Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Deep neural network approximations for Monte Carlo algorithms

Ph. Grohs and A. Jentzen and D. Salimova

Research Report No. 2019-50
September 2019

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

Deep neural network approximations for Monte Carlo algorithms

Philipp Grohs[1], Arnulf Jentzen[2], and Diyora Salimova[3]

[1]Faculty of Mathematics and Research Platform Data Science, University of Vienna, Austria, e-mail: philipp.grohs@univie.ac.at

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

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

September 3, 2019

Abstract

In the past few years deep artificial neural networks (DNNs) have been successfully employed in a large number of computational problems including, e.g., language processing, image recognition, fraud detection, and computational advertisement. Recently, it has also been proposed in the scientific literature to reformulate partial differential equations (PDEs) as stochastic learning problems and to employ DNNs together with stochastic gradient descent methods to approximate the solutions of such PDEs. There are also a few mathematical convergence results in the scientific literature which show that DNNs can approximate solutions of certain PDEs without the curse of dimensionality in the sense that the number of real parameters employed to describe the DNN grows at most polynomially both in the PDE dimension $d \in \mathbb{N}$ and the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$. One key argument in most of these results is, first, to employ a Monte Carlo approximation scheme which can approximate the solution of the PDE under consideration at a fixed space-time point without the curse of dimensionality and, thereafter, to prove then that DNNs are flexible enough to mimic the behaviour of the employed approximation scheme. Having this in mind, one could aim for a general abstract result which shows under suitable assumptions that if a certain function can be approximated by any kind of (Monte Carlo) approximation scheme without the curse of dimensionality, then the function can also be approximated with DNNs without the curse of dimensionality. It is a key contribution of this article to make a first step towards this direction. In particular, the main result of this paper, roughly speaking, shows that if a function can be approximated by means of some suitable discrete approximation scheme without the curse of dimensionality and if there exist DNNs which satisfy certain regularity properties and which approximate this discrete approximation scheme without the curse of dimensionality, then the function itself can also be approximated with DNNs without the curse of dimensionality. Moreover, for the number of real parameters used to describe such approximating DNNs we provide an explicit upper bound for the optimal exponent of the dimension $d \in \mathbb{N}$ of the function under consideration as well as an explicit lower bound for the optimal exponent of the prescribed approximation accuracy $\varepsilon > 0$. As an application of this result we derive that solutions of suitable Kolmogorov PDEs can be approximated with DNNs without the curse of dimensionality.

Contents

1 Introduction

In the past few years deep artificial neural networks (DNNs) have been successfully employed in a large number of computational problems including, e.g., language processing (cf., e.g., [12, 20, 26, 27, 34, 54]), image recognition (cf., e.g., [28, 35, 48, 50, 53]), fraud detection (cf., e.g., [11, 47]), and computational advertisement (cf., e.g., [52, 55]). Recently, it has also been proposed in [14, 23] to reformulate partial differential equations (PDEs) as stochastic learning problems and to employ DNNs together with stochastic gradient descent methods to approximate the solutions of such PDEs (cf., e.g., also [33, 37, 41, 51]). We refer, e.g., to [1, 2, 3, 4, 5, 8, 10, 13, 15, 17, 18, 19, 24, 25, 29, 31, 38, 39, 40, 44, 45, 49] for further developments and extensions of such deep learning based numerical approximation methods for PDEs. In particular, the references [2,8,15,31,40] deal with linear PDEs (and the stochastic differential equations (SDEs) related to them, respectively), the references [1, 10, 13, 17, 18, 25, 29] deal with semilinear PDEs (and the backward stochastic differential equations (BSDEs) related to them, respectively), the references [3, 38, 44, 45] deal with fully nonlinear PDEs (and the second-order backward stochastic differential equations (2BSDEs) related to them, respectively), the references [24,39,49] deal with certain specific subclasses of fully nonlinear PDEs (and the 2BSDEs related to them, respectively), and the references [4, 5, 19, 49] deal with free boundary PDEs (and the optimal stopping/option pricing problems related to them (see, e.g., [7, Chapter 1]), respectively). In the scientific literature there are also a few rigorous mathematical convergence results for such deep learning based numerical approximation methods for PDEs. For example, the references [24, 49] provide mathematical convergence results for such deep learning based numerical approximation methods for PDEs without any information on the convergence speed and, for instance, the references [9, 16, 21, 22, 30, 32, 36, 46] provide mathematical convergence results of such deep learning based numerical approximation methods for PDEs with dimension-independent convergence rates and error constants which are only polynomially dependent on the dimension. In particular, the latter references show that DNNs can approximate solutions of certain PDEs without the curse of dimensionality (cf. [6]) in the sense that the number of real parameters employed to describe the DNN grows at most polynomially both in the PDE dimension $d \in \mathbb{N}$ and the reciprocal of the prescribed approximation accuracy $\varepsilon > 0$ (cf., e.g., [42, Chapter 1] and [43, Chapter 9]). One key argument in most of these articles is, first, to employ a Monte Carlo approximation scheme which can approximate the solution of the PDE under consideration at a fixed space-time point without the curse of dimensionality and, thereafter, to prove then that DNNs are flexible enough to mimic the behaviour of the employed approximation scheme (cf., e.g., [32, Section 2 and (i)–(iii) in Section 1] and [21]). Having this in mind, one could aim for a general abstract result which shows under suitable assumptions that if a certain function can be approximated by any kind of (Monte Carlo) approximation scheme without the curse of dimensionality, then the function can also be approximated with DNNs without the curse of dimensionality.

It is a key contribution of this article to make a first step towards this direction. In particular, the main result of this paper, Theorem 2.3 below, roughly speaking, shows that if a function can be approximated by means of some suitable discrete approximation scheme without the curse of dimensionality (cf. (2.9) in Theorem 2.3 below) and if there exist DNNs which satisfy certain regularity properties and which approximate this discrete approximation scheme without the curse of dimensionality, then the function itself can also be approximated with DNNs without the curse of dimensionality. Moreover, for the number of real parameters used to describe such approximating DNNs we provide in Theorem 2.3 below an explicit upper bound for the optimal exponent of the dimension $d \in \mathbb{N}$ of the function under consideration as well as an explicit lower bound for the optimal exponent of the prescribed approximation accuracy $\varepsilon > 0$ (see (2.16) in Theorem 2.3 below).

In our applications of Theorem 2.3 we employ Theorem 2.3 to study in Theorem 4.5 below DNN approximations for PDEs. Theorem 4.5 can be considered as a special case of Theorem 2.3 with the function to be approximated to be equal to the solution of a suitable Kolmogorov PDE (cf. (4.42) below) at the final time $T \in (0,\infty)$ and the approximating scheme to be equal to the Monte Carlo Euler scheme. In particular, Theorem 4.5 shows that solutions of suitable Kolmogorov PDEs can be approximated with DNNs without the curse of dimensionality. For the number of real parameters used to describe such approximating DNNs Theorem 4.5 also provides an explicit upper bound for the optimal exponent of the dimension $d \in \mathbb{N}$ of the PDE under consideration as well as an explicit lower bound for the optimal exponent of the prescribed approximation accuracy $\varepsilon > 0$ (see (4.43) below). In order to illustrate the findings of Theorem 4.5 below, we now present in Theorem 1.1 below a special case of Theorem 4.5.

Theorem 1.1. *Let $\varphi_{0,d} \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $\varphi_{1,d} \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be functions, let $\|\cdot\| \colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to [0,\infty)$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\|x\| = (\sum_{i=1}^d |x_i|^2)^{1/2}$, let $A_d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $A_d(x) = (\max\{x_1, 0\}, \max\{x_2, 0\}, \ldots, \max\{x_d, 0\})$, let $\mathbf{N} = \cup_{L\in\mathbb{N}} \cup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, let $P \colon \mathbf{N} \to \mathbb{N}$ and $R \colon \mathbf{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, k \in \mathbb{N} \cap (0, L) \colon x_k = A_{l_k}(W_k x_{k-1} + B_k)$ that $P(\Phi) = \sum_{k=1}^L l_k(l_{k-1} + 1)$, $R(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, and $(R(\Phi))(x_0) = W_L x_{L-1} + B_L$, let $T, \kappa \in (0,\infty)$, $\mathfrak{d}_1 \in [1/2, \infty)$, $\mathfrak{d}_3 \in [4, \infty)$, $\mathfrak{e}, \mathfrak{d}_2, \mathfrak{d}_4, \mathfrak{d}_5, \mathfrak{d}_6 \in [0,\infty)$, $\theta \in [1,\infty)$, $(\phi_\varepsilon^{m,d})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $m \in \{0,1\}$, $x, y \in \mathbb{R}^d$ that $R(\phi_\varepsilon^{0,d}) \in C(\mathbb{R}^d, \mathbb{R})$, $R(\phi_\varepsilon^{1,d}) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $P(\phi_\varepsilon^{m,d}) \leq \kappa d^{2^{(-m)}\mathfrak{d}_3} \varepsilon^{-2^{(-m)}\mathfrak{e}}$, $|(R(\phi_\varepsilon^{0,d}))(x) - (R(\phi_\varepsilon^{0,d}))(y)| \leq \kappa d^{\mathfrak{d}_6}(1 + \|x\|^\theta + \|y\|^\theta)\|x - y\|$, $\|(R(\phi_\varepsilon^{1,d}))(x)\| \leq \kappa(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|x\|)$, $|\varphi_{0,d}(x)| \leq \kappa d^{\mathfrak{d}_6}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta)$, $\|\varphi_{1,d}(x) - \varphi_{1,d}(y)\| \leq \kappa\|x - y\|$, and*

$$\|\varphi_{m,d}(x) - (R(\phi_\varepsilon^{m,d}))(x)\| \leq \varepsilon\kappa d^{\mathfrak{d}_{(5-m)}}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta), \tag{1.1}$$

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

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\tfrac{\partial}{\partial x} u_d)(t,x)\, \varphi_{1,d}(x) + \sum_{i=1}^d (\tfrac{\partial^2}{\partial x_i^2} u_d)(t,x) \tag{1.2}$$

with $u_d(0,x) = \varphi_{0,d}(x)$ for $(t,x) \in (0,T) \times \mathbb{R}^d$. Then for every $p \in (0,\infty)$ there exist $c \in \mathbb{R}$ and $(\Psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $R(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $[\int_{[0,1]^d} |u_d(T,x) - (R(\Psi_{d,\varepsilon}))(x)|^p\, dx]^{1/p} \leq \varepsilon$, and

$$P(\Psi_{d,\varepsilon}) \leq c\varepsilon^{-(\mathfrak{e}+6)} d^{6[\mathfrak{d}_6 + (\mathfrak{d}_1+\mathfrak{d}_2)(\theta+1)] + \mathfrak{d}_3 + \mathfrak{e}\max\{\mathfrak{d}_5 + \theta(\mathfrak{d}_1+\mathfrak{d}_2), \mathfrak{d}_4 + \mathfrak{d}_6 + 2\theta(\mathfrak{d}_1+\mathfrak{d}_2)\}}. \tag{1.3}$$

Theorem 1.1 is an immediate consequence of Corollary 4.6 in Section 4 below. Corollary 4.6, in turn, is a special case of Theorem 4.5. Let us add some comments regarding the mathematical objects appearing in Theorem 1.1. The set $\mathbf{N}$ in Theorem 1.1 above is a set of tuples of pairs of real matrices and real vectors and this set represents the set of all DNNs (see also Definition 3.1 below). The functions $A_d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $d \in \mathbb{N}$, in Theorem 1.1 represent multidimensional rectifier functions. Theorem 1.1 is thus an approximation result for rectified DNNs. Moreover, for every DNN $\Phi \in \mathbf{N}$ in Theorem 1.1 above $P(\Phi) \in \mathbb{N}$ represents the number of real parameters which are used to describe the DNN Φ (see also Definition 3.1 below). In particular, for every DNN $\Phi \in \mathbf{N}$ in Theorem 1.1 one can think of $P(\Phi) \in \mathbb{N}$ as a number proportional to the amount of memory storage needed to store the DNN Φ. Furthermore, the function $R\colon \mathbf{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ from the set $\mathbf{N}$ of "all DNNs" to the union $\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ of continuous functions describes the realization functions associated to the DNNs (see also Definition 3.3 below). The real number $T > 0$ in Theorem 1.1 describes the time horizon under consideration and the real numbers $\kappa, \mathfrak{e}, \theta, \mathfrak{d}_1, \mathfrak{d}_2, \ldots, \mathfrak{d}_6 \in \mathbb{R}$ in Theorem 1.1 are constants used to formulate the assumptions in Theorem 1.1. The key assumption in Theorem 1.1 is the hypothesis that both the possibly nonlinear initial value functions $\varphi_{0,d}\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and the possibly nonlinear drift coefficient functions $\varphi_{1,d}\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, of the PDEs in (1.2) can be approximated by means of DNNs without the curse of dimensionality (see (1.1) above for details). Results related to Theorem 4.5 have been established in [21, Theorem 3.14], [32, Theorem 1.1], [30, Theorem 4.1], and [46, Corollary 2.2]. Theorem 3.14 in [21] proves a similar statement to (1.3) for a different class of PDEs than (1.2), that is, Theorem 3.14 in [21] deals with Black-Scholes PDEs with affine linear coefficient functions while in (1.2) the diffusion coefficient is constant and the drift coefficient may be nonlinear. Theorem 1.1 in [32] shows the existence of constants and exponents of $d \in \mathbb{N}$ and $\varepsilon > 0$ such that (1.3) holds but does not provide any explicit form for these exponents. Theorem 4.1 in [30] studies a different class of PDEs than (1.2) (the diffusion coefficient is chosen so that the second order term is the Laplacian and the drift coefficient is chosen to be zero but there is a nonlinearity depending on the PDE solution in the PDE in Theorem 4.1 in [30]) and provides an explicit exponent for $\varepsilon > 0$ and the existence of constants and exponents of $d \in \mathbb{N}$ such that (1.3) holds. Corollary 2.2 in [46] studies a more general class of Kolmogorov PDEs than (1.2) and shows the existence of constants and exponents of $d \in \mathbb{N}$ and $\varepsilon > 0$ such that (1.3) holds. Theorem 4.5 above extends these results by providing explicit exponents for $d \in \mathbb{N}$ and $\varepsilon > 0$ in terms of the used assumptions such that (1.3) holds and, in addition, Theorem 4.5 can be considered as a special case of the general DNN approximation result in Theorem 2.3 with the functions to be approximated to be equal to the solutions of the PDEs in (1.2) at the final time $T \in (0, \infty)$ and the approximating scheme to be equal to the Monte Carlo Euler scheme.

The remainder of this article is organized as follows. In Section 2 we present Theorem 2.3, which is the main result of this paper. The proof of Theorem 2.3 employs the elementary result in Lemma 2.2. Lemma 2.2 establishes suitable a priori bounds for random variables and follows from the well-known discrete Gronwall-type inequality in Lemma 2.1 below. In Section 3 we develop in Lemma 3.29 and Lemma 3.30 a few elementary results on representation flexibilities of DNNs. The proofs of Lemma 3.29 and Lemma 3.30 use results on a certain artificial neural network (ANN) calculus which we recall and extend in Subsections 3.1–3.7. In Section 4 in Theorem 4.5 we employ Lemma 3.29 and Lemma 3.30 to establish the existence of DNNs which approximate solutions of suitable Kolmogorov PDEs without the curse of dimensionality. In our proof of Theorem 4.5 we also employ error estimates for the Monte Carlo Euler method which we present in Proposition 4.4 in Section 4. The proof of Proposition 4.4, in turn, makes use of the elementary error estimate results in Lemmas 4.1–4.3 below.

2 Deep artificial neural network (DNN) approximations

In this section we show in Theorem 2.3 below that, roughly speaking, if a function can be approximated by means of some suitable discrete approximation scheme without the curse of dimensionality and if there exist DNNs which satisfy certain regularity properties and which approximate this discrete approximation scheme without the curse of dimensionality, then the function itself can also be approximated with DNNs without the curse of dimensionality. In our proof of Theorem 2.3 we employ the elementary a priori estimates for expectations of certain random variables in Lemma 2.2 below. Lemma 2.2, in turn, follows from the well-known discrete Gronwall-type inequality in Lemma 2.1 below.

2.1 A priori bounds for random variables

Lemma 2.1. *Let $\alpha \in [0, \infty)$, $\beta \in [0, \infty]$ and let $x \colon \mathbb{N}_0 \to \mathbb{R}$ satisfy for all $n \in \mathbb{N}$ that $x_n \leq \alpha x_{n-1} + \beta$. Then it holds for all $n \in \mathbb{N}$ that*

$$x_n \leq \alpha^n x_0 + \beta(1 + \alpha + \ldots + \alpha^{n-1}) \leq \alpha^n x_0 + \beta e^{\alpha}. \tag{2.1}$$

Proof of Lemma 2.1. We prove (2.1) by induction on $n \in \mathbb{N}$. For the base case $n = 1$ note that the hypothesis that $\forall\, k \in \mathbb{N} \colon x_k \leq \alpha x_{k-1} + \beta$ ensures that

$$x_1 \leq \alpha x_0 + \beta = \alpha^1 x_0 + \beta \leq \alpha^1 x_0 + \beta e^{\alpha}. \tag{2.2}$$

This establishes (2.1) in the base case $n = 1$. For the induction step $\mathbb{N} \ni (n-1) \to n \in \mathbb{N} \cap [2, \infty)$ observe that the hypothesis that $\forall\, k \in \mathbb{N} \colon x_k \leq \alpha x_{k-1} + \beta$ implies that for all $n \in \mathbb{N} \cap [2, \infty)$ with $x_{n-1} \leq \alpha^{n-1} x_0 + \beta(1 + \alpha + \ldots + \alpha^{n-2})$ it holds that

$$\begin{split} x_n &\leq \alpha x_{n-1} + \beta \leq \alpha^n x_0 + \alpha\beta(1 + \alpha + \ldots + \alpha^{n-2}) + \beta \\ &= \alpha^n x_0 + \beta(1 + \alpha + \ldots + \alpha^{n-1}) \leq \alpha^n x_0 + \beta e^{\alpha}. \end{split} \tag{2.3}$$

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

Lemma 2.2. *Let $N \in \mathbb{N}$, $p \in [1, \infty)$, $\alpha, \beta, \gamma \in [0, \infty)$ and let $X_n \colon \Omega \to \mathbb{R}$, $n \in \{0, 1, \ldots, N\}$, and $Z_n \colon \Omega \to \mathbb{R}$, $n \in \{0, 1, \ldots, N-1\}$, be random variables which satisfy for all $n \in \{1, 2, \ldots, N\}$ that*

$$|X_n| \leq \alpha |X_{n-1}| + \beta\big[\gamma + |Z_{n-1}|\big]. \tag{2.4}$$

Then it holds that

$$\big(\mathbb{E}[|X_N|^p]\big)^{1/p} \leq \alpha^N \big(\mathbb{E}[|X_0|^p]\big)^{1/p} + e^{\alpha}\beta\Big[\gamma + \sup\nolimits_{i \in \{0,1,\ldots,N-1\}} \big(\mathbb{E}[|Z_i|^p]\big)^{1/p}\Big]. \tag{2.5}$$

Proof of Lemma 2.2. First, note that (2.4) implies for all $n \in \{1, 2, \ldots, N\}$ that

$$\begin{aligned}(\mathbb{E}[|X_n|^p])^{1/p} &\leq \alpha(\mathbb{E}[|X_{n-1}|^p])^{1/p} + \beta\Big[\gamma + (\mathbb{E}[|Z_{n-1}|^p])^{1/p}\Big] \\ &\leq \alpha(\mathbb{E}[|X_{n-1}|^p])^{1/p} + \beta\Big[\gamma + \sup\nolimits_{i\in\{0,1,\ldots,N-1\}} (\mathbb{E}[|Z_i|^p])^{1/p}\Big].\end{aligned} \tag{2.6}$$

Lemma 2.1 (with $\alpha = \alpha$, $\beta = \beta\,[\gamma + \sup_{i\in\{0,1,\ldots,N-1\}}(\mathbb{E}[|Z_i|^p])^{1/p}]$ in the notation of Lemma 2.1) hence establishes for all $n \in \{1, 2, \ldots, N\}$ that

$$(\mathbb{E}[|X_n|^p])^{1/p} \leq \alpha^n(\mathbb{E}[|X_0|^p])^{1/p} + e^\alpha\beta\Big[\gamma + \sup\nolimits_{i\in\{0,1,\ldots,N-1\}} (\mathbb{E}[|Z_i|^p])^{1/p}\Big]. \tag{2.7}$$

The proof of Lemma 2.2 is thus completed. □

2.2 A DNN approximation result for Monte Carlo algorithms

Theorem 2.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathfrak{n}_0 \in (0, \infty)$, $\mathfrak{n}_1, \mathfrak{n}_2, \mathfrak{e}, \mathfrak{d}_0, \mathfrak{d}_1, \ldots, \mathfrak{d}_6 \in [0, \infty)$, $\mathfrak{C}, p, \theta \in [1, \infty)$, $(M_N)_{N\in\mathbb{N}} \subseteq \mathbb{N}$, let $Z^{N,d,m}_n \colon \Omega \to \mathbb{R}^d$, $n \in \{0, 1, \ldots, N-1\}$, $m \in \{1, 2, \ldots, M_N\}$, $d, N \in \mathbb{N}$, be random variables, let $f_{N,d} \in C(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)$, $d, N \in \mathbb{N}$, and $Y^{N,d,x}_n = (Y^{N,d,m,x}_n)_{m\in\{1,2,\ldots,M_N\}} \colon \Omega \to \mathbb{R}^{M_N d}$, $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $d, N \in \mathbb{N}$, satisfy for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $n \in \{1, 2, \ldots, N\}$, $\omega \in \Omega$ that $Y^{N,d,m,x}_0(\omega) = x$ and*

$$Y^{N,d,m,x}_n(\omega) = f_{N,d}\big(Z^{N,d,m}_{n-1}(\omega), Y^{N,d,m,x}_{n-1}(\omega)\big), \tag{2.8}$$

let $\|\cdot\| : (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to [0, \infty)$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\|x\| = (\sum_{i=1}^d |x_i|^2)^{1/2}$, for every $d \in \mathbb{N}$ let $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be a probability measure on $\mathbb{R}^d$, let $g_{N,d} \in C(\mathbb{R}^{Nd}, \mathbb{R})$, $d, N \in \mathbb{N}$, and $u_d \in C(\mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, satisfy for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $n \in \{0, 1, \ldots, N-1\}$ that

$$\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|u_d(x) - g_{M_N,d}(Y^{N,d,x}_N)\big|^p \,\nu_d(dx)\right]\right)^{1/p} \leq \mathfrak{C}d^{\mathfrak{d}_0} N^{-\mathfrak{n}_0}, \tag{2.9}$$

$$\big(\mathbb{E}\big[\|Z^{N,d,m}_n\|^{2p\theta}\big]\big)^{1/(2p\theta)} \leq \mathfrak{C}d^{\mathfrak{d}_1}, \quad \text{and} \quad \left[\int_{\mathbb{R}^d} \|x\|^{2p\theta} \,\nu_d(dx)\right]^{1/(2p\theta)} \leq \mathfrak{C}d^{\mathfrak{d}_1+\mathfrak{d}_2}, \tag{2.10}$$

let $\mathbf{N}$ be a set, let $\mathcal{P} \colon \mathbf{N} \to \mathbb{N}$, $\mathcal{D} \colon \mathbf{N} \to \cup_{L=2}^\infty \mathbb{N}^L$, and $\mathcal{R} \colon \mathbf{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ be functions, let $\mathfrak{N}_{d,\varepsilon} \subseteq \mathbf{N}$, $\varepsilon \in (0, 1]$, $d \in \mathbb{N}$, let $(\mathbf{f}^{N,d}_{\varepsilon,z})_{(N,d,\varepsilon,z)\in\mathbb{N}^2\times(0,1]\times\mathbb{R}^d} \subseteq \mathbf{N}$, $(\mathbf{g}^{N,d}_\varepsilon)_{(N,d,\varepsilon)\in\mathbb{N}^2\times(0,1]} \subseteq \mathbf{N}$, $(\mathfrak{I}_d)_{d\in\mathbb{N}} \subseteq \mathbf{N}$, assume for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, y, z \in \mathbb{R}^d$ that $\mathfrak{N}_{d,\varepsilon} \subseteq \{\Phi \in \mathbf{N} \colon \mathcal{R}(\Phi) \in C(\mathbb{R}^d, \mathbb{R}^d)\}$, $\mathfrak{I}_d \in \mathfrak{N}_{d,\varepsilon}$, $(\mathcal{R}(\mathfrak{I}_d))(x) = x$, $\mathcal{P}(\mathfrak{I}_d) \leq \mathfrak{C}d^{\mathfrak{d}_3}$, $\mathcal{R}(\mathbf{f}^{N,d}_{\varepsilon,z}) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $(\mathbb{R}^d \ni \mathfrak{z} \mapsto (\mathcal{R}(\mathbf{f}^{N,d}_{\varepsilon,\mathfrak{z}}))(x) \in \mathbb{R}^d)$ is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable, and

$$\|f_{N,d}(z, x) - (\mathcal{R}(\mathbf{f}^{N,d}_{\varepsilon,z}))(x)\| \leq \varepsilon\mathfrak{C}d^{\mathfrak{d}_4}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta), \tag{2.11}$$

$$\|(\mathcal{R}(\mathbf{f}^{N,d}_{\varepsilon,z}))(x)\| \leq \big(1 + \tfrac{\mathfrak{e}}{N}\big)\|x\| + \mathfrak{C}d^{\mathfrak{d}_2}(d^{\mathfrak{d}_1} + \|z\|), \tag{2.12}$$

$$\|f_{N,d}(z, x) - f_{N,d}(z, y)\| \leq \mathfrak{C}\|x - y\|, \tag{2.13}$$

assume for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi \in \mathfrak{N}_{d,\varepsilon}$ *that there exist* $(\phi_z)_{z\in\mathbb{R}^d} \subseteq \mathfrak{N}_{d,\varepsilon}$ *such that for all* $x, z, \mathfrak{z} \in \mathbb{R}^d$ *it holds that* $(\mathcal{R}(\phi_z))(x) = (\mathcal{R}(\mathbf{f}^{N,d}_{\varepsilon,z}))((\mathcal{R}(\Phi))(x))$, $\mathcal{P}(\phi_z) \leq \mathcal{P}(\Phi) + \mathfrak{C}N^{\mathfrak{n}_1} d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{e}}$, *and* $\mathcal{D}(\phi_z) = \mathcal{D}(\phi_{\mathfrak{z}})$, *assume for all* $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x = (x_i)_{i\in\{1,2,\dots,N\}} \in \mathbb{R}^{Nd}$, $y = (y_i)_{i\in\{1,2,\dots,N\}} \in \mathbb{R}^{Nd}$ *that* $\mathcal{R}(\mathbf{g}^{N,d}_\varepsilon) \in C(\mathbb{R}^{Nd}, \mathbb{R})$ *and*

$$|g_{N,d}(x) - (\mathcal{R}(\mathbf{g}^{N,d}_\varepsilon))(x)| \leq \varepsilon \mathfrak{C} d^{\mathfrak{d}_5} \left[d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \tfrac{1}{N} \sum_{i=1}^N \|x_i\|^\theta \right], \tag{2.14}$$

$$|(\mathcal{R}(\mathbf{g}^{N,d}_\varepsilon))(x) - (\mathcal{R}(\mathbf{g}^{N,d}_\varepsilon))(y)| \leq \frac{\mathfrak{C}d^{\mathfrak{d}_6}}{N} \left[\sum_{i=1}^N (d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x_i\|^\theta + \|y_i\|^\theta) \|x_i - y_i\| \right], \tag{2.15}$$

and assume for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi_1, \Phi_2, \dots, \Phi_{M_N} \in \mathfrak{N}_{d,\varepsilon}$ *with* $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_{M_N})$ *that there exists* $\varphi \in \mathbf{N}$ *such that for all* $x \in \mathbb{R}^d$ *it holds that* $\mathcal{R}(\varphi) \in C(\mathbb{R}^d, \mathbb{R})$, $(\mathcal{R}(\varphi))(x) = (\mathcal{R}(\mathbf{g}^{M_N,d}_\varepsilon))((\mathcal{R}(\Phi_1))(x), (\mathcal{R}(\Phi_2))(x), \dots, (\mathcal{R}(\Phi_{M_N}))(x))$, *and* $\mathcal{P}(\varphi) \leq \mathfrak{C}N^{\mathfrak{n}_2}(N^{\mathfrak{n}_1+1} d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{e}} + \mathcal{P}(\Phi_1))$. *Then there exist* $c \in \mathbb{R}$ *and* $(\Psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ *it holds that* $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $[\int_{\mathbb{R}^d} |u_d(x) - (\mathcal{R}(\Psi_{d,\varepsilon}))(x)|^p \, \nu_d(dx)]^{1/p} \leq \varepsilon$, *and*

$$\mathcal{P}(\Psi_{d,\varepsilon}) \leq c d^{\frac{\mathfrak{d}_0(\mathfrak{n}_1+\mathfrak{n}_2+1)}{\mathfrak{n}_0} + \mathfrak{d}_3 + \mathfrak{e}\max\{\mathfrak{d}_5+\theta(\mathfrak{d}_1+\mathfrak{d}_2), \mathfrak{d}_4+\mathfrak{d}_6+2\theta(\mathfrak{d}_1+\mathfrak{d}_2)\}} \varepsilon^{-\frac{(\mathfrak{n}_1+\mathfrak{n}_2+1)}{\mathfrak{n}_0} - \mathfrak{e}}. \tag{2.16}$$

Proof of Theorem 2.3. Throughout this proof let $\gamma = 46e^{\mathfrak{C}}\mathfrak{C}^2(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{2\theta}$, let $\delta = \max\{\mathfrak{d}_5+\theta(\mathfrak{d}_1+\mathfrak{d}_2), \mathfrak{d}_4+\mathfrak{d}_6+2\theta(\mathfrak{d}_1+\mathfrak{d}_2)\}$, let $X^{N,d,x,\varepsilon}_n = (X^{N,d,m,x,\varepsilon}_n)_{m\in\{1,2,\dots,M_N\}} \colon \Omega \to \mathbb{R}^{M_N d}$, $n \in \{0,1,\dots,N\}$, $\varepsilon \in (0,1]$, $x \in \mathbb{R}^d$, $d, N \in \mathbb{N}$, be the random variables which satisfy for all $N, d \in \mathbb{N}$, $m \in \{1,2,\dots,M_N\}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $n \in \{1,2,\dots,N\}$, $\omega \in \Omega$ that $X^{N,d,m,x,\varepsilon}_0(\omega) = x$ and

$$X^{N,d,m,x,\varepsilon}_n(\omega) = \Big(\mathcal{R}\Big(\mathbf{f}^{N,d}_{\varepsilon, Z^{N,d,m}_{n-1}(\omega)}\Big)\Big)\big(X^{N,d,m,x,\varepsilon}_{n-1}(\omega)\big), \tag{2.17}$$

and let $(\mathcal{N}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbb{N}$ and $(\mathcal{E}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq (0,1]$ satisfy for all $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ that

$$\mathcal{N}_{d,\varepsilon} = \min\Big(\mathbb{N} \cap \big[\big(\tfrac{2\mathfrak{C}d^{\mathfrak{d}_0}}{\varepsilon}\big)^{1/\mathfrak{n}_0}, \infty\big)\Big) \qquad \text{and} \qquad \mathcal{E}_{d,\varepsilon} = \tfrac{\varepsilon}{\gamma d^\delta}. \tag{2.18}$$

Note that for all $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $n \in \{0,1,2,\dots,N\}$ it holds that

$$\big(\mathbb{R}^d \ni x \mapsto X^{N,d,x,\varepsilon}_n \in \mathbb{R}^{M_N d}\big) \in C(\mathbb{R}^d, \mathbb{R}^{M_N d}). \tag{2.19}$$

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

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|u_d(x) - (\mathcal{R}(\mathbf{g}^{M_N,d}_\varepsilon))(X^{N,d,x,\varepsilon}_N)\big|^p \, \nu_d(dx)\right]\right)^{1/p} \\
&\leq \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|u_d(x) - g_{M_N,d}(Y^{N,d,x}_N)\big|^p \, \nu_d(dx)\right]\right)^{1/p} \\
&+ \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|g_{M_N,d}(Y^{N,d,x}_N) - (\mathcal{R}(\mathbf{g}^{M_N,d}_\varepsilon))(Y^{N,d,x}_N)\big|^p \, \nu_d(dx)\right]\right)^{1/p} \\
&+ \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|(\mathcal{R}(\mathbf{g}^{M_N,d}_\varepsilon))(Y^{N,d,x}_N) - (\mathcal{R}(\mathbf{g}^{M_N,d}_\varepsilon))(X^{N,d,x,\varepsilon}_N)\big|^p \, \nu_d(dx)\right]\right)^{1/p}.
\end{aligned} \tag{2.20}$$

Next observe that (2.14) ensures for all $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|g_{M_N,d}(Y_N^{N,d,x}) - (\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(Y_N^{N,d,x})\right|^p \nu_d(dx)\right]\right)^{1/p} \\
&\le \varepsilon \mathfrak{C} d^{\mathfrak{d}_5} \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \tfrac{1}{M_N}\sum_{m=1}^{M_N} \|Y_N^{N,d,m,x}\|^\theta\right|^p \nu_d(dx)\right]\right)^{1/p} \\
&\le \varepsilon \mathfrak{C} d^{\mathfrak{d}_5} \left[d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \tfrac{1}{M_N}\sum_{m=1}^{M_N} \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \|Y_N^{N,d,m,x}\|^{p\theta} \nu_d(dx)\right]\right)^{1/p}\right].
\end{aligned} \tag{2.21}$$

In addition, note that (2.11) and (2.12) assure that for all $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x, z \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\|f_{N,d}(z,x)\| &\le \|f_{N,d}(z,x) - (\mathcal{R}(\mathbf{f}_{\varepsilon,z}^{N,d}))(x)\| + \|(\mathcal{R}(\mathbf{f}_{\varepsilon,z}^{N,d}))(x)\| \\
&\le \varepsilon \mathfrak{C} d^{\mathfrak{d}_4}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta) + \left(1 + \tfrac{\mathfrak{C}}{N}\right)\|x\| + \mathfrak{C} d^{\mathfrak{d}_2}(d^{\mathfrak{d}_1} + \|z\|).
\end{aligned} \tag{2.22}$$

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

$$\|f_{N,d}(z,x)\| \le \left(1 + \tfrac{\mathfrak{C}}{N}\right)\|x\| + \mathfrak{C} d^{\mathfrak{d}_2}(d^{\mathfrak{d}_1} + \|z\|). \tag{2.23}$$

Hence, we obtain that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $n \in \{1, 2, \ldots, N\}$ it holds that

$$\begin{aligned}
\|Y_n^{N,d,m,x}\| &= \left\|f_{N,d}\big(Z_{n-1}^{N,d,m}, Y_{n-1}^{N,d,m,x}\big)\right\| \\
&\le \left(1 + \tfrac{\mathfrak{C}}{N}\right)\|Y_{n-1}^{N,d,m,x}\| + \mathfrak{C} d^{\mathfrak{d}_2}\big[d^{\mathfrak{d}_1} + \|Z_{n-1}^{N,d,m}\|\big].
\end{aligned} \tag{2.24}$$

Moreover, note that (2.12) assures that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $n \in \{1, 2, \ldots, N\}$ it holds that

$$\begin{aligned}
\|X_n^{N,d,m,x,\varepsilon}\| &= \left\|\Big(\mathcal{R}\Big(\mathbf{f}^{N,d}_{\varepsilon, Z_{n-1}^{N,d,m}}\Big)\Big)\big(X_{n-1}^{N,d,m,x,\varepsilon}\big)\right\| \\
&\le \left(1 + \tfrac{\mathfrak{C}}{N}\right)\|X_{n-1}^{N,d,m,x,\varepsilon}\| + \mathfrak{C} d^{\mathfrak{d}_2}\big[d^{\mathfrak{d}_1} + \|Z_{n-1}^{N,d,m}\|\big].
\end{aligned} \tag{2.25}$$

Lemma 2.2 (with $N = n$, $p = 2p\theta$, $\alpha = (1 + \frac{\mathfrak{C}}{N})$, $\beta = \mathfrak{C} d^{\mathfrak{d}_2}$, $\gamma = d^{\mathfrak{d}_1}$, $Z_i = \|Z_i^{N,d,m}\|$ for $N, d \in \mathbb{N}$, $n \in \{1, 2, \ldots, N\}$, $m \in \{1, 2, \ldots, M_N\}$, $i \in \{0, 1, \ldots, n-1\}$ in the notation of Lemma 2.2), (2.24), and (2.10) therefore demonstrate that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $n \in \{1, 2, \ldots, N\}$ it holds that

$$\begin{aligned}
&\max\left\{\big(\mathbb{E}\big[\|Y_n^{N,d,m,x}\|^{2p\theta}\big]\big)^{1/(2p\theta)}, \big(\mathbb{E}\big[\|X_n^{N,d,m,x,\varepsilon}\|^{2p\theta}\big]\big)^{1/(2p\theta)}\right\} \\
&\le \left(1 + \tfrac{\mathfrak{C}}{N}\right)^n\|x\| + e^{(1+\frac{\mathfrak{C}}{N})}\mathfrak{C} d^{\mathfrak{d}_2}\left[d^{\mathfrak{d}_1} + \sup\nolimits_{i\in\{0,1,\ldots,n-1\}} \left(\mathbb{E}\left[\|Z_i^{N,d,m}\|^{2p\theta}\right]\right)^{1/(2p\theta)}\right] \\
&\le e^{\mathfrak{C}}\|x\| + e^{\mathfrak{C}+1}\mathfrak{C} d^{\mathfrak{d}_2}\big[d^{\mathfrak{d}_1} + \mathfrak{C} d^{\mathfrak{d}_1}\big] \le e^{\mathfrak{C}}\|x\| + 2e^{\mathfrak{C}+1}\mathfrak{C}^2 d^{\mathfrak{d}_1+\mathfrak{d}_2} \\
&\le 2e^{\mathfrak{C}+1}\mathfrak{C}^2\big[\|x\| + d^{\mathfrak{d}_1+\mathfrak{d}_2}\big].
\end{aligned} \tag{2.26}$$

This and the fact that $\forall\, a,b\in\mathbb{R}\colon |a+b|^\theta\le 2^{\theta-1}(|a|^\theta+|b|^\theta)$ prove for all $N,d\in\mathbb{N}$, $m\in\{1,2,\ldots,M_N\}$, $x\in\mathbb{R}^d$, $\varepsilon\in(0,1]$, $n\in\{1,2,\ldots,N\}$ that

$$\begin{aligned}
&\max\bigl\{\mathbb{E}\bigl[\|Y_n^{N,d,m,x}\|^{2p\theta}\bigr],\mathbb{E}\bigl[\|X_n^{N,d,m,x,\varepsilon}\|^{2p\theta}\bigr]\bigr\}\\
&\le \bigl(2e^{\mathfrak{C}+1}\mathfrak{C}^2\bigl[\|x\|+d^{\mathfrak{d}_1+\mathfrak{d}_2}\bigr]\bigr)^{2p\theta}=\bigl(2e^{\mathfrak{C}+1}\mathfrak{C}^2\bigr)^{2p\theta}\bigl[\|x\|+d^{\mathfrak{d}_1+\mathfrak{d}_2}\bigr]^{2p\theta}\\
&\le 2^{2p(\theta-1)}\bigl(2e^{\mathfrak{C}+1}\mathfrak{C}^2\bigr)^{2p\theta}\bigl[\|x\|^\theta+d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\bigr]^{2p}\\
&\le (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^{2p\theta}\bigl[\|x\|^\theta+d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\bigr]^{2p}.
\end{aligned}\tag{2.27}$$

This and (2.10) establish that for all $N,d\in\mathbb{N}$, $m\in\{1,2,\ldots,M_N\}$, $\varepsilon\in(0,1]$ it holds that

$$\begin{aligned}
&\max\left\{\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\|Y_N^{N,d,m,x}\|^{2p\theta}\,\nu_d(dx)\right]\right)^{1/(2p)},\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\|X_N^{N,d,m,x,\varepsilon}\|^{2p\theta}\,\nu_d(dx)\right]\right)^{1/(2p)}\right\}\\
&\le (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^\theta\left(\int_{\mathbb{R}^d}\bigl[\|x\|^\theta+d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\bigr]^{2p}\,\nu_d(dx)\right)^{1/(2p)}\\
&\le (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^\theta\left[\left(\int_{\mathbb{R}^d}\|x\|^{2p\theta}\,\nu_d(dx)\right)^{1/(2p)}+d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\right]\\
&\le (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^\theta\bigl[\mathfrak{C}^\theta d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}+d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\bigr]\le 2(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^\theta d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}.
\end{aligned}\tag{2.28}$$

Hence, we obtain that for all $N,d\in\mathbb{N}$, $m\in\{1,2,\ldots,M_N\}$ it holds that

$$\begin{aligned}
\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\|Y_N^{N,d,m,x}\|^{p\theta}\,\nu_d(dx)\right]\right)^{1/p}&\le\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\|Y_N^{N,d,m,x}\|^{2p\theta}\,\nu_d(dx)\right]\right)^{1/(2p)}\\
&\le 2(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^\theta d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}.
\end{aligned}\tag{2.29}$$

Combining this and (2.21) demonstrates that for all $N,d\in\mathbb{N}$, $\varepsilon\in(0,1]$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\bigl|g_{M_N,d}(Y_N^{N,d,x})-(\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(Y_N^{N,d,x})\bigr|^p\,\nu_d(dx)\right]\right)^{1/p}\\
&\le\varepsilon\mathfrak{C}d^{\mathfrak{d}_5}\bigl[d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}+2(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^\theta d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\bigr]\le 3\varepsilon\mathfrak{C}(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^\theta d^{\mathfrak{d}_5+\theta(\mathfrak{d}_1+\mathfrak{d}_2)}.
\end{aligned}\tag{2.30}$$

In addition, observe that (2.15) ensures that for all $N,d\in\mathbb{N}$, $x\in\mathbb{R}^d$, $\varepsilon\in(0,1]$ it holds that

$$\begin{aligned}
&\bigl|(\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(Y_N^{N,d,x})-(\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(X_N^{N,d,x,\varepsilon})\bigr|\\
&\le\frac{\mathfrak{C}d^{\mathfrak{d}_6}}{M_N}\left[\sum_{m=1}^{M_N}\bigl(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}+\|Y_N^{N,d,m,x}\|^\theta+\|X_N^{N,d,m,x,\varepsilon}\|^\theta\bigr)\|Y_N^{N,d,m,x}-X_N^{N,d,m,x,\varepsilon}\|\right].
\end{aligned}\tag{2.31}$$

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

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\bigl|(\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(Y_N^{N,d,x})-(\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(X_N^{N,d,x,\varepsilon})\bigr|^p\,\nu_d(dx)\right]\right)^{1/p}\\
&\le\frac{\mathfrak{C}d^{\mathfrak{d}_6}}{M_N}\sum_{m=1}^{M_N}\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\bigl(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}+\|Y_N^{N,d,m,x}\|^\theta+\|X_N^{N,d,m,x,\varepsilon}\|^\theta\bigr)^p\right.\right.\\
&\qquad\qquad\left.\left.\cdot\,\|Y_N^{N,d,m,x}-X_N^{N,d,m,x,\varepsilon}\|^p\,\nu_d(dx)\right]\right)^{1/p}.
\end{aligned}\tag{2.32}$$

Hölder's inequality hence assures for all $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|(\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(Y_N^{N,d,x}) - (\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(X_N^{N,d,x,\varepsilon})\big|^p \, \nu_d(dx)\right]\right)^{1/p} \\
&\leq \frac{\mathfrak{C} d^{\mathfrak{d}_6}}{M_N} \sum_{m=1}^{M_N} \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|Y_N^{N,d,m,x}\|^\theta + \|X_N^{N,d,m,x,\varepsilon}\|^\theta\big)^{2p} \, \nu_d(dx)\right]\right)^{1/(2p)} \\
&\qquad \cdot \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \|Y_N^{N,d,m,x} - X_N^{N,d,m,x,\varepsilon}\|^{2p} \, \nu_d(dx)\right]\right)^{1/(2p)}.
\end{aligned} \tag{2.33}$$

Moreover, note that (2.28) implies that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|Y_N^{N,d,m,x}\|^\theta + \|X_N^{N,d,m,x,\varepsilon}\|^\theta\big)^{2p} \, \nu_d(dx)\right]\right)^{1/(2p)} \\
&\leq d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \|Y_N^{N,d,m,x}\|^{2p\theta} \, \nu_d(dx)\right]\right)^{1/(2p)} \\
&+ \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \|X_N^{N,d,m,x,\varepsilon}\|^{2p\theta} \, \nu_d(dx)\right]\right)^{1/(2p)} \\
&\leq d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + 4(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^\theta d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} \leq 5(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^\theta d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}.
\end{aligned} \tag{2.34}$$

Next observe that (2.13) and (2.11) prove that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $n \in \{1, 2, \ldots, N\}$ it holds that

$$\begin{aligned}
&\|Y_n^{N,d,m,x} - X_n^{N,d,m,x,\varepsilon}\| \\
&= \left\| f_{N,d}\big(Z_{n-1}^{N,d,m}, Y_{n-1}^{N,d,m,x}\big) - \Big(\mathcal{R}\Big(\mathbf{f}^{N,d}_{\varepsilon, Z_{n-1}^{N,d,m}}\Big)\Big)\big(X_{n-1}^{N,d,m,x,\varepsilon}\big) \right\| \\
&\leq \left\| f_{N,d}\big(Z_{n-1}^{N,d,m}, Y_{n-1}^{N,d,m,x}\big) - f_{N,d}\big(Z_{n-1}^{N,d,m}, X_{n-1}^{N,d,m,x,\varepsilon}\big) \right\| \\
&+ \left\| f_{N,d}\big(Z_{n-1}^{N,d,m}, X_{n-1}^{N,d,m,x,\varepsilon}\big) - \Big(\mathcal{R}\Big(\mathbf{f}^{N,d}_{\varepsilon, Z_{n-1}^{N,d,m}}\Big)\Big)\big(X_{n-1}^{N,d,m,x,\varepsilon}\big) \right\| \\
&\leq \mathfrak{C}\|Y_{n-1}^{N,d,m,x} - X_{n-1}^{N,d,m,x,\varepsilon}\| + \varepsilon \mathfrak{C} d^{\mathfrak{d}_4} \Big(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|X_{n-1}^{N,d,m,x,\varepsilon}\|^\theta\Big).
\end{aligned} \tag{2.35}$$

Lemma 2.2 (with $N = N$, $p = 2p$, $\alpha = \mathfrak{C}$, $\beta = \varepsilon \mathfrak{C} d^{\mathfrak{d}_4}$, $\gamma = d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}$, $Z_n = \|X_n^{N,d,m,x,\varepsilon}\|^\theta$, $X_n = \|Y_n^{N,d,m,x} - X_n^{N,d,m,x,\varepsilon}\|$ for $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $n \in \{1, 2, \ldots, N\}$ in the notation of Lemma 2.2) and (2.27) hence ensure for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ that

$$\begin{aligned}
&\Big(\mathbb{E}\big[\|Y_N^{N,d,m,x} - X_N^{N,d,m,x,\varepsilon}\|^{2p}\big]\Big)^{1/(2p)} \\
&\leq e^{\mathfrak{C}} \varepsilon \mathfrak{C} d^{\mathfrak{d}_4} \left[d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \sup\nolimits_{i \in \{0,1,\ldots,N-1\}} \Big(\mathbb{E}\big[\|X_i^{N,d,m,x,\varepsilon}\|^{2p\theta}\big]\Big)^{1/(2p)}\right] \\
&\leq \varepsilon e^{\mathfrak{C}} \mathfrak{C} d^{\mathfrak{d}_4} \left[d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \Big((4e^{\mathfrak{C}+1}\mathfrak{C}^2)^{2p\theta}\big[\|x\|^\theta + d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\big]^{2p}\Big)^{1/(2p)}\right] \\
&= \varepsilon e^{\mathfrak{C}} \mathfrak{C} d^{\mathfrak{d}_4} \big[d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^\theta\big[\|x\|^\theta + d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\big]\big] \\
&\leq 2\varepsilon e^{\mathfrak{C}} \mathfrak{C} d^{\mathfrak{d}_4} (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^\theta \big[\|x\|^\theta + d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\big].
\end{aligned} \tag{2.36}$$

This and (2.10) demonstrate that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $\varepsilon \in (0, 1]$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \|Y_N^{N,d,m,x} - X_N^{N,d,m,x,\varepsilon}\|^{2p}\, \nu_d(dx)\right]\right)^{1/(2p)} \\
&\leq 2\varepsilon e^{\mathfrak{C}} \mathfrak{C} d^{\mathfrak{d}_4} (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^{\theta} \left[\int_{\mathbb{R}^d} \big(\|x\|^{\theta} + d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\big)^{2p}\, \nu_d(dx)\right]^{1/(2p)} \\
&\leq 2\varepsilon e^{\mathfrak{C}} \mathfrak{C} d^{\mathfrak{d}_4} (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^{\theta} \left(\left[\int_{\mathbb{R}^d} \|x\|^{2p\theta}\, \nu_d(dx)\right]^{1/(2p)} + d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\right) \\
&\leq 2\varepsilon e^{\mathfrak{C}} \mathfrak{C} d^{\mathfrak{d}_4} (4e^{\mathfrak{C}+1}\mathfrak{C}^2)^{\theta} \big(\mathfrak{C}^{\theta} d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}\big) \\
&\leq 4\varepsilon e^{\mathfrak{C}} \mathfrak{C} (4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{\theta} d^{\mathfrak{d}_4+\theta(\mathfrak{d}_1+\mathfrak{d}_2)}.
\end{aligned} \tag{2.37}$$

Combining this with (2.33) and (2.34) establishes that for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|(\mathcal{R}(\mathbf{g}_{\varepsilon}^{M_N,d}))(Y_N^{N,d,x}) - (\mathcal{R}(\mathbf{g}_{\varepsilon}^{M_N,d}))(X_N^{N,d,x,\varepsilon})\big|^p\, \nu_d(dx)\right]\right)^{1/p} \\
&\leq \mathfrak{C} d^{\mathfrak{d}_6} \cdot 5(4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{\theta} d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} \cdot 4\varepsilon e^{\mathfrak{C}} \mathfrak{C} (4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{\theta} d^{\mathfrak{d}_4+\theta(\mathfrak{d}_1+\mathfrak{d}_2)} \\
&\leq 20\varepsilon e^{\mathfrak{C}} \mathfrak{C}^2 (4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{2\theta} d^{\mathfrak{d}_4+\mathfrak{d}_6+2\theta(\mathfrak{d}_1+\mathfrak{d}_2)}.
\end{aligned} \tag{2.38}$$

This, (2.9), (2.20), and (2.30) prove for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \big|u_d(x) - (\mathcal{R}(\mathbf{g}_{\varepsilon}^{M_N,d}))(X_N^{N,d,x,\varepsilon})\big|^p\, \nu_d(dx)\right]\right)^{1/p} \\
&\leq \mathfrak{C} d^{\mathfrak{d}_0} N^{-\mathfrak{n}_0} + 3\varepsilon \mathfrak{C} (4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{\theta} d^{\mathfrak{d}_5+\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + 20\varepsilon e^{\mathfrak{C}} \mathfrak{C}^2 (4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{2\theta} d^{\mathfrak{d}_4+\mathfrak{d}_6+2\theta(\mathfrak{d}_1+\mathfrak{d}_2)} \\
&\leq \mathfrak{C} d^{\mathfrak{d}_0} N^{-\mathfrak{n}_0} + 23\varepsilon e^{\mathfrak{C}} \mathfrak{C}^2 (4e^{\mathfrak{C}+1}\mathfrak{C}^3)^{2\theta} d^{\delta} \\
&= \mathfrak{C} d^{\mathfrak{d}_0} N^{-\mathfrak{n}_0} + \tfrac{\varepsilon \gamma d^{\delta}}{2}.
\end{aligned} \tag{2.39}$$

Combining this and (2.18) assures that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that

$$\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \Big|u_d(x) - \Big(\mathcal{R}\Big(\mathbf{g}_{\mathcal{E}_{d,\varepsilon}}^{M_{\mathcal{N}_{d,\varepsilon}},d}\Big)\Big)\Big(X_{\mathcal{N}_{d,\varepsilon}}^{\mathcal{N}_{d,\varepsilon},d,x,\mathcal{E}_{d,\varepsilon}}\Big)\Big|^p\, \nu_d(dx)\right]\right)^{1/p} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{2.40}$$

This and, e.g., [32, Corollary 2.4] establish that there exists $\mathfrak{w} = (\mathfrak{w}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \colon \mathbb{N}\times(0, 1] \to \Omega$ which satisfies for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\left[\int_{\mathbb{R}^d} \Big|u_d(x) - \Big(\mathcal{R}\Big(\mathbf{g}_{\mathcal{E}_{d,\varepsilon}}^{M_{\mathcal{N}_{d,\varepsilon}},d}\Big)\Big)\Big(X_{\mathcal{N}_{d,\varepsilon}}^{\mathcal{N}_{d,\varepsilon},d,x,\mathcal{E}_{d,\varepsilon}}(\mathfrak{w}_{d,\varepsilon})\Big)\Big|^p\, \nu_d(dx)\right]^{1/p} \leq \varepsilon. \tag{2.41}$$

Next note that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $\omega \in \Omega$ it holds that $X_0^{N,d,m,x,\varepsilon}(\omega) = (\mathcal{R}(\mathfrak{I}_d))(x)$. The assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathfrak{I}_d \in \mathfrak{N}_{d,\varepsilon}$ and (2.17) hence ensure that there exist $(\Phi_n^{N,d,m,\varepsilon,\omega})_{m\in\{1,2,\ldots,M_N\}} \subseteq \mathfrak{N}_{d,\varepsilon}$, $\omega \in \Omega$, $n \in \{0, 1, 2, \ldots, N\}$, $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, which satisfy for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $n \in \{0, 1, 2, \ldots, N\}$, $\omega \in \Omega$, $m \in$

$\{1, 2, \ldots, M_N\}$, $x \in \mathbb{R}^d$ that $\mathcal{P}(\Phi_n^{N,d,m,\varepsilon,\omega}) \leq \mathcal{P}(\mathfrak{I}_d) + n\mathfrak{C}N^{\mathfrak{n}_1} d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}}$, $\mathcal{D}(\Phi_n^{N,d,m,\varepsilon,\omega}) = \mathcal{D}(\Phi_n^{N,d,1,\varepsilon,\omega})$, and

$$(\mathcal{R}(\Phi_n^{N,d,m,\varepsilon,\omega}))(x) = \Big(\mathcal{R}\Big(\mathbf{f}^{N,d}_{\varepsilon, Z^{N,d,m}_{n-1}(\omega)}\Big)\Big)\big(X^{N,d,m,x,\varepsilon}_{n-1}(\omega)\big) = X^{N,d,m,x,\varepsilon}_n(\omega). \tag{2.42}$$

The assumption that for all $d \in \mathbb{N}$ it holds that $\mathcal{P}(\mathfrak{I}_d) \leq \mathfrak{C}d^{\mathfrak{d}_3}$ therefore implies for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, M_N\}$, $\varepsilon \in (0, 1]$, $\omega \in \Omega$ that

$$\mathcal{P}(\Phi_N^{N,d,m,\varepsilon,\omega}) \leq \mathfrak{C}d^{\mathfrak{d}_3} + N\mathfrak{C}N^{\mathfrak{n}_1}d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}} \leq \mathfrak{C}d^{\mathfrak{d}_3} + \mathfrak{C}N^{\mathfrak{n}_1+1}d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}}. \tag{2.43}$$

Therefore, we obtain that there exist $\Psi^{N,d,\varepsilon,\omega} \in \mathbf{N}$, $\omega \in \Omega$, $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, which satisfy for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $\omega \in \Omega$, $x \in \mathbb{R}^d$ that $\mathcal{R}(\Psi^{N,d,\varepsilon,\omega}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\Psi^{N,d,\varepsilon,\omega}) \leq \mathfrak{C}N^{\mathfrak{n}_2}(N^{\mathfrak{n}_1+1}d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}} + \mathfrak{C}d^{\mathfrak{d}_3} + \mathfrak{C}N^{\mathfrak{n}_1+1}d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}})$, and

$$\begin{aligned}
&(\mathcal{R}(\Psi^{N,d,\varepsilon,\omega}))(x)\\
&= (\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d})\big((\mathcal{R}(\Phi_N^{N,d,1,\varepsilon,\omega}))(x), (\mathcal{R}(\Phi_N^{N,d,2,\varepsilon,\omega}))(x), \ldots, (\mathcal{R}(\Phi_N^{N,d,M_N,\varepsilon,\omega}))(x)\big)\\
&= (\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))\big(X_N^{N,d,1,x,\varepsilon}(\omega), X_N^{N,d,2,x,\varepsilon}(\omega), \ldots, X_N^{N,d,M_N,x,\varepsilon}(\omega)\big)\\
&= (\mathcal{R}(\mathbf{g}_\varepsilon^{M_N,d}))(X_N^{N,d,x,\varepsilon}(\omega)).
\end{aligned} \tag{2.44}$$

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

$$\begin{aligned}
\mathcal{P}(\Psi^{N,d,\varepsilon,\omega}) &\leq \mathfrak{C}^2 N^{\mathfrak{n}_2} d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}}(2N^{\mathfrak{n}_1+1} + 1)\\
&\leq 3\mathfrak{C}^2 N^{\mathfrak{n}_1+\mathfrak{n}_2+1} d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}}.
\end{aligned} \tag{2.45}$$

This and (2.18) demonstrate that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that

$$\begin{aligned}
&\mathcal{P}\big(\Psi^{\mathcal{N}_{d,\varepsilon},d,\mathcal{E}_{d,\varepsilon},\mathfrak{w}_{d,\varepsilon}}\big)\\
&\leq 3\mathfrak{C}^2 2^{\mathfrak{n}_1+\mathfrak{n}_2+1}\big(\tfrac{2\mathfrak{C}d^{\mathfrak{d}_0}}{\varepsilon}\big)^{\frac{(\mathfrak{n}_1+\mathfrak{n}_2+1)}{\mathfrak{n}_0}} d^{\mathfrak{d}_3}\varepsilon^{-\mathfrak{c}}\gamma^{\mathfrak{c}}d^{\mathfrak{c}\delta}\\
&\leq 3\mathfrak{C}^2 2^{\mathfrak{n}_1+\mathfrak{n}_2+1}(2\mathfrak{C})^{\frac{\mathfrak{n}_1+\mathfrak{n}_2+1}{\mathfrak{n}_0}}\gamma^{\mathfrak{c}} d^{\frac{\mathfrak{d}_0(\mathfrak{n}_1+\mathfrak{n}_2+1)}{\mathfrak{n}_0}+\mathfrak{d}_3+\mathfrak{c}\delta}\varepsilon^{-\frac{(\mathfrak{n}_1+\mathfrak{n}_2+1)}{\mathfrak{n}_0}-\mathfrak{c}}.
\end{aligned} \tag{2.46}$$

Combining this, (2.41), and (2.44) establishes (2.16). The proof of Theorem 2.3 is thus completed. □

3 Artificial neural network (ANN) calculus

In this section we establish in Lemma 3.29 and Lemma 3.30 below a few elementary results on representation flexibilities of ANNs. In our proofs of Lemma 3.29 and Lemma 3.30 we use results from a certain ANN calculus which we recall and extend in Subsections 3.1–3.7. In particular, Definition 3.1 below is [22, Definitions 2.1], Definition 3.2 below is [22, Definitions 2.2], Definition 3.3 below is [22, Definitions 2.3], Definition 3.4 below is [22, Definitions 2.5], and Definition 3.5 below is [22, Definitions 2.17].

3.1 ANNs

Definition 3.1 (ANNs)**.** *We denote by* $\mathbf{N}$ *the set given by*

$$\mathbf{N} = \cup_{L\in\mathbb{N}} \cup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\right) \tag{3.1}$$

and we denote by $\mathcal{P},\mathcal{L},\mathcal{I},\mathcal{O}\colon \mathbf{N}\to\mathbb{N}$*,* $\mathcal{H}\colon\mathbf{N}\to\mathbb{N}_0$*, and* $\mathcal{D}\colon\mathbf{N}\to\cup_{L=2}^{\infty}\mathbb{N}^L$ *the functions which satisfy for all* $L\in\mathbb{N}$*,* $l_0,l_1,\ldots,l_L\in\mathbb{N}$*,* $\Phi\in(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$ *that* $\mathcal{P}(\Phi)=\sum_{k=1}^{L} l_k(l_{k-1}+1)$*,* $\mathcal{L}(\Phi)=L$*,* $\mathcal{I}(\Phi)=l_0$*,* $\mathcal{O}(\Phi)=l_L$*,* $\mathcal{H}(\Phi)=L-1$*, and* $\mathcal{D}(\Phi)=(l_0,l_1,\ldots,l_L)$*.*

3.2 Realizations of ANNs

Definition 3.2 (Multidimensional versions)**.** *Let* $d\in\mathbb{N}$ *and let* $\psi\colon\mathbb{R}\to\mathbb{R}$ *be a function. Then we denote by* $\mathfrak{M}_{\psi,d}\colon\mathbb{R}^d\to\mathbb{R}^d$ *the function which satisfies for all* $x=(x_1,\ldots,x_d)\in\mathbb{R}^d$ *that*

$$\mathfrak{M}_{\psi,d}(x) = (\psi(x_1),\ldots,\psi(x_d))\,. \tag{3.2}$$

Definition 3.3 (Realizations associated to ANNs)**.** *Let* $a\in C(\mathbb{R},\mathbb{R})$*. Then we denote by* $\mathcal{R}_a\colon\mathbf{N}\to\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k,\mathbb{R}^l)$ *the function which satisfies for all* $L\in\mathbb{N}$*,* $l_0,l_1,\ldots,l_L\in\mathbb{N}$*,* $\Phi=((W_1,B_1),(W_2,B_2),\ldots,(W_L,B_L))\in(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$*,* $x_0\in\mathbb{R}^{l_0}, x_1\in\mathbb{R}^{l_1},\ldots,x_{L-1}\in\mathbb{R}^{l_{L-1}}$ *with* $\forall\, k\in\mathbb{N}\cap(0,L)\colon x_k=\mathfrak{M}_{a,l_k}(W_k x_{k-1}+B_k)$ *that*

$$\mathcal{R}_a(\Phi)\in C(\mathbb{R}^{l_0},\mathbb{R}^{l_L}) \qquad \text{and} \qquad (\mathcal{R}_a(\Phi))(x_0)=W_L x_{L-1}+B_L \tag{3.3}$$

(cf. Definition 3.1 and Definition 3.2).

3.3 Compositions of ANNs

Definition 3.4 (Compositions of ANNs)**.** *We denote by* $(\cdot)\bullet(\cdot)\colon\{(\Phi_1,\Phi_2)\in\mathbf{N}\times\mathbf{N}\colon\mathcal{I}(\Phi_1)=\mathcal{O}(\Phi_2)\}\to\mathbf{N}$ *the function which satisfies for all* $L,\mathscr{L}\in\mathbb{N}$*,* $l_0,l_1,\ldots,l_L$*,* $\mathfrak{l}_0,\mathfrak{l}_1,\ldots,\mathfrak{l}_{\mathscr{L}}\in\mathbb{N}$*,* $\Phi_1=((W_1,B_1),(W_2,B_2),\ldots,(W_L,B_L))\in(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$*,* $\Phi_2=((\mathscr{W}_1,\mathscr{B}_1),(\mathscr{W}_2,\mathscr{B}_2),\ldots,(\mathscr{W}_{\mathscr{L}},\mathscr{B}_{\mathscr{L}}))\in(\times_{k=1}^{\mathscr{L}}(\mathbb{R}^{\mathfrak{l}_k\times\mathfrak{l}_{k-1}}\times\mathbb{R}^{\mathfrak{l}_k}))$ *with* $l_0=\mathcal{I}(\Phi_1)=\mathcal{O}(\Phi_2)=\mathfrak{l}_{\mathscr{L}}$ *that*

$$\Phi_1\bullet\Phi_2 = \begin{cases} \begin{array}{r}\big((\mathscr{W}_1,\mathscr{B}_1),(\mathscr{W}_2,\mathscr{B}_2),\ldots,(\mathscr{W}_{\mathscr{L}-1},\mathscr{B}_{\mathscr{L}-1}),(W_1\mathscr{W}_{\mathscr{L}},W_1\mathscr{B}_{\mathscr{L}}+B_1),\\ (W_2,B_2),(W_3,B_3),\ldots,(W_L,B_L)\big)\end{array} & : L>1<\mathscr{L} \\ \big((W_1\mathscr{W}_1,W_1\mathscr{B}_1+B_1),(W_2,B_2),(W_3,B_3),\ldots,(W_L,B_L)\big) & : L>1=\mathscr{L} \\ \big((\mathscr{W}_1,\mathscr{B}_1),(\mathscr{W}_2,\mathscr{B}_2),\ldots,(\mathscr{W}_{\mathscr{L}-1},\mathscr{B}_{\mathscr{L}-1}),(W_1\mathscr{W}_{\mathscr{L}},W_1\mathscr{B}_{\mathscr{L}}+B_1)\big) & : L=1<\mathscr{L} \\ (W_1\mathscr{W}_1,W_1\mathscr{B}_1+B_1) & : L=1=\mathscr{L} \end{cases} \tag{3.4}$$

(cf. Definition 3.1).

3.4 Parallelizations of ANNs with the same length

Definition 3.5 (Parallelizations of ANNs with the same length)**.** *Let $n \in \mathbb{N}$. Then we denote by*

$$\mathbf{P}_n \colon \big\{(\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n \colon \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)\big\} \to \mathbf{N} \tag{3.5}$$

the function which satisfies for all $L \in \mathbb{N}$, $(l_{1,0}, l_{1,1}, \dots, l_{1,L}), (l_{2,0}, l_{2,1}, \dots, l_{2,L}), \dots, (l_{n,0}, l_{n,1}, \dots, l_{n,L}) \in \mathbb{N}^{L+1}$, $\Phi_1 = ((W_{1,1}, B_{1,1}), (W_{1,2}, B_{1,2}), \dots, (W_{1,L}, B_{1,L})) \in (\times_{k=1}^{L} (\mathbb{R}^{l_{1,k} \times l_{1,k-1}} \times \mathbb{R}^{l_{1,k}}))$, $\Phi_2 = ((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \dots, (W_{2,L}, B_{2,L})) \in (\times_{k=1}^{L} (\mathbb{R}^{l_{2,k} \times l_{2,k-1}} \times \mathbb{R}^{l_{2,k}}))$, $\dots$, $\Phi_n = ((W_{n,1}, B_{n,1}), (W_{n,2}, B_{n,2}), \dots, (W_{n,L}, B_{n,L})) \in (\times_{k=1}^{L} (\mathbb{R}^{l_{n,k} \times l_{n,k-1}} \times \mathbb{R}^{l_{n,k}}))$ that

$$\begin{aligned} \mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n) = \Bigg(\Bigg(&\begin{pmatrix} W_{1,1} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,1} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,1} \end{pmatrix}, \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ B_{3,1} \\ \vdots \\ B_{n,1} \end{pmatrix} \Bigg), \\ \Bigg(&\begin{pmatrix} W_{1,2} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,2} \end{pmatrix}, \begin{pmatrix} B_{1,2} \\ B_{2,2} \\ B_{3,2} \\ \vdots \\ B_{n,2} \end{pmatrix} \Bigg), \dots, \\ \Bigg(&\begin{pmatrix} W_{1,L} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,L} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,L} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,L} \end{pmatrix}, \begin{pmatrix} B_{1,L} \\ B_{2,L} \\ B_{3,L} \\ \vdots \\ B_{n,L} \end{pmatrix} \Bigg) \Bigg) \end{aligned} \tag{3.6}$$

(cf. Definition 3.1).

3.5 Linear transformations of ANNs

Definition 3.6 (Identity matrix)**.** *Let $n \in \mathbb{N}$. Then we denote by $\mathrm{I}_n \in \mathbb{R}^{n \times n}$ the identity matrix in $\mathbb{R}^{n \times n}$.*

Definition 3.7 (ANNs with a vector input)**.** *Let $n \in \mathbb{N}$, $B \in \mathbb{R}^n$. Then we denote by $\mathfrak{B}_B \in (\mathbb{R}^{n \times n} \times \mathbb{R}^n)$ the pair given by $\mathfrak{B}_B = (\mathrm{I}_n, B)$ (cf. Definition 3.6).*

Lemma 3.8. *Let $n \in \mathbb{N}$, $B \in \mathbb{R}^n$. Then*

(i) it holds that $\mathfrak{B}_B \in \mathbf{N}$,

(ii) it holds that $\mathcal{D}(\mathfrak{B}_B) = (n, n) \in \mathbb{N}^2$,

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\mathfrak{B}_B) \in C(\mathbb{R}^n, \mathbb{R}^n)$, and

(iv) it holds for all $a \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^n$ that

$$(\mathcal{R}_a(\mathfrak{B}_B))(x) = x + B \tag{3.7}$$

(cf. Definition 3.1, Definition 3.3, and Definition 3.7).

Proof of Lemma 3.8. Note that the fact that $\mathfrak{B}_B \in (\mathbb{R}^{n\times n} \times \mathbb{R}^n)$ ensures that $\mathfrak{B}_B \in \mathbf{N}$ and $\mathcal{D}(\mathfrak{B}_B) = (n,n) \in \mathbb{N}^2$. This establishes items (i)–(ii). The fact that $\mathfrak{B}_B = (\mathrm{I}_n, B)$ (cf. Definition 3.6) and (3.3) therefore prove that for all $a \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathfrak{B}_B) \in C(\mathbb{R}^n,\mathbb{R}^n)$ and

$$(\mathcal{R}_a(\mathfrak{B}_B))(x) = x + B. \tag{3.8}$$

This establishes items (iii)–(iv). The proof of Lemma 3.8 is thus completed. □

Lemma 3.9. *Let $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

(i) it holds for all $B \in \mathbb{R}^{\mathcal{O}(\Phi)}$ that $\mathcal{D}(\mathfrak{B}_B \bullet \Phi) = \mathcal{D}(\Phi)$,

(ii) it holds for all $B \in \mathbb{R}^{\mathcal{O}(\Phi)}$, $a \in C(\mathbb{R},\mathbb{R})$ that $\mathcal{R}_a(\mathfrak{B}_B \bullet \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$,

(iii) it holds for all $B \in \mathbb{R}^{\mathcal{O}(\Phi)}$, $a \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that

$$(\mathcal{R}_a(\mathfrak{B}_B \bullet \Phi))(x) = (\mathcal{R}_a(\Phi))(x) + B, \tag{3.9}$$

(iv) it holds for all $B \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that $\mathcal{D}(\Phi \bullet \mathfrak{B}_B) = \mathcal{D}(\Phi)$,

(v) it holds for all $B \in \mathbb{R}^{\mathcal{I}(\Phi)}$, $a \in C(\mathbb{R},\mathbb{R})$ that $\mathcal{R}_a(\Phi \bullet \mathfrak{B}_B) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$, and

(vi) it holds for all $B \in \mathbb{R}^{\mathcal{I}(\Phi)}$, $a \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^{\mathcal{O}(\Phi)}$ that

$$(\mathcal{R}_a(\Phi \bullet \mathfrak{B}_B))(x) = (\mathcal{R}_a(\Phi))(x + B) \tag{3.10}$$

(cf. Definition 3.3, Definition 3.4, and Definition 3.7).

Proof of Lemma 3.9. Note that Lemma 3.8 demonstrates that for all $n \in \mathbb{N}$, $B \in \mathbb{R}^n$, $a \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{D}(\mathfrak{B}_B) = (n,n)$, $\mathcal{R}_a(\mathfrak{B}_B) \in C(\mathbb{R}^n,\mathbb{R}^n)$, and

$$(\mathcal{R}_a(\mathfrak{B}_B))(x) = x + B. \tag{3.11}$$

Combining this and, e.g., [22, Proposition 2.6] establishes items (i)–(vi). The proof of Lemma 3.9 is thus completed. □

Definition 3.10 (ANNs with a matrix input)**.** *Let $m,n \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$. Then we denote by $\mathfrak{W}_W \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m)$ the pair given by $\mathfrak{W}_W = (W,0)$.*

Lemma 3.11. *Let $m,n \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$. Then*

(i) it holds that $\mathfrak{W}_W \in \mathbf{N}$,

(ii) it holds that $\mathcal{D}(\mathfrak{W}_W) = (n,m) \in \mathbb{N}^2$,

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathcal{R}_a(\mathfrak{W}_W) \in C(\mathbb{R}^n, \mathbb{R}^m)$*, and*

(iv) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$*,* $x \in \mathbb{R}^n$ *that*

$$(\mathcal{R}_a(\mathfrak{W}_W))(x) = Wx \tag{3.12}$$

(cf. Definition 3.1, Definition 3.3, and Definition 3.10).

Proof of Lemma 3.11. Note that the fact that $\mathfrak{W}_W \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m)$ ensures that $\mathfrak{W}_W \in \mathbf{N}$ and $\mathcal{D}(\mathfrak{W}_W) = (n, m) \in \mathbb{N}^2$. This establishes items (i)–(ii). Next observe that the fact that $\mathfrak{W}_W = (W, 0) \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m)$ and (3.3) prove that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathfrak{W}_W) \in C(\mathbb{R}^n, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathfrak{W}_W))(x) = Wx. \tag{3.13}$$

This establishes items (iii)–(iv). The proof of Lemma 3.11 is thus completed. □

Lemma 3.12. *Let* $a \in C(\mathbb{R}, \mathbb{R})$*,* $\Phi \in \mathbf{N}$ *(cf. Definition 3.1). Then*

(i) it holds for all $m \in \mathbb{N}$*,* $W \in \mathbb{R}^{m\times \mathcal{O}(\Phi)}$ *that* $\mathcal{R}_a(\mathfrak{W}_W \bullet \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^m)$*,*

(ii) it holds for all $m \in \mathbb{N}$*,* $W \in \mathbb{R}^{m\times \mathcal{O}(\Phi)}$*,* $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ *that*

$$(\mathcal{R}_a(\mathfrak{W}_W \bullet \Phi))(x) = W\big((\mathcal{R}_a(\Phi))(x)\big), \tag{3.14}$$

(iii) it holds for all $n \in \mathbb{N}$*,* $W \in \mathbb{R}^{\mathcal{I}(\Phi)\times n}$ *that* $\mathcal{R}_a(\Phi \bullet \mathfrak{W}_W) \in C(\mathbb{R}^n, \mathbb{R}^{\mathcal{O}(\Phi)})$*, and*

(iv) it holds for all $n \in \mathbb{N}$*,* $W \in \mathbb{R}^{\mathcal{I}(\Phi)\times n}$*,* $x \in \mathbb{R}^n$ *that*

$$(\mathcal{R}_a(\Phi \bullet \mathfrak{W}_W))(x) = (\mathcal{R}_a(\Phi))(Wx) \tag{3.15}$$

(cf. Definition 3.3, Definition 3.4, and Definition 3.10).

Proof of Lemma 3.12. Note that Lemma 3.11 demonstrates that for all $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathfrak{W}_W) \in C(\mathbb{R}^n, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathfrak{W}_W))(x) = Wx. \tag{3.16}$$

Combining this and, e.g., [22, Proposition 2.6] establishes items (i)–(iv). The proof of Lemma 3.12 is thus completed. □

Definition 3.13 (Scalar multiplications of ANNs)**.** *We denote by* $(\cdot) \circledast (\cdot) : \mathbb{R} \times \mathbf{N} \to \mathbf{N}$ *the function which satisfies for all* $\lambda \in \mathbb{R}$*,* $\Phi \in \mathbf{N}$ *that*

$$\lambda \circledast \Phi = \mathfrak{W}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)}} \bullet \Phi \tag{3.17}$$

(cf. Definition 3.1, Definition 3.4, Definition 3.6, and Definition 3.10).

Lemma 3.14. *Let* $\lambda \in \mathbb{R}$*,* $\Phi \in \mathbf{N}$ *(cf. Definition 3.1). Then*

(i) it holds that $\mathcal{D}(\lambda \circledast \Phi) = \mathcal{D}(\Phi)$*,*

(ii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\lambda \circledast \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$, and

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that

$$(\mathcal{R}_a(\lambda \circledast \Phi))(x) = \lambda\big((\mathcal{R}_a(\Phi))(x)\big) \tag{3.18}$$

(cf. Definition 3.3 and Definition 3.13).

Proof of Lemma 3.14. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$ satisfy that $L = \mathcal{L}(\Phi)$ and $(l_0, l_1, \ldots, l_L) = \mathcal{D}(\Phi)$. Note that item (ii) in Lemma 3.11 proves that

$$\mathcal{D}(\mathfrak{W}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)}}) = (\mathcal{O}(\Phi), \mathcal{O}(\Phi)) \tag{3.19}$$

(cf. Definition 3.6 and Definition 3.10). Combining this and, e.g., [22, item (i) in Proposition 2.6] assures that

$$\mathcal{D}(\lambda \circledast \Phi) = \mathcal{D}(\mathfrak{W}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)}} \bullet \Phi) = (l_0, l_1, \ldots, l_{L-1}, \mathcal{O}(\Phi)) = \mathcal{D}(\Phi). \tag{3.20}$$

This establishes item (i). Moreover, observe that items (i)–(ii) in Lemma 3.12 demonstrate that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ it holds that $\mathcal{R}_a(\lambda \circledast \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$ and

$$\begin{aligned}(\mathcal{R}_a(\lambda \circledast \Phi))(x) &= (\mathcal{R}_a(\mathfrak{W}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)}} \bullet \Phi))(x) \\ &= \lambda \mathrm{I}_{\mathcal{O}(\Phi)}\big((\mathcal{R}_a(\Phi))(x)\big) = \lambda\big((\mathcal{R}_a(\Phi))(x)\big).\end{aligned} \tag{3.21}$$

This establishes items (ii)–(iii). The proof of Lemma 3.14 is thus completed. □

3.6 Representations of the identities with rectifier functions

Definition 3.15. *We denote by $\mathfrak{I} = (\mathfrak{I}_d)_{d \in \mathbb{N}} \colon \mathbb{N} \to \mathbf{N}$ the function which satisfies for all $d \in \mathbb{N}$ that*

$$\mathfrak{I}_1 = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right), \left(\begin{pmatrix} 1 & -1 \end{pmatrix}, 0\right)\right) \in \big((\mathbb{R}^{2\times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1\times 2} \times \mathbb{R}^1)\big) \tag{3.22}$$

and

$$\mathfrak{I}_d = \mathbf{P}_d(\mathfrak{I}_1, \mathfrak{I}_1, \ldots, \mathfrak{I}_1) \tag{3.23}$$

(cf. Definition 3.1 and Definition 3.5).

Lemma 3.16. *Let $d \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$. Then*

(i) it holds that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d) \in \mathbb{N}^3$,

(ii) it holds that $\mathcal{R}_a(\mathfrak{I}_d) \in C(\mathbb{R}^d, \mathbb{R}^d)$, and

(iii) it holds for all $x \in \mathbb{R}^d$ that

$$(\mathcal{R}_a(\mathfrak{I}_d))(x) = x \tag{3.24}$$

(cf. Definition 3.1, Definition 3.3, and Definition 3.15).

Proof of Lemma 3.16. Throughout this proof let $L = 2$, $l_0 = 1$, $l_1 = 2$, $l_2 = 1$. Note that (3.22) ensures that

$$\mathcal{D}(\mathfrak{I}_1) = (1, 2, 1) = (l_0, l_1, l_2). \tag{3.25}$$

This and, e.g., [22, Lemma 2.18] prove that

$$\begin{aligned} &\mathbf{P}_d(\mathfrak{I}_1, \mathfrak{I}_1, \dots, \mathfrak{I}_1) \\ &\in \big(\times_{k=1}^{L} \big(\mathbb{R}^{(dl_k)\times(dl_{k-1})} \times \mathbb{R}^{(dl_k)}\big)\big) = \big(\big(\mathbb{R}^{(2d)\times d} \times \mathbb{R}^{2d}\big) \times \big(\mathbb{R}^{d\times(2d)} \times \mathbb{R}^{d}\big)\big) \end{aligned} \tag{3.26}$$

(cf. Definition 3.5). Hence, we obtain that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d) \in \mathbb{N}^3$. This establishes item (i). Next note that (3.22) assures that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_a(\mathfrak{I}_1))(x) = a(x) - a(-x) = \max\{x, 0\} - \max\{-x, 0\} = x. \tag{3.27}$$

Combining this and, e.g., [22, Proposition 2.19] demonstrates that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\mathfrak{I}_d) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and

$$\begin{aligned} (\mathcal{R}_a(\mathfrak{I}_d))(x) &= \big(\mathcal{R}_a\big(\mathbf{P}_d(\mathfrak{I}_1, \mathfrak{I}_1, \dots, \mathfrak{I}_1)\big)\big)(x_1, x_2, \dots, x_d) \\ &= \big((\mathcal{R}_a(\mathfrak{I}_1))(x_1), (\mathcal{R}_a(\mathfrak{I}_1))(x_2), \dots, (\mathcal{R}_a(\mathfrak{I}_1))(x_d)\big) \\ &= (x_1, x_2, \dots, x_d) = x. \end{aligned} \tag{3.28}$$

This establishes items (ii)–(iii). The proof of Lemma 3.16 is thus completed. ☐

3.7 Sums of ANNs with the same length

Definition 3.17. *Let $m, n \in \mathbb{N}$. Then we denote by $\mathfrak{S}_{m,n} \in (\mathbb{R}^{m\times(nm)} \times \mathbb{R}^m)$ the pair given by*

$$\mathfrak{S}_{m,n} = \mathfrak{W}_{(\mathrm{I}_m \ \mathrm{I}_m \ \dots \ \mathrm{I}_m)} \tag{3.29}$$

(cf. Definition 3.6 and Definition 3.10).

Lemma 3.18. *Let $m, n \in \mathbb{N}$. Then*

(i) it holds that $\mathfrak{S}_{m,n} \in \mathbf{N}$,

(ii) it holds that $\mathcal{D}(\mathfrak{S}_{m,n}) = (nm, m) \in \mathbb{N}^2$,

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$, and

(iv) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $x_1, x_2, \dots, x_n \in \mathbb{R}^m$ that

$$(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \dots, x_n) = \textstyle\sum_{k=1}^{n} x_k \tag{3.30}$$

(cf. Definition 3.1, Definition 3.3, and Definition 3.17).

Proof of Lemma 3.18. Note that the fact that $\mathfrak{S}_{m,n} \in (\mathbb{R}^{m\times(nm)} \times \mathbb{R}^m)$ ensures that $\mathfrak{S}_{m,n} \in \mathbf{N}$ and $\mathcal{D}(\mathfrak{S}_{m,n}) = (nm, m) \in \mathbb{N}^2$. This establishes items (i)–(ii). Next observe that items (iii)–(iv) in Lemma 3.11 prove that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$ and

$$\begin{aligned}(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \ldots, x_n) &= \big(\mathcal{R}_a\big(\mathfrak{W}_{(\mathrm{I}_m\ \ \mathrm{I}_m\ \ \ldots\ \ \mathrm{I}_m)}\big)\big)(x_1, x_2, \ldots, x_n) \\ &= (\mathrm{I}_m\ \ \mathrm{I}_m\ \ \ldots\ \ \mathrm{I}_m)(x_1, x_2, \ldots, x_n) = \textstyle\sum_{k=1}^n x_k\end{aligned} \tag{3.31}$$

(cf. Definition 3.6 and Definition 3.10). This establishes items (iii)–(iv). The proof of Lemma 3.18 is thus completed. □

Lemma 3.19. *Let $m, n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \{\Psi \in \mathbf{N} \colon \mathcal{O}(\Psi) = nm\}$ (cf. Definition 3.1). Then*

(i) it holds that $\mathcal{R}_a(\mathfrak{S}_{m,n} \bullet \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^m)$ and

(ii) it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$, $y_1, y_2, \ldots, y_n \in \mathbb{R}^m$ with $(\mathcal{R}_a(\Phi))(x) = (y_1, y_2, \ldots, y_n)$ that

$$\big(\mathcal{R}_a(\mathfrak{S}_{m,n} \bullet \Phi)\big)(x) = \textstyle\sum_{k=1}^n y_k \tag{3.32}$$

(cf. Definition 3.3, Definition 3.4, and Definition 3.17).

Proof of Lemma 3.19. Note that Lemma 3.18 ensures that for all $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \ldots, x_n) = \textstyle\sum_{k=1}^n x_k. \tag{3.33}$$

Combining this and, e.g., [22, item (v) in Proposition 2.6] establishes items (i)–(ii). The proof of Lemma 3.19 is thus completed. □

Lemma 3.20. *Let $n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

(i) it holds that $\mathcal{R}_a(\Phi \bullet \mathfrak{S}_{\mathcal{I}(\Phi),n}) \in C(\mathbb{R}^{n\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$ and

(ii) it holds for all $x_1, x_2, \ldots, x_n \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that

$$\big(\mathcal{R}_a(\Phi \bullet \mathfrak{S}_{\mathcal{I}(\Phi),n})\big)(x_1, x_2, \ldots, x_n) = (\mathcal{R}_a(\Phi))(\textstyle\sum_{k=1}^n x_k) \tag{3.34}$$

(cf. Definition 3.3, Definition 3.4, and Definition 3.17).

Proof of Lemma 3.20. Note that Lemma 3.18 demonstrates that for all $m \in \mathbb{N}$, $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \ldots, x_n) = \textstyle\sum_{k=1}^n x_k. \tag{3.35}$$

Combining this and, e.g., [22, item (v) in Proposition 2.6] establishes items (i)–(ii). The proof of Lemma 3.20 is thus completed. □

Definition 3.21. *Let $m, n \in \mathbb{N}$, $A \in \mathbb{R}^{m\times n}$. Then we denote by $A^* \in \mathbb{R}^{n\times m}$ the transpose of A.*

Definition 3.22. *Let* $m, n \in \mathbb{N}$*. Then we denote by* $\mathfrak{T}_{m,n} \in (\mathbb{R}^{(nm)\times m} \times \mathbb{R}^{nm})$ *the pair given by*

$$\mathfrak{T}_{m,n} = \mathfrak{W}_{(\mathrm{I}_m \;\; \mathrm{I}_m \;\; \dots \;\; \mathrm{I}_m)^*} \tag{3.36}$$

(cf. Definition 3.6, Definition 3.10, and Definition 3.21).

Lemma 3.23. *Let* $m, n \in \mathbb{N}$*. Then*

(i) it holds that $\mathfrak{T}_{m,n} \in \mathbf{N}$*,*

(ii) it holds that $\mathcal{D}(\mathfrak{T}_{m,n}) = (m, nm) \in \mathbb{N}^2$*,*

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathcal{R}_a(\mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{nm})$*, and*

(iv) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$*,* $x \in \mathbb{R}^m$ *that*

$$(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) = (x, x, \dots, x) \tag{3.37}$$

(cf. Definition 3.1, Definition 3.3, and Definition 3.22).

Proof of Lemma 3.23. Note that the fact that $\mathfrak{T}_{m,n} \in (\mathbb{R}^{(nm)\times m} \times \mathbb{R}^{nm})$ ensures that $\mathfrak{T}_{m,n} \in \mathbf{N}$ and $\mathcal{D}(\mathfrak{T}_{m,n}) = (m, nm) \in \mathbb{N}^2$. This establishes items (i)–(ii). Next observe that items (iii)–(iv) in Lemma 3.11 prove that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{nm})$ and

$$\begin{aligned}(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) &= \big(\mathcal{R}_a\big(\mathfrak{W}_{(\mathrm{I}_m \;\; \mathrm{I}_m \;\; \dots \;\; \mathrm{I}_m)^*}\big)\big)(x) \\ &= (\mathrm{I}_m \;\; \mathrm{I}_m \;\; \dots \;\; \mathrm{I}_m)^* x = (x, x, \dots, x)\end{aligned} \tag{3.38}$$

(cf. Definition 3.6 and Definition 3.10). This establishes items (iii)–(iv). The proof of Lemma 3.23 is thus completed. ☐

Lemma 3.24. *Let* $n \in \mathbb{N}$*,* $a \in C(\mathbb{R}, \mathbb{R})$*,* $\Phi \in \mathbf{N}$ *(cf. Definition 3.1). Then*

(i) it holds that $\mathcal{R}_a(\mathfrak{T}_{\mathcal{O}(\Phi),n} \bullet \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{n\mathcal{O}(\Phi)})$ *and*

(ii) it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ *that*

$$\big(\mathcal{R}_a(\mathfrak{T}_{\mathcal{O}(\Phi),n} \bullet \Phi)\big)(x) = \big((\mathcal{R}_a(\Phi))(x), (\mathcal{R}_a(\Phi))(x), \dots, (\mathcal{R}_a(\Phi))(x)\big) \tag{3.39}$$

(cf. Definition 3.3, Definition 3.4, and Definition 3.22).

Proof of Lemma 3.24. Note that Lemma 3.23 ensures that for all $m \in \mathbb{N}$, $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{nm})$ and

$$(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) = (x, x, \dots, x). \tag{3.40}$$

Combining this and, e.g., [22, item (v) in Proposition 2.6] establishes items (i)–(ii). The proof of Lemma 3.24 is thus completed. ☐

Lemma 3.25. *Let* $m, n \in \mathbb{N}$*,* $a \in C(\mathbb{R}, \mathbb{R})$*,* $\Phi \in \{\Psi \in \mathbf{N} \colon \mathcal{I}(\Psi) = nm\}$ *(cf. Definition 3.1). Then* *(i) it holds that* $\mathcal{R}_a(\Phi \bullet \mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{\mathcal{O}(\Phi)})$ *and*

(ii) it holds for all $x \in \mathbb{R}^m$ *that*

$$\big(\mathcal{R}_a(\Phi \bullet \mathfrak{T}_{m,n})\big)(x) = (\mathcal{R}_a(\Phi))(x, x, \ldots, x) \tag{3.41}$$

(cf. Definition 3.3, Definition 3.4, and Definition 3.22).

Proof of Lemma 3.25. Observe that Lemma 3.23 demonstrates that for all $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{nm})$ and

$$(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) = (x, x, \ldots, x). \tag{3.42}$$

Combining this and, e.g., [22, item (v) in Proposition 2.6] establishes items (i)–(ii). The proof of Lemma 3.25 is thus completed. □

Definition 3.26 (Sums of ANNs with the same length)**.** *Let* $n \in \mathbb{N}$, $\Phi_1, \Phi_2, \ldots, \Phi_n \in \mathbf{N}$ *satisfy for all* $k \in \{1, 2, \ldots, n\}$ *that* $\mathcal{L}(\Phi_k) = \mathcal{L}(\Phi_1)$, $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_1)$, *and* $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_1)$. *Then we denote by* $\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k$ *(we denote by* $\Phi_1 \oplus \Phi_2 \oplus \ldots \oplus \Phi_n$*) the tuple given by*

$$\oplus_{k\in\{1,2,\ldots,n\}} \Phi_k = \big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n} \bullet \big[\mathbf{P}_n(\Phi_1, \Phi_2, \ldots, \Phi_n)\big] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big) \in \mathbf{N} \tag{3.43}$$

(cf. Definition 3.1, Definition 3.4, Definition 3.5, Definition 3.17, and Definition 3.22).

Definition 3.27 (Dimensions of ANNs)**.** *Let* $n \in \mathbb{N}_0$. *Then we denote by* $\mathbb{D}_n \colon \mathbf{N} \to \mathbb{N}_0$ *the function which satisfies for all* $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi \in (\times_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$ *that*

$$\mathbb{D}_n(\Phi) = \begin{cases} l_n & : n \leq L \\ 0 & : n > L \end{cases} \tag{3.44}$$

(cf. Definition 3.1).

Lemma 3.28. *Let* $n \in \mathbb{N}$, $\Phi_1, \Phi_2, \ldots, \Phi_n \in \mathbf{N}$ *satisfy for all* $k \in \{1, 2, \ldots, n\}$ *that* $\mathcal{L}(\Phi_k) = \mathcal{L}(\Phi_1)$, $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_1)$, *and* $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_1)$ *(cf. Definition 3.1). Then*

(i) it holds that $\mathcal{L}(\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k) = \mathcal{L}(\Phi_1)$,

(ii) it holds that

$$\begin{aligned} &\mathcal{D}(\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k) \\ &= \big(\mathcal{I}(\Phi_1), \textstyle\sum_{k=1}^{n} \mathbb{D}_1(\Phi_k), \sum_{k=1}^{n} \mathbb{D}_2(\Phi_k), \ldots, \sum_{k=1}^{n} \mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k), \mathcal{O}(\Phi_1)\big), \end{aligned} \tag{3.45}$$

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathcal{R}_a(\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k) \in C(\mathbb{R}^{\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$, *and*

(iv) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_1)}$ *that*

$$\big(\mathcal{R}_a(\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k)\big)(x) = \sum_{k=1}^{n} (\mathcal{R}_a(\Phi_k))(x) \tag{3.46}$$

(*cf. Definition 3.3, Definition 3.26, and Definition 3.27*).

Proof of Lemma 3.28. First, note that, e.g., [22, Lemma 2.18] proves that

$$\begin{aligned}&\mathcal{D}\big(\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)\big)\\&=\big(\textstyle\sum_{k=1}^n\mathbb{D}_0(\Phi_k),\sum_{k=1}^n\mathbb{D}_1(\Phi_k),\ldots,\sum_{k=1}^n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k),\sum_{k=1}^n\mathbb{D}_{\mathcal{L}(\Phi_1)}(\Phi_k)\big)\\&=\big(n\mathcal{I}(\Phi_1),\textstyle\sum_{k=1}^n\mathbb{D}_1(\Phi_k),\sum_{k=1}^n\mathbb{D}_2(\Phi_k),\ldots,\sum_{k=1}^n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k),n\mathcal{O}(\Phi_1)\big)\end{aligned}\tag{3.47}$$

(cf. Definition 3.5). Moreover, observe that item (ii) in Lemma 3.18 ensures that

$$\mathcal{D}\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n}\big)=(n\mathcal{O}(\Phi_1),\mathcal{O}(\Phi_1))\tag{3.48}$$

(cf. Definition 3.17). This, (3.47), and, e.g., [22, item (i) in Proposition 2.6] demonstrate that

$$\begin{aligned}&\mathcal{D}\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n}\bullet\big[\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)\big]\big)\\&=\big(n\mathcal{I}(\Phi_1),\textstyle\sum_{k=1}^n\mathbb{D}_1(\Phi_k),\sum_{k=1}^n\mathbb{D}_2(\Phi_k),\ldots,\sum_{k=1}^n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k),\mathcal{O}(\Phi_1)\big).\end{aligned}\tag{3.49}$$

Next note that item (ii) in Lemma 3.23 assures that

$$\mathcal{D}\big(\mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)=(\mathcal{I}(\Phi_1),n\mathcal{I}(\Phi_1))\tag{3.50}$$

(cf. Definition 3.22). Combining this, (3.49), and, e.g., [22, item (i) in Proposition 2.6] proves that

$$\begin{aligned}&\mathcal{D}(\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k)\\&=\mathcal{D}\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n}\bullet\big[\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)\big]\bullet\mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\\&=\big(\mathcal{I}(\Phi_1),\textstyle\sum_{k=1}^n\mathbb{D}_1(\Phi_k),\sum_{k=1}^n\mathbb{D}_2(\Phi_k),\ldots,\sum_{k=1}^n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k),\mathcal{O}(\Phi_1)\big).\end{aligned}\tag{3.51}$$

This establishes items (i)–(ii). Next observe that Lemma 3.25 and (3.47) ensure that for all $a\in C(\mathbb{R},\mathbb{R})$, $x\in\mathbb{R}^{\mathcal{I}(\Phi_1)}$ it holds that $\mathcal{R}_a([\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)]\bullet\mathfrak{T}_{\mathcal{I}(\Phi_1),n})\in C(\mathbb{R}^{\mathcal{I}(\Phi_1)},\mathbb{R}^{n\mathcal{O}(\Phi_1)})$ and

$$\begin{aligned}&\big(\mathcal{R}_a\big([\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)]\bullet\mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\big)(x)\\&=\big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)\big)\big)(x,x,\ldots,x).\end{aligned}\tag{3.52}$$

Combining this with, e.g., [22, item (ii) in Proposition 2.19] proves that for all $a\in C(\mathbb{R},\mathbb{R})$, $x\in\mathbb{R}^{\mathcal{I}(\Phi_1)}$ it holds that

$$\begin{aligned}&\big(\mathcal{R}_a\big([\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)]\bullet\mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\big)(x)\\&=\big((\mathcal{R}_a(\Phi_1))(x),(\mathcal{R}_a(\Phi_2))(x),\ldots,(\mathcal{R}_a(\Phi_n))(x)\big)\in\mathbb{R}^{n\mathcal{O}(\Phi_1)}.\end{aligned}\tag{3.53}$$

Lemma 3.19, (3.48), and, e.g., [22, Lemma 2.8] therefore demonstrate that for all $a\in C(\mathbb{R},\mathbb{R})$, $x\in\mathbb{R}^{\mathcal{I}(\Phi_1)}$ it holds that $\mathcal{R}_a(\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k)\in C(\mathbb{R}^{\mathcal{I}(\Phi_1)},\mathbb{R}^{\mathcal{O}(\Phi_1)})$ and

$$\begin{aligned}&\big(\mathcal{R}_a(\oplus_{k\in\{1,2,\ldots,n\}}\Phi_k)\big)(x)\\&=\big(\mathcal{R}_a\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n}\bullet[\mathbf{P}_n(\Phi_1,\Phi_2,\ldots,\Phi_n)]\bullet\mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\big)(x)=\sum_{k=1}^n(\mathcal{R}_a(\Phi_k))(x).\end{aligned}\tag{3.54}$$

This establishes items (iii)–(iv). The proof of Lemma 3.28 is thus completed. □

3.8 ANN representation results

Lemma 3.29. *Let $n \in \mathbb{N}$, $h_1, h_2, \ldots, h_n \in \mathbb{R}$, $\Phi_1, \Phi_2, \ldots, \Phi_n \in \mathbf{N}$ satisfy that $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_n)$, let $A_k \in \mathbb{R}^{\mathcal{I}(\Phi_1)\times(n\mathcal{I}(\Phi_1))}$, $k \in \{1, 2, \ldots, n\}$, satisfy for all $k \in \{1, 2, \ldots, n\}$, $x = (x_i)_{i\in\{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\phi_1)}$ that $A_k x = x_k$, and let $\Psi \in \mathbf{N}$ satisfy that*

$$\Psi = \oplus_{k\in\{1,2,\ldots,n\}}(h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k})) \tag{3.55}$$

(cf. Definition 3.1, Definition 3.10, Definition 3.13, and Definition 3.26). Then

(i) it holds that

$$\mathcal{D}(\Psi) = (n\mathcal{I}(\Phi_1), n\mathbb{D}_1(\Phi_1), n\mathbb{D}_2(\Phi_1), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_1), \mathcal{O}(\Phi_1)), \tag{3.56}$$

(ii) it holds that $\mathcal{P}(\Psi) \le n^2\mathcal{P}(\Phi_1)$,

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^{n\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$, and

(iv) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $x = (x_k)_{k\in\{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_1)}$ that

$$(\mathcal{R}_a(\Psi))(x) = \sum_{k=1}^{n} h_k(\mathcal{R}_a(\Phi_k))(x_k) \tag{3.57}$$

(cf. Definition 3.3 and Definition 3.27).

Proof of Lemma 3.29. First, note that item (ii) in Lemma 3.11 ensures for all $k \in \{1, 2, \ldots, n\}$ that

$$\mathcal{D}(\mathfrak{W}_{A_k}) = (n\mathcal{I}(\Phi_1), \mathcal{I}(\Phi_1)) \in \mathbb{N}^2. \tag{3.58}$$

This and, e.g., [22, item (i) in Proposition 2.6] prove for all $k \in \{1, 2, \ldots, n\}$ that

$$\mathcal{D}(\Phi_k \bullet \mathfrak{W}_{A_k}) = (n\mathcal{I}(\Phi_1), \mathbb{D}_1(\Phi_k), \mathbb{D}_2(\Phi_k), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_k)}(\Phi_k)). \tag{3.59}$$

Item (i) in Lemma 3.14 therefore demonstrates for all $k \in \{1, 2, \ldots, n\}$ that

$$\begin{aligned} \mathcal{D}(h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k})) &= \mathcal{D}(\Phi_k \bullet \mathfrak{W}_{A_k}) \\ &= (n\mathcal{I}(\Phi_1), \mathbb{D}_1(\Phi_k), \mathbb{D}_2(\Phi_k), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_k)-1}(\Phi_k), \mathcal{O}(\Phi_k)) \\ &= (n\mathcal{I}(\Phi_1), \mathbb{D}_1(\Phi_1), \mathbb{D}_2(\Phi_1), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_1), \mathcal{O}(\Phi_1)). \end{aligned} \tag{3.60}$$

Combining this with item (ii) in Lemma 3.28 ensures that

$$\begin{aligned} \mathcal{D}(\Psi) &= \mathcal{D}\big(\oplus_{k\in\{1,2,\ldots,n\}} (h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k}))\big) \\ &= (n\mathcal{I}(\Phi_1), n\mathbb{D}_1(\Phi_1), n\mathbb{D}_2(\Phi_1), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_1), \mathcal{O}(\Phi_1)). \end{aligned} \tag{3.61}$$

This establishes item (i). Hence, we obtain that

$$\mathcal{P}(\Psi) \le n^2\mathcal{P}(\Phi_1). \tag{3.62}$$

This establishes item (ii). Moreover, observe that items (iii)–(iv) in Lemma 3.12 assure for all $k \in \{1, 2, \ldots, n\}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x = (x_i)_{i\in\{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_1)}$ that $\mathcal{R}_a(\Phi_k \bullet \mathfrak{W}_{A_k}) \in C(\mathbb{R}^{n\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_k)})$ and

$$\big(\mathcal{R}_a(\Phi_k \bullet \mathfrak{W}_{A_k})\big)(x) = (\mathcal{R}_a(\Phi))(A_k x) = (\mathcal{R}_a(\Phi))(x_k). \tag{3.63}$$

Combining this with items (ii)–(iii) in Lemma 3.14 proves for all $k \in \{1, 2, \ldots, n\}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x = (x_i)_{i \in \{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_1)}$ that $\mathcal{R}_a(h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k})) \in C(\mathbb{R}^{n\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$ and

$$\big(\mathcal{R}_a(h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k}))\big)(x) = h_k(\mathcal{R}_a(\Phi))(x_k). \tag{3.64}$$

Items (iii)–(iv) in Lemma 3.28 and (3.60) hence ensure for all $a \in C(\mathbb{R}, \mathbb{R})$, $x = (x_i)_{i \in \{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_1)}$ that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^{n\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$ and

$$\begin{aligned}(\mathcal{R}_a(\Psi))(x) &= \big(\mathcal{R}_a\big(\oplus_{k \in \{1,2,\ldots,n\}} (h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k}))\big)\big)(x) \\ &= \sum_{k=1}^{n} \big(\mathcal{R}_a(h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k}))\big)(x) = \sum_{k=1}^{n} h_k(\mathcal{R}_a(\Phi_k))(x_k).\end{aligned} \tag{3.65}$$

This establishes items (iii)–(iv). The proof of Lemma 3.29 is thus completed. □

Lemma 3.30. *Let $a \in C(\mathbb{R}, \mathbb{R})$, $L_1, L_2 \in \mathbb{N}$, $\mathbb{I}, \Phi_1, \Phi_2 \in \mathbf{N}$, $d, \mathfrak{i}, l_{1,0}, l_{1,1}, \ldots, l_{1,L_1}, l_{2,0}, l_{2,1}, \ldots, l_{2,L_2} \in \mathbb{N}$ satisfy for all $k \in \{1, 2\}$, $x \in \mathbb{R}^d$ that $2 \le \mathfrak{i} \le 2d$, $l_{2,L_2-1} \le l_{1,L_1-1} + \mathfrak{i}$, $\mathcal{D}(\mathbb{I}) = (d, \mathfrak{i}, d)$, $(\mathcal{R}_a(\mathbb{I}))(x) = x$, $\mathcal{I}(\Phi_k) = \mathcal{O}(\Phi_k) = d$, and $\mathcal{D}(\Phi_k) = (l_{k,0}, l_{k,1}, \ldots, l_{k,L_k})$ (cf. Definition 3.1 and Definition 3.3). Then there exists $\Psi \in \mathbf{N}$ such that*

(i) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

(ii) it holds for all $x \in \mathbb{R}^d$ that

$$(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\Phi_2))(x) + \big((\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))\big)(x), \tag{3.66}$$

(iii) it holds that

$$\mathbb{D}_{\mathcal{L}(\Psi)-1}(\Psi) \le l_{1,L_1-1} + \mathfrak{i}, \tag{3.67}$$

and

(iv) it holds that $\mathcal{P}(\Psi) \le \mathcal{P}(\Phi_2) + \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\Phi_1)\big]^2$

(cf. Definition 3.4 and Definition 3.27).

Proof of Lemma 3.30. To prove items (i)–(iv) we distinguish between the case $L_1 = 1$ and the case $L_1 \in \mathbb{N} \cap [2, \infty)$. We first prove items (i)–(iv) in the case $L_1 = 1$. Note that, e.g., [22, Proposition 2.30] (with $a = a$, $d = d$, $\mathfrak{L} = L_2$, $(\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}}) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2})$, $\psi = \Phi_2$, $\phi_n = \Phi_1$ for $n \in \mathbb{N}_0$ in the notation of [22, Proposition 2.30]) implies that there exists $\Psi \in \mathbf{N}$ such that

(I) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

(II) it holds for all $x \in \mathbb{R}^d$ that

$$(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\Phi_2))(x) + \big((\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))\big)(x), \tag{3.68}$$

and

(III) it holds that $\mathcal{D}(\Psi) = \mathcal{D}(\Phi_2)$.

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

$$\mathbb{D}_{\mathcal{L}(\Psi)-1}(\Psi) = \mathbb{D}_{\mathcal{L}(\Phi_2)-1}(\Phi_2) = l_{2,L_2-1} \leq l_{1,L_1-1} + \mathfrak{i}. \tag{3.69}$$

Moreover, note that (III) assures that

$$\mathcal{P}(\Psi) = \mathcal{P}(\Phi_2) \leq \mathcal{P}(\Phi_2) + \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\Phi_1)\right]^2. \tag{3.70}$$

Combining this with (I) and (3.69) establishes items (i)–(iv) in the case $L_1 = 1$. We now prove items (i)–(iv) in the case $L_1 \in \mathbb{N} \cap [2, \infty)$. Observe that, e.g., [22, Proposition 2.28] (with $a = a$, $L_1 = L_1$, $L_2 = L_2$, $\mathbb{I} = \mathbb{I}$, $\Phi_1 = \Phi_1$, $\Phi_2 = \Phi_2$, $d = d$, $\mathfrak{i} = \mathfrak{i}$, $(l_{1,0}, l_{1,1}, \ldots, l_{1,L_1}) = (l_{1,0}, l_{1,1}, \ldots, l_{1,L_1})$, $(l_{2,0}, l_{2,1}, \ldots, l_{2,L_2}) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2})$ in the notation of [22, Proposition 2.28]) proves that there exists $\Psi \in \mathbf{N}$ such that

(a) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

(b) it holds for all $x \in \mathbb{R}^d$ that

$$(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\Phi_2))(x) + \big((\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))\big)(x), \tag{3.71}$$

(c) it holds that

$$\mathcal{D}(\Psi) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1} + \mathfrak{i}, l_{1,2} + \mathfrak{i}, \ldots, l_{1,L_1-1} + \mathfrak{i}, l_{1,L_1}), \tag{3.72}$$

and

(d) it holds that $\mathcal{P}(\Psi) \leq \mathcal{P}(\Phi_2) + \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\Phi_1)\right]^2$.

This establishes items (i)–(iv) in the case $L_1 \in \mathbb{N} \cap [2, \infty)$. The proof of Lemma 3.30 is thus completed. □

4 Kolmogorov partial differential equations (PDEs)

In this section we establish in Theorem 4.5 below the existence of DNNs which approximate solutions of suitable Kolmogorov PDEs without the curse of dimensionality. Moreover, in Corollary 4.6 below we specialize Theorem 4.5 to the case where for every $d \in \mathbb{N}$ we have that the probability measure ν_d appearing in Theorem 4.5 is the uniform distribution on the d-dimensional unit cube $[0,1]^d$. In addition, in Corollary 4.7 below we specialize Theorem 4.5, roughly speaking, to the case where the constants $\kappa \in (0, \infty)$, $\mathfrak{e}, \mathfrak{d}_1, \mathfrak{d}_2, \ldots, \mathfrak{d}_6 \in [0, \infty)$, which we use to specify the regularity hypotheses in Theorem 4.5, coincide. Corollary 4.7 follows immediately from Theorem 4.5 and is a slight generalization of [32, Theorem 6.3] and [32, Theorem 1.1], respectively. In our proof of Theorem 4.5 we employ the DNN representation results in Lemmas 3.29–3.30 from Section 3 above as well as essentially well-known error estimates for the Monte Carlo Euler method which we establish in Proposition 4.4 in this section below. The proof Proposition 4.4, in turn, employs the elementary error estimate results in Lemmas 4.1–4.3 below.

4.1 Error analysis for the Monte Carlo Euler method

Lemma 4.1 (Weak perturbation error)**.** *Let $d, m \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, $T \in (0, \infty)$, $L_0, L_1, l \in [0, \infty)$, $h \in (0, T]$, $B \in \mathbb{R}^{d\times m}$, 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 $f_0 : \mathbb{R}^d \to \mathbb{R}$ and $f_1 : \mathbb{R}^d \to \mathbb{R}^d$ be functions, let $\chi : [0, T] \to [0, T]$ be a function, assume for all $t \in [0, T]$, $x, y \in \mathbb{R}^d$ that*

$$|f_0(x) - f_0(y)| \leq L_0\left(1 + \int_0^1 [r\|x\| + (1-r)\|y\|]^l \, dr\right)\|x - y\|, \tag{4.1}$$

$$\|f_1(x) - f_1(y)\| \leq L_1\|x - y\|, \tag{4.2}$$

and $\chi(t) = \max(\{0, h, 2h, \ldots\} \cap [0, t])$, and let $X, Y : [0, T] \times \Omega \to \mathbb{R}^d$ be stochastic processes with continuous sample paths which satisfy for all $t \in [0, T]$ that $Y_t = \xi + \int_0^t f_1\big(Y_{\chi(s)}\big)\, ds + BW_t$ and

$$X_t = \xi + \int_0^t f_1(X_s)\, ds + BW_t. \tag{4.3}$$

Then it holds that

$$\begin{aligned} \big|\mathbb{E}[f_0(X_T)] - \mathbb{E}[f_0(Y_T)]\big| &\leq (h/T)^{1/2} e^{(l+3+2L_1+[lL_1+2L_1+2]T)} \max\{1, L_0\} \\ &\cdot \Big[\|\xi\| + 2 + \max\{1, \|f_1(0)\|\}\max\{1, T\} + \sqrt{(2\max\{l, 1\} - 1)\,\mathrm{Trace}(B^*B)T}\Big]^{1+l}. \end{aligned} \tag{4.4}$$

Proof of Lemma 4.1. First, note that (4.2) proves that for all $x \in \mathbb{R}^d$ it holds that

$$\|f_1(x)\| \leq \|f_1(x) - f_1(0)\| + \|f_1(0)\| \leq L_1\|x\| + \|f_1(0)\|. \tag{4.5}$$

This, (4.1), (4.2), and, e.g., [32, Proposition 4.6] (with $d = d$, $m = m$, $\xi = \xi$, $T = T$, $c = L_1$, $C = \|f_1(0)\|$, $\varepsilon_0 = 0$, $\varepsilon_1 = 0$, $\varepsilon_2 = 0$, $\varsigma_0 = 0$, $\varsigma_1 = 0$, $\varsigma_2 = 0$, $L_0 = L_0$, $L_1 = L_1$, $l = l$, $h = h$, $B = B$, $p = 2$, $q = 2$, $\|\cdot\| = \|\cdot\|$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W = W$, $\phi_0 = f_0$, $f_1 = f_1$, $\phi_2 = (\mathbb{R}^d \ni x \mapsto x \in \mathbb{R}^d)$, $\chi = \chi$, $f_0 = f_0$, $\phi_1 = \phi_1$, $\varpi_r = (\mathbb{E}[\|BW_T\|^r])^{1/r}$, $X = X$, $Y = Y$ for $r \in (0, \infty)$ in the notation of [32, Proposition 4.6]) establish that

$$\begin{aligned} \big|\mathbb{E}[f_0(X_T)] - \mathbb{E}[f_0(Y_T)]\big| &\leq (h/T)^{1/2} e^{(l+3+2L_1+[lL_1+L_1+L_1+2]T)} \max\{1, L_0\} \\ &\cdot \Big[\|\xi\| + 2 + \max\{1, \|f_1(0)\|\}\max\{1, T\} + \big(\mathbb{E}\big[\|BW_T\|^{\max\{2,2l\}}\big]\big)^{1/\max\{2,2l\}}\Big]^{1+l}. \end{aligned} \tag{4.6}$$

Combining this with, e.g., [32, Lemma 4.2] (with $d = d$, $m = m$, $T = T$, $p = \max\{2, 2l\}$, $B = B$, $\|\cdot\| = \|\cdot\|$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W = W$ in the notation of [32, Lemma 4.2]) ensures that

$$\begin{aligned} \big|\mathbb{E}[f_0(X_T)] - \mathbb{E}[f_0(Y_T)]\big| &\leq (h/T)^{1/2} e^{(l+3+2L_1+[lL_1+2L_1+2]T)} \max\{1, L_0\} \\ &\cdot \Big[\|\xi\| + 2 + \max\{1, \|f_1(0)\|\}\max\{1, T\} + \sqrt{(2\max\{l, 1\} - 1)\,\mathrm{Trace}(B^*B)T}\Big]^{1+l}. \end{aligned} \tag{4.7}$$

The proof of Lemma 4.1 is thus completed. ∎

Lemma 4.2. *Let $d, m \in \mathbb{N}$, $T, \kappa \in (0,\infty)$, $\theta, \mathfrak{d}_0, \mathfrak{d}_1 \in [0,\infty)$, $h \in (0,T]$, $B \in \mathbb{R}^{d\times m}$, $p \in [1,\infty)$, let $\nu \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure on $\mathbb{R}^d$, 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 $f_0 \colon \mathbb{R}^d \to \mathbb{R}$ and $f_1 \colon \mathbb{R}^d \to \mathbb{R}^d$ be functions, let $\chi \colon [0,T] \to [0,T]$ be a function, assume for all $t \in [0,T]$, $x, y \in \mathbb{R}^d$ that*

$$|f_0(x) - f_0(y)| \le \kappa d^{\mathfrak{d}_0}(1 + \|x\|^\theta + \|y\|^\theta)\|x-y\|, \tag{4.8}$$

$$\|f_1(x) - f_1(y)\| \le \kappa\|x-y\|, \qquad \operatorname{Trace}(B^*B) \le \kappa d^{2\mathfrak{d}_1}, \tag{4.9}$$

$$\|f_1(0)\| \le \kappa d^{\mathfrak{d}_1}, \qquad \left[\int_{\mathbb{R}^d} \|z\|^{p(1+\theta)}\,\nu(dz)\right]^{1/(p(1+\theta))} \le \kappa d^{\mathfrak{d}_1}, \tag{4.10}$$

and $\chi(t) = \max(\{0, h, 2h, \ldots\} \cap [0,t])$, and let $X^x \colon [0,T]\times\Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, and $Y^x \colon [0,T]\times\Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, be stochastic processes with continuous sample paths which satisfy for all $x \in \mathbb{R}^d$, $t \in [0,T]$ that $Y^x_t = x + \int_0^t f_1\big(Y^x_{\chi(s)}\big)\,ds + BW_t$ and

$$X^x_t = x + \int_0^t f_1(X^x_s)\,ds + BW_t. \tag{4.11}$$

Then it holds that

$$\begin{aligned}\left[\int_{\mathbb{R}^d} \big|\mathbb{E}[f_0(X^x_T)] - \mathbb{E}[f_0(Y^x_T)]\big|^p\,\nu(dx)\right]^{1/p} &\le 2^{4\theta+5}|\max\{1,T\}|^{\theta+1} \\ &\cdot |\max\{\kappa,\theta,1\}|^{\theta+3} e^{(6\max\{\kappa,\theta,1\}+5|\max\{\kappa,\theta,1\}|^2T)} d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)}(h/T)^{1/2}.\end{aligned} \tag{4.12}$$

Proof of Lemma 4.2. Throughout this proof let $\iota = \max\{\kappa, \theta, 1\}$. Note that (4.8) proves that for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned}|f_0(x) - f_0(y)| &\le \kappa d^{\mathfrak{d}_0}(1 + \|x\|^\theta + \|y\|^\theta)\|x-y\| \\ &\le \kappa d^{\mathfrak{d}_0}\left(1 + 2(\theta+1)\int_0^1 \big[r\|x\| + (1-r)\|y\|\big]^\theta\,dr\right)\|x-y\| \\ &\le 2\kappa(\theta+1)d^{\mathfrak{d}_0}\left(1 + \int_0^1 \big[r\|x\| + (1-r)\|y\|\big]^\theta\,dr\right)\|x-y\|.\end{aligned} \tag{4.13}$$

Lemma 4.1 (with $d = d$, $m = m$, $\xi = x$, $T = T$, $L_0 = 2\kappa(\theta+1)d^{\mathfrak{d}_0}$, $L_1 = \kappa$, $l = \theta$, $h = h$, $B = B$, $\|\cdot\| = \|\cdot\|$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W = W$, $f_0 = f_0$, $f_1 = f_1$, $\chi = \chi$, $X = X^x$, $Y = Y^x$ for $x \in \mathbb{R}^d$ in the notation of Lemma 4.1), (4.10), and (4.9) hence ensure that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}&\big|\mathbb{E}[f_0(X^x_T)] - \mathbb{E}[f_0(Y^x_T)]\big| \le (h/T)^{1/2} e^{(\theta+3+2\kappa+[\theta\kappa+2\kappa+2]T)} \max\{1, 2\kappa(\theta+1)d^{\mathfrak{d}_0}\} \\ &\cdot \left[\|x\| + 2 + \max\{1, \|f_1(0)\|\}\max\{1,T\} + \sqrt{(2\max\{\theta,1\}-1)\operatorname{Trace}(B^*B)T}\right]^{1+\theta} \\ &\le (h/T)^{1/2} e^{(\theta+3+2\kappa+[\theta\kappa+2\kappa+2]T)} \max\{1, 2\kappa(\theta+1)d^{\mathfrak{d}_0}\} \\ &\cdot \left[\|x\| + 2 + \max\{1, \kappa d^{\mathfrak{d}_1}\}\max\{1,T\} + \sqrt{(2\max\{\theta,1\}-1)\kappa d^{2\mathfrak{d}_1}T}\right]^{1+\theta}.\end{aligned} \tag{4.14}$$

Therefore, we obtain that for all $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\left|\mathbb{E}[f_0(X_T^x)] - \mathbb{E}[f_0(Y_T^x)]\right| \\
&\le 2\iota(\iota+1)d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\|x\| + 2 + \iota d^{\mathfrak{d}_1}\max\{1,T\} + \sqrt{(2\iota-1)\kappa d^{2\mathfrak{d}_1}T}\right]^{1+\theta} \\
&\le 4\iota^2 d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\|x\| + 2 + \iota d^{\mathfrak{d}_1}\max\{1,T\} + \sqrt{2\iota\kappa d^{2\mathfrak{d}_1}T}\right]^{1+\theta} \\
&\le 4\iota^2 d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\|x\| + 2 + 3\iota d^{\mathfrak{d}_1}\max\{1,T\}\right]^{1+\theta} \\
&\le 4\iota^2 d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\|x\| + 5\iota d^{\mathfrak{d}_1}\max\{1,T\}\right]^{1+\theta}.
\end{aligned}
\tag{4.15}
$$

This establishes that

$$
\begin{aligned}
&\left[\int_{\mathbb{R}^d}\left|\mathbb{E}[f_0(X_T^x)] - \mathbb{E}[f_0(Y_T^x)]\right|^p \nu(dx)\right]^{1/p} \qquad (4.16)\\
&\le 4\iota^2 d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\int_{\mathbb{R}^d}\left[\|x\| + 5\iota d^{\mathfrak{d}_1}\max\{1,T\}\right]^{p(1+\theta)}\nu(dx)\right]^{1/p} \\
&\le 4\iota^2 d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\int_{\mathbb{R}^d}\left[2^\theta\|x\|^{1+\theta} + 2^\theta(5\iota d^{\mathfrak{d}_1}\max\{1,T\})^{1+\theta}\right]^p \nu(dx)\right]^{1/p} \\
&\le 2^{\theta+2}\iota^2 d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\left[\int_{\mathbb{R}^d}\|x\|^{p(1+\theta)}\nu(dx)\right]^{1/p} + (5\iota d^{\mathfrak{d}_1}\max\{1,T\})^{1+\theta}\right].
\end{aligned}
$$

Combining this and (4.10) assures that

$$
\begin{aligned}
&\left[\int_{\mathbb{R}^d}\left|\mathbb{E}[f_0(X_T^x)] - \mathbb{E}[f_0(Y_T^x)]\right|^p \nu(dx)\right]^{1/p} \\
&\le 2^{\theta+2}\iota^2 d^{\mathfrak{d}_0}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)}\left[\kappa^{1+\theta}d^{\mathfrak{d}_1(1+\theta)} + (5\iota d^{\mathfrak{d}_1}\max\{1,T\})^{1+\theta}\right] \\
&\le 2^{\theta+2}(5\iota\max\{1,T\})^{\theta+1}\iota^2 d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)} \\
&\le 2^{\theta+2+3(\theta+1)}|\max\{1,T\}|^{\theta+1}\iota^{2+\theta+1}d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)}(h/T)^{1/2}e^{(6\iota+5\iota^2 T)} \\
&\le 2^{4\theta+5}|\max\{1,T\}|^{\theta+1}\iota^{\theta+3}e^{(6\iota+5\iota^2 T)}d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)}(h/T)^{1/2}.
\end{aligned}
\tag{4.17}
$$

The proof of Lemma 4.2 is thus completed. □

Lemma 4.3 (Monte Carlo error)**.** *Let $d, M, n \in \mathbb{N}$, $T, \kappa, \theta \in (0,\infty)$, $\mathfrak{d}_0, \mathfrak{d}_1 \in [0,\infty)$, $B \in \mathbb{R}^{d\times n}$, $p \in [2,\infty)$, let $\nu\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure on $\mathbb{R}^d$, 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 $W^m\colon [0,T] \to \mathbb{R}^n$, $m \in \{1,2,\dots,M\}$, be independent standard Brownian motions, let $f_0\colon \mathbb{R}^d \to \mathbb{R}$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$-measurable, let $f_1\colon \mathbb{R}^d \to \mathbb{R}^d$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable, let $\chi\colon [0,T] \to [0,T]$ be $\mathcal{B}([0,T])/\mathcal{B}([0,T])$-measurable, assume for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that*

$$
|f_0(x)| \le \kappa d^{\mathfrak{d}_0}(d^{\mathfrak{d}_1\theta} + \|x\|^\theta), \qquad \|f_1(x)\| \le \kappa(d^{\mathfrak{d}_1} + \|x\|), \tag{4.18}
$$

$$
\mathrm{Trace}(B^*B) \le \kappa d^{2\mathfrak{d}_1}, \qquad \left[\int_{\mathbb{R}^d}\|z\|^{p\theta}\nu(dz)\right]^{1/(p\theta)} \le \kappa d^{\mathfrak{d}_1}, \tag{4.19}
$$

and $\chi(t) \leq t$, and let $Y^{m,x}\colon [0,T]\times\Omega \to \mathbb{R}^d$, $m \in \{1,2,\dots,M\}$, $x \in \mathbb{R}^d$, be stochastic processes with continuous sample paths which satisfy for all $x \in \mathbb{R}^d$, $m \in \{1,2,\dots,M\}$, $t \in [0,T]$ that

$$Y_t^{m,x} = x + \int_0^t f_1\big(Y_{\chi(s)}^{m,x}\big)\,ds + BW_t^m, \tag{4.20}$$

Then it holds that

$$\begin{aligned} &\bigg(\mathbb{E}\bigg[\int_{\mathbb{R}^d} \Big|\mathbb{E}[f_0(Y_T^{1,x})] - \tfrac{1}{M}\Big[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\Big]\Big|^p\,\nu(dx)\bigg]\bigg)^{1/p} \\ &\leq 2^{\theta+2}p\kappa(p\theta+p+1)^\theta(\kappa T+1)^\theta e^{\kappa\theta T}(\kappa^\theta+1)d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2}. \end{aligned} \tag{4.21}$$

Proof of Lemma 4.3. Throughout this proof let $\iota = \max\{\theta,1\}$. Note that (4.18) and, e.g., [32, Lemma 4.1] (with $d = d$, $m = n$, $\xi = \xi$, $p = q$, $c = \kappa$, $C = \kappa d^{\mathfrak{d}_1}$, $T = T$, $B = B$, $\|\cdot\| = \|\cdot\|$, $(\Omega,\mathcal{F},\mathbb{P}) = (\Omega,\mathcal{F},\mathbb{P})$, $W = W^1$, $\mu = f_1$, $\chi = \chi$, $X = Y^{1,x}$ for $q \in [1,\infty)$, $x \in \mathbb{R}^d$ in the notation of [32, Lemma 4.1]) prove that for all $q \in [1,\infty)$, $x \in \mathbb{R}^d$ it holds that

$$\big(\mathbb{E}\big[\|Y_T^{1,x}\|^q\big]\big)^{1/q} \leq \Big(\|x\| + \kappa d^{\mathfrak{d}_1}T + \big(\mathbb{E}\big[\|BW_T^1\|^q\big]\big)^{1/q}\Big)e^{\kappa T}. \tag{4.22}$$

This, (4.19), and, e.g., [32, Lemma 4.2] (with $d = d$, $m = n$, $T = T$, $p = q$, $B = B$, $\|\cdot\| = \|\cdot\|$, $(\Omega,\mathcal{F},\mathbb{P}) = (\Omega,\mathcal{F},\mathbb{P})$, $W = W^1$ for $q \in [1,\infty)$ in the notation of [32, Lemma 4.2]) ensure that for all $q \in [1,\infty)$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \Big(\mathbb{E}\big[\|Y_T^{1,x}\|^q\big]\Big)^{1/q} &\leq \Big(\|x\| + \kappa d^{\mathfrak{d}_1}T + \sqrt{\max\{1,q-1\}\operatorname{Trace}(B^*B)T}\Big)e^{\kappa T} \\ &\leq \Big(\|x\| + \kappa d^{\mathfrak{d}_1}T + \sqrt{\max\{1,q-1\}\kappa d^{2\mathfrak{d}_1}T}\Big)e^{\kappa T} \\ &\leq \big(\|x\| + \kappa d^{\mathfrak{d}_1}T + q\max\{\kappa T,1\}d^{\mathfrak{d}_1}\big)e^{\kappa T} \\ &\leq \big(\|x\| + (q+1)\max\{\kappa T,1\}d^{\mathfrak{d}_1}\big)e^{\kappa T}. \end{aligned} \tag{4.23}$$

Combining this with (4.18) and Hölder's inequality establishes for all $x \in \mathbb{R}^d$ that

$$\begin{aligned} \Big(\mathbb{E}\big[|f_0(Y_T^{1,x})|^p\big]\Big)^{1/p} &\leq \kappa d^{\mathfrak{d}_0+\mathfrak{d}_1\theta} + \kappa d^{\mathfrak{d}_0}\Big(\mathbb{E}\big[\|Y_T^{1,x}\|^{p\theta}\big]\Big)^{1/p} \\ &\leq \kappa d^{\mathfrak{d}_0+\mathfrak{d}_1\theta} + \kappa d^{\mathfrak{d}_0}\Big(\mathbb{E}\big[\|Y_T^{1,x}\|^{p\iota}\big]\Big)^{\theta/(p\iota)} \\ &\leq \kappa d^{\mathfrak{d}_0+\mathfrak{d}_1\theta} + \kappa d^{\mathfrak{d}_0}\big(\|x\| + (p\iota+1)\max\{\kappa T,1\}d^{\mathfrak{d}_1}\big)^\theta e^{\kappa\theta T} \\ &\leq \kappa d^{\mathfrak{d}_0+\mathfrak{d}_1\theta} + \kappa d^{\mathfrak{d}_0}\big(\|x\| + (p\iota+1)(\kappa T+1)d^{\mathfrak{d}_1}\big)^\theta e^{\kappa\theta T}. \end{aligned} \tag{4.24}$$

The fact that $\forall\, y,z \in \mathbb{R}, \alpha \in [0,\infty)\colon |y+z|^\alpha \leq 2^\alpha(|y|^\alpha + |z|^\alpha)$ hence proves for all $x \in \mathbb{R}^d$ that

$$\begin{aligned} \Big(\mathbb{E}\big[|f_0(Y_T^{1,x})|^p\big]\Big)^{1/p} &\leq \kappa d^{\mathfrak{d}_0+\mathfrak{d}_1\theta} + 2^\theta\kappa d^{\mathfrak{d}_0}\big(\|x\|^\theta + (p\iota+1)^\theta(\kappa T+1)^\theta d^{\mathfrak{d}_1\theta}\big)e^{\kappa\theta T} \\ &\leq 2^\theta\kappa(p\iota+1)^\theta(\kappa T+1)^\theta d^{\mathfrak{d}_0}e^{\kappa\theta T}\big(\|x\|^\theta + 2d^{\mathfrak{d}_1\theta}\big) \\ &\leq 2^{\theta+1}\kappa(p\iota+1)^\theta(\kappa T+1)^\theta d^{\mathfrak{d}_0}e^{\kappa\theta T}\big(\|x\|^\theta + d^{\mathfrak{d}_1\theta}\big). \end{aligned} \tag{4.25}$$

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

$$\mathbb{E}\big[|f_0(Y_T^{1,x})|\big] \leq \Big(\mathbb{E}\big[|f_0(Y_T^{1,x})|^p\big]\Big)^{1/p} < \infty. \tag{4.26}$$

Combining this with, e.g., [21, Corollary 2.5] (with $p = p$, $d = 1$, $n = M$, $\|\cdot\| = \|\cdot\|$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $X_i = f_0(Y^{i,x})$ for $i \in \{1, 2, \ldots, M\}$, $x \in \mathbb{R}^d$ in the notation of [21, Corollary 2.5]) and (4.25) assures for all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\Big|\mathbb{E}[f_0(Y_T^{1,x})] - \tfrac{1}{M}\Big[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\Big]\Big|^p\Big]\Big)^{1/p} \\
&\leq 2M^{-1/2}\sqrt{(p-1)}\Big(\mathbb{E}\Big[\big|f_0(Y_T^{1,x}) - \mathbb{E}[f_0(Y_T^{1,x})]\big|^p\Big]\Big)^{1/p} \\
&\leq 4M^{-1/2}\sqrt{(p-1)}\Big(\mathbb{E}\big[|f_0(Y_T^{1,x})|^p\big]\Big)^{1/p} \\
&\leq 2^{\theta+3}M^{-1/2}\sqrt{(p-1)}\kappa(p\iota+1)^\theta(\kappa T+1)^\theta d^{\mathfrak{d}_0}e^{\kappa\theta T}\big(\|x\|^\theta + d^{\mathfrak{d}_1\theta}\big).
\end{aligned} \tag{4.27}$$

This and the fact that $\sqrt{p-1} \leq p/2$ establish that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\Big|\mathbb{E}[f_0(Y_T^{1,x})] - \tfrac{1}{M}\Big[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\Big]\Big|^p \nu(dx)\right]\right)^{1/p} \qquad (4.28)\\
&\leq 2^{\theta+2}M^{-1/2}p\kappa(p\iota+1)^\theta(\kappa T+1)^\theta d^{\mathfrak{d}_0}e^{\kappa\theta T}\left(\int_{\mathbb{R}^d}\big(\|x\|^\theta + d^{\mathfrak{d}_1\theta}\big)^p \nu(dx)\right)^{1/p} \\
&\leq 2^{\theta+2}M^{-1/2}p\kappa(p\iota+1)^\theta(\kappa T+1)^\theta d^{\mathfrak{d}_0}e^{\kappa\theta T}\left(d^{\mathfrak{d}_1\theta} + \left[\int_{\mathbb{R}^d}\|x\|^{p\theta}\,\nu(dx)\right]^{1/p}\right).
\end{aligned}$$

Combining this and (4.19) demonstrates that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\Big|\mathbb{E}[f_0(Y_T^{1,x})] - \tfrac{1}{M}\Big[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\Big]\Big|^p \nu(dx)\right]\right)^{1/p} \\
&\leq 2^{\theta+2}M^{-1/2}p\kappa(p\iota+1)^\theta(\kappa T+1)^\theta d^{\mathfrak{d}_0}e^{\kappa\theta T}\left[d^{\mathfrak{d}_1\theta} + \kappa^\theta d^{\mathfrak{d}_1\theta}\right] \\
&\leq 2^{\theta+2}p\kappa(p\iota+1)^\theta(\kappa T+1)^\theta e^{\kappa\theta T}(\kappa^\theta+1)d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2} \\
&\leq 2^{\theta+2}p\kappa(p\theta+p+1)^\theta(\kappa T+1)^\theta e^{\kappa\theta T}(\kappa^\theta+1)d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2}.
\end{aligned} \tag{4.29}$$

The proof of Lemma 4.3 is thus completed. □

Proposition 4.4. *Let $d, M, n \in \mathbb{N}$, $T, \kappa, \theta \in (0, \infty)$, $\mathfrak{d}_0, \mathfrak{d}_1 \in [0, \infty)$, $h \in (0, T]$, $B \in \mathbb{R}^{d\times n}$, $p \in [2, \infty)$, let $\nu \colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be a probability measure on $\mathbb{R}^d$, 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 $W^m \colon [0, T] \times \Omega \to \mathbb{R}^n$, $m \in \{1, 2, \ldots, M\}$, be independent standard Brownian motions, let $f_0 \colon \mathbb{R}^d \to \mathbb{R}$ and $f_1 \colon \mathbb{R}^d \to \mathbb{R}^d$ be functions, let $\chi \colon [0, T] \to [0, T]$ be a function, assume for all $t \in [0, T]$, $x, y \in \mathbb{R}^d$ that*

$$|f_0(x) - f_0(y)| \leq \kappa d^{\mathfrak{d}_0}(1 + \|x\|^\theta + \|y\|^\theta)\|x - y\|, \tag{4.30}$$

$$|f_0(x)| \leq \kappa d^{\mathfrak{d}_0}(d^{\mathfrak{d}_1\theta} + \|x\|^\theta), \qquad \operatorname{Trace}(B^*B) \leq \kappa d^{2\mathfrak{d}_1}, \tag{4.31}$$

$$\|f_1(x) - f_1(y)\| \leq \kappa\|x - y\|, \qquad \|f_1(x)\| \leq \kappa(d^{\mathfrak{d}_1} + \|x\|), \tag{4.32}$$

$$\left[\int_{\mathbb{R}^d}\|z\|^{p(1+\theta)}\,\nu(dz)\right]^{1/(p(1+\theta))} \leq \kappa d^{\mathfrak{d}_1}, \tag{4.33}$$

and $\chi(t) = \max(\{0, h, 2h, \ldots\} \cap [0,t])$, *and let* $X^x\colon [0,T] \times \Omega \to \mathbb{R}^d$, $x \in \mathbb{R}^d$, *and* $Y^{m,x}\colon [0,T] \times \Omega \to \mathbb{R}^d$, $m \in \{1, 2, \ldots, M\}$, $x \in \mathbb{R}^d$, *be stochastic processes with continuous sample paths which satisfy for all* $x \in \mathbb{R}^d$, $m \in \{1, 2, \ldots, M\}$, $t \in [0,T]$ *that* $X_t^x = x + \int_0^t f_1(X_s^x)\, ds + BW_t^1$ *and*

$$Y_t^{m,x} = x + \int_0^t f_1\big(Y_{\chi(s)}^{m,x}\big)\, ds + BW_t^m. \tag{4.34}$$

Then it holds that

$$\begin{aligned} &\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|\mathbb{E}[f_0(X_T^x)] - \tfrac{1}{M}\left[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\right]\right|^p \nu(dx)\right]\right)^{1/p} \\ &\le 2^{4\theta+5} |\max\{1,T\}|^{\theta+1} |\max\{\kappa,\theta,1\}|^{2\theta+3} e^{(6\max\{\kappa,\theta,1\}+5|\max\{\kappa,\theta,1\}|^2 T)} \\ &\quad \cdot p(p\theta+p+1)^\theta d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)} \big((h/T)^{1/2} + M^{-1/2}\big). \end{aligned} \tag{4.35}$$

Proof of Proposition 4.4. Throughout this proof let $\iota = \max\{\kappa, \theta, 1\}$. Note that the triangle inequality proves that

$$\begin{aligned} &\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|\mathbb{E}[f_0(X_T^x)] - \tfrac{1}{M}\left[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\right]\right|^p \nu(dx)\right]\right)^{1/p} \\ &\le \left(\int_{\mathbb{R}^d} \left|\mathbb{E}[f_0(X_T^x)] - \mathbb{E}[f_0(Y_T^{1,x})]\right|^p \nu(dx)\right)^{1/p} \\ &+ \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|\mathbb{E}[f_0(Y_T^{1,x})] - \tfrac{1}{M}\left[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\right]\right|^p \nu(dx)\right]\right)^{1/p}. \end{aligned} \tag{4.36}$$

Next note that (4.31)–(4.33) and Lemma 4.2 (with $d = d$, $m = n$, $T = T$, $\kappa = \kappa$, $\theta = \theta$, $\mathfrak{d}_0 = \mathfrak{d}_0$, $\mathfrak{d}_1 = \mathfrak{d}_1$, $h = h$, $B = B$, $p = p$, $\nu = \nu$, $\|\cdot\| = \|\cdot\|$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W = W^1$, $f_0 = f_0$, $f_1 = f_1$, $\chi = \chi$, $X^x = X^x$, $Y^x = Y^{1,x}$ for $x \in \mathbb{R}^d$ in the notation of Lemma 4.2) demonstrates that

$$\begin{aligned} &\left(\int_{\mathbb{R}^d} \left|\mathbb{E}[f_0(X_T^x)] - \mathbb{E}[f_0(Y_T^{1,x})]\right|^p \nu(dx)\right)^{1/p} \\ &\le 2^{4\theta+5} |\max\{1,T\}|^{\theta+1} |\max\{\kappa,\theta,1\}|^{\theta+3} e^{(6\max\{\kappa,\theta,1\}+5|\max\{\kappa,\theta,1\}|^2 T)} d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)} (h/T)^{1/2} \\ &= 2^{4\theta+5} |\max\{1,T\}|^{\theta+1} \iota^{\theta+3} e^{(6\iota+5\iota^2 T)} d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)} (h/T)^{1/2}. \end{aligned} \tag{4.37}$$

Moreover, observe that Hölder's inequality and (4.33) imply that

$$\left[\int_{\mathbb{R}^d} \|z\|^{(p\theta)}\, \nu(dz)\right]^{1/p\theta} \le \left[\int_{\mathbb{R}^d} \|z\|^{p(1+\theta)}\, \nu(dz)\right]^{1/(p(1+\theta))} \le \kappa d^{\mathfrak{d}_1}. \tag{4.38}$$

Lemma 4.3 (with $d = d$, $M = M$, $n = n$, $T = T$, $\kappa = \kappa$, $\theta = \theta$, $\mathfrak{d}_0 = \mathfrak{d}_0$, $\mathfrak{d}_1 = \mathfrak{d}_1$, $B = B$, $p = p$, $\nu = \nu$, $\|\cdot\| = \|\cdot\|$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W^m = W^m$, $f_0 = f_0$, $f_1 = f_1$, $\chi = \chi$, $Y^{m,x} = Y^{m,x}$ for $m \in \{1, 2, \ldots, M\}$, $x \in \mathbb{R}^d$ in the notation of Lemma 4.3), (4.31), and (4.32) hence establish that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\left|\mathbb{E}[f_0(Y_T^{1,x})]-\tfrac{1}{M}\left[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\right]\right|^p\nu(dx)\right]\right)^{1/p}\\
&\le 2^{\theta+2}p\kappa(p\theta+p+1)^\theta(\kappa T+1)^\theta e^{\kappa\theta T}(\kappa^\theta+1)d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2}\\
&\le 2^{\theta+2}p\kappa(p\theta+p+1)^\theta|\max\{1,T\}|^\theta(\kappa+1)^\theta e^{\kappa\theta T}(\kappa^\theta+1)d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2}\\
&\le 2^{\theta+2}p\iota(p\theta+p+1)^\theta|\max\{1,T\}|^\theta(2\iota)^\theta e^{\iota\theta T}(\iota^\theta+1)d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2}\\
&\le 2^{2\theta+3}p\iota^{2\theta+1}(p\theta+p+1)^\theta|\max\{1,T\}|^\theta e^{\iota\theta T}d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2}.
\end{aligned}\tag{4.39}$$

This, (4.36), and (4.37) assure that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\left|\mathbb{E}[f_0(X_T^{x})]-\tfrac{1}{M}\left[\textstyle\sum_{m=1}^M f_0(Y_T^{m,x})\right]\right|^p\nu(dx)\right]\right)^{1/p}\\
&\le 2^{4\theta+5}|\max\{1,T\}|^{\theta+1}\iota^{\theta+3}e^{(6\iota+5\iota^2T)}d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)}(h/T)^{1/2}\\
&\quad+2^{2\theta+3}p\iota^{2\theta+1}(p\theta+p+1)^\theta|\max\{1,T\}|^\theta e^{\iota\theta T}d^{\mathfrak{d}_0+\mathfrak{d}_1\theta}M^{-1/2}\\
&\le 2^{4\theta+5}|\max\{1,T\}|^{\theta+1}\iota^{2\theta+3}e^{(6\iota+5\iota^2T)}d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)}\\
&\quad\cdot\left((h/T)^{1/2}+p(p\theta+p+1)^\theta M^{-1/2}\right)\\
&\le 2^{4\theta+5}|\max\{1,T\}|^{\theta+1}\iota^{2\theta+3}e^{(6\iota+5\iota^2T)}d^{\mathfrak{d}_0+\mathfrak{d}_1(\theta+1)}p(p\theta+p+1)^\theta\\
&\quad\cdot\left((h/T)^{1/2}+M^{-1/2}\right).
\end{aligned}\tag{4.40}$$

The proof of Proposition 4.4 is thus completed. □

4.2 DNN approximations for Kolmogorov PDEs

Theorem 4.5. *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, let $\|\cdot\|\colon(\cup_{d\in\mathbb{N}}\mathbb{R}^d)\to[0,\infty)$ satisfy for all $d\in\mathbb{N}$, $x=(x_1,x_2,\ldots,x_d)\in\mathbb{R}^d$ that $\|x\|=(\sum_{i=1}^d|x_i|^2)^{1/2}$, for every $d\in\mathbb{N}$ let $\nu_d\colon\mathcal{B}(\mathbb{R}^d)\to[0,1]$ be a probability measure on $\mathbb{R}^d$, let $\varphi_{0,d}\colon\mathbb{R}^d\to\mathbb{R}$, $d\in\mathbb{N}$, and $\varphi_{1,d}\colon\mathbb{R}^d\to\mathbb{R}^d$, $d\in\mathbb{N}$, be functions, let $T,\kappa\in(0,\infty)$, $\mathfrak{e},\mathfrak{d}_1,\mathfrak{d}_2,\ldots,\mathfrak{d}_6\in[0,\infty)$, $\theta\in[1,\infty)$, $p\in[2,\infty)$, $(\phi_\varepsilon^{m,d})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]}\subseteq\mathbf{N}$, $a\in C(\mathbb{R},\mathbb{R})$ satisfy for all $x\in\mathbb{R}$ that $a(x)=\max\{x,0\}$, assume for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$, $m\in\{0,1\}$, $x,y\in\mathbb{R}^d$ that $\mathcal{R}_a(\phi_\varepsilon^{0,d})\in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{R}_a(\phi_\varepsilon^{1,d})\in C(\mathbb{R}^d,\mathbb{R}^d)$, $\operatorname{Trace}(A_d)\le\kappa d^{2\mathfrak{d}_1}$, $[\int_{\mathbb{R}^d}\|x\|^{2p\theta}\,\nu_d(dx)]^{1/(2p\theta)}\le\kappa d^{\mathfrak{d}_1+\mathfrak{d}_2}$, $\mathcal{P}(\phi_\varepsilon^{m,d})\le\kappa d^{2^{(-m)}\mathfrak{d}_3}\varepsilon^{-2^{(-m)}\mathfrak{e}}$, $|(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x)-(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(y)|\le\kappa d^{\mathfrak{d}_6}(1+\|x\|^\theta+\|y\|^\theta)\|x-y\|$, $\|(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\|\le\kappa(d^{\mathfrak{d}_1+\mathfrak{d}_2}+\|x\|)$, $|\varphi_{0,d}(x)|\le\kappa d^{\mathfrak{d}_6}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}+\|x\|^\theta)$, $\|\varphi_{1,d}(x)-\varphi_{1,d}(y)\|\le\kappa\|x-y\|$, and*

$$\|\varphi_{m,d}(x)-(\mathcal{R}_a(\phi_\varepsilon^{m,d}))(x)\|\le\varepsilon\kappa d^{\mathfrak{d}_{(5-m)}}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)}+\|x\|^\theta),\tag{4.41}$$

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

$$(\tfrac{\partial}{\partial t}u_d)(t,x)=(\tfrac{\partial}{\partial x}u_d)(t,x)\,\varphi_{1,d}(x)+\sum_{i,j=1}^d A_{d,i,j}\,(\tfrac{\partial^2}{\partial x_i\partial x_j}u_d)(t,x)\tag{4.42}$$

with $u_d(0,x) = \varphi_{0,d}(x)$ *for* $(t,x) \in (0,T) \times \mathbb{R}^d$ *(cf. Definition 3.1 and Definition 3.3). Then there exist* $c \in \mathbb{R}$ *and* $(\Psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ *it holds that* $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}(\Psi_{d,\varepsilon}))(x)|^p \, \nu_d(dx)]^{1/p} \le \varepsilon$, *and*

$$\mathcal{P}(\Psi_{d,\varepsilon}) \le c d^{6[\mathfrak{d}_6+(\mathfrak{d}_1+\mathfrak{d}_2)(\theta+1)]+\max\{4,\mathfrak{d}_3\}+\mathfrak{c}\max\{\mathfrak{d}_5+\theta(\mathfrak{d}_1+\mathfrak{d}_2),\mathfrak{d}_4+\mathfrak{d}_6+2\theta(\mathfrak{d}_1+\mathfrak{d}_2)\}} \varepsilon^{-(\mathfrak{c}+6)}. \tag{4.43}$$

Proof of Theorem 4.5. Throughout this proof 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 $(\Omega, \mathcal{F}, \mathbb{P})$ 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 $Z_n^{N,d,m}\colon \Omega \to \mathbb{R}^d$, $n \in \{0, 1, \dots, N-1\}$, $m \in \{1, 2, \dots, N\}$, $d, N \in \mathbb{N}$, be the random variables which satisfy for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \dots, N\}$, $n \in \{0, 1, \dots, N-1\}$ that

$$Z_n^{N,d,m} = \mathcal{A}_d W^{d,m}_{\frac{(n+1)T}{N}} - \mathcal{A}_d W^{d,m}_{\frac{nT}{N}}, \tag{4.44}$$

let $f_{N,d}\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, $d, N \in \mathbb{N}$, satisfy for all $N, d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$ that

$$f_{N,d}(x,y) = x + y + \tfrac{T}{N}\varphi_{1,d}(y), \tag{4.45}$$

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

$$X_t^{d,x} = x + \int_0^t \varphi_{1,d}(X_s^{d,x}) \, ds + \mathcal{A}_d W_t^{d,1} \tag{4.46}$$

(cf., e.g., [32, item (i) in Theorem 3.1] (with $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $T = T$, $d = d$, $m = d$, $B = \mathcal{A}_d$, $\mu = \varphi_{1,d}$ for $d \in \mathbb{N}$ in the notation of [32, Theorem 3.1])), let $Y_n^{N,d,x} = (Y_n^{N,d,m,x})_{m\in\{1,2,\dots,N\}}\colon \Omega \to \mathbb{R}^{Nd}$, $n \in \{0, 1, \dots, N\}$, $x \in \mathbb{R}^d$, $d, N \in \mathbb{N}$, satisfy for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \dots, N\}$, $x \in \mathbb{R}^d$, $n \in \{1, 2, \dots, N\}$ that $Y_0^{N,d,m,x} = x$ and

$$Y_n^{N,d,m,x} = f_{N,d}\big(Z_{n-1}^{N,d,m}, Y_{n-1}^{N,d,m,x}\big), \tag{4.47}$$

let $g_{N,d}\colon \mathbb{R}^{Nd} \to \mathbb{R}$, $d, N \in \mathbb{N}$, satisfy for all $N, d \in \mathbb{N}$, $x = (x_i)_{i\in\{1,2,\dots,N\}} \in \mathbb{R}^{Nd}$ that

$$g_{N,d}(x) = \frac{1}{N} \sum_{i=1}^N \varphi_{0,d}(x_i), \tag{4.48}$$

and let $\mathfrak{N}_{d,\varepsilon} \subseteq \mathbf{N}$, $\varepsilon \in (0,1]$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\begin{aligned} &\mathfrak{N}_{d,\varepsilon} \\ &= \Big\{ \Phi \in \mathbf{N}\colon \big[\big(\mathcal{R}_a(\Phi) \in C(\mathbb{R}^d, \mathbb{R}^d) \big) \wedge \big(\mathbb{D}_{\mathcal{L}(\Phi)-1}(\Phi) \le \mathbb{D}_{\mathcal{L}(\phi_\varepsilon^{1,d})-1}(\phi_\varepsilon^{1,d}) + 2d \big) \big] \Big\} \end{aligned} \tag{4.49}$$

(cf. Definition 3.27). Note that (4.44) and, e.g., [32, Lemma 4.2] (with $d = d$, $m = d$, $T = T$, $p = 2p\theta$, $B = \mathcal{A}_d$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W = W^{d,m}$ for $d, m \in \mathbb{N}$ in the notation of [32, Lemma 4.2]) ensure that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \dots, N\}$, $n \in \{0, 1, \dots, N-1\}$ it holds that

$$\begin{aligned} &\big(\mathbb{E}\big[\|Z_n^{N,d,m}\|^{2p\theta}\big]\big)^{1/(2p\theta)} = \left(\mathbb{E}\left[\left\| \mathcal{A}_d W^{d,m}_{\frac{(n+1)T}{N}} - \mathcal{A}_d W^{d,m}_{\frac{nT}{N}} \right\|^{2p\theta} \right] \right)^{1/(2p\theta)} \\ &\le \left(\mathbb{E}\left[\left\| \mathcal{A}_d W^{d,m}_{\frac{(n+1)T}{N}} \right\|^{2p\theta} \right] \right)^{1/(2p\theta)} + \left(\mathbb{E}\left[\left\| \mathcal{A}_d W^{d,m}_{\frac{nT}{N}} \right\|^{2p\theta} \right] \right)^{1/(2p\theta)} \\ &\le 2\sqrt{(2p\theta - 1)\operatorname{Trace}(\mathcal{A}_d^* \mathcal{A}_d) T} = 2\sqrt{2(2p\theta-1)\operatorname{Trace}(A_d) T}. \end{aligned} \tag{4.50}$$

This and the assumption that $\forall\, d \in \mathbb{N}\colon \operatorname{Trace}(A_d) \le \kappa d^{2\mathfrak{d}_1}$ assure for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \dots, N\}$, $n \in \{0, 1, \dots, N-1\}$ that

$$\big(\mathbb{E}\big[\|Z_n^{N,d,m}\|^{2p\theta}\big]\big)^{1/(2p\theta)} \le 4p\theta\sqrt{\kappa T} d^{\mathfrak{d}_1}. \tag{4.51}$$

Moreover, observe that Lemma 3.16 (with $d = d$, $a = a$ for $d \in \mathbb{N}$ in the notation of Lemma 3.16) ensures that there exist $\mathfrak{I}_d \in \mathbf{N}$, $d \in \mathbb{N}$, such that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d)$, $\mathcal{R}_a(\mathfrak{I}_d) \in C(\mathbb{R}^d, \mathbb{R}^d)$, and $(\mathcal{R}_a(\mathfrak{I}_d))(x) = x$. This and (4.49) assure for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $\mathfrak{I}_d \in \mathfrak{N}_{d,\varepsilon}$ and

$$\mathcal{P}(\mathfrak{I}_d) = 2d(d+1) + d(2d+1) = 2d^2 + 2d + 2d^2 + d = 4d^2 + 3d \le 7d^2. \tag{4.52}$$

Next note that Lemma 3.14 demonstrates that for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathcal{D}(\frac{T}{N} \circledast \phi_\varepsilon^{1,d}) = \mathcal{D}(\phi_\varepsilon^{1,d})$, $\mathcal{R}_a(\frac{T}{N} \circledast \phi_\varepsilon^{1,d}) \in C(\mathbb{R}^d, \mathbb{R}^d)$, and

$$\mathcal{R}_a\big(\tfrac{T}{N} \circledast \phi_\varepsilon^{1,d}\big) = \tfrac{T}{N}\mathcal{R}_a(\phi_\varepsilon^{1,d}) \tag{4.53}$$

(cf. Definition 3.13). This, the fact that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d)$, and Lemma 3.30 (with $a = a$, $L_1 = \mathcal{L}(\frac{T}{N} \circledast \phi_\varepsilon^{1,d})$, $L_2 = 2$, $\mathbb{I} = \mathfrak{I}_d$, $\Phi_1 = \frac{T}{N} \circledast \phi_\varepsilon^{1,d}$, $\Phi_2 = \mathfrak{I}_d$, $d = d$, $\mathfrak{i} = 2d$, $(l_{1,0}, l_{1,1}, \dots, l_{1,L_1}) = \mathcal{D}(\frac{T}{N} \circledast \phi_\varepsilon^{1,d})$, $(l_{2,0}, l_{2,1}, l_{2,L_2}) = (d, 2d, d)$ for $d, N \in \mathbb{N}$, $\varepsilon \in (0, 1]$ in the notation of Lemma 3.30) establish that there exist $\mathbf{f}_\varepsilon^{N,d} \in \mathbf{N}$, $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, such that for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\mathbf{f}_\varepsilon^{N,d}) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and

$$(\mathcal{R}_a(\mathbf{f}_\varepsilon^{N,d}))(x) = x + \big(\mathcal{R}_a\big(\tfrac{T}{N} \circledast \phi_\varepsilon^{1,d}\big)\big)(x) = x + \tfrac{T}{N}(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x). \tag{4.54}$$

Items (ii)–(iii) in Lemma 3.9 hence ensure that there exist $\mathbf{f}_{\varepsilon,z}^{N,d} \in \mathbf{N}$, $z \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, which satisfy for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $z, x \in \mathbb{R}^d$ that $\mathcal{R}_a(\mathbf{f}_{\varepsilon,z}^{N,d}) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and

$$(\mathcal{R}_a(\mathbf{f}_{\varepsilon,z}^{N,d}))(x) = (\mathcal{R}_a(\mathbf{f}_\varepsilon^{N,d}))(x) + z = z + x + \tfrac{T}{N}(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x). \tag{4.55}$$

This, (4.45), and (4.41) imply for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, z \in \mathbb{R}^d$ that $(\mathbb{R}^d \ni \mathfrak{z} \mapsto (\mathcal{R}_a(\mathbf{f}_{\varepsilon,\mathfrak{z}}^{N,d}))(x) \in \mathbb{R}^d)$ is $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable and

$$\begin{aligned} \|f_{N,d}(z, x) - (\mathcal{R}_a(\mathbf{f}_{\varepsilon,z}^{N,d}))(x)\| &= \tfrac{T}{N}\|\varphi_{1,d}(x) - (\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\| \\ &\le \tfrac{T\varepsilon\kappa d^{\mathfrak{d}_4}}{N}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta) \\ &\le \varepsilon T\kappa d^{\mathfrak{d}_4}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta). \end{aligned} \tag{4.56}$$

Next note that (4.55) and the assumption that $\forall\, \varepsilon \in (0, 1], d \in \mathbb{N}, x \in \mathbb{R}^d\colon \|(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\| \le \kappa(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|x\|)$ prove for all $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, z \in \mathbb{R}^d$ that

$$\begin{aligned} \|(\mathcal{R}_a(\mathbf{f}_{\varepsilon,z}^{N,d}))(x)\| &\le \|z\| + \|x\| + \tfrac{T}{N}\|(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\| \\ &\le \|z\| + \|x\| + \tfrac{T\kappa}{N}(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|x\|) \\ &= \big(1 + \tfrac{T\kappa}{N}\big)\|x\| + \tfrac{T\kappa d^{\mathfrak{d}_1+\mathfrak{d}_2}}{N} + \|z\| \\ &\le \big(1 + \tfrac{T\kappa}{N}\big)\|x\| + (T\kappa + 1)(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|z\|) \\ &\le \big(1 + \tfrac{T\kappa}{N}\big)\|x\| + (T\kappa + 1)d^{\mathfrak{d}_2}(d^{\mathfrak{d}_1} + \|z\|). \end{aligned} \tag{4.57}$$

In addition, observe that (4.45) and the assumption that $\forall\, d \in \mathbb{N}, x, y \in \mathbb{R}^d\colon \|\varphi_{1,d}(x) - \varphi_{1,d}(y)\| \leq \kappa\|x-y\|$ imply that for all $N, d \in \mathbb{N}$, $x, z \in \mathbb{R}^d$ it holds that

$$\begin{aligned}\|f_{N,d}(z,x) - f_{N,d}(z,y)\| &= \|x + \tfrac{T}{N}\varphi_{1,d}(x) - y - \tfrac{T}{N}\varphi_{1,d}(y)\| \\ &\leq \|x-y\| + \tfrac{T}{N}\|\varphi_{1,d}(x) - \varphi_{1,d}(y)\| \\ &\leq \big(1 + \tfrac{T\kappa}{N}\big)\|x-y\| \leq (1+T\kappa)\|x-y\|.\end{aligned} \tag{4.58}$$

Moreover, note that (4.53), the fact that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d)$, and Lemma 3.30 (with $a = a$, $L_1 = \mathcal{L}(\frac{T}{N} \circledast \phi_\varepsilon^{1,d})$, $L_2 = \mathcal{L}(\Phi)$, $\mathbb{I} = \mathfrak{I}_d$, $\Phi_1 = \frac{T}{N} \circledast \phi_\varepsilon^{1,d}$, $\Phi_2 = \Phi$, $d = d$, $\mathfrak{i} = 2d$, $(l_{1,0}, l_{1,1}, \ldots, l_{1,L_1}) = \mathcal{D}(\frac{T}{N} \circledast \phi_\varepsilon^{1,d}) = \mathcal{D}(\phi_\varepsilon^{1,d})$, $(l_{2,0}, l_{2,1}, \ldots, l_{2,L_2}) = \mathcal{D}(\Phi)$ for $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi \in \mathfrak{N}_{d,\varepsilon}$ in the notation of Lemma 3.30) prove that for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi \in \mathfrak{N}_{d,\varepsilon}$ there exists $\hat{\Phi} \in \mathbf{N}$ such that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\hat{\Phi}) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\mathbb{D}_{\mathcal{L}(\hat{\Phi})-1}(\hat{\Phi}) \leq \mathbb{D}_{\mathcal{L}(\phi_\varepsilon^{1,d})-1}(\phi_\varepsilon^{1,d}) + 2d$, $\mathcal{P}(\hat{\Phi}) \leq \mathcal{P}(\Phi) + [\frac{1}{2}\mathcal{P}(\mathfrak{I}_d) + \mathcal{P}(\phi_\varepsilon^{1,d})]^2$, and

$$\begin{aligned}(\mathcal{R}_a(\hat{\Phi}))(x) &= (\mathcal{R}_a(\Phi))(x) + \big(\big(\mathcal{R}_a\big(\tfrac{T}{N} \circledast \phi_\varepsilon^{1,d}\big)\big) \circ (\mathcal{R}_a(\Phi))\big)(x) \\ &= (\mathcal{R}_a(\Phi))(x) + \tfrac{T}{N}\big((\mathcal{R}_a(\phi_\varepsilon^{1,d})) \circ (\mathcal{R}_a(\Phi))\big)(x).\end{aligned} \tag{4.59}$$

This, (4.49), (4.52), and the fact that $\forall\, d \in \mathbb{N}, \varepsilon \in (0,1]\colon \mathcal{P}(\phi_\varepsilon^{1,d}) \leq \kappa d^{2^{(-1)}\mathfrak{d}_3}\varepsilon^{-2^{(-1)}\mathfrak{e}}$ demonstrate that for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi \in \mathfrak{N}_{d,\varepsilon}$ there exists $\hat{\Phi} \in \mathfrak{N}_{d,\varepsilon}$ such that for all $x \in \mathbb{R}^d$ it holds that

$$\mathcal{P}(\hat{\Phi}) \leq \mathcal{P}(\Phi) + (4d^2 + \kappa d^{2^{(-1)}\mathfrak{d}_3}\varepsilon^{-2^{(-1)}\mathfrak{e}})^2 \leq \mathcal{P}(\Phi) + (\kappa+4)^2 d^{\max\{4,\mathfrak{d}_3\}}\varepsilon^{-\mathfrak{e}} \tag{4.60}$$

and

$$(\mathcal{R}_a(\hat{\Phi}))(x) = (\mathcal{R}_a(\Phi))(x) + \tfrac{T}{N}\big((\mathcal{R}_a(\phi_\varepsilon^{1,d})) \circ (\mathcal{R}_a(\Phi))\big)(x). \tag{4.61}$$

Items (i)–(iii) in Lemma 3.9 and (4.55) hence ensure that for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi \in \mathfrak{N}_{d,\varepsilon}$ there exist $(\hat{\Phi}_z)_{z \in \mathbb{R}^d} \subseteq \mathfrak{N}_{d,\varepsilon}$ such that for all $x, z, \mathfrak{z} \in \mathbb{R}^d$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\hat{\Phi}_z))(x) &= z + (\mathcal{R}_a(\Phi))(x) + \tfrac{T}{N}\big((\mathcal{R}_a(\phi_\varepsilon^{1,d})) \circ (\mathcal{R}_a(\Phi))\big)(x) \\ &= (\mathcal{R}_a(\mathbf{f}_{\varepsilon,z}^{N,d}))\big((\mathcal{R}_a(\Phi))(x)\big),\end{aligned} \tag{4.62}$$

$$\mathcal{P}(\hat{\Phi}_z) \leq \mathcal{P}(\Phi) + (\kappa+4)^2 d^{\max\{4,\mathfrak{d}_3\}}\varepsilon^{-\mathfrak{e}}, \tag{4.63}$$

and $\mathcal{D}(\hat{\Phi}_z) = \mathcal{D}(\hat{\Phi}_\mathfrak{z})$. In the next step we observe that Lemma 3.29 (with $n = N$, $h_m = 1/N$, $\phi_m = \phi_\varepsilon^{0,d}$, $a = a$ for $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $m \in \{1, 2, \ldots, N\}$ in the notation of Lemma 3.29) demonstrates that there exist $\mathbf{g}_\varepsilon^{N,d} \in \mathbf{N}$, $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, which satisfy for all $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x = (x_i)_{i \in \{1,2,\ldots,N\}} \in \mathbb{R}^{Nd}$ that $\mathcal{R}_a(\mathbf{g}_\varepsilon^{N,d}) \in C(\mathbb{R}^{Nd}, \mathbb{R})$ and

$$(\mathcal{R}_a(\mathbf{g}_\varepsilon^{N,d}))(x) = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x_i). \tag{4.64}$$

This, (4.48), and (4.41) ensure that for all $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x = (x_i)_{i \in \{1,2,\ldots,N\}} \in \mathbb{R}^{Nd}$ it holds that

$$\begin{aligned}&|g_{N,d}(x) - (\mathcal{R}_a(\mathbf{g}_\varepsilon^{N,d}))(x)| \leq \frac{1}{N}\sum_{i=1}^{N}|\varphi_{0,d}(x_i) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x_i)| \\ &\leq \frac{\varepsilon\kappa d^{\mathfrak{d}_5}}{N}\sum_{i=1}^{N}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x_i\|^\theta) = \varepsilon\kappa d^{\mathfrak{d}_5}\left[d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\|x_i\|^\theta\right].\end{aligned} \tag{4.65}$$

Moreover, note that (4.64) and the assumption that $\forall \varepsilon \in (0,1], d \in \mathbb{N}, x, y \in \mathbb{R}^d \colon |(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(y)| \leq \kappa d^{\mathfrak{d}_6}(1 + \|x\|^\theta + \|y\|^\theta)\|x - y\|$ imply that for all $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x = (x_i)_{i \in \{1,2,\ldots,N\}} \in \mathbb{R}^{Nd}$, $y = (y_i)_{i \in \{1,2,\ldots,N\}} \in \mathbb{R}^{Nd}$ it holds that

$$\begin{aligned} &|(\mathcal{R}_a(\mathbf{g}_\varepsilon^{N,d}))(x) - (\mathcal{R}_a(\mathbf{g}_\varepsilon^{N,d}))(y)| \\ &\leq \frac{1}{N} \sum_{i=1}^N |(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x_i) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(y_i)| \\ &\leq \frac{\kappa d^{\mathfrak{d}_6}}{N} \left[\textstyle\sum_{i=1}^N (1 + \|x_i\|^\theta + \|y_i\|^\theta)\|x_i - y_i\| \right]. \end{aligned} \tag{4.66}$$

Next observe that the fact that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d)$ and, e.g., [22, Proposition 2.16] (with $\Psi = \mathfrak{I}_d$, $\Phi_1 = \phi_\varepsilon^{0,d}$, $\Phi_2 \in \{\Phi \in \mathbf{N} \colon \mathcal{I}(\Phi) = \mathcal{O}(\Phi) = d\}$, $\mathfrak{i} = 2d$ in the notation of [22, Proposition 2.16]) prove that for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi_1, \Phi_2, \ldots, \Phi_N \in \{\Phi \in \mathbf{N} \colon \mathcal{I}(\Phi) = \mathcal{O}(\Phi) = d\}$ with $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_N)$ there exist $\Psi_1, \Psi_2, \ldots, \Psi_N \in \mathbf{N}$ such that for all $i \in \{1, 2, \ldots, N\}$ it holds that $\mathcal{R}_a(\Psi_i) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{D}(\Psi_i) = \mathcal{D}(\Psi_1)$, $\mathcal{P}(\Psi_i) \leq 2(\mathcal{P}(\phi_\varepsilon^{0,d}) + \mathcal{P}(\Phi_i))$, and

$$\mathcal{R}_a(\Psi_i) = [\mathcal{R}_a(\phi_\varepsilon^{0,d})] \circ [\mathcal{R}_a(\mathfrak{I}_d)] \circ [\mathcal{R}_a(\Phi_i)] = [\mathcal{R}_a(\phi_\varepsilon^{0,d})] \circ [\mathcal{R}_a(\Phi_i)]. \tag{4.67}$$

This, (4.64), and Lemma 3.28 assure that for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi_1, \Phi_2, \ldots, \Phi_N \in \{\Phi \in \mathbf{N} \colon \mathcal{I}(\Phi) = \mathcal{O}(\Phi) = d\}$ with $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_N)$ there exists $\Psi \in \mathbf{N}$ such that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\Psi) \leq 2N^2(\mathcal{P}(\phi_\varepsilon^{0,d}) + \mathcal{P}(\Phi_1))$, and

$$\begin{aligned} (\mathcal{R}_a(\Psi))(x) &= \frac{1}{N} \sum_{i=1}^N (\mathcal{R}_a(\phi_\varepsilon^{0,d}))\big((\mathcal{R}_a(\Phi_i))(x)\big) \\ &= (\mathcal{R}_a(\mathbf{g}_\varepsilon^{N,d}))\big((\mathcal{R}_a(\Phi_1))(x), (\mathcal{R}_a(\Phi_2))(x), \ldots, (\mathcal{R}_a(\Phi_N))(x)\big). \end{aligned} \tag{4.68}$$

The assumption that $\forall d \in \mathbb{N}, \varepsilon \in (0,1] \colon \mathcal{P}(\phi_\varepsilon^{0,d}) \leq \kappa d^{\mathfrak{d}_3} \varepsilon^{-\mathfrak{e}}$ hence ensures that for every $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\Phi_1, \Phi_2, \ldots, \Phi_N \in \{\Phi \in \mathbf{N} \colon \mathcal{I}(\Phi) = \mathcal{O}(\Phi) = d\}$ with $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_N)$ there exists $\Psi \in \mathbf{N}$ such that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R})$, $(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\mathbf{g}_\varepsilon^{N,d}))((\mathcal{R}_a(\Phi_1))(x), (\mathcal{R}_a(\Phi_2))(x), \ldots, (\mathcal{R}_a(\Phi_N))(x))$, and

$$\mathcal{P}(\Psi) \leq 2N^2(\kappa d^{\mathfrak{d}_3} \varepsilon^{-\mathfrak{e}} + \mathcal{P}(\Phi_1)) \leq 2 \max\{\kappa, 1\} N^2 (d^{\mathfrak{d}_3} \varepsilon^{-\mathfrak{e}} + \mathcal{P}(\Phi_1)). \tag{4.69}$$

Furthermore, note that (4.41) and the assumption that $\forall d \in \mathbb{N}, \varepsilon \in (0,1], x, y \in \mathbb{R}^d \colon |(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(y)| \leq \kappa d^{\mathfrak{d}_6}(1 + \|x\|^\theta + \|y\|^\theta)\|x - y\|$ demonstrate for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} &|\varphi_{0,d}(x) - \varphi_{0,d}(y)| \\ &\leq |\varphi_{0,d}(x) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x)| + |\varphi_{0,d}(y) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(y)| \\ &\quad + |(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(y)| \\ &\leq \varepsilon \kappa d^{\mathfrak{d}_5}(d^{\theta(\mathfrak{d}_1 + \mathfrak{d}_2)} + \|x\|^\theta) + \varepsilon \kappa d^{\mathfrak{d}_5}(d^{\theta(\mathfrak{d}_1 + \mathfrak{d}_2)} + \|y\|^\theta) \\ &\quad + \kappa d^{\mathfrak{d}_6}(1 + \|x\|^\theta + \|y\|^\theta)\|x - y\|. \end{aligned} \tag{4.70}$$

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

$$|\varphi_{0,d}(x) - \varphi_{0,d}(y)| \leq \kappa d^{\mathfrak{d}_6}(1 + \|x\|^\theta + \|y\|^\theta)\|x - y\|. \tag{4.71}$$

Next observe that the assumption that $\forall\, d \in \mathbb{N}, \varepsilon \in (0,1], x \in \mathbb{R}^d \colon \|(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\| \leq \kappa(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|x\|)$ and (4.41) ensure for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} \|\varphi_{1,d}(x)\| &\leq \|\varphi_{1,d}(x) - (\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\| + \|(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\| \\ &\leq \varepsilon\kappa d^{\mathfrak{d}_4}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta) + \kappa(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|x\|). \end{aligned} \tag{4.72}$$

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

$$\|\varphi_{1,d}(x)\| \leq \kappa(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|x\|). \tag{4.73}$$

In the next step we note that the Hölder's inequality, the assumption that $\forall\, d \in \mathbb{N} \colon [\int_{\mathbb{R}^d} \|x\|^{2p\theta}\, \nu_d(dx)]^{1/(2p\theta)} \leq \kappa d^{\mathfrak{d}_1+\mathfrak{d}_2}$, and the assumption that $\theta \in [1,\infty)$ assure that for all $d \in \mathbb{N}$ it holds that

$$\left[\int_{\mathbb{R}^d} \|x\|^{p(1+\theta)}\, \nu_d(dx)\right]^{1/(p(1+\theta))} \leq \left[\int_{\mathbb{R}^d} \|x\|^{2p\theta}\, \nu_d(dx)\right]^{1/(2p\theta)} \leq \kappa d^{\mathfrak{d}_1+\mathfrak{d}_2}. \tag{4.74}$$

Next note that (4.47), (4.45), and (4.44) imply that for all $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, N\}$, $x \in \mathbb{R}^d$, $n \in \{1, 2, \ldots, N\}$ it holds that

$$\begin{aligned} Y_n^{N,d,m,x} &= Z_{n-1}^{N,d,m} + Y_{n-1}^{N,d,m,x} + \tfrac{T}{N}\varphi_{1,d}(Y_{n-1}^{N,d,m,x}) \\ &= Y_{n-1}^{N,d,m,x} + \tfrac{T}{N}\varphi_{1,d}(Y_{n-1}^{N,d,m,x}) + \mathcal{A}_d W^{d,m}_{\frac{nT}{N}} - \mathcal{A}_d W^{d,m}_{\frac{(n-1)T}{N}}. \end{aligned} \tag{4.75}$$

The assumption that $\forall\, d \in \mathbb{N}, x \in \mathbb{R}^d \colon |\varphi_{0,d}(x)| \leq \kappa d^{\mathfrak{d}_6}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta)$, the assumption that $\forall\, d \in \mathbb{N} \colon \operatorname{Trace}(A_d) \leq \kappa d^{2\mathfrak{d}_1}$, the assumption that $\forall\, d \in \mathbb{N}, x, y \in \mathbb{R}^d \colon \|\varphi_{1,d}(x) - \varphi_{1,d}(y)\| \leq \kappa\|x - y\|$, (4.71), (4.73), (4.74), (4.46), and Proposition 4.4 (with $d = d$, $M = N$, $n = d$, $T = T$, $\kappa = \kappa$, $\theta = \theta$, $\mathfrak{d}_0 = \mathfrak{d}_6$, $\mathfrak{d}_1 = \mathfrak{d}_1 + \mathfrak{d}_2$, $h = T/N$, $B = \mathcal{A}_d$, $p = p$, $\nu = \nu_d$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $W^m = W^{d,m}$, $f_0 = \varphi_{0,d}$, $f_1 = \varphi_{1,d}$ for $N, d \in \mathbb{N}$ in the notation of Proposition 4.4) hence establish that for all $N, d \in \mathbb{N}$ it holds that

$$\begin{aligned} &\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|\mathbb{E}[\varphi_{0,d}(X_T^{d,x})] - \tfrac{1}{N}\left[\textstyle\sum_{i=1}^N \varphi_{0,d}(Y_N^{N,d,i,x})\right]\right|^p \nu_d(dx)\right]\right)^{1/p} \\ &\leq 2^{4\theta+5}|\max\{1,T\}|^{\theta+1}|\max\{\kappa,\theta,1\}|^{2\theta+3}e^{(6\max\{\kappa,\theta,1\}+5|\max\{\kappa,\theta,1\}|^2T)} \\ &\quad \cdot p(p\theta + p + 1)^\theta d^{\mathfrak{d}_6+(\mathfrak{d}_1+\mathfrak{d}_2)(\theta+1)}\big(N^{-1/2} + N^{-1/2}\big) \\ &= 2^{4\theta+6}|\max\{1,T\}|^{\theta+1}|\max\{\kappa,\theta\}|^{2\theta+3}e^{(6\max\{\kappa,\theta\}+5|\max\{\kappa,\theta\}|^2T)} \\ &\quad \cdot p(p\theta + p + 1)^\theta d^{\mathfrak{d}_6+(\mathfrak{d}_1+\mathfrak{d}_2)(\theta+1)}N^{-1/2}. \end{aligned} \tag{4.76}$$

This, the fact that $\forall\, d \in \mathbb{N}, x \in \mathbb{R}^d \colon |\varphi_{0,d}(x)| \leq \kappa d^{\mathfrak{d}_6}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta)$, (4.73), (4.48), and, e.g., [32, Theorem 3.1] (with $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $T = T$, $d = d$, $m = d$,

$B = \mathcal{A}_d$, $\varphi = \varphi_{0,d}$, $\mu = \varphi_{1,d}$ for $d \in \mathbb{N}$ in the notation of [32, Theorem 3.1]) prove for all $N, d \in \mathbb{N}$ that

$$\begin{aligned}
&\left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|u_d(T,x) - g_{N,d}(Y_N^{N,d,x})\right|^p \nu_d(dx)\right]\right)^{1/p} \\
&= \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|\mathbb{E}[\varphi_{0,d}(X_T^{d,x})] - g_{N,d}(Y_N^{N,d,x})\right|^p \nu_d(dx)\right]\right)^{1/p} \\
&= \left(\mathbb{E}\left[\int_{\mathbb{R}^d} \left|\mathbb{E}[\varphi_{0,d}(X_T^{d,x})] - \tfrac{1}{N}\left[\textstyle\sum_{i=1}^N \varphi_{0,d}(Y_N^{N,d,i,x})\right]\right|^p \nu_d(dx)\right]\right)^{1/p} \\
&\leq 2^{4\theta+6}|\max\{1,T\}|^{\theta+1}|\max\{\kappa,\theta\}|^{2\theta+3}e^{(6\max\{\kappa,\theta\}+5|\max\{\kappa,\theta\}|^2T)} \\
&\quad \cdot p(p\theta+p+1)^\theta d^{\mathfrak{d}_6+(\mathfrak{d}_1+\mathfrak{d}_2)(\theta+1)}N^{-1/2}.
\end{aligned} \tag{4.77}$$

Combining this, (4.47), (4.51), (4.52), (4.56), (4.57), (4.58), (4.62), (4.63), (4.65), (4.66), (4.69), and Theorem 2.3 (with $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $\mathfrak{n}_0 = 1/2$, $\mathfrak{n}_1 = 0$, $\mathfrak{n}_2 = 2$, $\mathfrak{e} = \mathfrak{e}$, $\mathfrak{d}_0 = \mathfrak{d}_6 + (\mathfrak{d}_1 + \mathfrak{d}_2)(\theta+1)$, $\mathfrak{d}_1 = \mathfrak{d}_1$, $\mathfrak{d}_2 = \mathfrak{d}_2$, $\mathfrak{d}_3 = \max\{4, \mathfrak{d}_3\}$, $\mathfrak{d}_4 = \mathfrak{d}_4$, $\mathfrak{d}_5 = \mathfrak{d}_5$, $\mathfrak{d}_6 = \mathfrak{d}_6$, $\mathfrak{C} = 2^{4\theta+6}|\max\{1,T\}|^{\theta+1}|\max\{\kappa,\theta\}|^{2\theta+3}e^{(6\max\{\kappa,\theta\}+5|\max\{\kappa,\theta\}|^2T)}p(p\theta+p+1)^\theta$, $p = p$, $\theta = \theta$, $M_N = N$, $Z_n^{N,d,m} = Z_n^{N,d,m}$, $f_{N,d} = f_{N,d}$, $Y_l^{N,d,x} = Y_l^{N,d,x}$, $\|\cdot\| = \|\cdot\|$, $\nu_d = \nu_d$, $g_{N,d} = g_{N,d}$, $u_d(x) = u_d(T,x)$, $\mathbf{N} = \mathbf{N}$, $\mathcal{P} = \mathcal{P}$, $\mathcal{D} = \mathcal{D}$, $\mathcal{R} = \mathcal{R}_a$, $\mathfrak{N}_{d,\varepsilon} = \mathfrak{N}_{d,\varepsilon}$, $\mathbf{f}_{\varepsilon,z}^{N,d} = \mathbf{f}_{\varepsilon,z}^{N,d}$, $\mathbf{g}_\varepsilon^{N,d} = \mathbf{g}_\varepsilon^{N,d}$, $\mathfrak{I}_d = \mathfrak{I}_d$ for $N, d \in \mathbb{N}$, $m \in \{1, 2, \ldots, N\}$, $n \in \{0, 1, \ldots, N-1\}$, $l \in \{0, 1, \ldots, N\}$, $\varepsilon \in (0,1]$, $x, z \in \mathbb{R}^d$ in the notation of Theorem 2.3) establish (4.43). The proof of Theorem 4.5 is thus completed. □

Corollary 4.6. *Let $\varphi_{0,d}\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $\varphi_{1,d}\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be functions, let $\|\cdot\| : (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to [0,\infty)$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\|x\| = (\sum_{i=1}^d |x_i|^2)^{1/2}$, let $T, \kappa \in (0,\infty)$, $\mathfrak{e}, \mathfrak{d}_1, \mathfrak{d}_2, \ldots, \mathfrak{d}_6 \in [0,\infty)$, $\theta \in [1,\infty)$, $p \in [2,\infty)$, $(\phi_\varepsilon^{m,d})_{(m,d,\varepsilon)\in\{0,1\}\times\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$, $a \in C(\mathbb{R},\mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $m \in \{0,1\}$, $x, y \in \mathbb{R}^d$ that $\mathcal{R}_a(\phi_\varepsilon^{0,d}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{R}_a(\phi_\varepsilon^{1,d}) \in C(\mathbb{R}^d,\mathbb{R}^d)$, $\mathcal{P}(\phi_\varepsilon^{m,d}) \leq \kappa d^{2^{(-m)}\mathfrak{d}_3}\varepsilon^{-2^{(-m)}\mathfrak{e}}$, $|(\mathcal{R}_a(\phi_\varepsilon^{0,d}))(x) - (\mathcal{R}_a(\phi_\varepsilon^{0,d}))(y)| \leq \kappa d^{\mathfrak{d}_6}(1 + \|x\|^\theta + \|y\|^\theta)\|x-y\|$, $\|(\mathcal{R}_a(\phi_\varepsilon^{1,d}))(x)\| \leq \kappa(d^{\mathfrak{d}_1+\mathfrak{d}_2} + \|x\|)$, $|\varphi_{0,d}(x)| \leq \kappa d^{\mathfrak{d}_6}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta)$, $\|\varphi_{1,d}(x) - \varphi_{1,d}(y)\| \leq \kappa\|x-y\|$, and*

$$\|\varphi_{m,d}(x) - (\mathcal{R}_a(\phi_\varepsilon^{m,d}))(x)\| \leq \varepsilon\kappa d^{\mathfrak{d}_{(5-m)}}(d^{\theta(\mathfrak{d}_1+\mathfrak{d}_2)} + \|x\|^\theta), \tag{4.78}$$

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

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\tfrac{\partial}{\partial x}u_d)(t,x)\,\varphi_{1,d}(x) + \sum_{i=1}^d (\tfrac{\partial^2}{\partial x_i^2}u_d)(t,x) \tag{4.79}$$

with $u_d(0,x) = \varphi_{0,d}(x)$ for $(t,x) \in (0,T)\times\mathbb{R}^d$ (cf. Definition 3.1 and Definition 3.3). Then there exist $c \in \mathbb{R}$ and $(\Psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, $[\int_{[0,1]^d} |u_d(T,x) - (\mathcal{R}(\Psi_{d,\varepsilon}))(x)|^p\,dx]^{1/p} \leq \varepsilon$, and

$$\begin{aligned}
&\mathcal{P}(\Psi_{d,\varepsilon}) \leq c\varepsilon^{-(\mathfrak{e}+6)} \\
&\cdot d^{6[\mathfrak{d}_6+(\max\{\mathfrak{d}_1,1/2\}+\mathfrak{d}_2)(\theta+1)]+\max\{4,\mathfrak{d}_3\}+\mathfrak{e}\max\{\mathfrak{d}_5+\theta(\max\{\mathfrak{d}_1,1/2\}+\mathfrak{d}_2),\mathfrak{d}_4+\mathfrak{d}_6+2\theta(\max\{\mathfrak{d}_1,1/2\}+\mathfrak{d}_2)\}}.
\end{aligned} \tag{4.80}$$

Proof of Corollary 4.6. 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

$$\nu_d(B) = \lambda_d(B \cap [0,1]^d). \tag{4.81}$$

Observe that (4.81) implies that for all $d \in \mathbb{N}$ it holds that ν_d is a probability measure on $\mathbb{R}^d$. This and (4.81) ensure that for all $d \in \mathbb{N}$, $g \in C(\mathbb{R}^d, \mathbb{R})$ it holds that

$$\int_{\mathbb{R}^d} |g(x)| \, \nu_d(dx) = \int_{[0,1]^d} |g(x)| \, dx. \tag{4.82}$$

Combining this with, e.g., [21, Lemma 3.15] demonstrates that for all $d \in \mathbb{N}$ it holds that

$$\int_{\mathbb{R}^d} \|x\|^{2p\theta} \, \nu_d(dx) = \int_{[0,1]^d} \|x\|^{2p\theta} \, dx \leq d^{p\theta}. \tag{4.83}$$

This assures for all $d \in \mathbb{N}$ that

$$\left[\int_{\mathbb{R}^d} \|x\|^{2p\theta} \, \nu_d(dx) \right]^{1/(2p\theta)} \leq d^{1/2} \leq \max\{\kappa, 1\} d^{\max\{\mathfrak{d}_1, 1/2\} + \mathfrak{d}_2}. \tag{4.84}$$

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

$$\operatorname{Trace}(\mathrm{I}_d) \leq d \leq \max\{\kappa, 1\} d^{2\max\{\mathfrak{d}_1, 1/2\}} \tag{4.85}$$

(cf. Definition (3.6)). This, (4.84), and Theorem 4.5 (with $A_d = \mathrm{I}_d$, $\|\cdot\| = \|\cdot\|$, $\nu_d = \nu_d$, $\varphi_{0,d} = \varphi_{0,d}$, $\varphi_{1,d} = \varphi_{1,d}$, $T = T$, $\kappa = \max\{\kappa, 1\}$, $\mathfrak{e} = \mathfrak{e}$, $\mathfrak{d}_1 = \max\{\mathfrak{d}_1, 1/2\}$, $\mathfrak{d}_2 = \mathfrak{d}_2$, $\mathfrak{d}_3 = \mathfrak{d}_3$, $\mathfrak{d}_4 = \mathfrak{d}_4$, $\mathfrak{d}_5 = \mathfrak{d}_5$, $\mathfrak{d}_6 = \mathfrak{d}_6$, $\theta = \theta$, $p = p$, $\phi^{0,d}_\varepsilon = \phi^{0,d}_\varepsilon$, $\phi^{1,d}_\varepsilon = \phi^{1,d}_\varepsilon$, $a = a$, $u_d = u_d$ for $d \in \mathbb{N}$ in the notation of Theorem 4.5) establish (4.80). The proof of Corollary 4.6 is thus completed. ☐

Corollary 4.7. *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, let $\|\cdot\| : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \to [0, \infty)$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ that $\|x\| = (\sum_{i=1}^d |x_i|^2)^{1/2}$, for every $d \in \mathbb{N}$ let $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure on $\mathbb{R}^d$, let $\varphi_{0,d} \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, and $\varphi_{1,d} \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be functions, let $T, \kappa, p \in (0, \infty)$, $\theta \in [1, \infty)$, $(\phi^{m,d}_\varepsilon)_{(m,d,\varepsilon) \in \{0,1\} \times \mathbb{N} \times (0,1]} \subseteq \mathbf{N}$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $m \in \{0,1\}$, $x, y \in \mathbb{R}^d$ that $\mathcal{R}_a(\phi^{0,d}_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{R}_a(\phi^{1,d}_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $|\varphi_{0,d}(x)| + \operatorname{Trace}(A_d) \leq \kappa d^\kappa (1 + \|x\|^\theta)$, $[\int_{\mathbb{R}^d} \|x\|^{2\max\{p,2\}\theta} \, \nu_d(dx)]^{1/(2\max\{p,2\}\theta)} \leq \kappa d^\kappa$, $\mathcal{P}(\phi^{m,d}_\varepsilon) \leq \kappa d^\kappa \varepsilon^{-\kappa}$, $|(\mathcal{R}_a(\phi^{0,d}_\varepsilon))(x) - (\mathcal{R}_a(\phi^{0,d}_\varepsilon))(y)| \leq \kappa d^\kappa (1 + \|x\|^\theta + \|y\|^\theta) \|x - y\|$, $\|(\mathcal{R}_a(\phi^{1,d}_\varepsilon))(x)\| \leq \kappa (d^\kappa + \|x\|)$, $\|\varphi_{1,d}(x) - \varphi_{1,d}(y)\| \leq \kappa \|x - y\|$, and*

$$\|\varphi_{m,d}(x) - (\mathcal{R}_a(\phi^{m,d}_\varepsilon))(x)\| \leq \varepsilon \kappa d^\kappa (1 + \|x\|^\theta), \tag{4.86}$$

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

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\tfrac{\partial}{\partial x} u_d)(t,x) \, \varphi_{1,d}(x) + \sum_{i,j=1}^d A_{d,i,j} \, (\tfrac{\partial^2}{\partial x_i \partial x_j} u_d)(t,x) \tag{4.87}$$

with $u_d(0,x) = \varphi_{0,d}(x)$ *for* $(t,x) \in (0,T) \times \mathbb{R}^d$ *(cf. Definition 3.1 and Definition 3.3). Then there exist* $c \in \mathbb{R}$ *and* $(\Psi_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ *such that for all* $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ *it holds that* $\mathcal{P}(\Psi_{d,\varepsilon}) \le c\, d^c \varepsilon^{-c}$, $\mathcal{R}_a(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, *and*

$$\left[\int_{\mathbb{R}^d} |u_d(T,x) - (\mathcal{R}_a(\Psi_{d,\varepsilon}))(x)|^p \, \nu_d(dx)\right]^{1/p} \le \varepsilon. \tag{4.88}$$

Acknowledgments

This project has been partially supported through the research grant 200020_175699 funded by the Swiss National Science Foundation.

References

[1] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep splitting method for parabolic PDEs. *arXiv:1907.03452* (2019), 40 pages.

[2] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv:1806.00421* (2018), 56 pages.

[3] Beck, C., E, W., and Jentzen, A. Machine Learning Approximation Algorithms for High-Dimensional Fully Nonlinear Partial Differential Equations and Second-order Backward Stochastic Differential Equations. *J. Nonlinear Sci. 29*, 4 (2019), 1563–1619.

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

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

[6] Bellman, R. *Dynamic programming.* Princeton University Press, Princeton, N. J., 1957.

[7] Bensoussan, A., and Lions, J.-L. *Applications of variational inequalities in stochastic control*, vol. 12 of *Studies in Mathematics and its Applications.* North-Holland Publishing Co., Amsterdam-New York, 1982.

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

[9] Berner, J., Grohs, P., and Jentzen, A. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *arXiv:1809.03062* (2018), 35 pages.

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

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

[12] 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.

[13] Dockhorn, T. A discussion on solving partial differential equations using neural networks. *arXiv:1904.07200* (2019), 9 pages.

[14] 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.

[15] 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.

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

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

[18] Fujii, M., Takahashi, A., and Takahashi, M. Asymptotic expansion as prior knowledge in deep learning method for high dimensional BSDEs. *Asia-Pacific Financial Markets* (Mar 2019).

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

[20] 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.

[21] 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.

[22] Grohs, P., Hornung, F., Jentzen, A., and Zimmermann, P. Space-time error estimates for deep neural network approximations for differential equations. *arXiv:1908.03833* (2019), 86 pages.

[23] 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.

[24] Han, J., and Long, J. Convergence of the Deep BSDE Method for Coupled FBSDEs. *arXiv:1811.01165* (2018), 26 pages.

[25] 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`.

[26] 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.

[27] 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.

[28] 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.

[29] Huré, C., Pham, H., and Warin, X. Some machine learning schemes for high-dimensional nonlinear PDEs. *arXiv:1902.01599* (2019), 33 pages.

[30] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *arXiv:1901.10854* (2019), 29 pages.

[31] Jacquier, A., and Oumgari, M. Deep PPDEs for rough local stochastic volatility. *arXiv:1906.02551* (2019), 21 pages.

[32] Jentzen, A., Salimova, D., and Welti, T. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *arXiv:1809.07321* (2018), 48 pages.

[33] Jianyu, L., Siwei, L., Yingjian, Q., and Yaping, H. Numerical solution of elliptic partial differential equation using radial basis function neural networks. *Neural Networks 16*, 5 (2003), 729 – 734.

[34] 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.

[35] 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.

[36] Kutyniok, G., Petersen, P., Raslan, M., and Schneider, R. A theoretical analysis of deep neural networks and parametric PDEs. *arXiv:1904.00377* (2019), 43 pages.

[37] Lagaris, I. E., Likas, A., and Fotiadis, D. I. Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks 9 (5)* (1998), 987–1000.

[38] Long, Z., Lu, Y., Ma, X., and Dong, B. PDE-Net: Learning PDEs from Data. In *Proceedings of the 35th International Conference on Machine Learning* (2018), pp. 3208–3216.

[39] Lye, K. O., Mishra, S., and Ray, D. Deep learning observables in computational fluid dynamics. *arXiv:1903.03040* (2019), 57 pages.

[40] Magill, M., Qureshi, F., and de Haan, H. W. Neural networks trained to solve differential equations learn general representations. In *Advances in Neural Information Processing Systems* (2018), pp. 4071–4081.

[41] Meade, Jr., A. J., and Fernández, A. A. The numerical solution of linear ordinary differential equations by feedforward neural networks. *Math. Comput. Modelling 19*, 12 (1994), 1–25.

[42] Novak, E., and Woźniakowski, H. *Tractability of multivariate problems. Vol. 1: Linear information*, vol. 6 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2008.

[43] Novak, E., and Woźniakowski, H. *Tractability of multivariate problems. Volume II: Standard information for functionals*, vol. 12 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2010.

[44] Pham, H., and Warin, X. Neural networks-based backward scheme for fully nonlinear PDEs. *arXiv:1908.00412* (2019), 15 pages.

[45] Raissi, M. Deep hidden physics models: Deep learning of nonlinear partial differential equations. *J. Mach. Learn. Res. 19* (2018), 25:1–25:24.

[46] Reisinger, C., and Zhang, Y. Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. *arXiv:1903.06652* (2019), 34 pages.

[47] 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.

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

[49] Sirignano, J., and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *J. Comput. Phys. 375* (2018), 1339–1364.

[50] 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.

[51] Uchiyama, T., and Sonehara, N. Solving inverse problems in nonlinear PDEs by recurrent neural networks. In *IEEE International Conference on Neural Networks* (1993), IEEE, pp. 99–102.

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

[53] 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.

[54] 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.

[55] 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.