappa e^{\kappa^2T}\big)^p}{2}\left[1+\Big(\kappa d^\kappa T+\sqrt{2(p\iota-1)\kappa d^\kappa T}\Big)^{p\kappa}+\eta d^\eta\right].
\end{aligned}\tag{166}$$

This, (152), and (159) prove that for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ it holds that

$$\int_{\mathbb{R}^d}\mathbb{E}\left[\left|u_d(T,x)-\frac{1}{\mathfrak{M}_{d,\varepsilon}}\left[\sum_{m=1}^{\mathfrak{M}_{d,\varepsilon}}\Phi^{0,d}_{\mathcal{D}_{d,\varepsilon}}\Big(Y^{\mathcal{D}_{d,\varepsilon},d,m,x}_T\Big)\right]\right|^p\right]\nu_d(dx)\le\tfrac{\varepsilon^p}{4}+\tfrac{\varepsilon^p}{2}<\varepsilon^p.\tag{167}$$

Corollary 2.4 therefore assures that for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ there exists $\mathfrak{w}_{d,\varepsilon}\in\Omega$ such that

$$\left[\int_{\mathbb{R}^d}\left|u_d(T,x)-\frac{1}{\mathfrak{M}_{d,\varepsilon}}\left[\sum_{m=1}^{\mathfrak{M}_{d,\varepsilon}}\big(\mathcal{R}\phi^{0,d}_{\mathcal{D}_{d,\varepsilon}}\big)\Big(Y^{\mathcal{D}_{d,\varepsilon},d,m,x}_T(\mathfrak{w}_{d,\varepsilon})\Big)\right]\right|^p\nu_d(dx)\right]^{1/p}<\varepsilon.\tag{168}$$

Moreover, note that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
\mathfrak{M}_{d,\varepsilon} &\leq \left(\tfrac{2^{(\kappa+4)}p\kappa d^\kappa e^{\kappa^2 T}}{\varepsilon}\right)^2 \left[1 + \left(\kappa d^\kappa T + \sqrt{2(p\iota-1)\kappa d^\kappa T}\right)^{p\kappa} + \eta d^\eta\right]^{2/p} + 1 \\
&= 2^{2(\kappa+4)}p^2\kappa^2 d^{2\kappa} e^{2\kappa^2 T} \left[1 + \left(\kappa d^\kappa T + \sqrt{2(p\iota-1)\kappa d^\kappa T}\right)^{p\kappa} + \eta d^\eta\right]^{2/p} \varepsilon^{-2} + 1 \\
&\leq 2^{2(\kappa+4)}p^2\kappa^2 d^{2\kappa} e^{2\kappa^2 T} \left[1 + \left(\kappa d^\kappa T + \sqrt{2(p\iota-1)\kappa d^\kappa T}\right)^{p\kappa} + \eta d^\eta\right] \varepsilon^{-2} + 1 \\
&\leq 2^{2(\kappa+4)}p^2\kappa^2 d^{2\kappa} e^{2\kappa^2 T} \left[1 + \left(\iota d^\iota T + p\iota d^\iota \sqrt{T}\right)^{p\kappa} + \eta d^\eta\right] \varepsilon^{-2} + 1. \qquad (169)
\end{aligned}$$

Hence, we obtain that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
\mathfrak{M}_{d,\varepsilon} &\leq 2^{2(\kappa+4)}p^2\kappa^2 d^{2\kappa} e^{2\kappa^2 T}\left[1 + \left|2p\iota d^\iota \max\{1,T\}\right|^{p\kappa} + \eta d^\eta\right]\varepsilon^{-2} + 1 \\
&\leq 2^{2(\kappa+4)}p^2\kappa^2 d^{2\kappa} e^{2\kappa^2 T} d^{\max\{p\kappa\iota,\eta\}}\left[1 + |2p\iota \max\{1,T\}|^{p\kappa} + \eta\right]\varepsilon^{-2} + 1 \\
&\leq 2^{2(\kappa+4)}p^2\kappa^2 e^{2\kappa^2 T}\left[1 + |2p\iota \max\{1,T\}|^{p\kappa} + \eta\right] d^{p\kappa\iota+\eta+2\kappa}\varepsilon^{-2} + 1 \qquad (170) \\
&\leq 2^{2(\kappa+4)}p^2\iota^2 e^{2\kappa^2 T}\left[1 + |2p\iota \max\{1,T\}|^{p\kappa} + \eta\right] d^{p\kappa\iota+\eta+2\kappa}\varepsilon^{-2} + 1 \\
&\leq 2^{2(\kappa+4)}p^2\iota^2 e^{2\kappa^2 T}\left[2 + |2p\iota \max\{1,T\}|^{p\kappa} + \eta\right] d^{p\kappa\iota+\eta+2\kappa}\varepsilon^{-2}.
\end{aligned}$$

Next observe that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
\|f_{1,d}(0)\|_{\mathbb{R}^d} &\leq \|f_{1,d}(0) - (\mathcal{R}\phi^{1,d}_\varepsilon)(0)\|_{\mathbb{R}^d} + \|(\mathcal{R}\phi^{1,d}_\varepsilon)(0)\|_{\mathbb{R}^d} \\
&\leq \varepsilon\kappa d^\kappa + \kappa d^\kappa \leq 2\kappa d^\kappa \leq 2\iota d^\kappa.
\end{aligned} \qquad (171)$$

Therefore, we obtain that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
&\mathcal{D}_{d,\varepsilon} = \varepsilon\left[\max\{2\kappa d^\kappa,1\} + T^{-1/2}\right]^{-1} e^{-(3+3\kappa+[\kappa^2+2\kappa\iota+2]T)}\left|\max\{1,2\kappa(\kappa+1)d^\kappa\}\right|^{-1} 2^{-(2\iota+1)} \\
&\cdot\left(\left[2 + \max\{1,\kappa d^\kappa,\|f_{1,d}(0)\|_{\mathbb{R}^d}\}\max\{1,T\} + \sqrt{2(2\iota-1)\kappa d^\kappa T}\right]^{p\iota+p\kappa} + \eta d^\eta\right)^{-1/p} \\
&\geq \varepsilon\left[\max\{2\kappa d^\kappa,1\} + T^{-1/2}\right]^{-1} e^{-(3+3\kappa+[\kappa^2+2\kappa\iota+2]T)}\left|\max\{1,2\kappa(\kappa+1)d^\kappa\}\right|^{-1} 2^{-(2\iota+1)} \\
&\cdot\left(\left[2 + \max\{1,\kappa d^\kappa,2\iota d^\kappa\}\max\{1,T\} + \sqrt{2(2\iota-1)\kappa d^\kappa T}\right]^{p\iota+p\kappa} + \eta d^\eta\right)^{-1/p} \\
&\geq \varepsilon\left[2\max\{2\kappa d^\kappa,1,T^{-1/2}\}\right]^{-1} e^{-(3+3\iota+[3\iota^2+2]T)}\left|\max\{1,2\kappa(\kappa+1)d^\kappa\}\right|^{-1} 2^{-(2\iota+1)} \\
&\cdot\left(\left[2 + 2\iota d^\kappa\max\{1,T\} + 2\iota d^\kappa\sqrt{T}\right]^{p\iota+p\kappa} + \eta d^\eta\right)^{-1/p}. \qquad (172)
\end{aligned}$$

This proves that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
\mathcal{D}_{d,\varepsilon} &\geq \varepsilon\left|4\iota d^\kappa\max\{1,T^{-1/2}\}\right|^{-1} e^{-(3\iota^2+3)(T+1)}\left|4\iota^2 d^\kappa\right|^{-1} 2^{-(2\iota+1)} \\
&\quad\cdot\left(\left[6\iota d^\kappa\max\{1,T\}\right]^{p\iota+p\kappa} + \eta d^\eta\right)^{-1/p} \\
&\geq \varepsilon\left|4\iota d^\kappa\max\{1,T^{-1/2}\}\right|^{-1} e^{-(3\iota^2+3)(T+1)}\left|4\iota^2 d^\kappa\right|^{-1} 2^{-(2\iota+1)} \\
&\quad\cdot\left(\left[6\iota\max\{1,T\}\right]^{p\iota+p\kappa} + \eta\right)^{-1/p} d^{[-\kappa(p\iota+p\kappa)-\eta]/p} \\
&\geq \left|4\iota\max\{1,T^{-1/2}\}\right|^{-1} e^{-(3\iota^2+3)(T+1)}\left|4\iota^2\right|^{-1} 2^{-(2\iota+1)} \\
&\quad\cdot\left([6\iota\max\{1,T\}]^{p\iota+p\kappa} + \eta\right)^{-1/p} d^{-(2\kappa+\kappa(\kappa+\iota)+\eta)}\varepsilon.
\end{aligned} \qquad (173)$$

Hence, we obtain that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned} \mathcal{D}_{d,\varepsilon} \geq {} & |\min\{1, T^{1/2}\}| e^{-(3\iota^2+3)(T+1)} \, \iota^{-3} 2^{-(2\iota+5)} \\ & \cdot \big([6\iota \max\{1,T\}]^{p\iota+p\kappa} + \eta \big)^{-1/p} d^{-(\kappa(2+\kappa+\iota)+\eta)} \varepsilon. \end{aligned} \tag{174}$$

Moreover, note that Proposition 5.3 ensures that for all $\delta \in (0,1]$, $d \in \mathbb{N}$, $t \in [0,T]$, $\omega \in \Omega$ there exist $(\psi_{\delta,d,m,t,\omega})_{m\in\mathbb{N}} \subseteq \mathcal{N}$ such that for all $x \in \mathbb{R}^d$, $m \in \mathbb{N}$ it holds that $\mathcal{R}(\psi_{\delta,d,m,t,\omega}) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\mathcal{P}(\psi_{\delta,d,m,t,\omega}) \leq \mathcal{P}(\phi^{2,d}) + (\mathcal{P}(\phi^{2,d}) + \mathcal{P}(\phi^{1,d}_\delta))^3 [\chi_\delta(t)/\delta^2 + 1]$, $\mathcal{L}(\psi_{\delta,d,m,t,\omega}) = \mathcal{L}(\psi_{\delta,d,1,t,\omega})$, and

$$(\mathcal{R}\psi_{\delta,d,m,t,\omega})(x) = Y^{\delta,d,m,x}_t(\omega). \tag{175}$$

This demonstrates that for all $\delta \in (0,1]$, $d, m \in \mathbb{N}$, $\omega \in \Omega$ it holds that

$$\begin{aligned} \mathcal{P}(\psi_{\delta,d,m,T,\omega}) & \leq \big(\mathcal{P}(\phi^{2,d}) + \mathcal{P}(\phi^{1,d}_\delta) \big)^3 \big[\chi_\delta(T)/\delta^2 + 2 \big] \\ & \leq \big(\kappa d^\kappa \delta^{-\kappa} \big)^3 \big[T/\delta^2 + 2 \big] \leq \kappa^3 (T+2) d^{3\kappa} \delta^{-3\kappa-2}. \end{aligned} \tag{176}$$

Proposition 5.2 hence proves that for all $\delta \in (0,1]$, $d \in \mathbb{N}$, $\omega \in \Omega$ there exist $(\varphi_{\delta,d,m,\omega})_{m\in\mathbb{N}} \subseteq \mathcal{N}$ such that for all $x \in \mathbb{R}^d$, $m \in \mathbb{N}$ it holds that $\mathcal{R}(\varphi_{\delta,d,m,\omega}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\varphi_{\delta,d,m,\omega}) \leq \mathcal{P}(\phi^{2,d})(\mathcal{P}(\phi^{0,d}_\delta) + \mathcal{P}(\psi_{\delta,d,m,T,\omega}))$, $\mathcal{L}(\varphi_{\delta,d,m,\omega}) = \mathcal{L}(\varphi_{\delta,d,1,\omega})$, and

$$(\mathcal{R}\varphi_{\delta,d,m,\omega})(x) = \Phi^{0,d}_\delta(Y^{\delta,d,m,x}_T(\omega)). \tag{177}$$

This and (176) ensure that for all $\delta \in (0,1]$, $d, m \in \mathbb{N}$, $\omega \in \Omega$ it holds that

$$\begin{aligned} \mathcal{P}(\varphi_{\delta,d,m,\omega}) & \leq \kappa d^\kappa \big[\kappa d^\kappa \delta^{-\kappa} + \kappa^3 (T+2) d^{3\kappa} \delta^{-3\kappa-2} \big] \\ & \leq \kappa^2 d^{4\kappa} \delta^{-3\kappa-2} (T+2)[1+\kappa^2] \leq 2\iota^2 \kappa^2 (T+2) d^{4\kappa} \delta^{-3\kappa-2}. \end{aligned} \tag{178}$$

Lemma 5.1 and (177) therefore show that for all $M, d \in \mathbb{N}$, $\delta \in (0,1]$, $\omega \in \Omega$ there exists $\Psi_{M,d,\delta,\omega} \in \mathcal{N}$ such that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}(\Psi_{M,d,\delta,\omega}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\Psi_{M,d,\delta,\omega}) \leq 2M^2 \iota^2 \kappa^2 (T+2) d^{4\kappa} \delta^{-3\kappa-2}$, and

$$(\mathcal{R}\Psi_{M,d,\delta,\omega})(x) = \frac{1}{M} \sum_{m=1}^{M} \Phi^{0,d}_\delta(Y^{\delta,d,m,x}_T(\omega)). \tag{179}$$

This, (170), and (174) assure that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\omega \in \Omega$ it holds that

$$\begin{aligned} & \mathcal{P}(\Psi_{\mathfrak{M}_{d,\varepsilon},d,\mathcal{D}_{d,\varepsilon},\omega}) \leq 2|\mathfrak{M}_{d,\varepsilon}|^2 \iota^2 \kappa^2 (T+2) d^{4\kappa} |\mathcal{D}_{d,\varepsilon}|^{-3\kappa-2} \\ & \leq 2\Big(2^{2(\kappa+4)} p^2 \iota^2 e^{2\kappa^2 T} \big[2 + |2p\iota \max\{1,T\}|^{p\kappa} + \eta \big] d^{p\kappa\iota+\eta+2\kappa} \varepsilon^{-2} \Big)^2 \\ & \cdot \iota^2 \kappa^2 (T+2) d^{4\kappa} \Big[|\min\{1,T^{1/2}\}| e^{-(3\iota^2+3)(T+1)} \, \iota^{-3} 2^{-(2\iota+5)} \\ & \cdot \big([6\iota \max\{1,T\}]^{p\iota+p\kappa} + \eta \big)^{-1/p} d^{-(\kappa(2+\kappa+\iota)+\eta)} \varepsilon \Big]^{-3\kappa-2} \\ & = 2\Big(2^{2(\kappa+4)} p^2 \iota^2 e^{2\kappa^2 T} \big[2 + |2p\iota \max\{1,T\}|^{p\kappa} + \eta \big] \Big)^2 \iota^2 \kappa^2 (T+2) \\ & \cdot \Big[|\min\{1,T^{1/2}\}| e^{-(3\iota^2+3)(T+1)} \, \iota^{-3} 2^{-(2\iota+5)} \big([6\iota \max\{1,T\}]^{p\iota+p\kappa} + \eta \big)^{-1/p} \Big]^{-3\kappa-2} \\ & \cdot d^{2(p\kappa\iota+\eta+4\kappa)+(\kappa(2+\kappa+\iota)+\eta)(3\kappa+2)} \varepsilon^{-3\kappa-6}. \end{aligned} \tag{180}$$

Combining this and (168) finishes the proof of item (ii). The proof of Proposition 6.1 is thus completed. ∎

Corollary 6.2. *Let $T, \kappa, \eta, p \in (0,\infty)$, let $A_d = (a_{d,i,j})_{(i,j)\in\{1,\ldots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be symmetric positive semidefinite matrices, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm and let $\nu_d\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure on $\mathbb{R}^d$, let $f_{0,d}\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $f_{1,d}\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be functions, let $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $\mathbf{a}\colon \mathbb{R} \to \mathbb{R}$ be continuous functions which satisfy for all $d \in \mathbb{N}$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ that $\mathbf{A}_d(x) = (\mathbf{a}(x_1),\ldots,\mathbf{a}(x_d))$, let*

$$\mathcal{N} = \cup_{L\in\{2,3,4,\ldots\}} \cup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{n=1}^{L} (\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n})\right), \tag{181}$$

let $\mathcal{P}\colon \mathcal{N} \to \mathbb{N}$, $\mathcal{L}\colon \mathcal{N} \to \cup_{L\in\{2,3,4,\ldots\}}\mathbb{N}^{L+1}$, and $\mathcal{R}\colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k,\mathbb{R}^l)$ be the functions which satisfy for all $L \in \{2,3,4,\ldots\}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1,B_1),\ldots,(W_L,B_L)) \in (\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n}))$, $x_0 \in \mathbb{R}^{l_0}$, $\ldots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, n \in \mathbb{N} \cap [1,L)\colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ that $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1}+1)$, $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0},\mathbb{R}^{l_L})$, $\mathcal{L}(\Phi) = (l_0,l_1,\ldots,l_L)$, and

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{182}$$

let $(\phi_{\varepsilon}^{m,d})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$, $(\phi^{2,d})_{d\in\mathbb{N}} \subseteq \mathcal{N}$, and assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x, y \in \mathbb{R}^d$ that $\mathcal{R}(\phi_{\varepsilon}^{0,d}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{R}(\phi_{\varepsilon}^{1,d}), \mathcal{R}(\phi^{2,d}) \in C(\mathbb{R}^d,\mathbb{R}^d)$, $|f_{0,d}(x)| + \sum_{i,j=1}^{d} |a_{d,i,j}| \leq \kappa d^{\kappa}(1+\|x\|_{\mathbb{R}^d}^{\kappa})$, $\|f_{1,d}(x) - f_{1,d}(y)\|_{\mathbb{R}^d} \leq \kappa\|x-y\|_{\mathbb{R}^d}$, $\|(\mathcal{R}\phi_{\varepsilon}^{1,d})(x)\|_{\mathbb{R}^d} \leq \kappa(d^{\kappa} + \|x\|_{\mathbb{R}^d})$, $\mathcal{P}(\phi^{2,d}) + \sum_{m=0}^{1} \mathcal{P}(\phi_{\varepsilon}^{m,d}) \leq \kappa d^{\kappa}\varepsilon^{-\kappa}$, $|(\mathcal{R}\phi_{\varepsilon}^{0,d})(x) - (\mathcal{R}\phi_{\varepsilon}^{0,d})(y)| \leq \kappa d^{\kappa}(1+\|x\|_{\mathbb{R}^d}^{\kappa}+\|y\|_{\mathbb{R}^d}^{\kappa})\|x-y\|_{\mathbb{R}^d}$, $\mathcal{L}(\phi^{2,d}) \in \mathbb{N}^3$, $(\mathcal{R}\phi^{2,d})(x) = x$, $\int_{\mathbb{R}^d} \|z\|_{\mathbb{R}^d}^{p(2\kappa+1)}\,\nu_d(dz) \leq \eta d^{\eta}$, and

$$|f_{0,d}(x) - (\mathcal{R}\phi_{\varepsilon}^{0,d})(x)| + \|f_{1,d}(x) - (\mathcal{R}\phi_{\varepsilon}^{1,d})(x)\|_{\mathbb{R}^d} \leq \varepsilon\kappa d^{\kappa}(1 + \|x\|_{\mathbb{R}^d}^{\kappa}). \tag{183}$$

Then

(i) there exist unique at most polynomially growing functions $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, such that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $u_d(0,x) = f_{0,d}(x)$ and such that for all $d \in \mathbb{N}$ it holds that u_d is a viscosity solution of

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\tfrac{\partial}{\partial x}u_d)(t,x)\, f_{1,d}(x) + \sum_{i,j=1}^{d} a_{d,i,j}\,(\tfrac{\partial^2}{\partial x_i \partial x_j}u_d)(t,x) \tag{184}$$

for $(t,x) \in (0,T)\times\mathbb{R}^d$ and

(ii) there exist $(\psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$, $c \in \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\psi_{d,\varepsilon}) \leq c\, d^{c}\varepsilon^{-c}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, and

$$\left[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}\psi_{d,\varepsilon})(x)|^p\,\nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{185}$$

6.2 Rectified DNN approximations

Theorem 6.3. *Let $T, \kappa, \eta, p \in (0,\infty)$, let $A_d = (a_{d,i,j})_{(i,j)\in\{1,\ldots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be symmetric positive semidefinite matrices, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm and let $\nu_d\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability* *measure on* $\mathbb{R}^d$*, let* $f_{0,d}\colon \mathbb{R}^d \to \mathbb{R}$*,* $d \in \mathbb{N}$*, and* $f_{1,d}\colon \mathbb{R}^d \to \mathbb{R}^d$*,* $d \in \mathbb{N}$*, be functions, let* $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$*,* $d \in \mathbb{N}$*, be the functions which satisfy for all* $d \in \mathbb{N}$*,* $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ *that* $\mathbf{A}_d(x) = (\max\{x_1, 0\}, \ldots, \max\{x_d, 0\})$*, let*

$$\mathcal{N} = \cup_{L\in\{2,3,4,\dots\}} \cup_{(l_0,l_1,\dots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n})\right), \tag{186}$$

let $\mathcal{P}\colon \mathcal{N} \to \mathbb{N}$*,* $\mathcal{L}\colon \mathcal{N} \to \cup_{L\in\{2,3,4,\dots\}}\mathbb{N}^{L+1}$*, and* $\mathcal{R}\colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ *be the functions which satisfy for all* $L \in \{2, 3, 4, \dots\}$*,* $l_0, l_1, \dots, l_L \in \mathbb{N}$*,* $\Phi = ((W_1, B_1), \dots, (W_L, B_L)) \in (\times_{n=1}^{L}(\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n}))$*,* $x_0 \in \mathbb{R}^{l_0}$*,* $\dots$*,* $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ *with* $\forall\, n \in \mathbb{N} \cap [1, L)\colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ *that* $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1} + 1)$*,* $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$*,* $\mathcal{L}(\Phi) = (l_0, l_1, \dots, l_L)$*, and*

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{187}$$

let $(\phi^{m,d}_{\varepsilon})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$*, and assume for all* $d \in \mathbb{N}$*,* $\varepsilon \in (0, 1]$*,* $x, y \in \mathbb{R}^d$ *that* $\mathcal{R}(\phi^{0,d}_{\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$*,* $\mathcal{R}(\phi^{1,d}_{\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R}^d)$*,* $|f_{0,d}(x)| + \sum_{i,j=1}^{d} |a_{d,i,j}| \le \kappa d^{\kappa}(1 + \|x\|^{\kappa}_{\mathbb{R}^d})$*,* $\|f_{1,d}(x) - f_{1,d}(y)\|_{\mathbb{R}^d} \le \kappa\|x - y\|_{\mathbb{R}^d}$*,* $\|(\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \le \kappa(d^{\kappa} + \|x\|_{\mathbb{R}^d})$*,* $\sum_{m=0}^{1} \mathcal{P}(\phi^{m,d}_{\varepsilon}) \le \kappa d^{\kappa}\varepsilon^{-\kappa}$*,* $|(\mathcal{R}\phi^{0,d}_{\varepsilon})(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(y)| \le \kappa d^{\kappa}(1 + \|x\|^{\kappa}_{\mathbb{R}^d} + \|y\|^{\kappa}_{\mathbb{R}^d})\|x - y\|_{\mathbb{R}^d}$*,* $\int_{\mathbb{R}^d} \|z\|^{p(4\kappa+15)}_{\mathbb{R}^d}\, \nu_d(dz) \le \eta d^{\eta}$*, and*

$$|f_{0,d}(x) - (\mathcal{R}\phi^{0,d}_{\varepsilon})(x)| + \|f_{1,d}(x) - (\mathcal{R}\phi^{1,d}_{\varepsilon})(x)\|_{\mathbb{R}^d} \le \varepsilon\kappa d^{\kappa}(1 + \|x\|^{\kappa}_{\mathbb{R}^d}). \tag{188}$$

Then

(i) there exist unique at most polynomially growing functions $u_d\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$*,* $d \in \mathbb{N}$*, such that for all* $d \in \mathbb{N}$*,* $x \in \mathbb{R}^d$ *it holds that* $u_d(0, x) = f_{0,d}(x)$ *and such that for all* $d \in \mathbb{N}$ *it holds that* u_d *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u_d)(t, x) = (\tfrac{\partial}{\partial x}u_d)(t, x)\, f_{1,d}(x) + \sum_{i,j=1}^{d} a_{d,i,j}\, (\tfrac{\partial^2}{\partial x_i \partial x_j}u_d)(t, x) \tag{189}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$ *and*

(ii) there exist $(\psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$*,* $c \in \mathbb{R}$ *such that for all* $d \in \mathbb{N}$*,* $\varepsilon \in (0, 1]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon}) \le c\, d^c \varepsilon^{-c}$*,* $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$*, and*

$$\left[\int_{\mathbb{R}^d} |u_d(T, x) - (\mathcal{R}\psi_{d,\varepsilon})(x)|^p\, \nu_d(dx)\right]^{1/p} \le \varepsilon. \tag{190}$$

Proof of Theorem 6.3. Throughout this proof let $\mathbf{a}\colon \mathbb{R} \to \mathbb{R}$ be the function which satisfies for all $x \in \mathbb{R}$ that

$$\mathbf{a}(x) = \max\{x, 0\} \tag{191}$$

and let $(\phi^{2,d})_{d\in\mathbb{N}} \subseteq \mathcal{N}$ satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $\mathcal{R}(\psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\mathcal{L}(\phi^{2,d}) = (d, 2d, d)$, and $(\mathcal{R}\phi^{2,d})(x) = x$ (cf. Lemma 5.4). Observe that for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\mathbf{A}_d(x) = (\mathbf{a}(x_1), \dots, \mathbf{a}(x_d)). \tag{192}$$

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

$$\mathcal{P}(\phi^{2,d}) = 2d(d+1) + d(2d+1) = 2d^2 + 2d + 2d^2 + d = 4d^2 + 3d \leq 7d^2. \quad (193)$$

This proves that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned} \mathcal{P}(\phi^{2,d}) + \sum_{m=0}^{1} \mathcal{P}(\phi_\varepsilon^{m,d}) &\leq 7d^2 + \kappa d^\kappa \varepsilon^{-\kappa} \leq (\kappa+7) d^{\kappa+2} \varepsilon^{-\kappa} \\ &\leq (2\kappa+7) d^{2\kappa+7} \varepsilon^{-(2\kappa+7)}. \end{aligned} \quad (194)$$

Moreover, observe that Young's inequality assures that for all $\alpha \in [0,\infty)$ it holds that

$$\alpha^\kappa \leq \frac{\kappa}{2\kappa+7} \cdot \alpha^{2\kappa+7} + \frac{\kappa+7}{2\kappa+7} \leq \alpha^{2\kappa+7} + \frac{\kappa+7}{2\kappa+7}. \quad (195)$$

This ensures that for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$ it holds that

$$\kappa(1 + \|x\|_{\mathbb{R}^d}^\kappa) \leq \kappa(2 + \|x\|_{\mathbb{R}^d}^{2\kappa+7}) \leq (2\kappa+7)(1 + \|x\|_{\mathbb{R}^d}^{2\kappa+7}) \quad (196)$$

and

$$\begin{aligned} \kappa(1 + \|x\|_{\mathbb{R}^d}^\kappa + \|y\|_{\mathbb{R}^d}^\kappa) &\leq \kappa\left(1 + \frac{2(\kappa+7)}{2\kappa+7} + \|x\|_{\mathbb{R}^d}^{2\kappa+7} + \|y\|_{\mathbb{R}^d}^{2\kappa+7}\right) \\ &= \frac{\kappa(4\kappa+21)}{2\kappa+7} + \kappa(\|x\|_{\mathbb{R}^d}^{2\kappa+7} + \|y\|_{\mathbb{R}^d}^{2\kappa+7}) \\ &\leq 2\kappa + 7 + (2\kappa+7)(\|x\|_{\mathbb{R}^d}^{2\kappa+7} + \|y\|_{\mathbb{R}^d}^{2\kappa+7}) \\ &= (2\kappa+7)(1 + \|x\|_{\mathbb{R}^d}^{2\kappa+7} + \|y\|_{\mathbb{R}^d}^{2\kappa+7}). \end{aligned} \quad (197)$$

Combining this with (192), (194), the fact that $\mathbf{A}_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $\mathbf{a} \colon \mathbb{R} \to \mathbb{R}$ are continuous functions, and Corollary 6.2 (with $T = T$, $\kappa = 2\kappa + 7$, $\eta = \eta$, $p = p$, $(A_d)_{d\in\mathbb{N}} = (A_d)_{d\in\mathbb{N}}$, $(\nu_d)_{d\in\mathbb{N}} = (\nu_d)_{d\in\mathbb{N}}$, $(f_{0,d})_{d\in\mathbb{N}} = (f_{0,d})_{d\in\mathbb{N}}$, $(f_{1,d})_{d\in\mathbb{N}} = (f_{1,d})_{d\in\mathbb{N}}$, $(\mathbf{A}_d)_{d\in\mathbb{N}} = (\mathbf{A}_d)_{d\in\mathbb{N}}$, $\mathbf{a} = \mathbf{a}$, $\mathcal{N} = \mathcal{N}$, $\mathcal{P} = \mathcal{P}$, $\mathcal{L} = \mathcal{L}$, $\mathcal{R} = \mathcal{R}$, $(\phi_\varepsilon^{m,d})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} = (\phi_\varepsilon^{m,d})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]}$, $(\phi^{2,d})_{d\in\mathbb{N}} = (\phi^{2,d})_{d\in\mathbb{N}}$ in the notation of Corollary 6.2) establishes items (i)–(ii). The proof of Theorem 6.3 is thus completed. ☐

6.3 Rectified DNN approximations on the d-dimensional unit cube

Corollary 6.4. *Let $T, \kappa, p \in (0,\infty)$, let $A_d = (a_{d,i,j})_{(i,j)\in\{1,\dots,d\}^2} \in \mathbb{R}^{d\times d}$, $d \in \mathbb{N}$, be symmetric positive semidefinite matrices, for every $d \in \mathbb{N}$ let $\|\cdot\|_{\mathbb{R}^d} \colon \mathbb{R}^d \to [0,\infty)$ be the d-dimensional Euclidean norm, let $f_{0,d} \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $f_{1,d} \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be functions, let $\mathbf{A}_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be the functions which satisfy for all $d \in \mathbb{N}$, $x = (x_1,\dots,x_d) \in \mathbb{R}^d$ that $\mathbf{A}_d(x) = (\max\{x_1,0\},\dots,\max\{x_d,0\})$, let*

$$\mathcal{N} = \cup_{L\in\{2,3,4,\dots\}} \cup_{(l_0,l_1,\dots,l_L)\in\mathbb{N}^{L+1}} (\times_{n=1}^L (\mathbb{R}^{l_n\times l_{n-1}} \times \mathbb{R}^{l_n})), \quad (198)$$

let $\mathcal{P} \colon \mathcal{N} \to \mathbb{N}$ and $\mathcal{R} \colon \mathcal{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ be the functions which satisfy for all $L \in \{2,3,4,\dots\}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1,B_1),\dots,(W_L,B_L)) \in (\times_{n=1}^L (\mathbb{R}^{l_n\times l_{n-1}} \times$

$\mathbb{R}^{l_n})$), $x_0 \in \mathbb{R}^{l_0}$, …, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ *with* $\forall\, n \in \mathbb{N} \cap [1, L)\colon x_n = \mathbf{A}_{l_n}(W_n x_{n-1} + B_n)$ *that* $\mathcal{P}(\Phi) = \sum_{n=1}^{L} l_n(l_{n-1} + 1)$, $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, *and*

$$(\mathcal{R}\Phi)(x_0) = W_L x_{L-1} + B_L, \tag{199}$$

let $(\phi^{m,d}_\varepsilon)_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$ *, and assume for all* $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, y \in \mathbb{R}^d$ *that* $\mathcal{R}(\phi^{0,d}_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{R}(\phi^{1,d}_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $|f_{0,d}(x)| + \sum_{i,j=1}^{d} |a_{d,i,j}| \leq \kappa d^\kappa(1 + \|x\|^\kappa_{\mathbb{R}^d})$, $\|f_{1,d}(x) - f_{1,d}(y)\|_{\mathbb{R}^d} \leq \kappa\|x - y\|_{\mathbb{R}^d}$, $\|(\mathcal{R}\phi^{1,d}_\varepsilon)(x)\|_{\mathbb{R}^d} \leq \kappa(d^\kappa + \|x\|_{\mathbb{R}^d})$, $\sum_{m=0}^{1} \mathcal{P}(\phi^{m,d}_\varepsilon) \leq \kappa d^\kappa \varepsilon^{-\kappa}$, $|(\mathcal{R}\phi^{0,d}_\varepsilon)(x) - (\mathcal{R}\phi^{0,d}_\varepsilon)(y)| \leq \kappa d^\kappa(1 + \|x\|^\kappa_{\mathbb{R}^d} + \|y\|^\kappa_{\mathbb{R}^d})\|x - y\|_{\mathbb{R}^d}$, *and*

$$|f_{0,d}(x) - (\mathcal{R}\phi^{0,d}_\varepsilon)(x)| + \|f_{1,d}(x) - (\mathcal{R}\phi^{1,d}_\varepsilon)(x)\|_{\mathbb{R}^d} \leq \varepsilon\kappa d^\kappa(1 + \|x\|^\kappa_{\mathbb{R}^d}). \tag{200}$$

Then

(i) there exist unique at most polynomially growing functions $u_d\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, *such that for all* $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ *it holds that* $u_d(0, x) = f_{0,d}(x)$ *and such that for all* $d \in \mathbb{N}$ *it holds that* u_d *is a viscosity solution of*

$$(\tfrac{\partial}{\partial t}u_d)(t, x) = (\tfrac{\partial}{\partial x}u_d)(t, x)\, f_{1,d}(x) + \sum_{i,j=1}^{d} a_{d,i,j}\, (\tfrac{\partial^2}{\partial x_i \partial x_j}u_d)(t, x) \tag{201}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$ *and*

(ii) there exist $(\psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathcal{N}$, $c \in \mathbb{R}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ *it holds that* $\mathcal{P}(\psi_{d,\varepsilon}) \leq c\, d^c \varepsilon^{-c}$, $\mathcal{R}(\psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, *and*

$$\left[\int_{[0,1]^d} |u_d(T, x) - (\mathcal{R}\psi_{d,\varepsilon})(x)|^p\, dx\right]^{1/p} \leq \varepsilon. \tag{202}$$

Proof of Corollary 6.4. Throughout this proof for every $d \in \mathbb{N}$ let $\lambda_d\colon \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ be the Lebesgue-Borel measure on $\mathbb{R}^d$ and let $\nu_d\colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be the function which satisfies for all $B \in \mathcal{B}(\mathbb{R}^d)$ that

$$\nu_d(B) = \lambda_d(B \cap [0, 1]^d). \tag{203}$$

Observe that (203) implies that for all $d \in \mathbb{N}$ it holds that ν_d is a probability measure on $\mathbb{R}^d$. This and (203) ensure that for all $d \in \mathbb{N}$, $g \in C(\mathbb{R}^d, \mathbb{R})$ it holds that

$$\int_{\mathbb{R}^d} |g(x)|\, \nu_d(dx) = \int_{[0,1]^d} |g(x)|\, dx. \tag{204}$$

Combining this with, e.g., Grohs et al. [26, Lemma 3.12] demonstrates that for all $d \in \mathbb{N}$ it holds that

$$\begin{aligned}\int_{\mathbb{R}^d} \|x\|^{p(4\kappa+15)}_{\mathbb{R}^d}\, \nu_d(dx) &= \int_{[0,1]^d} \|x\|^{p(4\kappa+15)}_{\mathbb{R}^d}\, dx \leq d^{p(4\kappa+15)/2} \\ &\leq d^{p(2\kappa+8)} \leq p(2\kappa + 8)d^{p(2\kappa+8)}.\end{aligned} \tag{205}$$

Theorem 6.3 (with $T = T$, $\kappa = \kappa$, $\eta = p(2\kappa+8)$, $p = p$, $(A_d)_{d\in\mathbb{N}} = (A_d)_{d\in\mathbb{N}}$, $(\nu_d)_{d\in\mathbb{N}} = (\nu_d)_{d\in\mathbb{N}}$, $(f_{0,d})_{d\in\mathbb{N}} = (f_{0,d})_{d\in\mathbb{N}}$, $(f_{1,d})_{d\in\mathbb{N}} = (f_{1,d})_{d\in\mathbb{N}}$, $(\mathbf{A}_d)_{d\in\mathbb{N}} = (\mathbf{A}_d)_{d\in\mathbb{N}}$, $\mathcal{N} = \mathcal{N}$, $\mathcal{P} = \mathcal{P}$, $\mathcal{R} = \mathcal{R}$, $(\phi^{m,d}_\varepsilon)_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} = (\phi^{m,d}_\varepsilon)_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]}$, $(\phi^{2,d})_{d\in\mathbb{N}} = (\phi^{2,d})_{d\in\mathbb{N}}$ in the notation of Theorem 6.3) and (204) hence establish items (i)–(ii). The proof of Corollary 6.4 is thus completed. □

Acknowledgements

David Kofler is gratefully acknowledged for his useful comments regarding the a priori estimates in Subsections 4.1–4.2. This work has been partially supported through the research grant with the title “Higher order numerical approximation methods for stochastic partial differential equations” (Number 175699) from the Swiss National Science Foundation (SNSF). Furthermore, this work has been partially supported through the ETH Research Grant ETH-47 15-2 “Mild stochastic calculus and numerical approximations for nonlinear stochastic evolution equations with Lévy noise”.

References

[1] Bach, F. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research 18*, 19 (2017), 1–53.

[2] Barron, A. R. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory 39*, 3 (1993), 930–945.

[3] Barron, A. R. Approximation and estimation bounds for artificial neural networks. *Mach. Learn. 14*, 1 (1994), 115–133.

[4] 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), 56 pages.

[5] Beck, C., Jentzen, A., and E, W. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *arXiv:1709.05963* (2017), 56 pages.

[6] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *arXiv:1804.05394* (2018), 18 pages.

[7] Blum, E. K., and Li, L. K. Approximation theory and feedforward networks. *Neural networks 4*, 4 (1991), 511–515.

[8] Bölcskei, H., Grohs, P., Kutyniok, G., and Petersen, P. Optimal approximation with sparsely connected deep neural networks. *arXiv:1705.01714* (2017), 36 pages.

[9] Burger, M., and Neubauer, A. Error Bounds for Approximation with Neural Networks. *Journal of Approximation Theory 112*, 2 (2001), 235–250.

[10] Candès, E. J. Ridgelets: Theory and Applications, 1998. Ph.D. thesis, Stanford University.

[11] Chen, T., and Chen, H. Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. *IEEE Transactions on Neural Networks 6*, 4 (1995), 904–910.

[12] Chouiekh, A., and Haj, E. H. I. E. Convnets for fraud detection analysis. *Procedia Computer Science 127* (2018), 133–138.

[13] Chui, C. K., Li, X., and Mhaskar, H. N. Neural networks for localized approximation. *Math. Comp. 63*, 208 (1994), 607–623.

[14] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signal 2*, 4 (1989), 303–314.

[15] Dahl, G. E., Yu, D., Deng, L., and Acero, A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing 20*, 1 (2012), 30–42.

[16] DeVore, R., Oskolkov, K., and Petrushev, P. Approximation by feed-forward neural networks. *Ann. Numer. Math. 4* (1996), 261–287.

[17] E, W., Han, J., and Jentzen, A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat. 5*, 4 (2017), 349–380.

[18] E, W., and Wang, Q. Exponential convergence of the deep neural network approximation for analytic functions. *Science China Mathematics* (2018). Published online: September 6, 2018, https://doi.org/10.1007/s11425-018-9387-x.

[19] E, W., and Yu, B. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat. 6*, 1 (2018), 1–12.

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

[21] Eldan, R., and Shamir, O. The power of depth for feedforward neural networks. *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA* (2016), 907–940.

[22] Ellacott, S. Aspects of the numerical analysis of neural networks. *Acta Numer. 3* (1994), 145–202.

[23] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. *arXiv:1710.07030* (2017), 16 pages.

[24] Funahashi, K.-I. On the approximate realization of continuous mappings by neural networks. *Neural Networks 2*, 3 (1989), 183–192.

[25] Graves, A., Mohamed, A.-r., and Hinton, G. Speech recognition with deep recurrent neural networks. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing, ICASSP* (2013), pp. 6645–6649.

[26] Grohs, P., Hornung, F., Jentzen, A., and von Wurstemberger, P. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *arXiv:1809.02362* (2018), 124 pages.

[27] Hairer, M., Hutzenthaler, M., and Jentzen, A. Loss of regularity for Kolmogorov equations. *Ann. Probab. 43*, 2 (2015), 468–527.

[28] 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.

[29] Hartman, E. J., Keeler, J. D., and Kowalski, J. M. Layered neural networks with gaussian hidden units as universal approximations. *Neural computation 2*, 2 (1990), 210–215.

[30] Henry-Labordère, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. (November 15, 2017), 16 pages. Available at SSRN: `https://ssrn.com/abstract=3071506`.

[31] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine 29*, 6 (2012), 82–97.

[32] Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks 4*, 2 (1991), 251 – 257.

[33] Hornik, K. Some new results on neural network approximation. *Neural networks 6*, 8 (1993), 1069–1072.

[34] Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural Networks 2*, 5 (1989), 359–366.

[35] Hornik, K., Stinchcombe, M., and White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks 3*, 5 (1990), 551–560.

[36] Hu, B., Lu, Z., Li, H., and Chen, Q. Convolutional neural network architectures for matching natural language sentences. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (2014), pp. 2042–2050.

[37] Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 2261–2269.

[38] Hutzenthaler, M., and Jentzen, A. On a perturbation theory and on strong convergence rates for stochastic ordinary and partial differential equations with non-globally monotone coefficients. *arXiv:1401.0295* (2014), 41 pages.

[39] Kalchbrenner, N., Grefenstette, E., and Blunsom, P. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (2014), pp. 655–665.

[40] Khoo, Y., Lu, J., and Ying, L. Solving parametric PDE problems with artificial neural networks. *arXiv:1707.03351* (2017), 17 pages.

[41] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.

[42] Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks 6*, 6 (1993), 861–867.

[43] Mhaskar, H., and Micchelli, C. Degree of approximation by neural and translation networks with a single hidden layer. *Adv. Appl. Math. 16*, 2 (1995), 151–183.

[44] Mhaskar, H. N. Neural networks for optimal approximation of smooth and analytic functions. *Neural Comput. 8*, 1 (1996), 164–177.

[45] Mhaskar, H. N., and Poggio, T. Deep vs. shallow networks: An approximation theory perspective. *Analysis and Applications 14*, 06 (2016), 829–848.

[46] Mishra, S. A machine learning framework for data driven acceleration of computations of differential equations. *arXiv:1807.09519* (2018), 23 pages.

[47] Montanelli, H., and Du, Q. New error bounds for deep networks using sparse grids. *arXiv:1712.08688v3* (2017), 15 pages.

[48] Nabian, M. A., and Meidani, H. A Deep Neural Network Surrogate for High-Dimensional Random Partial Differential Equations. *arXiv:1806.02957* (2018), 23 pages.

[49] Nguyen-Thien, T., and Tran-Cong, T. Approximation of functions and their derivatives: A neural network implementation with applications. *Appl. Math. Model. 23*, 9 (1999), 687–704.

[50] Park, J., and Sandberg, I. W. Universal approximation using radial-basis-function networks. *Neural computation 3*, 2 (1991), 246–257.

[51] Perekrestenko, D., Grohs, P., Elbrächter, D., and Bölcskei, H. The universal approximation power of finite-width deep ReLU networks. *arXiv:1806.01528* (2018), 16 pages.

[52] Petersen, P., Raslan, M., and Voigtlaender, F. Topological properties of the set of functions generated by neural networks of fixed size. *arXiv:1806.08459* (2018), 45 pages.

[53] Petersen, P., and Voigtlaender, F. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *arXiv:1709.05289* (2017), 54 pages.

[54] Pinkus, A. Approximation theory of the MLP model in neural networks. *Acta Numer. 8* (1999), 143–195.

[55] Raissi, M. Forward-Backward Stochastic Neural Networks: Deep Learning of High-dimensional Partial Differential Equations. *arXiv:1804.07010* (2018), 17 pages.

[56] Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., and Beling, P. Deep learning detecting fraud in credit card transactions. In *2018 Systems and Information Engineering Design Symposium (SIEDS)* (2018), pp. 129–134.

[57] Schmitt, M. Lower bounds on the complexity of approximating continuous functions by sigmoidal neural networks. In *Proceedings of the 12th International Conference on Neural Information Processing Systems* (Cambridge, MA, USA, 1999), NIPS'99, MIT Press, pp. 328–334.

[58] Schwab, C., and Zech, J. Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ. *Analysis and Applications* (2018). Published online: August 30, 2018, https://doi.org/10.1142/S0219530518500203.

[59] Shaham, U., Cloninger, A., and Coifman, R. R. Provable approximation properties for deep neural networks. *Appl. Comput. Harmon. Anal. 44*, 3 (2018), 537–557.

[60] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* (2014), 14 pages.

[61] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *arXiv:1708.07469* (2017), 31 pages.

[62] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 1701–1708.

[63] Wang, R., Fu, B., Fu, G., and Wang, M. Deep & cross network for ad click predictions. In *Proceedings of the ADKDD'17* (2017).

[64] Wang, W., Yang, J., Xiao, J., Li, S., and Zhou, D. Face recognition based on deep learning. In *Human Centered Computing* (2015), pp. 812–820.

[65] Wu, C., Karanasou, P., Gales, M. J., and Sim, K. C. Stimulated deep neural network for speech recognition. In *Interspeech 2016* (2016), pp. 400–404.

[66] Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Networks 94* (2017), 103–114.

[67] Yarotsky, D. Universal approximations of invariant maps by neural networks. *arXiv:1804.10306* (2018), 64 pages.

[68] Zhai, S., Chang, K.-h., Zhang, R., and Zhang, Z. M. Deepintent: Learning attentions for online advertising with recurrent neural networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), pp. 1295–1304.