



A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations

Martin Hutzenthaler¹ · Arnulf Jentzen^{2,3} · Thomas Kruse⁴ · Tuan Anh Nguyen¹

Received: 10 August 2019 / Accepted: 22 December 2019 / Published online: 6 April 2020
© The Author(s) 2020

Abstract

Deep neural networks and other deep learning methods have very successfully been applied to the numerical approximation of high-dimensional nonlinear parabolic partial differential equations (PDEs), which are widely used in finance, engineering, and natural sciences. In particular, simulations indicate that algorithms based on deep learning overcome the curse of dimensionality in the numerical approximation of solutions of semilinear PDEs. For certain linear PDEs it has also been proved mathematically that deep neural networks overcome the curse of dimensionality in the numerical approximation of solutions of such linear PDEs. The key contribution of this article is to rigorously prove this for the first time for a class of nonlinear PDEs. More precisely, we prove in the case of semilinear heat equations with gradient-independent nonlinearities that the numbers of parameters of the employed deep neural networks grow at most polynomially in both the PDE dimension and the reciprocal of the prescribed approximation accuracy. Our proof relies on recently introduced full history recursive multilevel Picard approximations for semilinear PDEs.

This article is part of the section “Computational Approaches” edited by Siddhartha Mishra.

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

Arnulf Jentzen
arnulf.jentzen@sam.math.ethz.ch; ajentzen@uni-muenster.de

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

Tuan Anh Nguyen
tuan.nguyen@uni-due.de

¹ Faculty of Mathematics, University of Duisburg-Essen, 45117 Essen, Germany

² SAM, Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

³ Faculty of Mathematics and Computer Science, University of Münster, 48149 Münster, Germany

⁴ Institute of Mathematics, University of Gießen, 35392 Gießen, Germany

Keywords Curse of dimensionality · High-dimensional PDEs · Deep neural networks · Information based complexity · Tractability of multivariate problems · Multilevel Picard approximations

Mathematics Subject Classification 65C99 · 68T05

1 Introduction

Deep neural networks (DNNs) have revolutionized a number of computational problems; see, e.g., the references in Grohs et al. [13]. In 2017 deep learning-based approximation algorithms for certain parabolic partial differential equations (PDEs) have been proposed in Han et al. [6, 14] and based on these works there is now a series of deep learning-based numerical approximation algorithms for a large class of different kinds of PDEs in the scientific literature; see, e.g., [1, 2, 4, 9, 10, 11, 13, 15, 21–25]. There is empirical evidence that deep learning-based methods work exceptionally well for approximating solutions of high-dimensional PDEs and that these do not suffer from the *curse of dimensionality*; see, e.g., the simulations in [1, 2, 6, 14]. There exist, however, only few theoretical results which prove that DNN approximations of solutions of PDEs do not suffer from the curse of dimensionality: The recent articles [5, 10, 13, 20] prove rigorously that DNN approximations overcome the curse of dimensionality in the numerical approximation of solutions of certain linear PDEs.

The main result of this article, Theorem 4.1 below, proves for semilinear heat equations with gradient-independent nonlinearities that the number of parameters of the approximating DNN grows at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal of the prescribed accuracy $\varepsilon > 0$. Thereby, we establish for the first time that there exist DNN approximations of solutions of such PDEs which indeed overcome the curse of dimensionality. To illustrate the main result of this article we formulate in the following result, Theorem 1.1 below, a special case of Theorem 4.1.

Theorem 1.1 *Let $\mathbf{A}_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $d \in \mathbb{N} = \{1, 2, \dots\}$, and $\|\cdot\| : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow [0, \infty)$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $\mathbf{A}_d(x) = (\max\{x_1, 0\}, \dots, \max\{x_d, 0\})$ and $\|x\| = [\sum_{i=1}^d (x_i)^2]^{1/2}$, let $\mathbf{N} = \cup_{H \in \mathbb{N}} \cup_{(k_0, k_1, \dots, k_{H+1}) \in \mathbb{N}^{H+2}} \left[\prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}) \right]$, let $\mathcal{R} : \mathbf{N} \rightarrow (\cup_{k, l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ and $\mathcal{P} : \mathbf{N} \rightarrow \mathbb{N}$ satisfy for all $H \in \mathbb{N}$, $k_0, k_1, \dots, k_H, k_{H+1} \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n})$, $x_0 \in \mathbb{R}^{k_0}, \dots, x_H \in \mathbb{R}^{k_H}$ with $\forall n \in \mathbb{N} \cap [1, H] : x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ that*

$$\mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}}), \quad (\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1}, \quad \text{and} \quad \mathcal{P}(\Phi) = \sum_{n=1}^{H+1} k_n(k_{n-1} + 1),$$

let $T, \kappa \in (0, \infty)$, $f \in C(\mathbb{R}, \mathbb{R})$, $(g_{d,\varepsilon})_{d \in \mathbb{N}, \varepsilon \in (0,1]} \subseteq \mathbf{N}$, $(c_d)_{d \in \mathbb{N}} \subseteq (0, \infty)$, for every $d \in \mathbb{N}$ let $g_d \in C(\mathbb{R}^d, \mathbb{R})$, for every $d \in \mathbb{N}$ let $u_d \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, and assume for all $d \in \mathbb{N}$, $v, w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $t \in (0, T)$ that $|f(v) - f(w)| \leq \kappa|v - w|$, $\mathcal{R}(g_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $|(\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \kappa d^\kappa (1 + \|x\|^\kappa)$, $|g_d(x) - (\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \varepsilon \kappa d^\kappa (1 + \|x\|^\kappa)$, $\mathcal{P}(g_{d,\varepsilon}) \leq \kappa d^\kappa \varepsilon^{-\kappa}$, $|u_d(t, x)| \leq c_d(1 + \|x\|^{c_d})$, $u_d(0, x) = g_d(x)$, and

$$\left(\frac{\partial}{\partial t} u_d\right)(t, x) = (\Delta_x u_d)(t, x) + f(u_d(t, x)). \tag{1}$$

Then there exist $\eta \in (0, \infty)$ and $(u_{d,\varepsilon})_{d \in \mathbb{N}, \varepsilon \in (0,1]} \subseteq \mathbf{N}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathcal{R}(u_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(u_{d,\varepsilon}) \leq \eta d^n \varepsilon^{-\eta}$, and

$$\left[\int_{[0,1]^d} |u_d(T, x) - (\mathcal{R}(u_{d,\varepsilon}))(x)|^2 dx \right]^{\frac{1}{2}} \leq \varepsilon. \quad (2)$$

Theorem 1.1 is an immediate consequence of Corollary 4.2 in Sect. 4.2 below (with $T = 2T$, $u_d(t, x) = u_d(T - \frac{t}{2}, x)$, $f(v) = f(v)/2$ for $t \in [0, 2T]$, $x \in \mathbb{R}^d$, $v \in \mathbb{R}$ in the notation of Corollary 4.2). In the following we add a few comments on some of the mathematical objects which appear in Theorem 1.1. First, note that for all $d \in \mathbb{N}$ it holds that $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \rightarrow [0, \infty)$ in Theorem 1.1 is nothing else but the standard norm on \mathbb{R}^d . Theorem 1.1 shows under suitable assumptions that DNNs can overcome the curse of dimensionality in the numerical approximation of semilinear heat equations of the form (1) above and the functions $\mathbf{A}_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $d \in \mathbb{N}$, in Theorem 1.1 above specify the activation functions which we employ in the considered DNN approximations. The set \mathbf{N} in Theorem 1.1 above represents the set of all DNNs. Observe that for all $\Phi \in \mathbf{N}$ we have that the natural number $\mathcal{P}(\Phi)$ specifies the number of real parameters used to describe the DNN Φ . Moreover, note that for all $\Phi \in \mathbf{N}$ it holds that $\mathcal{R}(\Phi)$ is the realization function associated to the DNN Φ . The real number $T \in (0, \infty)$ specifies the time horizon $[0, T]$ of the PDEs in (1), the function $f : \mathbb{R} \rightarrow \mathbb{R}$ specifies the nonlinearity of the PDEs in (1), the functions $g_d : \mathbb{R}^d \rightarrow \mathbb{R}$, $d \in \mathbb{N}$, specify the initial conditions of the PDEs in (1), and the functions $u_d : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, $d \in \mathbb{N}$, specify the solutions of the PDEs in (1). The real numbers $\kappa \in (0, \infty)$ and $c_d \in (0, \infty)$, $d \in \mathbb{N}$, are constants which we use to specify suitable regularity assumptions on the nonlinearity $f : \mathbb{R} \rightarrow \mathbb{R}$, the initial conditions $g_d : \mathbb{R}^d \rightarrow \mathbb{R}$, $d \in \mathbb{N}$, and the PDE solutions $u_d : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, $d \in \mathbb{N}$. Theorem 1.1 establishes the existence of DNNs $u_{d,\varepsilon} \in \mathbf{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, which approximate the solutions $u_d : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, $d \in \mathbb{N}$, of (1) at time T without the curse of dimensionality.

Next we sketch the main steps in our proof of Theorem 1.1 above. Roughly speaking, the proof can be divided into four main steps. First, we approximate the solutions of the semilinear heat equations in (1) through solutions of PDEs whose initial conditions and nonlinearities can be exactly represented through suitable DNNs (cf. Lemma 2.3 below). Next we construct a suitable artificial probability space on which we approximate the solutions of these approximating PDEs by means of the in [7, 17] recently introduced full history recursive multilevel Picard (MLP) approximations (cf. Corollary 2.4 below and see also [3, 8, 12, 16, 18, 19] for further articles on MLP approximations). Thereafter, we prove that the MLP approximations of the approximating PDEs can be exactly represented by DNNs (cf. Lemma 3.10 below). We thus have constructed suitable random DNNs which approximate the solutions of (1) without the curse of dimensionality in the probabilistically strong sense. We are now in the position of the articles [5, 13, 20] to bring, e.g., [20, Corollary 2.4] into play to obtain the existence of a realization on the artificial probability space such that the error between the PDE solution of (1) and the realization of the constructed random DNNs is below the prescribed approximation accuracy $\varepsilon \in (0, 1]$ and this will allow us to complete the proof of Theorem 1.1 above. Our strategy of the proof of Theorem 1.1 is inspired by the procedure in [5, 13, 20] in the sense that we also construct suitable random DNNs to approximate the solutions of (1) on a suitable artificial probability space. The main difference of this work compared to [5, 13, 20] is that in this work we do not construct the approximating random DNNs through the plain vanilla Monte

Carlo method but through the recently introduced MLP approximation scheme which is an efficient nonlinear Monte Carlo algorithm and thereby allows us to overcome the curse of dimensionality in the case of nonlinear PDEs of the form (1).

The remainder of this article is organized as follows. In Sect. 2 we provide auxiliary results on MLP approximations ensuring that these approximations are stable against perturbations in the nonlinearity and the terminal condition of the PDE (1). In Sect. 3 we show that MLP approximations can be represented by DNNs and we provide bounds for the number of parameters of the representing DNN. In Sect. 4 we use the results of Sects. 2 and 3 to establish in Theorem 4.1 the main result of this article.

2 A stability result for full history recursive multilevel Picard (MLP) approximations

2.1 Setting

Setting 2.1 Let $d \in \mathbb{N}$, $T, L, \delta, B \in (0, \infty)$, $p, q \in [1, \infty)$, $f_1, f_2 \in C([0, T] \times \mathbb{R}^d \times \mathbb{R}, \mathbb{R})$, $g_1, g_2 \in C(\mathbb{R}^d, \mathbb{R})$, let $\|\cdot\| : \mathbb{R}^d \rightarrow [0, \infty)$ satisfy for all $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $\|x\| = [\sum_{i=1}^d (x_i)^2]^{1/2}$, assume for all $t \in [0, T]$, $x \in \mathbb{R}^d$, $w, v \in \mathbb{R}$, $i \in \{1, 2\}$ that

$$|f_i(t, x, w) - f_i(t, x, v)| \leq L|w - v|, \tag{3}$$

$$\max\{|f_i(t, x, 0)|, |g_i(x)|\} \leq B(1 + \|x\|)^p, \tag{4}$$

and

$$\max\{|f_1(t, x, v) - f_2(t, x, v)|, |g_1(x) - g_2(x)|\} \leq \delta((1 + \|x\|)^{pq} + |v|^q), \tag{5}$$

let $F_i : C([0, T] \times \mathbb{R}^d, \mathbb{R}) \rightarrow C([0, T] \times \mathbb{R}^d, \mathbb{R})$, $i \in \{1, 2\}$, satisfy for all $v \in C([0, T] \times \mathbb{R}^d, \mathbb{R})$, $t \in [0, T]$, $x \in \mathbb{R}^d$, $i \in \{1, 2\}$ that

$$(F_i(v))(t, x) = f_i(t, x, v(t, x)), \tag{6}$$

let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathbf{W} : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ be a standard Brownian motion with continuous sample paths, let $u_1, u_2 \in C([0, T] \times \mathbb{R}^d, \mathbb{R})$, assume for all $i \in \{1, 2\}$, $s \in [0, T]$, $x \in \mathbb{R}^d$ that

$$\mathbb{E} \left[|g_i(x + \mathbf{W}_{T-s})| + \int_s^T |(F_i(u_i))(t, x + \mathbf{W}_{t-s})| dt \right] < \infty \tag{7}$$

and

$$u_i(s, x) = \mathbb{E} \left[g_i(x + \mathbf{W}_{T-s}) + \int_s^T (F_i(u_i))(t, x + \mathbf{W}_{t-s}) dt \right], \tag{8}$$

let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, let $u^\theta : \Omega \rightarrow [0, 1]$, $\theta \in \Theta$, be independent random variables which are uniformly distributed on $[0, 1]$, let $U^\theta : [0, T] \times \Omega \rightarrow [0, T]$, $\theta \in \Theta$, satisfy for all $t \in [0, T]$, $\theta \in \Theta$ that $U_t^\theta = t + (T - t)u^\theta$, let $W^\theta : [0, T] \times \Omega \rightarrow \mathbb{R}^d$, $\theta \in \Theta$, be independent standard Brownian motions, assume that $(u^\theta)_{\theta \in \Theta}$, $(W^\theta)_{\theta \in \Theta}$, and \mathbf{W} are independent, and let $U_{n,M}^\theta : [0, T] \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$, $n, M \in \mathbb{Z}$, $\theta \in \Theta$, be functions which satisfy for all $n, M \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that $U_{-1,M}^\theta(t, x) = U_{0,M}^\theta(t, x) = 0$ and

$$\begin{aligned}
 U_{n,M}^\theta(t,x) &= \frac{1}{M^n} \sum_{i=1}^{M^n} g_2 \left(x + W_T^{(\theta,0,-i)} - W_t^{(\theta,0,-i)} \right) \\
 &+ \sum_{l=0}^{n-1} \frac{(T-t)}{M^{n-l}} \left[\sum_{i=1}^{M^{n-l}} \left(F_2 \left(U_{l,M}^{(\theta,l,i)} \right) - \mathbb{1}_{\mathbb{N}}(l) F_2 \left(U_{l-1,M}^{(\theta,l,i)} \right) \right) \right. \\
 &\quad \left. \left(\mathcal{U}_t^{(\theta,l,i)}, x + W_{\mathcal{U}_t^{(\theta,l,i)}}^{(\theta,l,i)} - W_t^{(\theta,l,i)} \right) \right]. \tag{9}
 \end{aligned}$$

2.2 An a priori estimate for solutions of partial differential equations (PDEs)

Lemma 2.2 (*q-th moment of the exact solution*) Assume Setting 2.1 and let $x \in \mathbb{R}^d$, $i \in \{1, 2\}$. Then it holds that

$$\sup_{t \in [0,T]} \left(\mathbb{E} \left[|u_i(t, x + \mathbf{W}_t)|^q \right] \right)^{\frac{1}{q}} \leq e^{LT} (T + 1) B \left[\sup_{t \in [0,T]} (\mathbb{E}[(1 + \|x + \mathbf{W}_t\|)^{pq}]^{\frac{1}{q}} \right]. \tag{10}$$

Proof of Lemma 2.2 Throughout this proof let $\mu_t : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $t \in [0, T]$, be the probability measures which satisfy for all $t \in [0, T]$, $B \in \mathcal{B}(\mathbb{R}^d)$ that

$$\mu_t(B) = \mathbb{P}(x + \mathbf{W}_t \in B). \tag{11}$$

The integral transformation theorem, (8), and the triangle inequality show for all $t \in [0, T]$ that

$$\begin{aligned}
 \left(\mathbb{E} \left[|u_i(t, x + \mathbf{W}_t)|^q \right] \right)^{\frac{1}{q}} &= \left(\int_{\mathbb{R}^d} |u_i(t, z)|^q \mu_t(dz) \right)^{\frac{1}{q}} \\
 &= \left(\int_{\mathbb{R}^d} \left| \mathbb{E} \left[g_i(z + \mathbf{W}_{T-t}) + \int_t^T (F_i(u_i))(s, z + \mathbf{W}_{s-t}) ds \right] \right|^q \mu_t(dz) \right)^{\frac{1}{q}} \\
 &\leq \left(\int_{\mathbb{R}^d} |\mathbb{E}[g_i(z + \mathbf{W}_{T-t})]|^q \mu_t(dz) \right)^{\frac{1}{q}} \\
 &\quad + \int_t^T \left(\int_{\mathbb{R}^d} |\mathbb{E}[(F_i(u_i))(s, z + \mathbf{W}_{s-t})]|^q \mu_t(dz) \right)^{\frac{1}{q}} ds.
 \end{aligned} \tag{12}$$

Next, Jensen’s inequality, Fubini’s theorem, (11), the fact that \mathbf{W} has independent and stationary increments, and (4) demonstrate that for all $t \in [0, T]$ it holds that

$$\begin{aligned}
 \int_{\mathbb{R}^d} |\mathbb{E}[g_i(z + \mathbf{W}_{T-t})]|^q \mu_t(dz) &\leq \int_{\mathbb{R}^d} \mathbb{E} \left[|g_i(z + \mathbf{W}_T - \mathbf{W}_t)|^q \right] \mu_t(dz) \\
 &= \mathbb{E} \left[|g_i(x + \mathbf{W}_t + \mathbf{W}_T - \mathbf{W}_t)|^q \right] = \mathbb{E}[|g_i(x + \mathbf{W}_T)|^q] \leq \mathbb{E}[B^q (1 + \|x + \mathbf{W}_T\|)^{pq}].
 \end{aligned} \tag{13}$$

Furthermore, Jensen’s inequality, Fubini’s theorem, (11), the fact that \mathbf{W} has independent and stationary increments, the triangle inequality, (3), and (4) demonstrate for all $t \in [0, T]$ that

$$\begin{aligned}
 & \int_t^T \left(\int_{\mathbb{R}^d} |\mathbb{E}[(F_i(u_i))(s, z + \mathbf{W}_{s-t})]|^q \mu_t(dz) \right)^{\frac{1}{q}} ds \\
 & \leq \int_t^T \left(\int_{\mathbb{R}^d} \mathbb{E} \left[|(F_i(u_i))(s, z + \mathbf{W}_s - \mathbf{W}_t)|^q \right] \mu_t(dz) \right)^{\frac{1}{q}} ds \\
 & = \int_t^T \left(\mathbb{E} \left[|(F_i(u_i))(s, x + \mathbf{W}_t + \mathbf{W}_s - \mathbf{W}_t)|^q \right] \right)^{\frac{1}{q}} ds \\
 & \leq \int_t^T \left(\mathbb{E} \left[|(F_i(0))(s, x + \mathbf{W}_s)|^q \right] \right)^{\frac{1}{q}} ds + \int_t^T \left(\mathbb{E} \left[|(F_i(u_i) - F_i(0))(s, x + \mathbf{W}_s)|^q \right] \right)^{\frac{1}{q}} ds \\
 & \leq T \sup_{s \in [0, T]} (\mathbb{E}[B^q(1 + \|x + \mathbf{W}_s\|)^{pq}])^{\frac{1}{q}} + \int_t^T \left(\mathbb{E} \left[L^q |u_i(s, x + \mathbf{W}_s)|^q \right] \right)^{\frac{1}{q}} ds. \tag{14}
 \end{aligned}$$

Combining this with (12) and (13) implies that for all $t \in [0, T]$ it holds that

$$\begin{aligned}
 & \left(\mathbb{E} \left[|u_i(t, x + \mathbf{W}_t)|^q \right] \right)^{\frac{1}{q}} \\
 & \leq (T + 1)B \sup_{s \in [0, T]} (\mathbb{E}[(1 + \|x + \mathbf{W}_s\|)^{pq}])^{\frac{1}{q}} + L \int_t^T \left(\mathbb{E} \left[|u_i(s, x + \mathbf{W}_s)|^q \right] \right)^{\frac{1}{q}} ds. \tag{15}
 \end{aligned}$$

Next, [17, Corollary 3.11] shows that

$$\sup_{s \in [0, T]} \sup_{y \in \mathbb{R}^d} \frac{|u_i(s, y)|}{(1 + \|y\|)^p} \leq \sup_{s \in [0, T]} \sup_{y \in \mathbb{R}^d} \frac{|u_i(s, y)|}{1 + \|y\|^p} < \infty. \tag{16}$$

This, the triangle inequality, and the fact that $\mathbb{E}[\|\mathbf{W}_T\|^{pq}] < \infty$ show that

$$\begin{aligned}
 & \int_0^T \left(\mathbb{E} \left[|u_i(s, x + \mathbf{W}_s)|^q \right] \right)^{\frac{1}{q}} ds \leq \left[\sup_{s \in [0, T]} \sup_{y \in \mathbb{R}^d} \frac{|u(s, y)|}{(1 + \|y\|)^p} \right] \int_0^T (\mathbb{E}[(1 + \|x + \mathbf{W}_s\|)^{pq}])^{\frac{1}{q}} ds \\
 & \leq \left[\sup_{s \in [0, T]} \sup_{y \in \mathbb{R}^d} \frac{|u(s, y)|}{(1 + \|y\|)^p} \right] T \left(1 + \|x\| + (\mathbb{E}[\|\mathbf{W}_T\|^{pq}])^{\frac{1}{pq}} \right)^p < \infty. \tag{17}
 \end{aligned}$$

This, Gronwall’s integral inequality, and (15) establish for all $t \in [0, T]$ that

$$\left(\mathbb{E} \left[|u_i(t, x + \mathbf{W}_t)|^q \right] \right)^{\frac{1}{q}} \leq e^{LT} (T + 1)B \sup_{s \in [0, T]} (\mathbb{E}[(1 + \|x + \mathbf{W}_s\|)^{pq}])^{\frac{1}{q}}. \tag{18}$$

The proof of Lemma 2.2 is thus completed. □

2.3 A stability result for solutions of PDEs

Lemma 2.3 *Assume Setting 2.1. Then it holds for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that*

$$\begin{aligned}
 & \mathbb{E} \left[|u_1(t, x + \mathbf{W}_t) - u_2(t, x + \mathbf{W}_t)| \right] \\
 & \leq \delta (e^{LT} (T + 1))^{q+1} (B^q + 1) \left(1 + \|x\| + (\mathbb{E}[\|\mathbf{W}_T\|^{pq}])^{\frac{1}{pq}} \right)^{pq}. \tag{19}
 \end{aligned}$$

Proof of Lemma 2.3 First, (8), the triangle inequality, and the fact that \mathbf{W} has stationary increments show for all $s \in [0, T]$, $z \in \mathbb{R}^d$ that

$$\begin{aligned}
 & |u_1(s, z) - u_2(s, z)| \\
 &= \left| \mathbb{E} \left[(g_1 - g_2)(z + \mathbf{W}_{T-s}) + \int_s^T (F_1(u_1) - F_1(u_2) + F_1(u_2) - F_2(u_2))(t, z + \mathbf{W}_{t-s}) dt \right] \right| \\
 &\leq \mathbb{E} \left[|(g_1 - g_2)(z + \mathbf{W}_T - \mathbf{W}_s)| \right] + \int_s^T \mathbb{E} \left[|(F_1(u_1) - F_1(u_2))(t, z + \mathbf{W}_t - \mathbf{W}_s)| \right] dt \\
 &\quad + \int_s^T \mathbb{E} \left[|(F_1(u_2) - F_2(u_2))(t, z + \mathbf{W}_t - \mathbf{W}_s)| \right] dt.
 \end{aligned} \tag{20}$$

This, Fubini’s theorem, the fact that \mathbf{W} has independent increments, and the Lipschitz condition in (3) ensure that for all $s \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
 & \mathbb{E} \left[|(u_1 - u_2)(s, x + \mathbf{W}_s)| \right] = \mathbb{E} \left[|u_1(s, z) - u_2(s, z)|_{z=x+\mathbf{W}_s} \right] \\
 &\leq \mathbb{E} \left[\mathbb{E} \left[|(g_1 - g_2)(z + \mathbf{W}_T - \mathbf{W}_s)| \right]_{z=x+\mathbf{W}_s} \right] \\
 &\quad + \int_s^T \mathbb{E} \left[\mathbb{E} \left[|(F_1(u_1) - F_1(u_2))(t, z + \mathbf{W}_t - \mathbf{W}_s)| \right]_{z=x+\mathbf{W}_s} \right] dt \\
 &\quad + \int_s^T \mathbb{E} \left[\mathbb{E} \left[|(F_1(u_2) - F_2(u_2))(t, z + \mathbf{W}_t - \mathbf{W}_s)| \right]_{z=x+\mathbf{W}_s} \right] dt \\
 &= \mathbb{E} \left[|(g_1 - g_2)(x + \mathbf{W}_T)| \right] + \int_s^T \mathbb{E} \left[|(F_1(u_1) - F_1(u_2))(t, x + \mathbf{W}_t)| \right] dt \\
 &\quad + \int_s^T \mathbb{E} \left[|(F_1(u_2) - F_2(u_2))(t, x + \mathbf{W}_t)| \right] dt \\
 &\leq \mathbb{E} \left[|(g_1 - g_2)(x + \mathbf{W}_T)| \right] + \int_s^T \mathbb{E} \left[L|(u_1 - u_2)(t, x + \mathbf{W}_t)| \right] dt \\
 &\quad + T \sup_{t \in [0, T]} \mathbb{E} \left[|(F_1(u_2) - F_2(u_2))(t, x + \mathbf{W}_t)| \right].
 \end{aligned} \tag{21}$$

This, Gronwall’s lemma, and Lemma 2.2 yield for all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
 & \sup_{t \in [0, T]} \mathbb{E} \left[|(u_1 - u_2)(t, x + \mathbf{W}_t)| \right] \\
 &\leq e^{LT} (T + 1) \sup_{t \in [0, T]} \max \left\{ \mathbb{E} \left[|(g_1 - g_2)(x + \mathbf{W}_T)| \right], \mathbb{E} \left[|(F_1(u_2) - F_2(u_2))(t, x + \mathbf{W}_t)| \right] \right\}.
 \end{aligned} \tag{22}$$

Furthermore, (5), the triangle inequality, and Lemma 2.2 imply for all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
 & \sup_{t \in [0, T]} \max \left\{ \mathbb{E} \left[|(g_1 - g_2)(x + \mathbf{W}_T)| \right], \mathbb{E} \left[|(F_1(u_2) - F_2(u_2))(t, x + \mathbf{W}_t)| \right] \right\} \\
 & \leq \delta \sup_{t \in [0, T]} \mathbb{E} [(1 + \|x + \mathbf{W}_t\|)^{pq} + |u_2(x + \mathbf{W}_t)|^q] \\
 & \leq \delta \sup_{t \in [0, T]} \mathbb{E} [(1 + \|x + \mathbf{W}_t\|)^{pq}] + \delta \sup_{t \in [0, T]} \mathbb{E} [|u_2(x + \mathbf{W}_t)|^q]. \tag{23} \\
 & \leq \delta \sup_{t \in [0, T]} \mathbb{E} [(1 + \|x + \mathbf{W}_t\|)^{pq}] + \delta (e^{LT}(T + 1)B)^q \sup_{t \in [0, T]} \mathbb{E} [(1 + \|x + \mathbf{W}_t\|)^{pq}] \\
 & \leq \delta (e^{LT}(T + 1))^q (B^q + 1) \sup_{t \in [0, T]} \mathbb{E} [(1 + \|x + \mathbf{W}_t\|)^{pq}].
 \end{aligned}$$

This, (22), and the triangle inequality yield that

$$\begin{aligned}
 & \sup_{t \in [0, T]} \mathbb{E} \left[|(u_1 - u_2)(t, x + \mathbf{W}_t)| \right] \\
 & \leq \delta (e^{LT}(T + 1))^{q+1} (B^q + 1) \sup_{t \in [0, T]} \mathbb{E} [(1 + \|x + \mathbf{W}_t\|)^{pq}] \tag{24} \\
 & \leq \delta (e^{LT}(T + 1))^{q+1} (B^q + 1) \left(1 + \|x\| + \left(\mathbb{E} [\|\mathbf{W}_T\|^{pq}] \right)^{\frac{1}{pq}} \right)^{pq}.
 \end{aligned}$$

This completes the proof of Lemma 2.3. □

2.4 A stability result for MLP approximations

Corollary 2.4 *Assume Setting 2.1, let $x \in \mathbb{R}^d$, $N, M \in \mathbb{N}$, and assume that $q \geq 2$. Then it holds that*

$$\begin{aligned}
 & \left(\mathbb{E} \left[\left| U_{N, M}^0(0, x) - u_1(0, x) \right|^2 \right] \right)^{\frac{1}{2}} \\
 & \leq (e^{LT}(T + 1))^{q+1} (B^q + 1) \left(\delta + \frac{e^{M/2}(1 + 2LT)^N}{M^{N/2}} \right) \left(1 + \|x\| + \left(\mathbb{E} [\|\mathbf{W}_T\|^{pq}] \right)^{\frac{1}{pq}} \right)^{pq}. \tag{25}
 \end{aligned}$$

Proof of Corollary 2.4 First, Lemma 2.2 implies that $\int_0^T \left(\mathbb{E} [|u_2(t, x + \mathbf{W}_t)|^2] \right)^{\frac{1}{2}} dt < \infty$. This, [17, Theorem 3.5] (with $\xi = x$, $F = F_2$, $g = g_2$, and $u = u_2$ in the notation of [17, Theorem 3.5]), (4), and the triangle inequality ensure that

$$\begin{aligned}
 & \left(\mathbb{E} \left[\left| U_{N,M}^0(0,x) - u_2(0,x) \right|^2 \right] \right)^{\frac{1}{2}} \\
 & \leq e^{LT} \left[\left(\mathbb{E} \left[|g_2(x + \mathbf{W}_T)|^2 \right] \right)^{\frac{1}{2}} + T \left(\frac{1}{T} \int_0^T \mathbb{E} \left[|(F_2(0))(t,x + \mathbf{W}_t)|^2 \right] dt \right)^{\frac{1}{2}} \right] \frac{e^{M/2}(1 + 2LT)^N}{M^{N/2}} \\
 & \leq e^{LT} (T + 1) \sup_{t \in [0,T]} \left(\mathbb{E} \left[B^2(1 + \|x + \mathbf{W}_t\|)^{2p} \right] \right)^{\frac{1}{2}} \frac{e^{M/2}(1 + 2LT)^N}{M^{N/2}} \\
 & \leq e^{LT} (T + 1) B \left(1 + \|x\| + \left(\mathbb{E} \left[\|\mathbf{W}_T\|^{2p} \right] \right)^{\frac{1}{2p}} \right)^p \frac{e^{M/2}(1 + 2LT)^N}{M^{N/2}}.
 \end{aligned} \tag{26}$$

Furthermore, Lemma 2.3 shows that

$$|u_2(0,x) - u_1(0,x)| \leq \delta (e^{LT}(T + 1))^{q+1} (B^q + 1) \left(1 + \|x\| + (\mathbb{E}[\|\mathbf{W}_T\|^{pq}])^{\frac{1}{pq}} \right)^{pq}. \tag{27}$$

This, the triangle inequality, (26), the fact that $B \leq B^q + 1$, the assumption that $q \geq 2$, and Jensen’s inequality show that

$$\begin{aligned}
 & \left(\mathbb{E} \left[\left| U_{N,M}^0(0,x) - u_1(0,x) \right|^2 \right] \right)^{\frac{1}{2}} \\
 & \leq \left(\mathbb{E} \left[\left| U_{N,M}^0(0,x) - u_2(0,x) \right|^2 \right] \right)^{\frac{1}{2}} + |u_2(0,x) - u_1(0,x)| \\
 & \leq (e^{LT}(T + 1))^{q+1} (B^q + 1) \left(\delta + \frac{e^{M/2}(1 + 2LT)^N}{M^{N/2}} \right) \left(1 + \|x\| + (\mathbb{E}[\|\mathbf{W}_T\|^{pq}])^{\frac{1}{pq}} \right)^{pq}.
 \end{aligned} \tag{28}$$

The proof of Corollary 2.4 is thus completed. □

3 Deep neural network representations for MLP approximations

The main result of this section, Lemma 3.10 below, shows that multilevel Picard approximations can be well represented by DNNs. The central tools for the proof of Lemma 3.10 are Lemmas 3.8 and 3.9 which show that DNNs are stable under compositions and summations. We formulate Lemmas 3.8 and 3.9 in terms of the operators defined in (33) and (34) below, whose properties are studied in Lemmas 3.3, 3.4, and 3.5.

3.1 A mathematical framework for deep neural networks

Setting 3.1 (Artificial neural networks) *Let $\|\cdot\|, \|\cdot\| : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow [0, \infty)$ and $\dim : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow \mathbb{N}$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $\|x\| = \sqrt{\sum_{i=1}^d (x_i)^2}$, $\|\|x\|\| = \max_{i \in [1,d] \cap \mathbb{N}} |x_i|$, and $\dim(x) = d$, let $\mathbf{A}_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that*

$$\mathbf{A}_d(x) = (\max\{x_1, 0\}, \dots, \max\{x_d, 0\}), \tag{29}$$

let $\mathbf{D} = \cup_{H \in \mathbb{N}} \mathbb{N}^{H+2}$, let

$$\mathbf{N} = \bigcup_{H \in \mathbb{N}} \bigcup_{(k_0, k_1, \dots, k_{H+1}) \in \mathbb{N}^{H+2}} \left[\prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}) \right], \tag{30}$$

let $\mathcal{D} : \mathbf{N} \rightarrow \mathbf{D}$ and $\mathcal{R} : \mathbf{N} \rightarrow (\cup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ satisfy for all $H \in \mathbb{N}$, $k_0, k_1, \dots, k_H, k_{H+1} \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n})$, $x_0 \in \mathbb{R}^{k_0}, \dots, x_H \in \mathbb{R}^{k_H}$ with $\forall n \in \mathbb{N} \cap [1, H] : x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ that

$$\mathcal{D}(\Phi) = (k_0, k_1, \dots, k_H, k_{H+1}), \quad \mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}}), \tag{31}$$

$$\text{and } (\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1}, \tag{32}$$

let $\odot : \mathbf{D} \times \mathbf{D} \rightarrow \mathbf{D}$ satisfy for all $H_1, H_2 \in \mathbb{N}$, $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{H_1}, \alpha_{H_1+1}) \in \mathbb{N}^{H_1+2}$, $\beta = (\beta_0, \beta_1, \dots, \beta_{H_2}, \beta_{H_2+1}) \in \mathbb{N}^{H_2+2}$ that

$$\alpha \odot \beta = (\beta_0, \beta_1, \dots, \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{H_1+1}) \in \mathbb{N}^{H_1+H_2+3}, \tag{33}$$

let $\boxplus : \mathbf{D} \times \mathbf{D} \rightarrow \mathbf{D}$ satisfy for all $H \in \mathbb{N}$, $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_H, \alpha_{H+1}) \in \mathbb{N}^{H+2}$, $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_H, \beta_{H+1}) \in \mathbb{N}^{H+2}$ that

$$\alpha \boxplus \beta = (\alpha_0, \alpha_1 + \beta_1, \dots, \alpha_H + \beta_H, \beta_{H+1}) \in \mathbb{N}^{H+2}, \tag{34}$$

and let $\mathbf{n}_n \in \mathbf{D}$, $n \in [3, \infty) \cap \mathbb{N}$, satisfy for all $n \in [3, \infty) \cap \mathbb{N}$ that

$$\mathbf{n}_n = (1, \underbrace{2, \dots, 2}_{(n-2)\text{-times}}, 1) \in \mathbb{N}^n. \tag{35}$$

Remark 3.2 The set \mathbf{N} can be viewed as the set of all artificial neural networks. For each network $\Phi \in \mathbf{N}$ the function $\mathcal{R}(\Phi)$ is the function represented by Φ and the vector $\mathcal{D}(\Phi)$ describes the layer dimensions of Φ .

3.2 Properties of operations associated to deep neural networks

Lemma 3.3 (\odot is associative) Assume Setting 3.1 and let $\alpha, \beta, \gamma \in \mathbf{D}$. Then it holds that $(\alpha \odot \beta) \odot \gamma = \alpha \odot (\beta \odot \gamma)$.

Proof of Lemma 3.3 Throughout this proof let $H_1, H_2, H_3 \in \mathbb{N}$, $(\alpha_i)_{i \in [0, H_1+1] \cap \mathbb{N}_0} \in \mathbb{N}^{H_1+2}$, $(\beta_i)_{i \in [0, H_2+1] \cap \mathbb{N}_0} \in \mathbb{N}^{H_2+2}$, $(\gamma_i)_{i \in [0, H_3+1] \cap \mathbb{N}_0} \in \mathbb{N}^{H_3+2}$ satisfy that

$$\begin{aligned} \alpha &= (\alpha_0, \alpha_1, \dots, \alpha_{H_1+1}), & \beta &= (\beta_0, \beta_1, \dots, \beta_{H_2+1}), & \text{and} \\ \gamma &= (\gamma_0, \gamma_1, \dots, \gamma_{H_3+1}). \end{aligned} \tag{36}$$

The definition of \odot in (33) then shows that

$$\begin{aligned}
 (\alpha \odot \beta) \odot \gamma &= (\beta_0, \beta_1, \beta_2, \dots, \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{H_1+1}) \odot (\gamma_0, \gamma_1, \dots, \gamma_{H_3+1}) \\
 &= (\gamma_0, \dots, \gamma_{H_3}, \gamma_{H_3+1} + \beta_0, \beta_1, \dots, \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{H_1+1}) \\
 &= (\alpha_0, \alpha_1, \dots, \alpha_{H_1+1}) \odot (\gamma_0, \gamma_1, \dots, \gamma_{H_3}, \gamma_{H_3+1} + \beta_0, \beta_1, \beta_2, \dots, \beta_{H_2+1}) \\
 &= \alpha \odot (\beta \odot \gamma).
 \end{aligned}
 \tag{37}$$

The proof of Lemma 3.3 is thus completed. □

Lemma 3.4 (*⊕ and associativity*) Assume Setting 3.1, let $H, k, l \in \mathbb{N}$, and let $\alpha, \beta, \gamma \in (\{k\} \times \mathbb{N}^H \times \{l\})$. Then

- (i) it holds that $\alpha \boxplus \beta \in (\{k\} \times \mathbb{N}^H \times \{l\})$,
- (ii) it holds that $\beta \boxplus \gamma \in (\{k\} \times \mathbb{N}^H \times \{l\})$, and
- (iii) it holds that $(\alpha \boxplus \beta) \boxplus \gamma = \alpha \boxplus (\beta \boxplus \gamma)$.

Proof of Lemma 3.4 Throughout this proof let $\alpha_i, \beta_i, \gamma_i \in \mathbb{N}$, $i \in [1, H] \cap \mathbb{N}$, satisfy that $\alpha = (k, \alpha_1, \alpha_2, \dots, \alpha_H, l)$, $\beta = (k, \beta_1, \beta_2, \dots, \beta_H, l)$, and $\gamma = (k, \gamma_1, \gamma_2, \dots, \gamma_H, l)$. The definition of \boxplus (see (34)) then shows that

$$\begin{aligned}
 \alpha \boxplus \beta &= (k, \alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_H + \beta_H, l) \in \{k\} \times \mathbb{N}^H \times \{l\}, \\
 \beta \boxplus \gamma &= (k, \beta_1 + \gamma_1, \beta_2 + \gamma_2, \dots, \beta_H + \gamma_H, l) \in \{k\} \times \mathbb{N}^H \times \{l\},
 \end{aligned}
 \tag{38}$$

and

$$\begin{aligned}
 (\alpha \boxplus \beta) \boxplus \gamma &= (k, (\alpha_1 + \beta_1) + \gamma_1, (\alpha_2 + \beta_2) + \gamma_2, \dots, (\alpha_H + \beta_H) + \gamma_H, l) \\
 &= (k, \alpha_1 + (\beta_1 + \gamma_1), \alpha_2 + (\beta_2 + \gamma_2), \dots, \alpha_H + (\beta_H + \gamma_H), l) = \alpha \boxplus (\beta \boxplus \gamma).
 \end{aligned}
 \tag{39}$$

The proof of Lemma 3.4 is thus completed. □

Lemma 3.5 (*Triangle inequality*) Assume Setting 3.1, let $H, k, l \in \mathbb{N}$, and let $\alpha, \beta \in \{k\} \times \mathbb{N}^H \times \{l\}$. Then it holds that $|||\alpha \boxplus \beta||| \leq |||\alpha||| + |||\beta|||$.

Proof of Lemma 3.5 Throughout this proof let $\alpha_i, \beta_i \in \mathbb{N}$, $i \in [1, H] \cap \mathbb{N}$, satisfy that $\alpha = (k, \alpha_1, \alpha_2, \dots, \alpha_H, l)$ and $\beta = (k, \beta_1, \beta_2, \dots, \beta_H, l)$. The definition of \boxplus (see (34)) then shows that $\alpha \boxplus \beta = (k, \alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_H + \beta_H, l)$. This together with the triangle inequality implies that

$$\begin{aligned}
 |||\alpha \boxplus \beta||| &= \sup\{|k|, |\alpha_1 + \beta_1|, |\alpha_2 + \beta_2|, \dots, |\alpha_H + \beta_H|, |l|\} \\
 &\leq \sup\{|k|, |\alpha_1|, |\alpha_2|, \dots, |\alpha_H|, |l|\} + \sup\{|k|, |\beta_1|, |\beta_2|, \dots, |\beta_H|, |l|\} \\
 &= |||\alpha||| + |||\beta|||.
 \end{aligned}
 \tag{40}$$

This completes the proof of Lemma 3.5. □

The following result, Lemma 3.6, can in a slightly modified variant, e.g., be found in [20, Lemma 5.4] in the scientific literature.

Lemma 3.6 (*Existence of DNNs with $H \in \mathbb{N}$ hidden layers for the identity in \mathbb{R}*) Assume Setting 3.1 and let $H \in \mathbb{N}$. Then it holds that $\text{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathfrak{n}_{H+2}\})$.

Proof of Lemma 3.6 Throughout this proof let $W_1 \in \mathbb{R}^{2 \times 1}$, $W_i \in \mathbb{R}^{2 \times 2}$, $i \in [2, H] \cap \mathbb{N}$, $W_{H+1} \in \mathbb{R}^{1 \times 2}$, $B_i \in \mathbb{R}^2$, $i \in [1, H] \cap \mathbb{N}$, $B_{H+1} \in \mathbb{R}^1$ satisfy that

$$\begin{aligned}
 W_1 &= \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \forall i \in [2, H] \cap \mathbb{N} : W_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad W_{H+1} = (1 \quad -1), \\
 \forall i \in [1, H] \cap \mathbb{N} : B_i &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad B_{H+1} = 0,
 \end{aligned}
 \tag{41}$$

let $\phi \in \mathbb{N}$ satisfy that $\phi = ((W_1, B_1), (W_2, B_2), \dots, (W_H, B_H), (W_{H+1}, B_{H+1}))$, for every $a \in \mathbb{R}$ let $a^+ \in [0, \infty)$ be the non-negative part of a , i.e., $a^+ = \max\{a, 0\}$, and let $x_0 \in \mathbb{R}$, $x_1, x_2, \dots, x_H \in \mathbb{R}^2$ satisfy for all $n \in \mathbb{N} \cap [1, H]$ that

$$x_n = \mathbf{A}_2(W_n x_{n-1} + B_n). \tag{42}$$

Note that (41) and the definition of \mathcal{D} (see (31)) imply that $\mathcal{D}(\phi) = n_{H+2}$. Furthermore, (41), (42), and an induction argument show that

$$\begin{aligned}
 x_1 &= \mathbf{A}_2(W_1 x_0 + B_1) = \mathbf{A}_2\left(\begin{pmatrix} x_0 \\ -x_0 \end{pmatrix}\right) = \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}, \\
 x_2 &= \mathbf{A}_2(W_2 x_1 + B_2) = \mathbf{A}_2(x_1) = \mathbf{A}_2\left(\begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}\right) = \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}, \\
 &\vdots \\
 x_H &= \mathbf{A}_2(W_H x_{H-1} + B_H) = \mathbf{A}_2(x_{H-1}) = \mathbf{A}_2\left(\begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}\right) = \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}.
 \end{aligned}
 \tag{43}$$

The definition of \mathcal{R} (see (32)) hence ensures that

$$(\mathcal{R}(\phi))(x_0) = W_{H+1} x_H + B_{H+1} = (1 \quad -1) \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix} = x_0^+ - (-x_0)^+ = x_0. \tag{44}$$

The fact that x_0 was arbitrary therefore proves that $\mathcal{R}(\phi) = \text{Id}_{\mathbb{R}}$. This and the fact that $\mathcal{D}(\phi) = n_{H+2}$ demonstrate that $\text{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathbb{N} : \mathcal{D}(\Phi) = n_{H+2}\})$. The proof of Lemma 3.6 is thus completed. \square

Lemma 3.7 (DNNs for affine transformations) Assume Setting 3.1 and let $d, m \in \mathbb{N}$, $\lambda \in \mathbb{R}$, $b \in \mathbb{R}^d$, $a \in \mathbb{R}^m$, $\Psi \in \mathbb{N}$ satisfy that $\mathcal{R}(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^m)$. Then it holds that

$$\lambda((\mathcal{R}(\Psi))(\cdot + b) + a) \in \mathcal{R}(\{\Phi \in \mathbb{N} : \mathcal{D}(\Phi) = \mathcal{D}(\Psi)\}). \tag{45}$$

Proof of Lemma 3.7 Throughout this proof let $H, k_0, k_1, \dots, k_{H+1} \in \mathbb{N}$ satisfy that

$$H + 2 = \dim(\mathcal{D}(\Psi)) \quad \text{and} \quad (k_0, k_1, \dots, k_H, k_{H+1}) = \mathcal{D}(\Psi), \tag{46}$$

let $((W_1, B_1), (W_2, B_2), \dots, (W_H, B_H), (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n})$ satisfy that

$$((W_1, B_1), (W_2, B_2), \dots, (W_H, B_H), (W_{H+1}, B_{H+1})) = \Psi, \tag{47}$$

let $\phi \in \mathbb{N}$ satisfy that

$$\phi = ((W_1, B_1 + W_1b), (W_2, B_2), \dots, (W_H, B_H), (\lambda W_{H+1}, \lambda B_{H+1} + \lambda a)), \tag{48}$$

and let $x_0, y_0 \in \mathbb{R}^{k_0}, x_1, y_1 \in \mathbb{R}^{k_1}, \dots, x_H, y_H \in \mathbb{R}^{k_H}$ satisfy for all $n \in \mathbb{N} \cap [1, H]$ that

$$x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n), y_n = \mathbf{A}_{k_n}(W_n y_{n-1} + B_n + \mathbb{1}_{\{1\}}(n)W_1b), \text{ and } x_0 = y_0 + b. \tag{49}$$

Then it holds that

$$y_1 = \mathbf{A}_{k_1}(W_1 y_0 + B_1 + W_1b) = \mathbf{A}_{k_1}(W_1(y_0 + b) + B_1) = \mathbf{A}_{k_1}(W_1 x_0 + B_1) = x_1. \tag{50}$$

This and an induction argument prove for all $i \in [2, H] \cap \mathbb{N}$ that

$$y_i = \mathbf{A}_{k_i}(W_i y_{i-1} + B_i) = \mathbf{A}_{k_i}(W_i x_{i-1} + B_i) = x_i. \tag{51}$$

The definition of \mathcal{R} (see (32)) hence shows that

$$\begin{aligned} (\mathcal{R}(\phi))(y_0) &= \lambda W_{H+1}y_H + \lambda B_{H+1} + \lambda a = \lambda W_{H+1}x_H + \lambda B_{H+1} + \lambda a \\ &= \lambda(W_{H+1}x_H + B_{H+1} + a) = \lambda((\mathcal{R}(\Psi))(x_0) + a) = \lambda((\mathcal{R}(\Psi))(y_0 + b) + a). \end{aligned} \tag{52}$$

This and the fact that y_0 was arbitrary prove that $\mathcal{R}(\phi) = \lambda((\mathcal{R}(\Psi))(\cdot + b) + a)$. This and the fact that $\mathcal{D}(\phi) = \mathcal{D}(\Psi)$ imply that $\lambda((\mathcal{R}(\Psi))(\cdot + b) + a) \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathcal{D}(\Psi)\})$. The proof of Lemma 3.7 is thus completed. \square

Lemma 3.8 (Composition) Assume Setting 3.1 and let $d_1, d_2, d_3 \in \mathbb{N}, f \in C(\mathbb{R}^{d_2}, \mathbb{R}^{d_3}), g \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_2}), \alpha, \beta \in \mathbf{D}$ satisfy that $f \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \alpha\})$ and $g \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \beta\})$. Then it holds that $(f \circ g) \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \alpha \circ \beta\})$.

Proof of Lemma 3.8 Throughout this proof let $H_1, H_2, \alpha_0, \dots, \alpha_{H_1+1}, \beta_0, \dots, \beta_{H_2+1} \in \mathbb{N}, \Phi_f, \Phi_g \in \mathbf{N}$ satisfy that

$$\begin{aligned} (\alpha_0, \alpha_1, \dots, \alpha_{H_1+1}) &= \alpha, \quad (\beta_0, \beta_1, \dots, \beta_{H_2+1}) = \beta, \quad \mathcal{R}(\Phi_f) = f, \\ \mathcal{D}(\Phi_f) &= \alpha, \quad \mathcal{R}(\Phi_g) = g, \quad \text{and } \mathcal{D}(\Phi_g) = \beta. \end{aligned} \tag{53}$$

Lemma 5.4 in [20] shows that there exists $\mathbb{1} \in \mathbf{N}$ such that $\mathcal{D}(\mathbb{1}) = d_2 n_3 = (d_2, 2d_2, d_2)$ and $\mathcal{R}(\mathbb{1}) = \text{Id}_{\mathbb{R}^{d_2}}$. Note that $2d_2 = \beta_{H_2+1} + \alpha_0$. This and [20, Proposition 5.2] (with $\phi_1 = \Phi_f, \phi_2 = \Phi_g$, and $\mathbb{1} = \mathbb{1}$ in the notation of [20, Proposition 5.2]) show that there exists $\Phi_{f \circ g} \in \mathbf{N}$ such that $\mathcal{R}(\Phi_{f \circ g}) = f \circ g$ and $\mathcal{D}(\Phi_{f \circ g}) = \mathcal{D}(\Phi_f) \circ \mathcal{D}(\Phi_g) = \alpha \circ \beta$. Hence, it holds that $f \circ g \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \alpha \circ \beta\})$. The proof of Lemma 3.8 is thus completed. \square

The following result, Lemma 3.9, can roughly speaking in a specialized form be found, e.g., in [20, Lemma 5.1].

Lemma 3.9 (Sum of DNNs of the same length) Assume Setting 3.1 and let $M, H, p, q \in \mathbb{N}, h_1, h_2, \dots, h_M \in \mathbb{R}, k_i \in \mathbf{D}, f_i \in C(\mathbb{R}^p, \mathbb{R}^q), i \in [1, M] \cap \mathbb{N}$, satisfy for all $i \in [1, M] \cap \mathbb{N}$ that

$$\dim(k_i) = H + 2 \quad \text{and } f_i \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = k_i\}). \tag{54}$$

Then it holds that

$$\sum_{i=1}^M h_i f_i \in \mathcal{R} \left(\left\{ \Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \prod_{i=1}^M k_i \right\} \right). \tag{55}$$

Proof of Lemma 3.9 Throughout this proof let $\phi_i \in \mathbf{N}$, $i \in [1, M] \cap \mathbb{N}$, and $k_{i,n} \in \mathbb{N}$, $i \in [1, M] \cap \mathbb{N}$, $n \in [0, H + 1] \cap \mathbb{N}_0$, satisfy for all $i \in [1, M] \cap \mathbb{N}$ that

$$\mathcal{D}(\phi_i) = k_i = (k_{i,0}, k_{i,1}, k_{i,2}, \dots, k_{i,H}, k_{i,H+1}) \quad \text{and} \quad \mathcal{R}(\phi_i) = f_i, \tag{56}$$

for every $i \in [1, M] \cap \mathbb{N}$ let $((W_{i,1}, B_{i,1}), \dots, (W_{i,H+1}, B_{i,H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{k_{i,n} \times k_{i,n-1}} \times \mathbb{R}^{k_{i,n}})$ satisfy that

$$\phi_i = ((W_{i,1}, B_{i,1}), \dots, (W_{i,H+1}, B_{i,H+1})), \tag{57}$$

let $k_n^{\boxplus} \in \mathbb{N}$, $n \in [1, H] \cap \mathbb{N}$, $k^{\boxplus} \in \mathbb{N}^{H+2}$ satisfy for all $n \in [1, H] \cap \mathbb{N}$ that

$$k_n^{\boxplus} = \sum_{i=1}^M k_{i,n} \quad \text{and} \quad k^{\boxplus} = (p, k_1^{\boxplus}, k_2^{\boxplus}, \dots, k_H^{\boxplus}, q), \tag{58}$$

let $W_1 \in \mathbb{R}^{k_1^{\boxplus} \times p}$, $B_1 \in \mathbb{R}^{k_1^{\boxplus}}$ satisfy that

$$W_1 = \begin{pmatrix} W_{1,1} \\ W_{2,1} \\ \vdots \\ W_{M,1} \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ \vdots \\ B_{M,1} \end{pmatrix}, \tag{59}$$

let $W_n \in \mathbb{R}^{k_n^{\boxplus} \times k_{n-1}^{\boxplus}}$, $B_n \in \mathbb{R}^{k_n^{\boxplus}}$, $n \in [2, H] \cap \mathbb{N}$, satisfy for all $n \in [2, H] \cap \mathbb{N}$ that

$$W_n = \begin{pmatrix} W_{1,n} & 0 & 0 & 0 \\ 0 & W_{2,n} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & W_{M,n} \end{pmatrix} \quad \text{and} \quad B_n = \begin{pmatrix} B_{1,n} \\ B_{2,n} \\ \vdots \\ B_{M,n} \end{pmatrix}, \tag{60}$$

let $W_{H+1} \in \mathbb{R}^{q \times k_H^{\boxplus}}$, $B_{H+1} \in \mathbb{R}^q$ satisfy that

$$W_{H+1} = (h_1 W_{1,H+1} \quad \dots \quad h_M W_{M,H+1}) \quad \text{and} \quad B_{H+1} = \sum_{i=1}^M h_i B_{i,H+1}, \tag{61}$$

let $x_0 \in \mathbb{R}^p$, $x_1 \in \mathbb{R}^{k_1^{\boxplus}}$, $x_2 \in \mathbb{R}^{k_2^{\boxplus}}$, ..., $x_H \in \mathbb{R}^{k_H^{\boxplus}}$, let $x_{1,0}, x_{2,0}, \dots, x_{M,0} \in \mathbb{R}^p$, $x_{i,n} \in \mathbb{R}^{k_{i,n}}$, $i \in [1, M] \cap \mathbb{N}$, $n \in [1, H] \cap \mathbb{N}$, satisfy for all $i \in [1, M] \cap \mathbb{N}$, $n \in [1, H] \cap \mathbb{N}$ that

$$\begin{aligned} x_0 &= x_{1,0} = x_{2,0} = \dots = x_{M,0}, \\ x_{i,n} &= \mathbf{A}_{k_{i,n}}(W_{i,n} x_{i,n-1} + B_{i,n}), \\ x_n &= \mathbf{A}_{k_n^{\boxplus}}(W_n x_{n-1} + B_n), \end{aligned} \tag{62}$$

and let $\psi \in \mathbf{N}$ satisfy that

$$\psi = ((W_1, B_1), (W_2, B_2), \dots, (W_H, B_H), (W_{H+1}, B_{H+1})). \tag{63}$$

First, the definitions of \mathcal{D} and \mathcal{R} (see (31) and (32)), (56), and the fact that $\forall i \in [1, M] \cap \mathbb{N} : f_i \in C(\mathbb{R}^p, \mathbb{R}^q)$ show for all $i \in [1, M] \cap \mathbb{N}$ that $k_i = (p, k_{i,1}, k_{i,2}, \dots, k_{i,H}, q)$. The definition of \mathcal{D} (see (31)), the definition of \boxplus (see (34)), and (58) then show that

$$\mathcal{D}(\psi) = (p, k_1^{\boxplus}, \dots, k_H^{\boxplus}, q) = \boxplus_{i=1}^M k_i. \tag{64}$$

Next, we prove by induction on $n \in [1, H] \cap \mathbb{N}$ that $x_n = (x_{1,n}, x_{2,n}, \dots, x_{M,n})$. First, (59) shows that

$$W_1 x_0 + B_1 = \begin{pmatrix} W_{1,1} \\ W_{2,1} \\ \vdots \\ W_{M,1} \end{pmatrix} x_0 + \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ \vdots \\ B_{M,1} \end{pmatrix} = \begin{pmatrix} W_{1,1}x_0 + B_{1,1} \\ W_{2,1}x_0 + B_{2,1} \\ \vdots \\ W_{M,1}x_0 + B_{M,1} \end{pmatrix}. \tag{65}$$

This implies that

$$x_1 = \mathbf{A}_{k_1^{\boxplus}}(W_1 x_0 + B_1) = \begin{pmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{M,1} \end{pmatrix}. \tag{66}$$

This proves the base case. Next, for the induction step let $n \in [2, H] \cap \mathbb{N}$ and assume that $x_{n-1} = (x_{1,n-1}, x_{2,n-1}, \dots, x_{M,n-1})$. Then (60) and the induction hypothesis ensure that

$$\begin{aligned} &W_n x_{n-1} + B_n \\ &= W_n \begin{pmatrix} x_{1,n-1} \\ x_{2,n-1} \\ \vdots \\ x_{M,n-1} \end{pmatrix} + B_n = \begin{pmatrix} W_{1,n} & 0 & 0 & 0 \\ 0 & W_{2,n} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & W_{M,n} \end{pmatrix} \begin{pmatrix} x_{1,n-1} \\ x_{2,n-1} \\ \vdots \\ x_{M,n-1} \end{pmatrix} + \begin{pmatrix} B_{1,n} \\ B_{2,n} \\ \vdots \\ B_{M,n} \end{pmatrix} \\ &= \begin{pmatrix} W_{1,n}x_{1,n-1} + B_{1,n} \\ W_{2,n}x_{2,n-1} + B_{2,n} \\ \vdots \\ W_{M,n}x_{M,n-1} + B_{M,n} \end{pmatrix}. \end{aligned} \tag{67}$$

This yields that

$$x_n = \mathbf{A}_{k_n} (W_n x_{n-1} + B_n) = \begin{pmatrix} x_{1,n} \\ x_{2,n} \\ \vdots \\ x_{M,n} \end{pmatrix}. \tag{68}$$

This proves the induction step. Induction now proves for all $n \in [1, H] \cap \mathbb{N}$ that $x_n = (x_{1,n}, x_{2,n}, \dots, x_{M,n})$. This, the definition of \mathcal{R} (see (32)), and (61) imply that

$$\begin{aligned} (\mathcal{R}(\psi))(x_0) &= W_{H+1} x_H + B_{H+1} \\ &= W_{H+1} \begin{pmatrix} x_{1,H} \\ x_{2,H} \\ \vdots \\ x_{M,H} \end{pmatrix} + B_{H+1} = (h_1 W_{1,H+1} \quad \dots \quad h_M W_{M,H+1}) \begin{pmatrix} x_{1,H} \\ x_{2,H} \\ \vdots \\ x_{M,H} \end{pmatrix} + \left[\sum_{i=1}^M h_i B_{i,H+1} \right] \\ &= \left[\sum_{i=1}^M h_i W_{i,H+1} x_{i,H} \right] + \left[\sum_{i=1}^M h_i B_{i,H+1} \right] = \sum_{i=1}^M h_i (W_{i,H+1} x_{i,H} + B_{i,H+1}) \\ &= \sum_{i=1}^M h_i (\mathcal{R}(\phi_i))(x_0). \end{aligned} \tag{69}$$

This, the fact that $x_0 \in \mathbb{R}^p$ was arbitrary, and (56) yield that

$$\mathcal{R}(\psi) = \sum_{i=1}^M h_i \mathcal{R}(\phi_i) = \sum_{i=1}^M h_i f_i. \tag{70}$$

This and (64) show that

$$\sum_{i=1}^M h_i f_i \in \mathcal{R} \left(\left\{ \Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \bigoplus_{i=1}^M k_i \right\} \right). \tag{71}$$

The proof of Lemma 3.9 is thus completed. □

3.3 Deep neural network representations for MLP approximations

Lemma 3.10 *Assume Setting 3.1, let $d, M \in \mathbb{N}$, $T, c \in (0, \infty)$, $f \in C(\mathbb{R}, \mathbb{R})$, $g \in C(\mathbb{R}^d, \mathbb{R})$, $\Phi_f, \Phi_g \in \mathbf{N}$ satisfy that $\mathcal{R}(\Phi_f) = f$, $\mathcal{R}(\Phi_g) = g$, and*

$$c \geq \max \{ 2, \|\mathcal{D}(\Phi_f)\|, \|\mathcal{D}(\Phi_g)\| \}, \tag{72}$$

let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, let $u^\theta : \Omega \rightarrow [0, 1]$, $\theta \in \Theta$, be independent random variables which are uniformly distributed on $[0, 1]$, let $\mathcal{U}^\theta : [0, T] \times \Omega \rightarrow [0, T]$, $\theta \in \Theta$, satisfy for all $t \in [0, T]$, $\theta \in \Theta$ that $\mathcal{U}_t^\theta = t + (T - t)u^\theta$, let $W^\theta : [0, T] \times \Omega \rightarrow \mathbb{R}^d$, $\theta \in \Theta$, be independent standard Brownian motions with continuous sample paths, assume that $(u^\theta)_{\theta \in \Theta}$ and $(W^\theta)_{\theta \in \Theta}$ are independent, let

$U_{n,M}^\theta : [0, T] \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{Z}$, $\theta \in \Theta$, satisfy for all $n \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that $U_{-1,M}^\theta(t, x) = U_{0,M}^\theta(t, x) = 0$ and

$$\begin{aligned}
 U_{n,M}^\theta(t, x) &= \frac{1}{M^n} \sum_{i=1}^{M^n} g\left(x + W_T^{(\theta,0,-i)} - W_t^{(\theta,0,-i)}\right) \\
 &+ \sum_{l=0}^{n-1} \frac{(T-t)}{M^{n-l}} \left[\sum_{i=1}^{M^{n-l}} \left(f \circ U_{l,M}^{(\theta,l,i)} - \mathbb{1}_{\mathbb{N}}(l) f \circ U_{l-1,M}^{(\theta,-l,i)} \right) \right. \\
 &\quad \left. \left(\mathcal{U}_t^{(\theta,l,i)}, x + W_{\mathcal{U}_t^{(\theta,l,i)}}^{(\theta,l,i)} - W_t^{(\theta,l,i)} \right) \right], \tag{73}
 \end{aligned}$$

and let $\omega \in \Omega$. Then for all $n \in \mathbb{N}_0$ there exists a family $(\Phi_{n,t}^\theta)_{\theta \in \Theta, t \in [0, T]} \subseteq \mathbf{N}$ such that

- (i) it holds for all $t_1, t_2 \in [0, T]$, $\theta_1, \theta_2 \in \Theta$ that

$$\mathcal{D}(\Phi_{n,t_1}^{\theta_1}) = \mathcal{D}(\Phi_{n,t_2}^{\theta_2}), \tag{74}$$

- (ii) it holds for all $t \in [0, T]$, $\theta \in \Theta$ that

$$\dim(\mathcal{D}(\Phi_{n,t}^\theta)) = n(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)), \tag{75}$$

- (iii) it holds for all $t \in [0, T]$, $\theta \in \Theta$ that

$$\|\mathcal{D}(\Phi_{n,t}^\theta)\| \leq c(3M)^n, \tag{76}$$

and

- (iv) it holds for all $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that

$$U_{n,M}^\theta(t, x, \omega) = (\mathcal{R}(\Phi_{n,t}^\theta))(x). \tag{77}$$

Proof of Lemma 3.10 We prove Lemma 3.10 by induction on $n \in \mathbb{N}_0$. For the base case $n = 0$ note that the fact that $\forall t \in [0, T], \theta \in \Theta : U_{0,M}^\theta(t, \cdot, \omega) = 0$, the fact that the function 0 can be represented by a network with depth $\dim(\mathcal{D}(\Phi_g))$, and (72) imply that there exists $(\Phi_{0,t}^\theta)_{\theta \in \Theta, t \in [0, T]} \subseteq \mathbf{N}$ such that it holds for all $t_1, t_2 \in [0, T]$, $\theta_1, \theta_2 \in \Theta$ that $\mathcal{D}(\Phi_{0,t_1}^{\theta_1}) = \mathcal{D}(\Phi_{0,t_2}^{\theta_2})$ and such that it holds for all $\theta \in \Theta, t \in [0, T]$ that $\dim(\mathcal{D}(\Phi_{0,t}^\theta)) = \dim(\mathcal{D}(\Phi_g))$, $\|\mathcal{D}(\Phi_{0,t}^\theta)\| \leq \|\mathcal{D}(\Phi_g)\| \leq c$, and $U_{0,M}^\theta(t, \cdot, \omega) = \mathcal{R}(\Phi_{0,t}^\theta)$. This proves the base case $n = 0$.

For the induction step from $n \in \mathbb{N}_0$ to $n + 1 \in \mathbb{N}$ let $n \in \mathbb{N}_0$ and assume that (i)–(iv) hold true for all $k \in [0, n] \cap \mathbb{N}_0$. The assumption that $g = \mathcal{R}(\Phi_g)$ and Lemma 3.7 (with $d = d, m = 1, \lambda = 1, a = 0, b = W_T^\theta(\omega) - W_t^\theta(\omega)$, and $\Psi = \Phi_g$ for $\theta \in \Theta, t \in [0, T]$ in the notation of Lemma 3.7) show for all $\theta \in \Theta, t \in [0, T]$ that

$$\begin{aligned}
 g(\cdot + W_T^\theta(\omega) - W_t^\theta(\omega)) &= (\mathcal{R}(\Phi_g))(\cdot + W_T^\theta(\omega) - W_t^\theta(\omega)) \\
 &\in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathcal{D}(\Phi_g)\}). \tag{78}
 \end{aligned}$$

Furthermore, Lemma 3.6 (with $H = (n + 1)(\dim(\mathcal{D}(\Phi_f)) - 1) - 1$ in the notation of Lemma 3.6) ensures that

$$\text{Id}_{\mathbb{R}} \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}\right\}\right). \tag{79}$$

This, (78), and Lemma 3.8 (with $d_1 = d$, $d_2 = 1$, $d_3 = 1$, $f = \text{Id}_{\mathbb{R}}$, $g = g(\cdot + W_T^\theta(\omega) - W_t^\theta(\omega))$, $\alpha = \mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1}$, and $\beta = \mathcal{D}(\Phi_g)$ for $\theta \in \Theta$, $t \in [0, T]$ in the notation of Lemma 3.8) show that for all $\theta \in \Theta$, $t \in [0, T]$ it holds that

$$g(\cdot + W_T^\theta(\omega) - W_t^\theta(\omega)) \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g)\right\}\right). \tag{80}$$

Next, the induction hypothesis implies for all $\theta \in \Theta$, $t \in [0, T]$, $l \in [0, n] \cap \mathbb{N}_0$ that

$$U_{l,M}^\theta(t, \cdot, \omega) = \mathcal{R}(\Phi_{l,t}^\theta) \quad \text{and} \quad \mathcal{D}(\Phi_{l,t}^\theta) = \mathcal{D}(\Phi_{l,0}^0). \tag{81}$$

This and Lemma 3.7 (with

$$\begin{aligned} d &= d, \quad m = 1, \quad a = 0, \quad b = W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \quad \text{and} \\ \Psi &= \Phi_{l, \mathcal{U}_t^\theta(\omega)}^\eta \quad \text{for} \quad \theta, \eta \in \Theta, \quad t \in [0, T], \quad l \in [0, n] \cap \mathbb{N}_0 \end{aligned} \tag{82}$$

in the notation of Lemma 3.7) imply that for all $\theta, \eta \in \Theta$, $t \in [0, T]$, $l \in [0, n] \cap \mathbb{N}_0$ it holds that

$$\begin{aligned} &U_{l,M}^\eta\left(\mathcal{U}_t^\theta(\omega), \cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \omega\right) \\ &= \left(\mathcal{R}(\Phi_{l, \mathcal{U}_t^\theta(\omega)}^\eta)\right)\left(\cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega)\right) \\ &\in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathcal{D}(\Phi_{l, \mathcal{U}_t^\theta(\omega)}^\eta)\right\}\right) = \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathcal{D}(\Phi_{l,0}^0)\right\}\right). \end{aligned} \tag{83}$$

Moreover, Lemma 3.6 (with $H = (n - l)(\dim(\mathcal{D}(\Phi_f)) - 1) - 1$ for $l \in [0, n - 1] \cap \mathbb{N}_0$ in the notation of Lemma 3.6) ensures for all $l \in [0, n - 1] \cap \mathbb{N}_0$ that

$$\text{Id}_{\mathbb{R}} \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1}\right\}\right). \tag{84}$$

This, (83), and Lemma 3.8 (with

$$\begin{aligned} d_1 &= d, \quad d_2 = 1, \quad d_3 = 1, \quad f = \text{Id}_{\mathbb{R}}, \quad \alpha = \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1}, \\ \beta &= \mathcal{D}(\Phi_{l,0}^0), \quad \text{and} \quad g = U_{l,M}^\eta\left(\mathcal{U}_t^\theta(\omega), \cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \omega\right) \\ &\text{for} \quad \eta, \theta \in \Theta, \quad t \in [0, T], \quad l \in [0, n - 1] \cap \mathbb{N}_0 \end{aligned} \tag{85}$$

in the notation of Lemma 3.8) prove for all $\eta, \theta \in \Theta$, $t \in [0, T]$, $l \in [0, n - 1] \cap \mathbb{N}_0$ that

$$\begin{aligned} &U_{l,M}^\eta\left(\mathcal{U}_t^\theta(\omega), \cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \omega\right) \\ &\in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l,0}^0)\right\}\right). \end{aligned} \tag{86}$$

This and Lemma 3.8 (with

$$\begin{aligned}
 d_1 = d, \quad d_2 = 1, \quad d_3 = 1, \quad f = f, \quad \alpha = \mathcal{D}(\Phi_f), \\
 \beta = \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l,0}^0), \quad \text{and} \quad g = U_{l,M}^\eta \left(\mathcal{U}_t^\theta(\omega), \cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \omega \right) \\
 \text{for } \eta, \theta \in \Theta, \quad t \in [0, T], \quad l \in [0, n-1] \cap \mathbb{N}_0
 \end{aligned} \tag{87}$$

in the notation of Lemma 3.8 assure for all $\eta, \theta \in \Theta, t \in [0, T], l \in [0, n-1] \cap \mathbb{N}_0$ that

$$\begin{aligned}
 & \left(f \circ U_{l,M}^\eta \right) \left(\mathcal{U}_t^\theta(\omega), \cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \omega \right) \\
 & \in \mathcal{R} \left(\left\{ \Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathcal{D}(\Phi_f) \odot \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l,0}^0) \right\} \right).
 \end{aligned} \tag{88}$$

Next, (83) (with $l = n$) and Lemma 3.8 (with

$$\begin{aligned}
 d_1 = d, \quad d_2 = 1, \quad d_3 = 1, \quad f = f, \quad \alpha = \mathcal{D}(\Phi_f), \quad \beta = \mathcal{D}(\Phi_{n,0}^0), \quad \text{and} \\
 g = \left(U_{n,M}^\eta \right) \left(\mathcal{U}_t^\theta(\omega), \cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \omega \right) \quad \text{for } \eta, \theta \in \Theta, \quad t \in [0, T]
 \end{aligned} \tag{89}$$

in the notation of Lemma 3.8) prove for all $\eta, \theta \in \Theta, t \in [0, T]$ that

$$\begin{aligned}
 & \left(f \circ U_{n,M}^\eta \right) \left(\mathcal{U}_t^\theta(\omega), \cdot + W_{\mathcal{U}_t^\theta(\omega)}^\theta(\omega) - W_t^\theta(\omega), \omega \right) \\
 & \in \mathcal{R} \left(\left\{ \Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi_{n,0}^0) \right\} \right).
 \end{aligned} \tag{90}$$

Furthermore, the definition of \odot in (33) and the fact that

$$\forall l \in [0, n] \cap \mathbb{N}_0 : \dim \left(\mathcal{D}(\Phi_{l,0}^0) \right) = l(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)) \tag{91}$$

in the induction hypothesis imply that

$$\begin{aligned}
 & \dim \left(\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g) \right) \\
 & = [(n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + 1] + \dim(\mathcal{D}(\Phi_g)) - 1 \\
 & = (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)),
 \end{aligned} \tag{92}$$

that

$$\begin{aligned}
 & \dim \left(\mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi_{n,0}^0) \right) = \dim(\mathcal{D}(\Phi_f)) + \dim \left(\mathcal{D}(\Phi_{n,0}^0) \right) - 1 \\
 & = \dim(\mathcal{D}(\Phi_f)) + [n(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g))] - 1 \\
 & = (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)),
 \end{aligned} \tag{93}$$

and for all $l \in [0, n-1] \cap \mathbb{N}_0$ that

$$\begin{aligned}
 & \dim\left(\mathcal{D}(\Phi_f) \odot \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l,0}^0)\right) \\
 &= \dim(\mathcal{D}(\Phi_f)) + \dim\left(\mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1}\right) + \dim\left(\mathcal{D}(\Phi_{l,0}^0)\right) - 2 \\
 &= \dim(\mathcal{D}(\Phi_f)) + [(n-l)(\dim(\mathcal{D}(\Phi_f)) - 1) + 1] \\
 &\quad + [l(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g))] - 2 \\
 &= \dim(\mathcal{D}(\Phi_f)) + n(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)) - 1 \\
 &= (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)).
 \end{aligned} \tag{94}$$

This shows, roughly speaking, that the functions in (80), (90), and (88) can be represented by networks with the same depth (i.e. number of layers): $(n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g))$. Hence, Lemma 3.9 and (73) imply that there exists a family $(\Phi_{n+1,t}^\theta)_{\theta \in \Theta, t \in [0,T]} \subseteq \mathbf{N}$ such that for all $\theta \in \Theta, t \in [0, T], x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
 & \left(\mathcal{R}(\Phi_{n+1,t}^\theta)\right)(x) \\
 &= \frac{1}{M^{n+1}} \sum_{i=1}^{M^{n+1}} g\left(x + W_T^{(\theta,0,-i)}(\omega) - W_i^{(\theta,0,-i)}(\omega)\right) \\
 &\quad + \frac{(T-t)}{M} \sum_{i=1}^M \left(f \circ U_{n,M}^{(\theta,n,i)}\right)\left(\mathcal{U}_i^{(\theta,n,i)}(\omega), x + W_{\mathcal{U}_i^{(\theta,n,i)}(\omega)}^{(\theta,n,i)}(\omega) - W_i^{(\theta,n,i)}(\omega), \omega\right) \\
 &\quad + \sum_{l=0}^{n-1} \frac{(T-t)}{M^{n+1-l}} \sum_{i=1}^{M^{n+1-l}} \left(f \circ U_{l,M}^{(\theta,l,i)}\right)\left(\mathcal{U}_i^{(\theta,l,i)}(\omega), x + W_{\mathcal{U}_i^{(\theta,l,i)}(\omega)}^{(\theta,l,i)}(\omega) - W_i^{(\theta,l,i)}(\omega), \omega\right) \\
 &\quad - \sum_{l=1}^n \frac{(T-t)}{M^{n+1-l}} \sum_{i=1}^{M^{n+1-l}} \left(f \circ U_{l-1,M}^{(\theta,l,i)}\right)\left(\mathcal{U}_i^{(\theta,l,i)}(\omega), x + W_{\mathcal{U}_i^{(\theta,l,i)}(\omega)}^{(\theta,l,i)}(\omega) - W_i^{(\theta,l,i)}(\omega), \omega\right) \\
 &= U_{n+1,M}^\theta(t, x, \omega),
 \end{aligned} \tag{95}$$

that

$$\dim\left(\mathcal{D}(\Phi_{n+1,t}^\theta)\right) = (n+1)(\dim(\mathcal{D}(\Phi_f)) - 1) + \dim(\mathcal{D}(\Phi_g)), \tag{96}$$

and that

$$\begin{aligned}
 \mathcal{D}(\Phi_{n+1,t}^\theta) &= \left(\boxplus_{i=1}^{M^{n+1}} \left[\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g)\right]\right) \boxplus \left(\boxplus_{i=1}^M \left(\mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi_{n,0}^0)\right)\right) \\
 &\quad \boxplus \left(\boxplus_{l=0}^{n-1} \boxplus_{i=1}^{M^{n+1-l}} \left[\left(\mathcal{D}(\Phi_f) \odot \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l,0}^0)\right)\right]\right) \\
 &\quad \boxplus \left(\boxplus_{l=1}^n \boxplus_{i=1}^{M^{n+1-l}} \left(\mathcal{D}(\Phi_f) \odot \mathfrak{n}_{(n-l+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l-1,0}^0)\right)\right).
 \end{aligned} \tag{97}$$

This shows for all $t_1, t_2 \in [0, T], \theta_1, \theta_2 \in \Theta$ that

$$\mathcal{D}(\Phi_{n+1,t_1}^{\theta_1}) = \mathcal{D}(\Phi_{n+1,t_2}^{\theta_2}). \tag{98}$$

Furthermore, (97), the triangle inequality (see Lemma 3.5), and the fact that

$$\forall l \in [0, n] \cap \mathbb{N}_0 : \|\mathcal{D}(\Phi_{l,0}^0)\| \leq c(3M)^l \tag{99}$$

in the induction hypothesis show for all $\theta \in \Theta, t \in [0, T]$ that

$$\begin{aligned} \|\mathcal{D}(\Phi_{n+1,t}^\theta)\| &\leq \sum_{i=1}^{M^{n+1}} \|\mathfrak{n}_{(n+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_g)\| + \sum_{i=1}^M \|\mathcal{D}(\Phi_f) \odot \mathcal{D}(\Phi_{n,0}^0)\| \\ &\quad + \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n+1-l}} \|\mathcal{D}(\Phi_f) \odot \mathfrak{n}_{(n-l)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l,0}^0)\| \\ &\quad + \sum_{l=1}^n \sum_{i=1}^{M^{n+1-l}} \|\mathcal{D}(\Phi_f) \odot \mathfrak{n}_{(n-l+1)(\dim(\mathcal{D}(\Phi_f))-1)+1} \odot \mathcal{D}(\Phi_{l-1,0}^0)\|. \end{aligned} \tag{100}$$

Note that for all $H_1, H_2, \alpha_0, \dots, \alpha_{H_1+1}, \beta_0, \dots, \beta_{H_2+1} \in \mathbb{N}, \alpha, \beta \in \mathbf{D}$ with $\alpha = (\alpha_0, \dots, \alpha_{H_1+1}), \beta = (\beta_0, \dots, \beta_{H_2+1}), \alpha_0 = \beta_{H_2+1} = 1$ it holds that $\|\alpha \odot \beta\| \leq \max\{\|\alpha\|, \|\beta\|, 2\}$ (see (33)). This, (100), the fact that $\forall H \in \mathbb{N} : \|\mathfrak{n}_{H+2}\| = 2$ (see (35), (72), and (99)) prove for all $\theta \in \Theta, t \in [0, T]$ that

$$\begin{aligned} &\|\mathcal{D}(\Phi_{n+1,t}^\theta)\| \\ &\leq \left[\sum_{i=1}^{M^{n+1}} c \right] + \left[\sum_{i=1}^M c(3M)^n \right] + \left[\sum_{l=0}^{n-1} \sum_{i=1}^{M^{n+1-l}} c(3M)^l \right] + \left[\sum_{l=1}^n \sum_{i=1}^{M^{n+1-l}} c(3M)^{l-1} \right] \\ &= M^{n+1}c + Mc(3M)^n + \left[\sum_{l=0}^{n-1} M^{n+1-l}c(3M)^l \right] + \left[\sum_{l=1}^n M^{n+1-l}c(3M)^{l-1} \right] \\ &\leq M^{n+1}c \left[1 + 3^n + \sum_{l=0}^{n-1} 3^l + \sum_{l=1}^n 3^{l-1} \right] = M^{n+1}c \left[1 + \sum_{l=0}^n 3^l + \sum_{l=1}^n 3^{l-1} \right] \\ &\leq cM^{n+1} \left[1 + 2 \sum_{l=0}^n 3^l \right] = cM^{n+1} \left[1 + 2 \frac{3^{n+1} - 1}{3 - 1} \right] = c(3M)^{n+1}. \end{aligned} \tag{101}$$

Combining (95), (96), (98), and (101) completes the induction step. Induction hence establishes (i)–(iv). The proof of Lemma 3.10 is thus completed. \square

3.4 Deep neural network approximations for the PDE nonlinearity

Lemma 3.11 (DNN interpolation) *Assume Setting 3.1, let $N \in \mathbb{N}, a_0, a_1, \dots, a_{N-1}, \zeta_0, \zeta_1, \dots, \zeta_N \in \mathbb{R}$ satisfy that $\zeta_0 < \zeta_1 < \dots < \zeta_N$, let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a function, assume for all $x \in (-\infty, \zeta_0]$ that $f(x) = f(\zeta_0)$, assume for all $n \in [0, N - 1] \cap \mathbb{Z}$,*

$x \in (\zeta_n, \zeta_{n+1}]$ that $f(x) = f(\zeta_n) + a_n(x - \zeta_n)$, and assume for all $x \in (\zeta_N, \infty)$ that $f(x) = f(\zeta_N)$. Then it holds that

$$f \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = (1, N + 1, 1)\}). \tag{102}$$

Proof of Lemma 3.11 Throughout this proof let $a_{-1} = 0$ and $a_N = 0$, let $c_n \in \mathbb{R}, n \in [0, N] \cap \mathbb{Z}$, be the real numbers which satisfy for all $n \in [0, N] \cap \mathbb{Z}$ that $c_n = a_n - a_{n-1}$, let $W_1 \in \mathbb{R}^{(N+1) \times 1}, B_1 \in \mathbb{R}^{N+1}, W_2 \in \mathbb{R}^{1 \times (N+1)}, B_2 \in \mathbb{R}, \Phi \in \mathbf{N}$ be given by

$$W_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} -\zeta_0 \\ -\zeta_1 \\ \vdots \\ -\zeta_N \end{pmatrix}, \quad W_2 = (c_0 \quad c_1 \quad \dots \quad c_N), \quad B_2 = f(\zeta_0), \tag{103}$$

and

$$\Phi = ((W_1, B_1), (W_2, B_2)), \tag{104}$$

and let $g : \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}$ that

$$g(x) = f(\zeta_0) + \sum_{k=0}^N c_k \max\{x - \zeta_k, 0\}. \tag{105}$$

First, observe that the fact that $\forall n \in [0, N - 1] \cap \mathbb{Z} : \zeta_n < \zeta_{n+1}$ and the fact that $\forall n \in [0, N] \cap \mathbb{Z} : a_n = \sum_{k=0}^n c_k$ then show for all $n \in [0, N - 1] \cap \mathbb{Z}, x \in (\zeta_n, \zeta_{n+1}]$ that

$$\begin{aligned} g(x) - g(\zeta_n) &= \left[\sum_{k=0}^N c_k (\max\{x - \zeta_k, 0\} - \max\{\zeta_n - \zeta_k, 0\}) \right] \\ &= \sum_{k=0}^n c_k [(x - \zeta_k) - (\zeta_n - \zeta_k)] = \sum_{k=0}^n c_k (x - \zeta_n) = a_n(x - \zeta_n). \end{aligned} \tag{106}$$

This shows for all $n \in [0, N - 1] \cap \mathbb{Z}$ that g is affine linear on the interval $(\zeta_n, \zeta_{n+1}]$. This, the fact that for all $n \in [0, N - 1] \cap \mathbb{Z}$ it holds that f is affine linear on the interval $(\zeta_n, \zeta_{n+1}]$, the fact that $\forall x \in (-\infty, \zeta_0] : f(x) = g(x) = f(\zeta_0)$, and an induction argument imply for all $x \in (-\infty, \zeta_N]$ that $f(x) = g(x)$. Furthermore, (105), the fact that $\forall n \in [0, N - 1] \cap \mathbb{Z} : \zeta_n < \zeta_{n+1}$, and the fact that $\sum_{k=0}^N c_k = 0$ imply for all $x \in (\zeta_N, \infty)$ that

$$\begin{aligned} g(x) - g(\zeta_N) &= \left[\sum_{k=0}^N c_k (\max\{x - \zeta_k, 0\} - \max\{\zeta_N - \zeta_k, 0\}) \right] \\ &= \sum_{k=0}^N c_k [(x - \zeta_k) - (\zeta_N - \zeta_k)] = \sum_{k=0}^N c_k (x - \zeta_N) = 0. \end{aligned} \tag{107}$$

This shows for all $x \in (\zeta_N, \infty)$ that $g(x) = g(\zeta_N)$. This, the fact that $\forall x \in (\zeta_N, \infty) : f(x) = f(\zeta_N)$, the fact that $\forall x \in (-\infty, \zeta_N] : f(x) = g(x)$, and (105) prove for all $x \in \mathbb{R}$ that

$$f(x) = g(x) = f(\zeta_0) + \sum_{k=0}^N c_k \max\{x - \zeta_k, 0\}. \tag{108}$$

Next, the definition of \mathcal{R} and \mathcal{D} (see (31) and (32)), (103), (104), and (108) imply that for all $x \in \mathbb{R}$ it holds that $\mathcal{D}(\Phi) = (1, N + 1, 1)$ and

$$\begin{aligned} (\mathcal{R}(\Phi))(x) &= W_2(\mathbf{A}_{N+1}(W_1x + B_1)) + B_2 \\ &= (c_0 \quad c_1 \quad \dots \quad c_N) \begin{pmatrix} \max\{x - \zeta_0, 0\} \\ \max\{x - \zeta_1, 0\} \\ \vdots \\ \max\{x - \zeta_N, 0\} \end{pmatrix} + f(\zeta_0) = f(\zeta_0) + \sum_{k=0}^N c_k \max\{x - \zeta_k, 0\} = f(x). \end{aligned} \tag{109}$$

This establishes (102). The proof of Lemma 3.11 is thus completed. \square

Lemma 3.12 *Let $L \in [0, \infty)$, $N \in \mathbb{N}$, $a \in \mathbb{R}$, $b \in (a, \infty)$, $\zeta_0, \zeta_1, \dots, \zeta_N \in \mathbb{R}$ satisfy for all $n \in [0, N] \cap \mathbb{Z}$ that $\zeta_n = a + \frac{(b-a)n}{N}$, let $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that*

$$|f(x) - f(y)| \leq L|x - y|, \tag{110}$$

and let $g : \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}$, $n \in [0, N - 1] \cap \mathbb{Z}$ that

$$g(x) = \begin{cases} f(\zeta_0) & : x \in (-\infty, \zeta_0] \\ \frac{f(\zeta_n)(\zeta_{n+1} - x) + f(\zeta_{n+1})(x - \zeta_n)}{\zeta_{n+1} - \zeta_n} & : x \in (\zeta_n, \zeta_{n+1}] \\ f(\zeta_N) & : x \in (\zeta_N, \infty). \end{cases} \tag{111}$$

Then

- (i) *it holds for all $n \in [0, N] \cap \mathbb{Z}$ that $g(\zeta_n) = f(\zeta_n)$,*
- (ii) *it holds for all $x, y \in \mathbb{R}$ that $|g(x) - g(y)| \leq L|x - y|$, and*
- (iii) *it holds for all $x \in [a, b]$ that $|g(x) - f(x)| \leq \frac{2L(b-a)}{N}$.*

Proof of Lemma 3.12 Throughout this proof let $r, \ell : \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R} \setminus (a, b)$ that

$$r(x) = \ell(x) = x \tag{112}$$

and for all $n \in [0, N - 1] \cap \mathbb{Z}$, $x \in (\zeta_n, \zeta_{n+1}]$ that

$$r(x) = \zeta_{n+1} \quad \text{and} \quad \ell(x) = \zeta_n. \tag{113}$$

Note that (111) implies (i). Next observe that for all $x, y \in (a, b]$ with $x \leq y$ and $\ell(y) < r(x)$ it holds that $r(x) = r(y)$ and $\ell(x) = \ell(y)$. This, (112), (111), and (110) show that for all $x, y \in \mathbb{R}$ with $x \leq y$ and $\ell(y) < r(x)$ it holds that $x, y \in (a, b]$, $r(x) = r(y)$, $\ell(x) = \ell(y)$, and

$$|g(x) - g(y)| = \left| \frac{f(r(x)) - f(\ell(x))}{r(x) - \ell(x)}(x - y) \right| \leq L|x - y|. \tag{114}$$

Furthermore, (111), (110), and the fact that $\forall x \in \mathbb{R} : \ell(x) \leq x \leq r(x)$ imply for all $x \in (a, b]$ that

$$\begin{aligned}
 |g(x) - g(r(x))| &= \left| \frac{f(\ell(x)) - f(r(x))}{\ell(x) - r(x)}(x - \ell(x)) + f(\ell(x)) - f(r(x)) \right| \\
 &= \left| \frac{(f(\ell(x)) - f(r(x)))(x - r(x))}{\ell(x) - r(x)} \right| \leq L|x - r(x)| = L(r(x) - x)
 \end{aligned}
 \tag{115}$$

and

$$\begin{aligned}
 g(x) - g(\ell(x)) &= \left| \frac{f(\ell(x)) - f(r(x))}{\ell(x) - r(x)}(x - \ell(x)) + f(\ell(x)) - f(\ell(x)) \right| \\
 &= \left| \frac{f(\ell(x)) - f(r(x))}{\ell(x) - r(x)}(x - \ell(x)) \right| \leq L|x - \ell(x)| = L(x - \ell(x)).
 \end{aligned}
 \tag{116}$$

This and (112) show for all $x \in \mathbb{R}$ that

$$|g(x) - g(r(x))| \leq L(r(x) - x) \quad \text{and} \quad |g(x) - g(\ell(x))| \leq L(x - \ell(x)).
 \tag{117}$$

The triangle inequality therefore shows for all $x, y \in \mathbb{R}$ with $x \leq y$ and $r(x) \leq \ell(y)$ that

$$\begin{aligned}
 |g(x) - g(y)| &\leq |g(x) - g(r(x))| + |g(r(x)) - g(\ell(y))| + |g(\ell(y)) - g(y)| \\
 &\leq L(r(x) - x) + L(\ell(y) - r(x)) + L(y - \ell(y)) = L(y - x) = L|y - x|.
 \end{aligned}
 \tag{118}$$

This and (114) show for all $x, y \in \mathbb{R}$ with $x \leq y$ that $|g(x) - g(y)| \leq L|x - y|$. Symmetry hence establishes (ii). Next, the fact that $\forall x \in \mathbb{R} : g(\ell(x)) = f(\ell(x))$, the triangle inequality, (110), (117), and the fact that $\forall x \in [a, b] : 0 \leq x - \ell(x) \leq (b - a)/N$ imply for all $x \in [a, b]$ that

$$\begin{aligned}
 |g(x) - f(x)| &= |g(x) - f(\ell(x)) + f(\ell(x)) - f(x)| \\
 &= |g(x) - g(\ell(x)) + f(\ell(x)) - f(x)| \\
 &\leq |g(x) - g(\ell(x))| + |f(\ell(x)) - f(x)| \leq 2L(x - \ell(x)) \leq 2L(b - a)/N.
 \end{aligned}
 \tag{119}$$

This establishes (iii). The proof of Lemma 3.12 is thus completed. □

Corollary 3.13 *Assume Setting 3.1, let $\epsilon \in (0, 1]$, $L \in [0, \infty)$, $q \in (1, \infty)$, and let $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leq L|x - y|$. Then there exists a function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that*

- (i) *it holds for all $x, y \in \mathbb{R}$ that $|g(x) - g(y)| \leq L|x - y|$,*
- (ii) *it holds for all $x \in \mathbb{R}$ that $|f(x) - g(x)| \leq \epsilon(1 + |x|^q)$, and*
- (iii) *it holds that*

$$g \in \mathcal{R} \left(\left\{ \Phi \in \mathbf{N} : \mathcal{D}(\Phi) \in \mathbb{N}^3 \text{ and } \|\mathcal{D}(\Phi)\| \leq \frac{4L \left(1 + (4L + 2|f(0)|)^{\frac{1}{q-1}} \right) + 2}{\epsilon^{\frac{q}{q-1}}} \right\} \right).
 \tag{120}$$

Proof of Corollary 3.13 Throughout this proof let $R \in \mathbb{R}$, $N \in \mathbb{N}$ satisfy that

$$R = \max\left(1, \left(\frac{4L + 2|f(0)|}{\epsilon}\right)^{\frac{1}{q-1}}\right) \quad \text{and} \quad N = \min\left\{n \in \mathbb{N} : \frac{4LR}{n} \leq \epsilon\right\}, \tag{121}$$

let $\xi_0, \xi_1, \dots, \xi_N \in \mathbb{R}$ be the real numbers which satisfy for all $n \in [0, N] \cap \mathbb{Z}$ that $\xi_n = R(-1 + \frac{2n}{N})$, and let $g : \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}$, $n \in [0, N - 1] \cap \mathbb{Z}$ that

$$g(x) = \begin{cases} f(\xi_0) & : x \in (-\infty, \xi_0] \\ \frac{f(\xi_n)(\xi_{n+1} - x) + f(\xi_{n+1})(x - \xi_n)}{\xi_{n+1} - \xi_n} & : x \in (\xi_n, \xi_{n+1}] \\ f(\xi_N) & : x \in (\xi_N, \infty). \end{cases} \tag{122}$$

By (ii) in Lemma 3.12 the function g satisfies (i). Next, it follows from (iii) in Lemma 3.12 that for all $x \in [-R, R]$ it holds $|g(x) - f(x)| \leq 4LR/N$. This and the fact that $4LR/N \leq \epsilon$ prove that for all $x \in [-R, R]$ it holds that $|g(x) - f(x)| \leq \epsilon \leq \epsilon(1 + |x|^q)$. Next, the triangle inequality, the fact that $f(R) = g(R)$, and the Lipschitz condition of f and g imply for all $x \in \mathbb{R}$ that

$$\begin{aligned} |f(x) - g(x)| &\leq |f(x) - f(0)| + |f(0)| + |g(x) - g(R)| + |g(R)| \\ &= |f(x) - f(0)| + |f(0)| + |g(x) - g(R)| + |f(R)| \\ &\leq |f(x) - f(0)| + |f(0)| + |g(x) - g(R)| + |f(R) - f(0)| + |f(0)| \tag{123} \\ &\leq L|x| + 2|f(0)| + L|x - R| + LR \\ &\leq 2L(|x| + R) + 2|f(0)|. \end{aligned}$$

This and (121) show for all $x \in \mathbb{R} \setminus [-R, R]$ that

$$\begin{aligned} \frac{|f(x) - g(x)|}{1 + |x|^q} &\leq \frac{2L(|x| + R) + 2|f(0)|}{1 + |x|^q} \leq \frac{4L|x| + 2|f(0)|}{1 + |x|^q} \tag{124} \\ &\leq \frac{4L}{|x|^{q-1}} + \frac{2|f(0)|}{|x|^q} \leq \frac{4L}{R^{q-1}} + \frac{2|f(0)|}{R^q} \leq \frac{4L + 2|f(0)|}{R^{q-1}} \leq \epsilon. \end{aligned}$$

This and the fact that $\forall x \in [-R, R] : |g(x) - f(x)| \leq \epsilon(1 + |x|^q)$ prove that for all $x \in \mathbb{R}$ it holds that $|g(x) - f(x)| \leq \epsilon(1 + |x|^q)$. This shows that g satisfies (ii). Next, (i) in Lemma 3.12 ensures that g satisfies for all $x \in (-\infty, -R]$ that $g(x) = g(-R)$, for all $n \in [0, N - 1] \cap \mathbb{Z}$, $x \in (\xi_n, \xi_{n+1}]$ that $g(x) = g(\xi_n) + \frac{g(\xi_{n+1}) - g(\xi_n)}{\xi_{n+1} - \xi_n}(x - \xi_n)$, and for all $x \in (R, \infty)$ that $g(x) = g(R)$. This and Lemma 3.11 (with $N = N$, $f = g$, $\xi_n = \xi_n$ for $n \in [0, N] \cap \mathbb{Z}$, and $a_n = (g(\xi_{n+1}) - g(\xi_n))/(\xi_{n+1} - \xi_n)$ for $n \in [0, N - 1] \cap \mathbb{Z}$ in the notation of Lemma 3.11) imply that

$$g \in \mathcal{R}(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = (1, N + 1, 1)\}). \tag{125}$$

Furthermore, if $N > 1$, then (121) implies that $\frac{4LR}{N-1} > \epsilon$. Hence, if $N > 1$ it holds that $N < \frac{4LR}{\epsilon} + 1$. This and (121) ensure that

$$N \leq \frac{4LR}{\epsilon} + 1 = \frac{4L \max\left(1, \left(\frac{4L + 2|f(0)|}{\epsilon}\right)^{\frac{1}{q-1}}\right) + \epsilon}{\epsilon}. \tag{126}$$

This and (125) imply that

$$\begin{aligned}
 |||\mathcal{D}(\Phi)||| &= N + 1 \leq \frac{4L \max\left(1, \left(\frac{4L+2|f(0)|}{\epsilon}\right)^{\frac{1}{q-1}}\right) + 2\epsilon}{\epsilon} \\
 &= \frac{4L \max\left(\epsilon^{\frac{1}{q-1}}, (4L + 2|f(0)|)^{\frac{1}{q-1}}\right) + 2\epsilon^{\frac{q}{q-1}}}{\epsilon^{\frac{q}{q-1}}} \tag{127} \\
 &\leq \frac{4L\left(1 + (4L + 2|f(0)|)^{\frac{1}{q-1}}\right) + 2}{\epsilon^{\frac{q}{q-1}}}.
 \end{aligned}$$

This establishes (iii). The proof of Corollary 3.13 is thus completed. \square

4 Deep neural network approximations for PDEs

4.1 Deep neural network approximations with specific polynomial convergence rates

Theorem 4.1 *Let $\|\cdot\|, |||\cdot||| : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow [0, \infty)$ and $\dim : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow \mathbb{N}$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $\|x\| = [\sum_{i=1}^d (x_i)^2]^{1/2}$, $|||x||| = \max_{i \in [1, d] \cap \mathbb{N}} |x_i|$, and $\dim(x) = d$, let $T, L, B, \beta \in [0, \infty)$, $p, \mathfrak{p} \in \mathbb{N}$, $q \in \mathbb{N} \cap [2, \infty)$, $\alpha \in [2, \infty)$, let $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leq L|x - y|$, for every $d \in \mathbb{N}$ let $g_d \in C(\mathbb{R}^d, \mathbb{R})$, for every $d \in \mathbb{N}$ let $\nu_d : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, for every $d \in \mathbb{N}$ let $\mathbf{A}_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfy for all $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that*

$$\mathbf{A}_d(x) = (\max\{x_1, 0\}, \dots, \max\{x_d, 0\}), \tag{128}$$

let $\mathbf{D} = \cup_{H \in \mathbb{N}} \mathbb{N}^{H+2}$, let

$$\mathbf{N} = \bigcup_{H \in \mathbb{N}} \bigcup_{(k_0, k_1, \dots, k_{H+1}) \in \mathbb{N}^{H+2}} \left[\prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}) \right], \tag{129}$$

let $\mathcal{P} : \mathbf{N} \rightarrow \mathbb{N}$, $\mathcal{D} : \mathbf{N} \rightarrow \mathbf{D}$, and $\mathcal{R} : \mathbf{N} \rightarrow (\cup_{k, l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ satisfy for all $H \in \mathbb{N}$, $k_0, k_1, \dots, k_H, k_{H+1} \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n})$, $x_0 \in \mathbb{R}^{k_0}, \dots, x_H \in \mathbb{R}^{k_H}$ with $\forall n \in \mathbb{N} \cap [1, H] : x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ that

$$\mathcal{P}(\Phi) = \sum_{n=1}^{H+1} k_n(k_{n-1} + 1), \quad \mathcal{D}(\Phi) = (k_0, k_1, \dots, k_H, k_{H+1}), \tag{130}$$

$$\mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}}), \quad \text{and} \quad (\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1}, \tag{131}$$

for every $\epsilon \in (0, 1]$, $d \in \mathbb{N}$ let $g_{d, \epsilon} \in \mathbf{N}$, and assume for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\epsilon \in (0, 1]$ that $\mathcal{R}(g_{d, \epsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $|(\mathcal{R}(g_{d, \epsilon}))(x)| \leq Bd^p(1 + \|x\|)^p$, $|g_d(x) - (\mathcal{R}(g_{d, \epsilon}))(x)| \leq \epsilon Bd^p(1 + \|x\|)^{pq}$, $|||\mathcal{D}(g_{d, \epsilon})||| \leq Bd^p \epsilon^{-\alpha}$, $\dim(\mathcal{D}(g_{d, \epsilon})) \leq Bd^p \epsilon^{-\beta}$, and $(\int_{\mathbb{R}^d} \|y\|^{2pq} \nu_d(dy))^{1/(2pq)} \leq Bd^p$. Then

- (i) *there exist unique $u_d \in C([0, T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, such that for every $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $s \in [0, T]$, every probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and every standard Brownian motion $\mathbf{W} : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ with continuous sample paths it holds that $\sup_{t \in [0, T]} \sup_{y \in \mathbb{R}^d} \left(\frac{|u_d(t, y)|}{1 + \|y\|^p} \right) < \infty$ and*

$$u_d(s, x) = \mathbb{E} \left[g_d(x + \mathbf{W}_{T-s}) + \int_s^T f(u_d(t, x + \mathbf{W}_{t-s})) dt \right] \tag{132}$$

and

- (ii) *there exist $(\Psi_{d,\varepsilon})_{d \in \mathbb{N}, \varepsilon \in (0, 1]} \subseteq \mathbf{N}$, $\eta \in (0, \infty)$, $C = (C_\gamma)_{\gamma \in (0, 1]} : (0, 1] \rightarrow (0, \infty)$ such that for all $d \in \mathbb{N}$, $\varepsilon, \gamma \in (0, 1]$ it holds that $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\Psi_{d,\varepsilon}) \leq C_\gamma d^\eta \varepsilon^{-(4+2\alpha+\beta+\gamma)}$, and*

$$\left[\int_{\mathbb{R}^d} |u_d(0, x) - (\mathcal{R}(\Psi_{d,\varepsilon}))(x)|^2 v_d(dx) \right]^{\frac{1}{2}} \leq \varepsilon. \tag{133}$$

Proof of Theorem 4.1 Throughout this proof assume without loss of generality that

$$B \geq \max \left\{ |f(0)| + 1, 4L \left(1 + (4L + 2|f(0)|)^{\frac{1}{q-1}} \right) + 2 \right\}. \tag{134}$$

Note that the triangle inequality, the fact that $\forall d \in \mathbb{N}, x \in \mathbb{R}^d, \varepsilon \in (0, 1] : |(\mathcal{R}(g_{d,\varepsilon}))(x)| \leq Bd^p(1 + \|x\|)^p$, and the fact that $\forall d \in \mathbb{N}, x \in \mathbb{R}^d, \varepsilon \in (0, 1] : |g_d(x) - (\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \varepsilon Bd^p(1 + \|x\|)^{pq}$ imply for all $d \in \mathbb{N}, x \in \mathbb{R}^d, \varepsilon \in (0, 1]$ that

$$|g_d(x)| \leq |g_d(x) - (\mathcal{R}(g_{d,\varepsilon}))(x)| + |(\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \varepsilon Bd^p(1 + \|x\|)^{pq} + Bd^p(1 + \|x\|)^p. \tag{135}$$

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

$$|g_d(x)| \leq Bd^p(1 + \|x\|)^p. \tag{136}$$

Corollary 3.11 in [17], the fact that f is globally Lipschitz continuous, and (136) hence establish (i). It thus remains to prove (ii). To this end note that Corollary 3.13 ensures that there exist $\mathfrak{f}_\varepsilon \in \mathbf{N}$, $\varepsilon \in (0, 1]$, which satisfy for all $v, w \in \mathbb{R}$, $\varepsilon \in (0, 1]$ that $\mathcal{R}(\mathfrak{f}_\varepsilon) \in C(\mathbb{R}, \mathbb{R})$, $|(\mathcal{R}(\mathfrak{f}_\varepsilon))(w) - (\mathcal{R}(\mathfrak{f}_\varepsilon))(v)| \leq L|w - v|$, $|f(v) - (\mathcal{R}(\mathfrak{f}_\varepsilon))(v)| \leq \varepsilon(1 + |v|^q)$, $\dim(\mathcal{D}(\mathfrak{f}_\varepsilon)) = 3$, and

$$\| \mathcal{D}(\mathfrak{f}_\varepsilon) \| \leq \left[4L \left(1 + (4L + 2|f(0)|)^{\frac{1}{q-1}} \right) + 2 \right] \varepsilon^{-\frac{q}{(q-1)}}. \tag{137}$$

Note that the fact that $B \geq 1 + |f(0)|$ implies for all $\varepsilon \in (0, 1]$ that

$$|(\mathcal{R}(\mathfrak{f}_\varepsilon))(0)| \leq |(\mathcal{R}(\mathfrak{f}_\varepsilon))(0) - f(0)| + |f(0)| \leq \varepsilon + |f(0)| \leq B. \tag{138}$$

Next let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, for every $d \in \mathbb{N}$ let $\mathbf{W}^d : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ be a standard Brownian motion with continuous sample paths, let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, let $u^\theta : \Omega \rightarrow [0, 1]$, $\theta \in \Theta$, be independent random variables which are uniformly distributed on $[0, 1]$, let $U^\theta : [0, T] \times \Omega \rightarrow [0, T]$, $\theta \in \Theta$, satisfy for all $t \in [0, T]$, $\theta \in \Theta$ that

$\mathcal{U}_t^\theta = t + (T - t)u^\theta$, for every $d \in \mathbb{N}$ let $W^{\theta,d} : [0, T] \times \Omega \rightarrow \mathbb{R}^d$, $\theta \in \Theta$, be independent standard Brownian motions with continuous sample paths, assume for every $d \in \mathbb{N}$ that $(u^\theta)_{\theta \in \Theta}$ and $(W^{\theta,d})_{\theta \in \Theta}$ are independent, and let $U_{n,M,d,\delta}^\theta : [0, T] \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}$, $n, M \in \mathbb{Z}$, $d \in \mathbb{N}$, $\delta \in (0, 1]$, $\theta \in \Theta$, satisfy for all $d, n, M \in \mathbb{N}$, $\delta \in (0, 1]$, $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that $U_{-1,M,d,\delta}^\theta(t, x) = U_{0,M,d,\delta}^\theta(t, x) = 0$ and

$$\begin{aligned}
 U_{n,M,d,\delta}^\theta(t, x) &= \frac{1}{M^n} \sum_{i=1}^{M^n} (\mathcal{R}(\mathfrak{g}_{d,\delta})) \left(x + W_T^{(\theta,0,-i),d} - W_t^{(\theta,0,-i),d} \right) \\
 &+ \sum_{l=0}^{n-1} \frac{(T-t)}{M^{n-l}} \left[\sum_{i=1}^{M^{n-l}} \left((\mathcal{R}(\mathfrak{f}_\delta) \circ U_{l,M,d,\delta}^{(\theta,l,i)} - \mathbb{1}_{\mathbb{N}}(l) (\mathcal{R}(\mathfrak{f}_\delta) \circ U_{l-1,M,d,\delta}^{(\theta,-l,i)} \right) \right. \\
 &\quad \left. \left(\mathcal{U}_t^{(\theta,l,i)}, x + W_{U_t^{(\theta,l,i),d}}^{(\theta,l,i),d} - W_t^{(\theta,l,i),d} \right) \right],
 \end{aligned}
 \tag{139}$$

let $c_d \in [1, \infty)$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$ that

$$c_d = (e^{LT}(T+1))^{q+1} ((Bd^p)^q + 1) \left[1 + \left(\int_{\mathbb{R}^d} \|x\|^{2pq} \nu_d(dx) \right)^{\frac{1}{(2pq)}} + (\mathbb{E}[\|\mathbf{W}_T^d\|^{pq}])^{\frac{1}{(pq)}} \right]^{pq},
 \tag{140}$$

let $k_{d,\varepsilon} \in \mathbb{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$k_{d,\varepsilon} = \max \left\{ \|\mathcal{D}(\mathfrak{f}_\varepsilon)\|, \|\mathcal{D}(\mathfrak{g}_{d,\varepsilon})\|, 2 \right\},
 \tag{141}$$

let $\tilde{C} = (\tilde{C}_\gamma)_{\gamma \in (0,1]} : (0, \infty) \rightarrow (0, \infty]$ satisfy for all $\gamma \in (0, \infty)$ that

$$\tilde{C}_\gamma = \sup_{n \in \mathbb{N} \cap [2, \infty)} \left[n(3n)^{2n} \left(\frac{\sqrt{e}(1+2LT)}{\sqrt{n-1}} \right)^{(n-1)(4+\gamma)} \right],
 \tag{142}$$

let $N_{d,\varepsilon} \in \mathbb{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$N_{d,\varepsilon} = \min \left\{ n \in \mathbb{N} \cap [2, \infty) : \left[c_d \left(\frac{\sqrt{e}(1+2LT)}{\sqrt{n}} \right)^n \right] \leq \frac{\varepsilon}{2} \right\},
 \tag{143}$$

and let $\delta_{d,\varepsilon} \in (0, 1]$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $\delta_{d,\varepsilon} = \frac{\varepsilon}{4Bd^p c_d}$.

Note that the fact that for all $d \in \mathbb{N}$ the random variable $\left\| \frac{\mathbf{W}_T^d}{\sqrt{T}} \right\|^2$ is chi-squared distributed with d degrees of freedom and Jensen's inequality imply that for all $d \in \mathbb{N}$ it holds that

$$(\mathbb{E}[\|\mathbf{W}_T^d\|^{pq}])^2 \leq \mathbb{E}[\|\mathbf{W}_T^d\|^{2pq}] = (2T)^{pq} \left[\frac{\Gamma(\frac{d}{2} + pq)}{\Gamma(\frac{d}{2})} \right] = (2T)^{pq} \left[\prod_{k=0}^{pq-1} \left(\frac{d}{2} + k \right) \right].
 \tag{144}$$

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

$$(\mathbb{E}[\|\mathbf{W}_T^d\|^{pq}])^{\frac{1}{pq}} = (\mathbb{E}[\|\mathbf{W}_T^d\|^{2pq}])^{\frac{2}{2pq}} \leq \sqrt{2T} \left(\prod_{k=0}^{pq-1} \left(\frac{d}{2} + k \right) \right)^{\frac{1}{2pq}} \leq \sqrt{2T} \left(\frac{d}{2} + pq - 1 \right). \tag{145}$$

This together with the fact that $\forall d \in \mathbb{N} : \left(\int_{\mathbb{R}^d} \|x\|^{2pq} \nu_d(dx) \right)^{\frac{1}{2pq}} \leq Bd^p$ implies that there exists $\bar{C} \in (0, \infty)$ such that for all $d \in \mathbb{N}$ it holds that

$$c_d \leq \bar{C} d^{pq} \left(\frac{1 + d^p + \sqrt{d}}{3} \right)^{pq} \leq \bar{C} d^{(p+1)pq}. \tag{146}$$

Next note that for all $\gamma \in (0, \infty)$ it holds that

$$\begin{aligned}
 \tilde{C}_\gamma &= \sup_{n \in \mathbb{N} \cap [2, \infty)} \left[n(3n)^{2n} \left(\frac{\sqrt{e}(1 + 2LT)}{\sqrt{n-1}} \right)^{(n-1)(4+\gamma)} \right] \\
 &= \sup_{n \in \mathbb{N} \cap [2, \infty)} \left[(\sqrt{e}(1 + 2LT))^{(n-1)(4+\gamma)} n^3 3^{2n} (n-1)^{-(n-1)\frac{\gamma}{2}} \left(\frac{n}{n-1} \right)^{2(n-1)} \right] \\
 &\leq \left[\sup_{n \in \mathbb{N} \cap [2, \infty)} \left[(\sqrt{e}(1 + 2LT))^{(n-1)(4+\gamma)} n^3 3^{2n} (n-1)^{-(n-1)\frac{\gamma}{2}} \right] \right] \left[\sup_{n \in \mathbb{N} \cap [2, \infty)} \left(\frac{n}{n-1} \right)^{2(n-1)} \right] \\
 &< \infty.
 \end{aligned} \tag{147}$$

The fact that for all $d \in \mathbb{N}, v \in \mathbb{R}, x \in \mathbb{R}^d, \varepsilon \in (0, 1]$ it holds that $|f(v) - (\mathcal{R}(\tilde{f}_\varepsilon))(v)| \leq \varepsilon(1 + |v|^q)$ and $|g_d(x) - (\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \varepsilon Bd^p(1 + \|x\|)^{pq}$ implies that for all $d \in \mathbb{N}, v \in \mathbb{R}, x \in \mathbb{R}^d, \varepsilon \in (0, 1]$ it holds that

$$\begin{aligned}
 \max \left\{ |f(v) - (\mathcal{R}(\tilde{f}_\varepsilon))(v)|, |g_d(x) - (\mathcal{R}(g_{d,\varepsilon}))(x)| \right\} &\leq \max \{ \varepsilon(1 + |v|^q), \varepsilon Bd^p(1 + \|x\|)^{pq} \} \\
 &\leq \varepsilon Bd^p ((1 + \|x\|)^{pq} + |v|^q).
 \end{aligned} \tag{148}$$

This, (136), (138), the fact that for all $d \in \mathbb{N}, v, w \in \mathbb{R}, x \in \mathbb{R}^d, \varepsilon \in (0, 1]$ it holds that $|f(v) - f(w)| \leq L, |(\mathcal{R}(\tilde{f}_\varepsilon))(v) - (\mathcal{R}(\tilde{f}_\varepsilon))(w)| \leq L|v - w|, |f(0)| \leq B, |(\mathcal{R}(g_{d,\varepsilon}))(x)| \leq Bd^p(1 + \|x\|)^p$, and Corollary 2.4 (with $f_1 = f, f_2 = \mathcal{R}(\tilde{f}_\delta), g_1 = g_d, g_2 = \mathcal{R}(g_{d,\delta}), L = L, \delta = \delta Bd^p, B = Bd^p, \mathbf{W} = \mathbf{W}^d$ in the notation of Corollary 2.4) imply that for all $d, N, M \in \mathbb{N}, \delta \in (0, 1]$ it holds that

$$\begin{aligned}
 &\left(\int_{\mathbb{R}^d} \mathbb{E} \left[\left| U_{N,M,d,\delta}^0(0, x) - u_d(0, x) \right|^2 \right] \nu_d(dx) \right)^{\frac{1}{2}} \\
 &\leq (e^{LT}(T + 1))^{q+1} ((Bd^p)^q + 1) \left(\delta Bd^p + \frac{e^{M/2}(1 + 2LT)^N}{M^{N/2}} \right) \\
 &\cdot \left[\int_{\mathbb{R}^d} \left(1 + \|x\| + (\mathbb{E}[\|\mathbf{W}_T^d\|^{pq}])^{\frac{1}{pq}} \right)^{2pq} \nu_d(dx) \right]^{\frac{1}{2}}.
 \end{aligned} \tag{149}$$

The triangle inequality hence proves for all $d, N, M \in \mathbb{N}$, $\delta \in (0, 1]$ that

$$\begin{aligned} & \left(\int_{\mathbb{R}^d} \mathbb{E} \left[\left| U_{N,M,d,\delta}^0(0, x) - u_d(0, x) \right|^2 \right] v_d(dx) \right)^{\frac{1}{2}} \\ & \leq (e^{LT}(T+1))^{q+1} ((Bd^p)^q + 1) \left(\delta B d^p + \frac{e^{M/2}(1+2LT)^N}{M^{N/2}} \right) \\ & \quad \cdot \left[1 + \left(\int_{\mathbb{R}^d} \|x\|^{2pq} v_d(dx) \right)^{\frac{1}{(2pq)}} + (\mathbb{E}[\|\mathbf{W}_T^d\|^{pq}])^{\frac{1}{(pq)}} \right]^{pq} \\ & = c_d \left(\delta B d^p + \frac{e^{M/2}(1+2LT)^N}{M^{N/2}} \right). \end{aligned} \tag{150}$$

This and Fubini’s theorem imply that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that

$$\begin{aligned} & \mathbb{E} \left[\int_{\mathbb{R}^d} \left| U_{N_{d,\varepsilon}, N_{d,\varepsilon}, d, \delta_{d,\varepsilon}}^0(0, x) - u_d(0, x) \right|^2 v_d(dx) \right] \\ & = \int_{\mathbb{R}^d} \mathbb{E} \left[\left| U_{N_{d,\varepsilon}, N_{d,\varepsilon}, d, \delta_{d,\varepsilon}}^0(0, x) - u_d(0, x) \right|^2 \right] v_d(dx) \\ & \leq \left(c_d \delta_{d,\varepsilon} B d^p + c_d \left(\frac{\sqrt{\varepsilon}(1+2LT)}{\sqrt{N_{d,\varepsilon}}} \right)^{N_{d,\varepsilon}} \right)^2 \leq \left(\frac{\varepsilon}{4} + \frac{\varepsilon}{2} \right)^2 < \varepsilon^2. \end{aligned} \tag{151}$$

This implies that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ there exists $\omega_{d,\varepsilon} \in \Omega$ such that

$$\int_{\mathbb{R}^d} \left| U_{N_{d,\varepsilon}, N_{d,\varepsilon}, d, \delta_{d,\varepsilon}}^0(0, x, \omega_{d,\varepsilon}) - u_d(0, x) \right|^2 v_d(dx) < \varepsilon^2. \tag{152}$$

Next, observe that Lemma 3.10 shows that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ there exists $\Psi_{d,\varepsilon} \in \mathbf{N}$ such that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $(\mathcal{R}(\Psi_{d,\varepsilon}))(x) = U_{N_{d,\varepsilon}, N_{d,\varepsilon}, d, \delta_{d,\varepsilon}}^0(0, x, \omega_{d,\varepsilon})$, $\|\mathcal{D}(\Psi_{d,\varepsilon})\| \leq k_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{N_{d,\varepsilon}}$, and

$$\dim(\mathcal{D}(\Psi_{d,\varepsilon})) = N_{d,\varepsilon} \left(\dim(\mathcal{D}(\check{f}_{\delta_{d,\varepsilon}})) - 1 \right) + \dim(\mathcal{D}(\mathfrak{g}_{d,\delta_{d,\varepsilon}})). \tag{153}$$

This and (152) prove (133). Moreover, (153) and (130) imply that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that

$$\begin{aligned} \mathcal{P}(\Psi_{d,\varepsilon}) & \leq \sum_{j=1}^{\dim(\mathcal{D}(\Psi_{d,\varepsilon}))} k_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{N_{d,\varepsilon}} (k_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{N_{d,\varepsilon}} + 1) \\ & \leq 2 \dim(\mathcal{D}(\Psi_{d,\varepsilon})) k_{d,\delta_{d,\varepsilon}}^2(3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \\ & = 2 \left(N_{d,\varepsilon} \left(\dim(\mathcal{D}(\check{f}_{\delta_{d,\varepsilon}})) - 1 \right) + \dim(\mathcal{D}(\mathfrak{g}_{d,\delta_{d,\varepsilon}})) \right) k_{d,\delta_{d,\varepsilon}}^2(3N_{d,\varepsilon})^{2N_{d,\varepsilon}}. \end{aligned} \tag{154}$$

In addition, it follows from (137) and (134) that for all $\varepsilon \in (0, 1]$ it holds that

$$\|\mathcal{D}(\check{f}_\varepsilon)\| \leq \left[4L \left(1 + (4L + 2|f(0)|)^{\frac{1}{q-1}} \right) + 2 \right] \varepsilon^{-\frac{\alpha}{(q-1)}} \leq B\varepsilon^{-2} \leq B\varepsilon^{-\alpha}. \tag{155}$$

Combining this with (154), the fact that $\forall \varepsilon \in (0, 1] : \dim(\mathcal{D}(\mathfrak{f}_\varepsilon)) = 3$, the fact that $\forall d \in \mathbb{N}, \varepsilon \in (0, 1] : \|\mathcal{D}(\mathfrak{g}_{d,\varepsilon})\| \leq d^p \varepsilon^{-\alpha} B$, and the fact that $\forall d \in \mathbb{N}, \varepsilon \in (0, 1] : \dim(\mathcal{D}(\mathfrak{g}_{d,\varepsilon})) \leq d^p \varepsilon^{-\beta} B$ implies that for all $d \in \mathbb{N}, \varepsilon \in (0, 1]$ it holds that $k_{d,\delta_{d,\varepsilon}} \leq d^p \delta_{d,\varepsilon}^{-\alpha} B$ and that

$$\begin{aligned} \mathcal{P}(\Psi_{d,\varepsilon}) &\leq 2 \left(N_{d,\varepsilon} \left(\dim(\mathcal{D}(\mathfrak{f}_{\delta_{d,\varepsilon}})) - 1 \right) + \dim(\mathcal{D}(\mathfrak{g}_{d,\delta_{d,\varepsilon}})) \right) \left(d^p \delta_{d,\varepsilon}^{-\alpha} B \right)^2 (3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \\ &\leq 2 \left(2N_{d,\varepsilon} + d^p (\delta_{d,\varepsilon})^{-\beta} B \right) \left(d^p \delta_{d,\varepsilon}^{-\alpha} B \right)^2 (3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \\ &\leq 4d^p \delta_{d,\varepsilon}^{-\beta} B d^{2p} \delta_{d,\varepsilon}^{-2\alpha} B^2 N_{d,\varepsilon} (3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \\ &= 4B^3 (4c_d B d^p)^{2\alpha+\beta} d^{3p} \varepsilon^{-(2\alpha+\beta)} N_{d,\varepsilon} (3N_{d,\varepsilon})^{2N_{d,\varepsilon}}. \end{aligned} \tag{156}$$

Furthermore, (143) ensures that for all $d \in \mathbb{N}, \varepsilon \in (0, 1]$ it holds that

$$\varepsilon \leq 2c_d \left(\frac{\sqrt{e}(1+2LT)}{\sqrt{N_{d,\varepsilon}} - 1} \right)^{N_{d,\varepsilon}-1}. \tag{157}$$

This together with (156) implies that for all $d \in \mathbb{N}, \varepsilon \in (0, 1], \gamma \in (0, 1]$ it holds that

$$\begin{aligned} \mathcal{P}(\Psi_{d,\varepsilon}) &\leq 4B^{2\alpha+\beta+3} (4c_d)^{2\alpha+\beta} d^{(2\alpha+\beta+3)p} \varepsilon^{-(2\alpha+\beta)} N_{d,\varepsilon} (3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \varepsilon^{4+\gamma} \varepsilon^{-(4+\gamma)} \\ &\leq 4B^{2\alpha+\beta+3} (4c_d)^{4+2\alpha+\beta+\gamma} d^{(2\alpha+\beta+3)p} N_{d,\varepsilon} (3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \left(\frac{\sqrt{e}(1+2LT)}{\sqrt{N_{d,\varepsilon}} - 1} \right)^{(N_{d,\varepsilon}-1)(4+\gamma)} \varepsilon^{-(4+2\alpha+\beta+\gamma)} \\ &\leq 4B^{2\alpha+\beta+3} (4c_d)^{5+2\alpha+\beta} d^{(2\alpha+\beta+3)p} \sup_{n \in \mathbb{N} \cap [2, \infty)} \left[n(3n)^{2n} \left(\frac{\sqrt{e}(1+2LT)}{\sqrt{n-1}} \right)^{(n-1)(4+\gamma)} \right] \varepsilon^{-(4+2\alpha+\beta+\gamma)} \\ &= 4B^{2\alpha+\beta+3} (4c_d)^{5+2\alpha+\beta} d^{(2\alpha+\beta+3)p} \tilde{C}_{\gamma,\varepsilon} \varepsilon^{-(4+2\alpha+\beta+\gamma)}. \end{aligned} \tag{158}$$

Combining this with (146) and (147) establishes that there exist $\eta \in (0, \infty), C = (C_\gamma)_{\gamma \in (0,1]} : (0, 1] \rightarrow (0, \infty)$ such that for all $d \in \mathbb{N}, \varepsilon, \gamma \in (0, 1]$ it holds that $\mathcal{P}(\Psi_{d,\varepsilon}) \leq C_\gamma d^\eta \varepsilon^{-(4+2\alpha+\beta+\gamma)}$. The proof of Theorem 4.1 is thus completed. \square

4.2 Deep neural network approximations with general polynomial convergence rates

Corollary 4.2 *Let $\|\cdot\| : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow [0, \infty)$ and $\mathbf{A}_d : \mathbb{R}^d \rightarrow \mathbb{R}^d, d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}, x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $\|x\| = [\sum_{i=1}^d (x_i)^2]^{1/2}$ and $\mathbf{A}_d(x) = (\max\{x_1, 0\}, \dots, \max\{x_d, 0\})$, let*

$$\mathbf{N} = \bigcup_{H \in \mathbb{N}} \bigcup_{(k_0, k_1, \dots, k_{H+1}) \in \mathbb{N}^{H+2}} \left[\prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}) \right], \tag{159}$$

let $\mathcal{R} : \mathbf{N} \rightarrow (\cup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ and $\mathcal{P} : \mathbf{N} \rightarrow \mathbb{N}$ satisfy for all $H \in \mathbb{N}$, $k_0, k_1, \dots, k_H, k_{H+1} \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n})$, $x_0 \in \mathbb{R}^{k_0}, \dots, x_H \in \mathbb{R}^{k_H}$ with $\forall n \in \mathbb{N} \cap [1, H] : x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ that

$$\mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}}), \quad (\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1}, \quad \text{and} \quad \mathcal{P}(\Phi) = \sum_{n=1}^{H+1} k_n(k_{n-1} + 1),$$

let $T, \kappa \in (0, \infty)$, $f \in C(\mathbb{R}, \mathbb{R})$, $(g_{d,\varepsilon})_{d \in \mathbb{N}, \varepsilon \in (0,1]} \subseteq \mathbf{N}$, $(c_d)_{d \in \mathbb{N}} \subseteq (0, \infty)$, for every $d \in \mathbb{N}$ let $g_d \in C(\mathbb{R}^d, \mathbb{R})$, for every $d \in \mathbb{N}$ let $u_d \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, and assume for all $d \in \mathbb{N}$, $v, w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $t \in (0, T)$ that

$$\mathcal{P}(g_{d,\varepsilon}) \leq \kappa d^\kappa \varepsilon^{-\kappa}, \quad |g_d(x) - (\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \varepsilon \kappa d^\kappa (1 + \|x\|^\kappa), \quad \mathcal{R}(g_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R}), \tag{160}$$

$$|(\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \kappa d^\kappa (1 + \|x\|^\kappa), \quad |f(v) - f(w)| \leq \kappa |v - w|, \quad |u_d(t, x)| \leq c_d (1 + \|x\|^{c_d}), \tag{161}$$

$$\left(\frac{\partial}{\partial t} u_d\right)(t, x) + \frac{1}{2}(\Delta_x u_d)(t, x) + f(u_d(t, x)) = 0, \quad \text{and} \quad u_d(T, x) = g_d(x). \tag{162}$$

Then there exist $(\Psi_{d,\varepsilon})_{d \in \mathbb{N}, \varepsilon \in (0,1]} \subseteq \mathbf{N}$, $\eta \in (0, \infty)$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\Psi_{d,\varepsilon}) \leq \eta d^\eta \varepsilon^{-\eta}$, and

$$\left[\int_{[0,1]^d} |u_d(0, x) - (\mathcal{R}(\Psi_{d,\varepsilon}))(x)|^2 dx \right]^{\frac{1}{2}} \leq \varepsilon. \tag{163}$$

Proof of Corollary 4.2 Throughout this proof assume without loss of generality that $\kappa \geq 2$, let $\|\cdot\| : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow [0, \infty)$ and $\dim : (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \rightarrow \mathbb{N}$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that $\|x\| = \max_{i \in [1,d] \cap \mathbb{N}} |x_i|$ and $\dim(x) = d$, let $\mathcal{D} : \mathbf{N} \rightarrow \mathbf{D}$ satisfy for all $H \in \mathbb{N}$, $k_0, k_1, \dots, k_H, k_{H+1} \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n})$ that

$$\mathcal{D}(\Phi) = (k_0, k_1, \dots, k_H, k_{H+1}), \tag{164}$$

and let $B = \max \left\{ 1, \kappa \left[\sup_{r \in [0, \infty)} \frac{1+r^\kappa}{(1+r)^\kappa} \right] \right\}$. The fact that $\forall d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d : |u_d(t, x)| \leq c_d (1 + \|x\|^{c_d})$, the fact that $\forall d \in \mathbb{N}$, $x \in \mathbb{R}^d : u_d(T, x) = g_d(x)$, the fact that $\forall v, w \in \mathbb{R} : |f(v) - f(w)| \leq \kappa |v - w|$, (162), and the Feynman–Kac-formula ensure that the functions u_d , $d \in \mathbb{N}$, satisfy (132). Next note that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$ it holds that

$$|(\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \kappa d^\kappa (1 + \|x\|^\kappa) \leq \kappa \left[\sup_{r \in [0, \infty)} \frac{1+r^\kappa}{(1+r)^\kappa} \right] d^\kappa (1 + \|x\|)^\kappa \leq B d^\kappa (1 + \|x\|)^\kappa, \tag{165}$$

$$|g_d(x) - (\mathcal{R}(g_{d,\varepsilon}))(x)| \leq \varepsilon \kappa d^\kappa (1 + \|x\|^\kappa) \leq \varepsilon \kappa \left[\sup_{r \in [0, \infty)} \frac{1 + r^\kappa}{(1 + r)^\kappa} \right] d^\kappa (1 + \|x\|)^\kappa \leq \varepsilon B d^\kappa (1 + \|x\|)^{2\kappa}, \tag{166}$$

$$|||\mathcal{D}(g_{d,\varepsilon})||| \leq \mathcal{P}(g_{d,\varepsilon}) \leq \kappa d^\kappa \varepsilon^{-\kappa} \leq B d^\kappa \varepsilon^{-\kappa}, \tag{167}$$

and

$$\dim(\mathcal{D}(g_{d,\varepsilon})) \leq \mathcal{P}(g_{d,\varepsilon}) \leq \kappa d^\kappa \varepsilon^{-\kappa} \leq B d^\kappa \varepsilon^{-\kappa}. \tag{168}$$

Moreover, observe that the fact that $\forall d \in \mathbb{N}, y \in [0, 1]^d : \|y\| \leq \sqrt{d}$ ensures that for all $d \in \mathbb{N}, \alpha \in (0, \infty)$ it holds that

$$\left(\int_{[0,1]^d} \|y\|^\alpha dy \right)^{\frac{1}{\alpha}} \leq \sqrt{d} \leq B d. \tag{169}$$

Combining this with (165)–(168) and Theorem 4.1 (with $\alpha = \kappa, \beta = \kappa, B = B, L = \kappa, p = \kappa, q = 2, \mathfrak{p} = 1,$ and $\gamma = \frac{1}{2}$ in the notation of Theorem 4.1) ensures that there exist $(\Psi_{d,\varepsilon})_{d \in \mathbb{N}, \varepsilon \in (0,1]} \subseteq \mathbf{N}, \eta \in (0, \infty)$ such that for all $d \in \mathbb{N}, \varepsilon \in (0, 1]$ it holds that $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R}), \mathcal{P}(\Psi_{d,\varepsilon}) \leq \eta d^\eta \varepsilon^{-\eta},$ and

$$\left[\int_{[0,1]^d} |u_d(0, x) - (\mathcal{R}(\Psi_{d,\varepsilon}))(x)|^2 dx \right]^{\frac{1}{2}} \leq \varepsilon. \tag{170}$$

The proof of Corollary 4.2 is thus completed. □

Acknowledgements Open Access funding provided by Projekt DEAL. This project has been partially supported through the German Research Foundation via research Grants HU1889/6-1 (MH) and UR313/1-1 (TK) and via RTG 2131 *High-dimensional Phenomena in Probability-Fluctuations and Discontinuity* (TAN). The second author acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC 2044-390685587, Mathematics Muenster: Dynamics-Geometry-Structure. Diyora Salimova and an anonymous referee are gratefully acknowledged for giving helpful comments.

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

References

1. Beck, C., Becker, S., Grohs, P., Jaafari, N., Jentzen, A.: Solving stochastic differential equations and Kolmogorov equations by means of deep learning. [arXiv:1806.00421](https://arxiv.org/abs/1806.00421), (2018)

2. Beck, C., E, W., 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**, 1563–1619 (2019)
3. Beck, C., Hornung, F., Hutzenthaler, M., Jentzen, A., Kruse, T.: Overcoming the curse of dimensionality in the numerical approximation of Allen–Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. [arXiv:1907.06729](https://arxiv.org/abs/1907.06729) (2019)
4. Becker, S., Cheridito, P., Jentzen, A.: Deep optimal stopping. *J. Mach. Learn. Res.* **20**(74), 1–25 (2019)
5. Berner, J., Grohs, P., 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](https://arxiv.org/abs/1809.03062), (2018)
6. E, W., Han, J., Jentzen, A.: Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat.* **5**(4), 349–380 (2017)
7. E, W., Hutzenthaler, M., Jentzen, A., Kruse, T.: Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. [arXiv:1607.03295](https://arxiv.org/abs/1607.03295), (2016)
8. E, W., Hutzenthaler, M., Jentzen, A., Kruse, T.: On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *J. Sci. Comput.* **79**(3), 1534–1571 (2019)
9. E, W., Yu, B.: The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat.* **6**(1), 1–12 (2018)
10. Elbrächter, D., Grohs, P., Jentzen, A., Schwab, C.: DNN expression rate analysis of high-dimensional PDEs: application to option pricing. [arXiv:1809.07669](https://arxiv.org/abs/1809.07669), (2018)
11. Fujii, M., Takahashi, A., Takahashi, M.: Asymptotic expansion as prior knowledge in deep learning method for high dimensional BSDEs. [arXiv:1710.07030](https://arxiv.org/abs/1710.07030), (2017)
12. Giles, M., Jentzen, A., Welti, T.: Generalised multilevel Picard approximations. [arXiv:1911.03188](https://arxiv.org/abs/1911.03188), (2019)
13. Grohs, P., Hornung, F., Jentzen, A., von Wurstemberger, P.: A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. *Mem. Am. Math. Soc.* [arxiv:1809.02362](https://arxiv.org/abs/1809.02362), (2020)
14. Han, J., Jentzen, A., E, W.: Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci.* **115**(34), 8505–8510 (2018)
15. Henry-Labordere, P.L.: Deep primal-dual algorithm for BSDEs: applications of machine learning to CVA and IM. Available at SSRN: <https://ssrn.com/abstract=3071506>, (2017)
16. Hutzenthaler, M., Jentzen, A., Kruse, T.: Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. [arXiv:1912.02571](https://arxiv.org/abs/1912.02571), (2019)
17. Hutzenthaler, M., Jentzen, A., Kruse, T., Nguyen, T.A., von Wurstemberger, P.: Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. [arXiv:1807.01212](https://arxiv.org/abs/1807.01212), (2018)
18. Hutzenthaler, M., Jentzen, A., von Wurstemberger, P.: Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *Electron. J. Probab.* [arXiv:1903.05985](https://arxiv.org/abs/1903.05985), (2019)
19. Hutzenthaler, M., Kruse, T.: Multi-level Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *SIAM J. Numer. Anal.* [arxiv:1711.01080](https://arxiv.org/abs/1711.01080) (2020)
20. Jentzen, A., Salimova, D., Welti, T.: A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. [arXiv:1809.07321](https://arxiv.org/abs/1809.07321), (2018)
21. Khoo, Y., Lu, J., Ying, L.: Solving parametric PDE problems with artificial neural networks. [arXiv:1707.03351](https://arxiv.org/abs/1707.03351), (2017)
22. Mishra, S.: A machine learning framework for data driven acceleration of computations of differential equations. [arXiv:1807.09519](https://arxiv.org/abs/1807.09519), (2018)
23. Nabian, M. A., Meidani, H.: A deep neural network surrogate for high-dimensional random partial differential equations. [arXiv:1806.02957](https://arxiv.org/abs/1806.02957), (2018)
24. Raissi, M.: Forward–backward stochastic neural networks: deep learning of high-dimensional partial differential equations. [arXiv:1804.07010](https://arxiv.org/abs/1804.07010), (2018)
25. Sirignano, J., Spiliopoulos, K.: Dgm: A deep learning algorithm for solving partial differential equations. *J. Comput. Phys.* **375**, 1339–1364 (2018)