Deep neural networks with ReLU, leaky ReLU, and softplus activation provably overcome the curse of dimensionality for Kolmogorov partial differential equations with Lipschitz nonlinearities in the L^p-sense

Julia Ackermann[1], Arnulf Jentzen[2,3], Thomas Kruse[4], Benno Kuckuck[5], and Joshua Lee Padgett[6,7]

[1] Department of Mathematics & Informatics, University of Wuppertal, Germany, e-mail: jackermann@uni-wuppertal.de

[2] School of Data Science and Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), China, e-mail: ajentzen@cuhk.edu.cn

[3] Applied Mathematics: Institute for Analysis and Numerics, University of Münster, Germany, e-mail: ajentzen@uni-muenster.de

[4] Department of Mathematics & Informatics, University of Wuppertal, Germany, e-mail: tkruse@uni-wuppertal.de

[5] Applied Mathematics: Institute for Analysis and Numerics, University of Münster, Germany, e-mail: bkuckuck@uni-muenster.de

[6] Department of Mathematical Sciences, University of Arkansas, Arkansas, USA, e-mail: padgett@uark.edu

[7] Center for Astrophysics, Space Physics, and Engineering Research, Baylor University, Texas, USA, e-mail: padgett@uark.edu

September 26, 2023

Abstract

In recent years, several deep learning-based methods for the approximation of high-dimensional partial differential equations (PDEs) have been proposed. The considerable interest that these methods have generated in the scientific literature is in large part due to numerical simulations which appear to demonstrate that such deep learning-based approximation methods seem to have the capacity to overcome the curse of dimensionality (COD) in the numerical approximation of PDEs in the sense that the number of computational operations they require to achieve a certain approximation accuracy $\varepsilon \in (0, \infty)$ grows at most polynomially in the PDE dimension $d \in \mathbb{N} = \{1, 2, 3, \dots\}$ and the reciprocal of ε. While there is thus far no mathematical result which proves that one of these methods is indeed capable of overcoming the COD in the numerical approximation of PDEs, there are now a number of rigorous mathematical results in the scientific literature which show that deep neural networks (DNNs) have the expressive power to approximate solutions of high-dimensional PDEs without the COD in the sense that the number of real parameters used to describe the approximating DNNs grows at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal $1/\varepsilon$ of the prescribed approximation accuracy $\varepsilon \in (0, \infty)$. More specifically, [Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A., *SN Part. Diff. Equ. Appl. 1*, 2 (2020)] proves that for every $T \in (0, \infty)$, $a \in \mathbb{R}$, $b \in [a, \infty)$ it holds that solutions $u_d \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, of semilinear heat equations with Lipschitz continuous nonlinearities can be approximated by DNNs with the rectified linear unit (ReLU) activation at the terminal time in the L^2-sense on $[a, b]^d$ without the COD provided that the initial value functions $\mathbb{R}^d \ni x \mapsto u_d(0, x) \in \mathbb{R}$, $d \in \mathbb{N}$, can be approximated by ReLU DNNs without the COD. It is the key contribution of this article to generalize this result by establishing this statement in the L^p-sense with $p \in (0, \infty)$ and by allowing the activation function to be more general covering the ReLU, the leaky ReLU, and the softplus activation functions as special cases.

Contents

1 Introduction

Finding approximate solutions to high-dimensional partial differential equations (PDEs) is one of the most challenging issues in computational mathematics. In recent years, several deep learning-based methods for this approximation problem have been proposed and have received significant attention in the scientific literature. Some of such deep learning-based approximation methods for PDEs are based on classical or strong formulations of PDEs (cf., for example, [13,42,60,63]), some are based on variational or weak formulations of PDEs (cf., for example, [2,23,25,64,65]), and some are based on suitable stochastic formulations of the Feynman–Kac type involving the associated forward stochastic differential equations (SDEs) or backward stochastic differential equations (BSDEs), respectively (cf., for example, [3,4,5, 16,17,19,26,27,37,38,39,43,53,58,59]).

In particular, we refer, for instance, to [19, 37, 38] for certain *deep BSDE approximations* for classes of semilinear parabolic PDEs. We refer, for example, to [63] for certain *deep Galerkin approximations* for general classes of PDEs. We refer, for instance, to [5] for certain *deep 2BSDE approximations* for a class of fully nonlinear parabolic second-order PDEs. We refer, for example, to [23] for certain *deep Ritz approximations* for a class of elliptic PDEs. We refer, for instance, to [26] for certain *deep BSDE approximations involving asymptotic expansions* for a class of nonlinear parabolic PDEs and extensions of such approximation techniques to reflected BSDEs. We refer, for example, to [39] for certain *deep primal-dual approximations* for a class of nonlinear parabolic PDEs and applications of such approximation methods to the pricing of counterparty risks and to the computation of initial margins. We refer, for instance, to [13] for certain *deep artificial neural network (ANN) approximations using collocation techniques* for advection and diffusion type PDEs in complex geometries. We refer, for example, to [42, 60] for certain *physics-informed neural network (PINN) approximations* for general classes of PDEs. We refer, for instance, to [4] for certain *deep ANN approximations based on discretizations of SDEs* for a class of linear Kolmogorov PDEs on an entire region. We refer, for example, to [16] for certain *Monte Carlo based deep ANN approximations* for a class of semilinear Kolmogorov PDEs. We refer, for instance, to [27, 43, 59] for certain *deep backward dynamic programming approximations* for classes of nonlinear parabolic PDEs. We refer, for example, to [53] for certain *deep BSDE based approximations* for a class of path-dependent PDEs arising in affine rough volatility models. We refer, for instance, to [3] for certain *deep splitting approximations* for a class of nonlinear parabolic PDEs. We refer, for example, to [58] for certain *iterative diffusion optimization approximations* for a class of Hamilton–Jacobi–Bellman PDEs. We refer, for instance, to [17] for certain *deep Runge–Kutta approximations* for a class of semilinear parabolic PDEs. For more extensive overviews on such and related deep learning-based methods for high-dimensional PDEs, we refer, for example, to the survey articles [10, 15, 20, 28].

The considerable interest in deep learning-based approximation methods for high-dimensional PDEs is in large part due to numerical simulations which appear to demonstrate that some of these deep learning-based approximation methods might have the capacity to overcome the curse of dimensionality (COD) (cf., e.g., Bellman [12] and Novak & Woźniakowski [57, Chapter 1]) in the numerical approximation of PDEs in the sense that the number of computational operations they require to achieve a certain approximation accuracy $\varepsilon \in (0, \infty)$ grows at most polynomially in the PDE dimension $d \in \mathbb{N} = \{1, 2, 3, \dots\}$

and the reciprocal of ε. In the last few years, a number of rigorous mathematical results have appeared in the scientific literature which show that deep ANNs have the expressive power to approximate solutions of high-dimensional PDEs without the COD in the sense that the number of real parameters used to describe the approximating deep ANNs grows at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal $1/\varepsilon$ of the prescribed approximation accuracy $\varepsilon \in (0, \infty)$; cf., e.g., [1, 14, 18, 24, 30, 31, 32, 33, 34, 36, 41, 46, 54, 55, 61].

While the articles [1, 14, 24, 30, 31, 33, 34, 36, 41, 54, 55, 61] prove such deep ANN approximation results for linear PDEs, the articles [18, 32, 46] establish deep ANN approximation results for certain nonlinear PDEs. In the article Hutzenthaler et al. [46] it is shown that deep ANNs with the rectified linear unit (ReLU) activation function $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$ can approximate solutions of semilinear heat PDEs at the terminal time in the L^2-sense without the COD provided that the initial value functions can be approximated by deep ReLU ANNs without the COD. The article Cioica-Licht et al. [18] extends the findings in Hutzenthaler et al. [46] in several ways. Specifically, in Cioica-Licht et al. [18] it is shown that for every $T \in (0, \infty)$ solutions $u_d\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, of certain semilinear Kolmogorov PDEs with Lipschitz continuous nonlinearities can be approximated by deep ANNs with ReLU activation at the terminal time in the L^2-sense without the COD provided that the initial value functions $\mathbb{R}^d \ni x \mapsto u_d(0, x) \in \mathbb{R}$, $d \in \mathbb{N}$, and the coefficients of the PDEs can be approximated by ANNs with ReLU activation without the COD. For $p > 2$ or for the leaky ReLU or the softplus activation, up to our best knowledge, there is no result in the literature that shows that deep ANNs can overcome the COD in the L^p-approximation of nonlinear PDEs.

It is the key contribution of the present article to show that for every $p \in (0, \infty)$ we have that solutions of semilinear heat PDEs with Lipschitz continuous nonlinearities can be approximated in the L^p-sense by deep ANNs with ReLU, leaky ReLU, or softplus activation without the COD. More precisely, we prove that for any of these types of activation functions and for every $T \in (0, \infty)$, $a \in \mathbb{R}$, $b \in (a, \infty)$ solutions $u_d\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, of semilinear heat equations with Lipschitz continuous nonlinearities can be approximated by deep ANNs in the L^p-sense with $p \in (0, \infty)$ on $[a, b]^d$ at the terminal time without the COD provided that the initial value functions $\mathbb{R}^d \ni x \mapsto u_d(0, x) \in \mathbb{R}$, $d \in \mathbb{N}$, can be approximated by ANNs without the COD (see Corollary 4.16 below for details). This extends the result in Hutzenthaler et al. [46] from L^2-approximation to L^p-approximation with $p \in (0, \infty)$ and from ReLU activation to ReLU, leaky ReLU, or softplus activation functions.

In order to illustrate the contribution of this article in more detail, we now present in the following result, Theorem 1.1 below, a special case of Theorem 4.1 in Section 4.1, which is the main result of this paper. Below Theorem 1.1 we add several explanatory sentences in which we aim to describe the used mathematical objects and the statement of Theorem 1.1 in words.

Theorem 1.1. *Let $T, \kappa, p \in (0, \infty)$, let $f \colon \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous, for every $d \in \mathbb{N}$ let $u_d \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T]$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = (\Delta_x u_d)(t,x) + f(u_d(t,x)), \tag{1.1}$$

let $\nu \in \{0,1\}$, $\alpha \in [0,\infty)\backslash\{1\}$, $\mathfrak{a}_0, \mathfrak{a}_1 \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{a}_0(x) = \max\{x, \alpha x\}$ and $\mathfrak{a}_1(x) = \ln(1 + \exp(x))$, for every $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ let $\mathbf{A}(x) \in \mathbb{R}^d$ satisfy $\mathbf{A}(x) = (\mathfrak{a}_\nu(x_1), \dots, \mathfrak{a}_\nu(x_d))$, let

$$\mathbf{N} = \cup_{L \in \mathbb{N}} \cup_{l_0, l_1, \dots, l_L \in \mathbb{N}} (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})), \tag{1.2}$$

for every $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), \dots, (W_L, B_L)) \in (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$ let $\mathcal{R}(\Phi) \colon \mathbb{R}^{l_0} \to \mathbb{R}^{l_L}$ and $\mathcal{P}(\Phi) \in \mathbb{N}$ satisfy for all $v_0 \in \mathbb{R}^{l_0}$, $v_1 \in \mathbb{R}^{l_1}$, $\dots$, $v_L \in \mathbb{R}^{l_L}$ with $\forall\, k \in \{1, 2, \dots, L\} \colon v_k = \mathbf{A}(W_k v_{k-1} + B_k)$ that

$$(\mathcal{R}(\Phi))(v_0) = W_L v_{L-1} + B_L \qquad \textit{and} \qquad \mathcal{P}(\Phi) = \textstyle\sum_{k=1}^L l_k(l_{k-1} + 1), \tag{1.3}$$

and assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that there exists $\mathbf{G} \in \mathbf{N}$ such that for all $t \in [0,T]$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{R}(\mathbf{G}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\mathbf{G}) \leq \kappa d^\kappa \varepsilon^{-\kappa}$, and

$$\varepsilon |u_d(t,x)| + |u_d(0,x) - (\mathcal{R}(\mathbf{G}))(x)| \leq \varepsilon \kappa d^\kappa (1 + \textstyle\sum_{k=1}^d |x_k|^\kappa). \tag{1.4}$$

Then there exists $c \in \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists $\mathbf{U} \in \mathbf{N}$ such that $\mathcal{R}(\mathbf{U}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\mathbf{U}) \leq c d^c \varepsilon^{-c}$, and

$$\left[\textstyle\int_{[0,1]^d} |u_d(T,x) - (\mathcal{R}(\mathbf{U}))(x)|^p \, dx\right]^{1/p} \leq \varepsilon. \tag{1.5}$$

Theorem 1.1 is an immediate consequence of Corollary 4.16 in Section 4.3 below. Corollary 4.16, in turn, follows from Theorem 4.1, which is the main result of this article (see Section 4 for details). In the following we provide some explanatory comments concerning the mathematical objects appearing in Theorem 1.1.

The function $\mathfrak{a}_\nu \colon \mathbb{R} \to \mathbb{R}$ in Theorem 1.1 serves as the activation function which we employ in the approximating ANNs in Theorem 1.1 and the function

$$(\cup_{d \in \mathbb{N}} \mathbb{R}^d) \ni x \mapsto \mathbf{A}(x) \in (\cup_{d \in \mathbb{N}} \mathbb{R}^d) \tag{1.6}$$

represents a suitable multidimensional version of the activation function $\mathfrak{a}_\nu \colon \mathbb{R} \to \mathbb{R}$ in Theorem 1.1. The function $\mathfrak{a}_\nu \colon \mathbb{R} \to \mathbb{R}$ may be the ReLU activation (corresponding to the case $\nu = \alpha = 0$ in Theorem 1.1), the leaky ReLU activation (corresponding to the case $\nu = 0$ and $\alpha \in (0,1)$ in Theorem 1.1), or the softplus activation (corresponding to the case $\nu = 1$ in Theorem 1.1).

The set $\mathbf{N} = \cup_{L \in \mathbb{N}} \cup_{l_0, l_1, \dots, l_L \in \mathbb{N}} (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$ in Theorem 1.1 represents the set of all ANNs which we employ to approximate the solutions of the PDEs under consideration. Observe that for every ANN $\Phi \in \mathbf{N}$ we have that

$$\mathcal{R}(\Phi) \in \cup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l) \tag{1.7}$$

is the realization function of the ANN Φ with the activation function $\mathfrak{a}_\nu\colon \mathbb{R} \to \mathbb{R}$. Moreover, we note that for every ANN $\Phi \in \mathbf{N}$ we have that $\mathcal{P}(\Phi) \in \mathbb{N}$ is the number of real numbers used to describe the ANN Φ. Very roughly speaking, $\mathcal{P}(\Phi)$ corresponds to the amount of memory that is needed on a computer to store the ANN $\Phi \in \mathbf{N}$.

The real number $T \in (0,\infty)$ in Theorem 1.1 specifies the time horizon of the PDEs (see (1.1) above) whose solutions we intend to approximate by deep ANNs in (1.5) in Theorem 1.1. The real number $\kappa \in (0,\infty)$ in Theorem 1.1 is a constant which we employ to formulate our regularity and approximation hypotheses in Theorem 1.1. The real number $p \in (0,\infty)$ in Theorem 1.1 is used to specify the way we measure the error between the exact solutions of the PDEs under consideration and their deep ANN approximations, that is, we measure the error between the exact solutions of the PDEs under consideration and their deep ANN approximations in the L^p-sense (see (1.5) for details).

In Theorem 1.1 we assume that the initial conditions of the PDEs (see (1.1)) whose solutions we intend to approximate by deep ANNs without the COD can be approximated by ANNs without the COD (see (1.4) above). The function $f\colon \mathbb{R} \to \mathbb{R}$ in Theorem 1.1 specifies the nonlinearity in the PDEs (see (1.1)) whose solutions we intend to approximate by deep ANNs in Theorem 1.1. The functions

$$u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}, \quad d \in \mathbb{N}, \tag{1.8}$$

in Theorem 1.1 describe the exact solutions of the PDEs in (1.1). Observe that the hypothesis in (1.4) in Theorem 1.1 also ensures that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $t \in [0,T]$, $x \in \mathbb{R}^d$ we have that $\varepsilon|u_d(t,x)| \le \varepsilon\kappa d^\kappa(1 + \sum_{k=1}^d |x_k|^\kappa)$. This, in turn, assures that for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ we have that

$$|u_d(t,x)| \le \kappa d^\kappa(1 + \textstyle\sum_{k=1}^d |x|^\kappa). \tag{1.9}$$

Note that (1.9) ensures that for all $d \in \mathbb{N}$ we have that the solution $u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ of (1.1) grows at most polynomially. These polynomial growth properties of the solutions assure that the solutions of (1.1) with the fixed initial value functions $\mathbb{R}^d \ni x \mapsto u_d(0,x) \in \mathbb{R}$, $d \in \mathbb{N}$, are unique.

Theorem 1.1 establishes that for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists an ANN $\mathbf{U}_{d,\varepsilon} \in \mathbf{N}$ such that the L^p-distance with respect to the Lebesgue measure on $[0,1]^d$ between the exact solution $u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ at time T of the PDE in (1.1) and the realization

$$\mathcal{R}(\mathbf{U}_{d,\varepsilon})\colon \mathbb{R}^d \to \mathbb{R} \tag{1.10}$$

of the ANN $\mathbf{U}_{d,\varepsilon}$ is bounded by ε and such that the number of parameters of the ANN $\mathbf{U}_{d,\varepsilon} \in \mathbf{N}$ grows at most polynomially in both the PDE dimension d and the reciprocal $1/\varepsilon$ of the prescribed approximation accuracy ε.

Although Theorem 1.1 is restricted to measuring the L^p-distance with respect to the Lebesgue measure on $[0,1]^d$, our more general deep ANN approximation results in Section 4 (see Theorem 4.1, Corollary 4.15, and Corollary 4.16 in Section 4) allow measuring the L^p-distance with respect to more general probability measures on $\mathbb{R}^d$. In particular, for all $a \in \mathbb{R}$, $b \in (a,\infty)$ we have that the more general deep ANN approximation results in Section 4 enable measuring the L^p-distance with respect to the uniform distribution on $[a,b]^d$. Furthermore, we note that Theorem 4.1 in Section 4 is formulated for general activation functions provided that with the considered general activation function there exists a shallow ANN representation for the identity function $\mathbb{R} \ni x \mapsto x \in \mathbb{R}$ on the real numbers and provided that with the considered general activation function the Lipschitz continuous nonlinearity can be approximated with appropriate convergence rates by ANNs.

Our proof of Theorem 1.1 is strongly based on employing suitable nonlinear Monte Carlo approximations, so-called *full history recursive multilevel Picard* (MLP) approximations, which provably approximate PDEs of the form (1.1) without the COD. MLP methods have been introduced in [21, 22, 47] and have been extended to more general situations in, e.g., [7, 8, 11, 29, 44, 45, 49, 50, 51, 52, 56]. In particular, the L^p-error with $p \in (0, \infty)$ for MLP approximations of certain semilinear PDEs has been analyzed in Hutzenthaler et al. [48] and in the present work we employ these findings to establish the desired ANN approximation results. For further references on MLP methods we refer, for example, to the survey articles [10, 20].

The remainder of this article is organized as follows: In Section 2 we review the necessary basic preparatory material on ANNs that we need in the later sections of this work. In Section 3 we study deep ANN representations for MLP approximations for PDEs of the form (1.1). The ANN representations for MLP approximations from Section 3 are then used in Section 4 to prove Theorem 1.1 and its above sketched generalizations.

2 Artificial neural network (ANN) calculus

In this section we review the necessary basic preparatory material on ANNs that we need in the later sections of this work. The conceptualities and the lemmas provided in this section are elementary or well-known in the literature and the specific material in this section mostly consists of slightly modified extracts from Grohs et al. [35, Section 2] and Grohs et al. [36, Section 3].

2.1 Structured description of ANNs

Definition 2.1 (ANNs)**.** We denote by $\mathbf{N}$ the set given by

$$\mathbf{N} = \textstyle\bigcup_{L\in\mathbb{N}} \bigcup_{l_0,l_1,\ldots,l_L\in\mathbb{N}} \big(\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \big) \tag{2.1}$$

and we denote by $\mathcal{P}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{L}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{I}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{O}\colon \mathbf{N} \to \mathbb{N}$, $\mathcal{H}\colon \mathbf{N} \to \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathcal{D}\colon \mathbf{N} \to \cup_{L\in\mathbb{N}}\mathbb{N}^L$, and $\mathbb{D}_n\colon \mathbf{N} \to \mathbb{N}_0$, $n \in \mathbb{N}_0$, the functions which satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $n \in \mathbb{N}_0$ that $\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1} + 1)$, $\mathcal{L}(\Phi) = L$, $\mathcal{I}(\Phi) = l_0$, $\mathcal{O}(\Phi) = l_L$, $\mathcal{H}(\Phi) = L - 1$, $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$, and

$$\mathbb{D}_n(\Phi) = \begin{cases} l_n & : n \leq L \\ 0 & : n > L \end{cases}. \tag{2.2}$$

Definition 2.2 (ANN)**.** We say that Φ is an ANN if and only if it holds that $\Phi \in \mathbf{N}$ (cf. Definition 2.1).

Definition 2.3 (Standard and maximum norms)**.** We denote by $\|\cdot\| \colon \cup_{d\in\mathbb{N}}\mathbb{R}^d \to \mathbb{R}$ and $|||\cdot||| \colon \cup_{d\in\mathbb{N}}\mathbb{R}^d \to \mathbb{R}$ the functions which 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 $|||x||| = \max_{i\in\{1,2,\dots,d\}}|x_i|$.

Lemma 2.4. *It holds for all $\Phi \in \mathbf{N}$ that $\mathcal{L}(\Phi) + |||\mathcal{D}(\Phi)||| \leq \mathcal{P}(\Phi)$ (cf. Definitions 2.1 and 2.3).*

Proof of Lemma 2.4. Observe that it holds for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$ that

$$\sum_{k=1}^{L} l_k l_{k-1} \geq \max\left\{\sum_{k=1}^{L} l_k, \sum_{k=1}^{L} l_{k-1}\right\} \geq \max\{l_0, l_1, \dots, l_L\}. \tag{2.3}$$

Hence, we obtain for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$ that

$$\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1}+1) = \sum_{k=1}^{L} l_k + \sum_{k=1}^{L} l_k l_{k-1} \geq L + \max\{l_0, l_1, \dots, l_L\} = \mathcal{L}(\Phi) + |||\mathcal{D}(\Phi)|||. \tag{2.4}$$

The proof of Lemma 2.4 is thus complete. □

Definition 2.5 (Multidimensional version)**.** Let $d \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$. Then we denote by $\mathfrak{M}_{a,d} \colon \mathbb{R}^d \to \mathbb{R}^d$ the function which satisfies for all $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ that

$$\mathfrak{M}_{a,d}(x) = \big(a(x_1), \dots, a(x_d)\big). \tag{2.5}$$

Definition 2.6 (Realization associated to an ANN)**.** Let $a \in C(\mathbb{R}, \mathbb{R})$. Then we denote by $\mathcal{R}_a \colon \mathbf{N} \to (\cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ the function which satisfies for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L)) \in \big(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big)$, $x_0 \in \mathbb{R}^{l_0}$, $x_1 \in \mathbb{R}^{l_1}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, k \in \{1, 2, \dots, L-1\} \colon x_k = \mathfrak{M}_{a,l_k}(W_k x_{k-1} + B_k)$ that

$$\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}) \qquad \text{and} \qquad (\mathcal{R}_a(\Phi))(x_0) = W_L x_{L-1} + B_L \tag{2.6}$$

(cf. Definitions 2.1 and 2.5).

2.2 Compositions of ANNs

Definition 2.7 (Composition of ANNs)**.** We denote by $(\cdot) \bullet (\cdot) \colon \{(\Phi_1, \Phi_2) \in \mathbf{N}\times\mathbf{N} \colon \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)\} \to \mathbf{N}$ the function which satisfies for all $L, \mathfrak{L} \in \mathbb{N}$, $l_0, l_1, \dots, l_L, \mathfrak{l}_0, \mathfrak{l}_1, \dots, \mathfrak{l}_{\mathfrak{L}} \in \mathbb{N}$, $\Phi_1 = ((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L)) \in \big(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big)$, $\Phi_2 = ((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}}, \mathscr{B}_{\mathfrak{L}})) \in \big(\times_{k=1}^{\mathfrak{L}}(\mathbb{R}^{\mathfrak{l}_k\times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k})\big)$ with $l_0 = \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2) = \mathfrak{l}_{\mathfrak{L}}$ that

$$\Phi_1 \bullet \Phi_2 = \begin{cases} \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1\mathscr{W}_{\mathfrak{L}}, W_1\mathscr{B}_{\mathfrak{L}} + B_1), \\ \qquad (W_2, B_2), (W_3, B_3), \dots, (W_L, B_L)\big) & : L > 1 < \mathfrak{L} \\ \big((W_1\mathscr{W}_1, W_1\mathscr{B}_1 + B_1), (W_2, B_2), (W_3, B_3), \dots, (W_L, B_L)\big) & : L > 1 = \mathfrak{L} \\ \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1\mathscr{W}_{\mathfrak{L}}, W_1\mathscr{B}_{\mathfrak{L}} + B_1)\big) & : L = 1 < \mathfrak{L} \\ \big((W_1\mathscr{W}_1, W_1\mathscr{B}_1 + B_1)\big) & : L = 1 = \mathfrak{L} \end{cases} \tag{2.7}$$

(cf. Definition 2.1).

2.3 Powers and extensions of ANNs

Definition 2.8 (Identity matrix)**.** Let $d \in \mathbb{N}$. Then we denote by $\mathrm{I}_d \in \mathbb{R}^{d\times d}$ the identity matrix in $\mathbb{R}^{d\times d}$.

Definition 2.9 (Powers of ANNs)**.** We denote by $(\cdot)^{\bullet n}\colon \{\Phi \in \mathbf{N}\colon \mathcal{I}(\Phi) = \mathcal{O}(\Phi)\} \to \mathbf{N}$, $n \in \mathbb{N}_0$, the functions which satisfy for all $n \in \mathbb{N}_0$, $\Phi \in \mathbf{N}$ with $\mathcal{I}(\Phi) = \mathcal{O}(\Phi)$ that

$$\Phi^{\bullet n} = \begin{cases} \big(\mathrm{I}_{\mathcal{O}(\Phi)}, (0,0,\dots,0)\big) \in \mathbb{R}^{\mathcal{O}(\Phi)\times\mathcal{O}(\Phi)} \times \mathbb{R}^{\mathcal{O}(\Phi)} & : n = 0 \\ \Phi \bullet (\Phi^{\bullet(n-1)}) & : n \in \mathbb{N} \end{cases} \tag{2.8}$$

(cf. Definitions 2.1, 2.7, and 2.8).

Definition 2.10 (Extension of ANNs)**.** Let $L \in \mathbb{N}$, $\Psi \in \mathbf{N}$ satisfy that $\mathcal{I}(\Psi) = \mathcal{O}(\Psi)$. Then we denote by $\mathcal{E}_{L,\Psi}\colon \{\Phi \in \mathbf{N}\colon (\mathcal{L}(\Phi) \leq L \text{ and } \mathcal{O}(\Phi) = \mathcal{I}(\Psi))\} \to \mathbf{N}$ the function which satisfies for all $\Phi \in \mathbf{N}$ with $\mathcal{L}(\Phi) \leq L$ and $\mathcal{O}(\Phi) = \mathcal{I}(\Psi)$ that

$$\mathcal{E}_{L,\Psi}(\Phi) = (\Psi^{\bullet(L-\mathcal{L}(\Phi))}) \bullet \Phi \tag{2.9}$$

(cf. Definitions 2.1, 2.7, and 2.9).

2.4 Parallelizations of ANNs

Definition 2.11 (Parallelization of ANNs)**.** Let $n \in \mathbb{N}$. Then we denote by

$$\mathbf{P}_n\colon \big\{(\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n\colon \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)\big\} \to \mathbf{N} \tag{2.10}$$

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

$$\begin{aligned} \mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n) = \Bigg(&\left(\begin{pmatrix} W_{1,1} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,1} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,1} \end{pmatrix}, \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ B_{3,1} \\ \vdots \\ B_{n,1} \end{pmatrix} \right), \\ &\left(\begin{pmatrix} W_{1,2} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,2} \end{pmatrix}, \begin{pmatrix} B_{1,2} \\ B_{2,2} \\ B_{3,2} \\ \vdots \\ B_{n,2} \end{pmatrix} \right), \dots, \\ &\left(\begin{pmatrix} W_{1,L} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,L} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,L} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,L} \end{pmatrix}, \begin{pmatrix} B_{1,L} \\ B_{2,L} \\ B_{3,L} \\ \vdots \\ B_{n,L} \end{pmatrix} \right) \Bigg) \end{aligned} \tag{2.11}$$

(cf. Definition 2.1).

2.5 Summations of ANNs

Definition 2.12 (Affine linear transformation ANN)**.** Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^m$. Then we denote by $\mathbf{A}_{W,B} \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m) \subseteq \mathbf{N}$ the ANN given by $\mathbf{A}_{W,B} = (W, B)$ (cf. Definitions 2.1 and 2.2).

Definition 2.13. Let $m, n \in \mathbb{N}$. Then we denote by $\mathfrak{S}_{m,n} \in (\mathbb{R}^{m\times(nm)} \times \mathbb{R}^m)$ the ANN given by $\mathfrak{S}_{m,n} = \mathbf{A}_{(\mathrm{I}_m\ \ \mathrm{I}_m\ \ \ldots\ \ \mathrm{I}_m),0}$ (cf. Definitions 2.2, 2.8, and 2.12).

Definition 2.14 (Matrix transpose)**.** Let $m, n \in \mathbb{N}$, $A \in \mathbb{R}^{m\times n}$. Then we denote by $A^* \in \mathbb{R}^{n\times m}$ the transpose of A.

Definition 2.15 (Transpose ANN)**.** Let $m, n \in \mathbb{N}$. Then we denote by $\mathfrak{T}_{m,n} \in (\mathbb{R}^{(nm)\times m} \times \mathbb{R}^{nm})$ the ANN given by $\mathfrak{T}_{m,n} = \mathbf{A}_{(\mathrm{I}_m\ \ \mathrm{I}_m\ \ \ldots\ \ \mathrm{I}_m)^*,0}$ (cf. Definitions 2.2, 2.8, 2.12, and 2.14).

Definition 2.16 (Sums of ANNs with the same length)**.** Let $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, $\Phi_u, \Phi_{u+1}, \ldots, \Phi_v \in \mathbf{N}$ satisfy for all $k \in \mathbb{Z} \cap [u, v]$ that $\mathcal{L}(\Phi_k) = \mathcal{L}(\Phi_u)$, $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_u)$, and $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_u)$ (cf. Definition 2.1). Then we denote by $\oplus_{k=u}^{v}\Phi_k$ (we denote by $\Phi_u \oplus \Phi_{u+1} \oplus \ldots \oplus \Phi_v$) the ANN given by

$$\overset{v}{\underset{k=u}{\oplus}} \Phi_k = \Big(\mathfrak{S}_{\mathcal{O}(\Phi_u),v-u+1} \bullet \big[\mathbf{P}_{v-u+1}(\Phi_u, \Phi_{u+1}, \ldots, \Phi_v)\big] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_u),v-u+1}\Big) \in \mathbf{N} \tag{2.12}$$

(cf. Definitions 2.2, 2.7, 2.11, 2.13, and 2.15).

Definition 2.17 (Sums of ANNs with different lengths)**.** Let $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, Φ_u, $\Phi_{u+1}, \ldots, \Phi_v, \Psi \in \mathbf{N}$ satisfy for all $k \in \mathbb{Z} \cap [u, v]$ that $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_u)$, $\mathcal{O}(\Phi_k) = \mathcal{I}(\Psi) = \mathcal{O}(\Psi)$, and $\mathcal{H}(\Psi) = 1$ (cf. Definition 2.1). Then we denote by $\boxplus_{k=u,\Psi}^{v}\Phi_k$ (we denote by $\Phi_u \boxplus_\Psi \Phi_{u+1} \boxplus_\Psi \ldots \boxplus_\Psi \Phi_v$) the ANN given by

$$\overset{v}{\underset{k=u,\Psi}{\boxplus}} \Phi_k = \overset{v}{\underset{k=u}{\oplus}} \mathcal{E}_{\max_{j\in\{u,u+1,\ldots,v\}} \mathcal{L}(\Phi_j),\Psi}(\Phi_k) \in \mathbf{N} \tag{2.13}$$

(cf. Definitions 2.2, 2.10, and 2.16).

2.6 Linear combinations of ANNs

2.6.1 Linear combinations of ANNs with the same length

Definition 2.18 (Scalar multiplications of ANNs)**.** We denote by $(\cdot) \circledast (\cdot) \colon \mathbb{R} \times \mathbf{N} \to \mathbf{N}$ the function which satisfies for all $\lambda \in \mathbb{R}$, $\Phi \in \mathbf{N}$ that $\lambda \circledast \Phi = \mathbf{A}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi$ (cf. Definitions 2.1, 2.7, 2.8, and 2.12).

Lemma 2.19. *Let $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, $n = v-u+1$, $h_u, h_{u+1}, \ldots, h_v \in \mathbb{R}$, $t_u, t_{u+1}, \ldots, t_v \in \mathbb{R}$, $\Phi_u, \Phi_{u+1}, \ldots, \Phi_v, \Psi \in \mathbf{N}$, $B_u, B_{u+1}, \ldots, B_v \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ satisfy $\mathcal{D}(\Phi_u) = \mathcal{D}(\Phi_{u+1}) = \ldots = \mathcal{D}(\Phi_v)$ and*

$$\Psi = \overset{v}{\underset{k=u}{\oplus}} \Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)},B_k}\big)\Big) \tag{2.14}$$

(cf. Definitions 2.1, 2.7, 2.8, 2.12, 2.16, and 2.18). Then

(i) it holds that

$$\begin{aligned}\mathcal{D}(\Psi) &= \big(\mathcal{I}(\Phi_u), \textstyle\sum_{k=u}^{v} \mathbb{D}_1(\Phi_u), \sum_{k=u}^{v} \mathbb{D}_2(\Phi_u), \ldots, \sum_{k=u}^{v} \mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big) \\ &= \big(\mathcal{I}(\Phi_u), n\mathbb{D}_1(\Phi_u), n\mathbb{D}_2(\Phi_u), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big),\end{aligned} \tag{2.15}$$

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

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

$$(\mathcal{R}_a(\Psi))(x) = \sum_{k=u}^{v} h_k (\mathcal{R}_a(\Phi_k))(t_k x + B_k) \tag{2.16}$$

(cf. Definition 2.6).

Proof of Lemma 2.19. First, observe that the hypothesis that $\mathcal{D}(\Phi_u) = \mathcal{D}(\Phi_{u+1}) = \ldots = \mathcal{D}(\Phi_v)$ ensures that for all $k \in \{u, u+1, \ldots, v\}$ it holds that

$$\mathcal{D}\big(\mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big) = \mathcal{D}\big(\mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_u)}, B_k}\big) = \big(\mathcal{I}(\Phi_u), \mathcal{I}(\Phi_u)\big) \in \mathbb{N}^2. \tag{2.17}$$

This and, e.g., Grohs et al. [35, item (i) in Proposition 2.6] assure that for all $k \in \{u, u+1, \ldots, v\}$ it holds that

$$\mathcal{D}\big(\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big) = \big(\mathcal{I}(\Phi_u), \mathbb{D}_1(\Phi_u), \mathbb{D}_2(\Phi_u), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_u)}(\Phi_u)\big). \tag{2.18}$$

Note that, e.g., Grohs et al. [36, item (i) in Lemma 3.14] proves that for all $k \in \{u, u+1, \ldots, v\}$ it holds that

$$\mathcal{D}\Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big) = \mathcal{D}\big(\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big). \tag{2.19}$$

Combining this, (2.18), and, e.g., Grohs et al. [36, item (ii) in Lemma 3.28] show that

$$\begin{aligned}\mathcal{D}(\Psi) &= \mathcal{D}\Big(\oplus_{k=u}^{v}\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k})\big)\Big) \\ &= \big(\mathcal{I}(\Phi_u), \textstyle\sum_{k=u}^{v} \mathbb{D}_1(\Phi_u), \sum_{k=u}^{v} \mathbb{D}_2(\Phi_u), \ldots, \sum_{k=u}^{v} \mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big) \\ &= \big(\mathcal{I}(\Phi_u), n\mathbb{D}_1(\Phi_u), n\mathbb{D}_2(\Phi_u), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big).\end{aligned} \tag{2.20}$$

This establishes item (i). Next, observe that, e.g., Grohs et al. [35, item (v) in Proposition 2.6] demonstrates that for all $k \in \{u, u+1, \ldots, v\}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ it holds that $\mathcal{R}_a(\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$ and

$$\Big(\mathcal{R}_a\big(\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big)(x) = (\mathcal{R}_a(\Phi_k))(t_k x + B_k) \tag{2.21}$$

(cf. Definition 2.6). Combining this and, e.g., Grohs et al. [36, Lemma 3.14] yields that for all $k \in \{u, u+1, \ldots, v\}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ it holds that

$$\mathcal{R}_a\Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)}) \tag{2.22}$$

and

$$\Big(\mathcal{R}_a\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k})\big)\Big)(x) = h_k(\mathcal{R}_a(\Phi_k))(t_k x + B_k). \tag{2.23}$$

Furthermore, note that, e.g., Grohs et al. [36, Lemma 3.28] and (2.19) establish that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$ and

$$\begin{aligned}(\mathcal{R}_a(\Psi))(x) &= \Big(\mathcal{R}_a\big(\oplus_{k=u}^{v}(h_k \circledast (\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}))\big)\Big)(x) \\ &= \sum_{k=u}^{v} \Big(\mathcal{R}_a\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{t_k \mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k})\big)\Big)(x) = \sum_{k=u}^{v} h_k(\mathcal{R}_a(\Phi_k))(t_k x + B_k).\end{aligned} \tag{2.24}$$

This establishes items (ii) and (iii). The proof of Lemma 2.19 is thus complete. □

2.6.2 Linear combinations of ANNs with different lengths

Lemma 2.20. *Let $L \in \mathbb{N}$, $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, $h_u, h_{u+1}, \ldots, h_v \in \mathbb{R}$, $\Phi_u, \Phi_{u+1}, \ldots, \Phi_v, \mathfrak{J}, \Psi \in \mathbf{N}$, $B_u, B_{u+1}, \ldots, B_v \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $j \in \mathbb{Z} \cap [u, v]$ that $L = \max_{k \in \mathbb{N} \cap [u,v]} \mathcal{L}(\Phi_k)$, $\mathcal{I}(\Phi_j) = \mathcal{I}(\Phi_u)$, $\mathcal{O}(\Phi_j) = \mathcal{I}(\mathfrak{J}) = \mathcal{O}(\mathfrak{J})$, $\mathcal{H}(\mathfrak{J}) = 1$, $\mathcal{R}_a(\mathfrak{J}) = \mathrm{id}_{\mathbb{R}}$, and*

$$\Psi = \mathop{\boxplus}_{k=u,\mathfrak{J}}^{v} \Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big) \tag{2.25}$$

(cf. Definitions 2.1, 2.6, 2.7, 2.8, 2.12, 2.17, and 2.18). Then

(i) it holds that

$$\begin{aligned}&\mathcal{D}(\Psi) \\ &= \Big(\mathcal{I}(\Phi_u), \sum_{k=u}^{v} \mathbb{D}_1\big(\mathcal{E}_{L,\mathfrak{J}}(\Phi_k)\big), \sum_{k=u}^{v} \mathbb{D}_2\big(\mathcal{E}_{L,\mathfrak{J}}(\Phi_k)\big), \ldots, \sum_{k=u}^{v} \mathbb{D}_{L-1}\big(\mathcal{E}_{L,\mathfrak{J}}(\Phi_k)\big), \mathcal{O}(\Phi_u)\Big),\end{aligned} \tag{2.26}$$

(ii) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$, and

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

$$(\mathcal{R}_a(\Psi))(x) = \sum_{k=u}^{v} h_k(\mathcal{R}_a(\Phi_k))(x + B_k) \tag{2.27}$$

(cf. Definition 2.10).

Proof of Lemma 2.20. Observe that item (i) in Lemma 2.19 establishes item (i). Moreover, note that, e.g., Grohs et al. [35, item (v) in Proposition 2.6] implies that for all $k \in \mathbb{Z} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ it holds that $\mathcal{R}_a(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$ and

$$\Big(\mathcal{R}_a\big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big)(x) = (\mathcal{R}_a(\Phi_k))(x + B_k). \tag{2.28}$$

This, e.g., Grohs et al. [36, Lemma 3.14], and, e.g., Grohs et al. [35, item (ii) in Lemma 2.14] ensure that for all $k \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ it holds that

$$\mathcal{R}_a\Big(\mathcal{E}_{L,\mathfrak{J}}\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k})\big)\Big) = \mathcal{R}_a\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k})\big) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)}) \tag{2.29}$$

and

$$\Big(\mathcal{R}_a\Big(\mathcal{E}_{L,\mathfrak{J}}\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)},B_k})\big)\Big)\Big)(x) = \Big(\mathcal{R}_a\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)},B_k})\big)\Big)(x) \\ = h_k(\mathcal{R}_a(\Phi_k))(x+B_k) \tag{2.30}$$

(cf. Definition 2.10). Combining this, e.g., Grohs et al. [36, Lemma 3.28], and (2.19) assures that for all $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$ and

$$\begin{aligned}(\mathcal{R}_a(\Psi))(x) &= \Bigg(\mathcal{R}_a\Bigg(\mathop{\boxplus}_{k=u,\mathfrak{J}}^{v}\Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)},B_k}\big)\Big)\Bigg)\Bigg)(x) \\ &= \Bigg(\mathcal{R}_a\Bigg(\mathop{\oplus}_{k=u}^{v}\mathcal{E}_{L,\mathfrak{J}}\Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)},B_k}\big)\Big)\Bigg)\Bigg)(x) \\ &= \sum_{k=u}^{v}\Big(\mathcal{R}_a\Big(\mathcal{E}_{L,\mathfrak{J}}\Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)},B_k}\big)\Big)\Big)\Big)(x) = \sum_{k=u}^{v} h_k(\mathcal{R}_a(\Phi_k))(x+B_k)\end{aligned} \tag{2.31}$$

(cf. Definition 2.16). This establishes items (ii) and (iii). The proof of Lemma 2.20 is thus complete. □

3 ANN representations for MLP approximations

In this section we study deep ANN representations for MLP approximations for PDEs of the form (1.1). Specifically, in Proposition 3.9 below we show that realizations of suitable deep ANNs with a given general activation function coincide with appropriate MLP approximations for PDEs of the form (1.1) provided that with the considered general activation function there exists a shallow ANN representation for the one-dimensional identity function $\mathbb{R} \ni x \mapsto x \in \mathbb{R}$.

In the elementary results in Section 3.1 and Section 3.2 we explicitly construct such shallow ANN representations for the one-dimensional identity function $\mathbb{R} \ni x \mapsto x \in \mathbb{R}$ in the case of several activation functions such as the ReLU (see Lemma 3.5), the leaky ReLU (see Lemma 3.5), the rectified power unit (RePU) (see Lemma 3.7), and the softplus (see Lemma 3.8) activation functions. In the special case of the ReLU activation results similar to Proposition 3.9 can, e.g., be found in Hutzenthaler et al. [46, Lemma 3.10] and Cioica-Licht et al. [18, Lemma 3.10].

3.1 Activation functions as ANNs

Definition 3.1 (Activation ANN)**.** Let $n \in \mathbb{N}$. Then we denote by $\mathfrak{i}_n \in ((\mathbb{R}^{n\times n} \times \mathbb{R}^n) \times (\mathbb{R}^{n\times n} \times \mathbb{R}^n)) \subseteq \mathbf{N}$ the ANN given by $\mathfrak{i}_n = ((\mathrm{I}_n, 0), (\mathrm{I}_n, 0))$ (cf. Definitions 2.1, 2.2, and 2.8).

Lemma 3.2. *Let $n \in \mathbb{N}$. Then*

(i) it holds that $\mathcal{D}(\mathfrak{i}_n) = (n, n, n) \in \mathbb{N}^3$,

(ii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\mathfrak{i}_n) \in C(\mathbb{R}^n, \mathbb{R}^n)$, and

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\mathfrak{i}_n) = \mathfrak{M}_{a,n}$

(cf. Definitions 2.1, 2.5, 2.6, and 3.1).

Proof of Lemma 3.2. Observe that the fact that $\mathfrak{i}_n \in ((\mathbb{R}^{n\times n} \times \mathbb{R}^n) \times (\mathbb{R}^{n\times n} \times \mathbb{R}^n)) \subseteq \mathbf{N}$ proves that $\mathcal{D}(\mathfrak{i}_n) = (n, n, n) \in \mathbb{N}^3$ (cf. Definitions 2.1 and 3.1). This establishes item (i). In addition, note that the fact that $\mathfrak{i}_n = ((\mathrm{I}_n, 0), (\mathrm{I}_n, 0)) \in ((\mathbb{R}^{n\times n} \times \mathbb{R}^n) \times (\mathbb{R}^{n\times n} \times \mathbb{R}^n))$ and (2.6) show that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathfrak{i}_n) \in C(\mathbb{R}^n, \mathbb{R}^n)$ and

$$(\mathcal{R}_a(\mathfrak{i}_n))(x) = \mathrm{I}_n(\mathfrak{M}_{a,n}(\mathrm{I}_n x + 0)) + 0 = \mathfrak{M}_{a,n}(x) \tag{3.1}$$

(cf. Definitions 2.5, 2.6, and 2.8). This establishes items (ii) and (iii). The proof of Lemma 3.2 is thus complete. ☐

Lemma 3.3. *Let $\Phi \in \mathbf{N}$ (cf. Definition 2.1). Then*

(i) it holds that

$$\mathcal{D}\big(\mathfrak{i}_{\mathcal{O}(\Phi)} \bullet \Phi\big) = \big(\mathcal{I}(\Phi), \mathbb{D}_1(\Phi), \mathbb{D}_2(\Phi), \ldots, \mathbb{D}_{\mathcal{L}(\Phi)-1}(\Phi), \mathcal{O}(\Phi), \mathcal{O}(\Phi)\big) \in \mathbb{N}^{\mathcal{L}(\Phi)+2}, \tag{3.2}$$

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

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that $(\mathcal{R}_a(\mathfrak{i}_{\mathcal{O}(\Phi)} \bullet \Phi))(x) = \mathfrak{M}_{a,\mathcal{O}(\Phi)}((\mathcal{R}_a(\Phi))(x))$,

(iv) it holds that

$$\mathcal{D}\big(\Phi \bullet \mathfrak{i}_{\mathcal{I}(\Phi)}\big) = \big(\mathcal{I}(\Phi), \mathcal{I}(\Phi), \mathbb{D}_1(\Phi), \mathbb{D}_2(\Phi), \ldots, \mathbb{D}_{\mathcal{L}(\Phi)-1}(\Phi), \mathcal{O}(\Phi)\big) \in \mathbb{N}^{\mathcal{L}(\Phi)+2}, \tag{3.3}$$

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

(vi) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that $(\mathcal{R}_a(\Phi \bullet \mathfrak{i}_{\mathcal{I}(\Phi)}))(x) = (\mathcal{R}_a(\Phi))(\mathfrak{M}_{a,\mathcal{I}(\Phi)}(x))$

(cf. Definitions 2.5, 2.6, 2.7, and 3.1).

Proof of Lemma 3.3. Observe that Lemma 3.2 demonstrates that for all $n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathfrak{i}_n) \in C(\mathbb{R}^n, \mathbb{R}^n)$ and

$$(\mathcal{R}_a(\mathfrak{i}_n))(x) = \mathfrak{M}_{a,n}(x) \tag{3.4}$$

(cf. Definitions 2.5, 2.6, and 3.1). Combining this and, e.g., Grohs et al. [35, Proposition 2.6] establishes items (i), (ii), (iii), (iv), (v), and (vi). The proof of Lemma 3.3 is thus complete. ☐

3.2 ANN representations for the one-dimensional identity function

Definition 3.4 (RePU monomial ANNs)**.** Let $\gamma \in \mathbb{N}_0$. Then we denote by $\mathbf{I}_\gamma \in \mathbf{N}$ the ANN given by

$$\mathbf{I}_\gamma = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right), \left(\begin{pmatrix} 1 & (-1)^\gamma \end{pmatrix}, 0\right)\right) \in ((\mathbb{R}^{2\times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1\times 2} \times \mathbb{R}^1)) \tag{3.5}$$

(cf. Definitions 2.1 and 2.2).

Lemma 3.5 (Shallow leaky ReLU ANN representation for the one-dimensional identity function)**.** *Let $\alpha \in [0, \infty)$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Psi \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, \alpha x\}$ and $\Psi = (1 + \alpha)^{-1} \circledast \mathbf{I}_1$ (cf. Definitions 2.1, 2.18, and 3.4). Then*

(i) it holds for all $\gamma \in \mathbb{N}_0$ that $\mathcal{D}(\mathbf{I}_\gamma) = (1, 2, 1) \in \mathbb{N}^3$,

(ii) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_a(\mathbf{I}_1))(x) = (1 + \alpha)x$,

(iii) it holds that $\mathcal{D}(\Psi) = (1, 2, 1) \in \mathbb{N}^3$,

(iv) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}, \mathbb{R})$, and

(v) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_a(\Psi))(x) = x$

(cf. Definition 2.6).

Proof of Lemma 3.5. Note that (3.5) yields that for all $\gamma \in \mathbb{N}_0$ it holds that $\mathcal{D}(\mathbf{I}_\gamma) = (1, 2, 1) \in \mathbb{N}^3$. This establishes item (i). Next, observe that (3.5) establishes that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\mathbf{I}_1))(x) = a(x) - a(-x) &= \max\{x, \alpha x\} - \max\{-x, -\alpha x\} \\ &= \max\{x, \alpha x\} + \min\{x, \alpha x\} = (1 + \alpha)x\end{aligned} \tag{3.6}$$

(cf. Definition 2.6). This establishes item (ii). Furthermore, note that item (i) in Lemma 2.19 (applied with $u \curvearrowleft 1$, $v \curvearrowleft 1$, $h_u \curvearrowleft (1 + \alpha)^{-1}$, $t_u \curvearrowleft 1$, $\Phi_u \curvearrowleft \mathbf{I}_1$, $B_u \curvearrowleft 0$, $\Psi \curvearrowleft \Psi$ in the notation of Lemma 2.19) establishes item (iii). Combining (3.6) and item (iii) in Lemma 2.19 (applied with $u \curvearrowleft 1$, $v \curvearrowleft 1$, $h_u \curvearrowleft (1 + \alpha)^{-1}$, $t_u \curvearrowleft 1$, $\Phi_u \curvearrowleft \mathbf{I}_1$, $B_u \curvearrowleft 0$, $\Psi \curvearrowleft \Psi$, $a \curvearrowleft a$ in the notation of Lemma 2.19) therefore proves items (iv) and (v). The proof of Lemma 3.5 is thus complete. ☐

Lemma 3.6 (Shallow power ANN representation for the one-dimensional identity function)**.** *Let $\gamma \in \mathbb{N} \backslash \{1\}$, $b_1, b_2, \ldots, b_\gamma \in \mathbb{R}$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = x^\gamma$ and $b_1 < b_2 < \ldots < b_\gamma$. Then*

(i) there exist unique $c_0, c_1, \ldots, c_\gamma \in \mathbb{R}$ which satisfy for all $k \in \{0, 1, \ldots, \gamma\}$ that $\mathbb{1}_{\{\gamma\}}(k)\, c_0 + \sum_{i=1}^{\gamma} c_i (b_i)^k = \mathbb{1}_{\{\gamma-1\}}(k)\, \gamma^{-1}$,

(ii) there exists a unique $\Psi \in \mathbf{N}$ which satisfies

$$\Psi = \mathbf{A}_{1,c_0} \bullet \left(\overset{\gamma}{\underset{i=1}{\oplus}} \Big(c_i \circledast \big(\mathfrak{i}_1 \bullet \mathbf{A}_{1,b_i} \big) \Big) \right), \tag{3.7}$$

(iii) it holds that $\mathcal{D}(\Psi) = (1, \gamma, 1) \in \mathbb{N}^3$,

(iv) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}, \mathbb{R})$, and

(v) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_a(\Psi))(x) = x$

(cf. Definitions 2.1, 2.6, 2.7, 2.12, 2.16, 2.18, and 3.1).

Proof of Lemma 3.6. Throughout this proof let $\mathbf{B} = (\mathbf{B}_{i,j})_{i,j\in\{1,2,\dots,\gamma+1\}} \in \mathbb{R}^{(\gamma+1)\times(\gamma+1)}$ satisfy for all $i, j \in \{1, 2, \dots, \gamma\}$ that $\mathbf{B}_{1,i+1} = 1$, $\mathbf{B}_{i,1} = 0$, $\mathbf{B}_{\gamma+1,1} = 1$, and $\mathbf{B}_{i+1,j+1} = (b_j)^i$ and let $\mathbf{D} = (\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_{\gamma+1})^* \in \mathbb{R}^{(\gamma+1)\times 1}$ satisfy for all $k \in \{1, 2, \dots, \gamma+1\}$ that $\mathbf{D}_k = \mathbb{1}_{\{\gamma\}}(k)\,\gamma^{-1}$ (cf. Definition 2.14). Observe that the assumption that $b_1 < b_2 < \ldots < b_\gamma$ and, e.g., Horn and Johnson [40, Eq. (0.9.11.2)] show that

$$\det(\mathbf{B}) = (-1)^{\gamma+1}\det\big((\mathbf{B}_{i,j+1})_{i,j\in\{1,2,\dots,\gamma\}}\big) = (-1)^{\gamma}\left[\textstyle\prod_{\substack{i,j\in\{1,2,\dots,\gamma\}\\ i<j}}(b_j - b_i)\right] \neq 0. \tag{3.8}$$

This demonstrates that there exists a unique $\mathbf{C} = (c_0, c_1, \dots, c_\gamma)^* \in \mathbb{R}^{(\gamma+1)\times 1}$ such that $\mathbf{B}\mathbf{C} = \mathbf{D}$. This establishes item (i). Moreover, note that Lemma 2.19 and item (i) establish item (ii). In addition, observe that, e.g., Grohs et al. [35, item (i) in Proposition 2.6], item (i) in Lemma 2.19, and item (i) in Lemma 3.3 yield that

$$\mathcal{D}(\Psi) = \big(\mathcal{I}(\mathfrak{i}_1), \textstyle\sum_{k=1}^{\gamma}\mathbb{D}_1(\mathfrak{i}_1), \mathcal{O}(\mathfrak{i}_1)\big) = (1, \gamma, 1) \tag{3.9}$$

(cf. Definitions 2.1 and 3.1). This establishes item (iii). Next, note that, e.g., Grohs et al. [35, item (v) in Proposition 2.6], item (iii) in Lemma 3.2, and item (iii) in Lemma 2.19 establish that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Psi))(x) &= c_0 + \Big(\mathcal{R}_a\Big(\oplus_{i=1}^{\gamma}\big(c_i \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,b_i})\big)\Big)\Big)(x)\\ &= c_0 + \sum_{i=1}^{\gamma} c_i(\mathcal{R}_a(\mathfrak{i}_1))(x + b_i) = c_0 + \sum_{i=1}^{\gamma} c_i(x + b_i)^{\gamma}\end{aligned} \tag{3.10}$$

(cf. Definitions 2.6, 2.7, 2.12, 2.16, and 2.18). This and the binomial theorem imply that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_a(\Psi))(x) = c_0 + \sum_{i=1}^{\gamma} c_i\left[\sum_{j=0}^{\gamma}\tbinom{\gamma}{j}x^{\gamma-j}(b_i)^j\right] = c_0 + \sum_{j=0}^{\gamma}\tbinom{\gamma}{j}\left[\sum_{i=1}^{\gamma}c_i(b_i)^j\right]x^{\gamma-j}. \tag{3.11}$$

Combining (3.11) and item (i) hence ensures that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Psi))(x) &= c_0 + \sum_{j=0}^{\gamma}\tbinom{\gamma}{j}\left[\sum_{i=1}^{\gamma}c_i(b_i)^j\right]x^{\gamma-j}\\ &= c_0 + \sum_{j=0}^{\gamma}\tbinom{\gamma}{j}\big[\mathbb{1}_{\{\gamma-1\}}(j)\,\gamma^{-1} - \mathbb{1}_{\{\gamma\}}(j)\,c_0\big]x^{\gamma-j} = x.\end{aligned} \tag{3.12}$$

This establishes items (iv) and (v). The proof of Lemma 3.6 is thus complete. ◻

Lemma 3.7 (Shallow RePU ANN representation for the one-dimensional identity function)**.** *Let $\gamma \in \mathbb{N}\backslash\{1\}$, $b_1, b_2, \dots, b_\gamma \in \mathbb{R}$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = (\max\{x, 0\})^\gamma$ and $b_1 < b_2 < \ldots < b_\gamma$. Then*

(i) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_a(\mathbf{I}_\gamma))(x) = x^\gamma$,

(ii) there exist unique $c_0, c_1, \dots, c_\gamma \in \mathbb{R}$ which satisfy for all $k \in \{0, 1, \dots, \gamma\}$ that $\mathbb{1}_{\{\gamma\}}(k)\,c_0 + \sum_{i=1}^{\gamma} c_i(b_i)^k = \mathbb{1}_{\{\gamma-1\}}(k)\,\gamma^{-1}$,

(iii) there exists a unique $\Psi \in \mathbf{N}$ *which satisfies*

$$\Psi = \mathbf{A}_{1,c_0} \bullet \left(\overset{\gamma}{\underset{i=1}{\oplus}} \Big(c_i \circledast \big(\mathbf{I}_\gamma \bullet \mathbf{A}_{1,b_i} \big) \Big) \right), \tag{3.13}$$

(iv) it holds that $\mathcal{D}(\Psi) = (1, 2\gamma, 1) \in \mathbb{N}^3$*,*

(v) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}, \mathbb{R})$*, and*

(vi) it holds for all $x \in \mathbb{R}$ *that* $(\mathcal{R}_a(\Psi))(x) = x$

(cf. Definitions 2.1, 2.6, 2.7, 2.12, 2.16, 2.18, and 3.4).

Proof of Lemma 3.7. First, observe that (3.5) assures that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\mathbf{I}_\gamma))(x) = a(x) + (-1)^\gamma a(-x) &= \big(\max\{x, 0\}\big)^\gamma + (-1)^\gamma \big(\max\{-x, 0\}\big)^\gamma \\ &= \big(\max\{x, 0\}\big)^\gamma + \big(\min\{x, 0\}\big)^\gamma = x^\gamma\end{aligned} \tag{3.14}$$

(cf. Definitions 2.6 and 3.4). This establishes item (i). Furthermore, note that item (i) in Lemma 3.6 establishes item (ii). Moreover, observe that Lemma 2.19 and item (ii) establish item (iii). In addition, note that item (i) in Lemma 2.19 and item (i) in Lemma 3.5 prove that

$$\mathcal{D}(\Psi) = \big(\mathcal{I}(\mathbf{I}_\gamma), \textstyle\sum_{k=1}^{\gamma} \mathbb{D}_1(\mathbf{I}_\gamma), \mathcal{O}(\mathbf{I}_\gamma)\big) = (1, 2\gamma, 1) \tag{3.15}$$

(cf. Definition 2.1). This establishes item (iv). Next, observe that item (i), e.g., Grohs et al. [35, item (v) in Proposition 2.6], and item (iii) in Lemma 2.19 show that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Psi))(x) &= c_0 + \Big(\mathcal{R}_a\Big(\oplus_{i=1}^{\gamma}\big(c_i \circledast (\mathbf{I}_\gamma \bullet \mathbf{A}_{1,b_i})\big)\Big)\Big)(x) \\ &= c_0 + \sum_{i=1}^{\gamma} c_i (\mathcal{R}_a(\mathbf{I}_\gamma))(x + b_i) = c_0 + \sum_{i=1}^{\gamma} c_i (x + b_i)^\gamma\end{aligned} \tag{3.16}$$

(cf. Definitions 2.7, 2.12, 2.16, and 2.18). This and the binomial theorem demonstrate that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_a(\Psi))(x) = c_0 + \sum_{i=1}^{\gamma} c_i \left[\sum_{j=0}^{\gamma} \tbinom{\gamma}{j} x^{\gamma-j} (b_i)^j\right] = c_0 + \sum_{j=0}^{\gamma} \tbinom{\gamma}{j} \left[\sum_{i=1}^{\gamma} c_i (b_i)^j\right] x^{\gamma-j}. \tag{3.17}$$

Combining (3.17) and item (ii) therefore yields that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Psi))(x) &= c_0 + \sum_{j=0}^{\gamma} \tbinom{\gamma}{j} \left[\sum_{i=1}^{\gamma} c_i (b_i)^j\right] x^{\gamma-j} \\ &= c_0 + \sum_{j=0}^{\gamma} \tbinom{\gamma}{j} \big[\mathbb{1}_{\{\gamma-1\}}(j)\, \gamma^{-1} - \mathbb{1}_{\{\gamma\}}(j)\, c_0\big] x^{\gamma-j} = x.\end{aligned} \tag{3.18}$$

This establishes items (v) and (vi). The proof of Lemma 3.7 is thus complete. ☐

Lemma 3.8 (Shallow softplus ANN representation for the one-dimensional identity function)**.** *Let $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \ln(1 + \exp(x))$. Then*

(i) it holds that $\mathcal{R}_a(\mathbf{I}_1) \in C(\mathbb{R}, \mathbb{R})$ and

(ii) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_a(\mathbf{I}_1))(x) = x$

(cf. Definitions 2.6 and 3.4).

Proof of Lemma 3.8. Note that (3.5) establishes that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\mathbf{I}_1))(x) = a(x) - a(-x) &= \ln\big(1 + \exp(x)\big) - \ln\big(1 + \exp(-x)\big) \\ &= \ln\bigg(\frac{1 + \exp(x)}{1 + \exp(-x)}\bigg) = \ln\big(\exp(x)\big) = x\end{aligned} \tag{3.19}$$

(cf. Definitions 2.6 and 3.4). This establishes items (i) and (ii). The proof of Lemma 3.8 is thus complete. □

3.3 ANN representations for MLP approximations

Proposition 3.9. *Let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, $d, M, \mathfrak{d} \in \mathbb{N}$, $T \in (0, \infty)$, $a \in C(\mathbb{R}, \mathbb{R})$, $\mathfrak{J}, \mathbf{F}, \mathbf{G} \in \mathbf{N}$ satisfy $\mathcal{D}(\mathfrak{J}) = (1, \mathfrak{d}, 1)$, $\mathcal{R}_a(\mathfrak{J}) = \mathrm{id}_{\mathbb{R}}$, $\mathcal{R}_a(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$, and $\mathcal{R}_a(\mathbf{G}) \in C(\mathbb{R}^d, \mathbb{R})$, for every $\theta \in \Theta$ let $\mathcal{U}^\theta \colon [0, T] \to [0, T]$ and $W^\theta \colon [0, T] \to \mathbb{R}^d$ be functions, for every $\theta \in \Theta$, $n \in \mathbb{N}_0$ let $U_n^\theta \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that*

$$\begin{aligned} U_n^\theta(t, x) &= \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n} \Bigg[\sum_{k=1}^{M^n} \big(\mathcal{R}_a(\mathbf{G})\big)\big(x + W_{T-t}^{(\theta,0,-k)}\big)\Bigg] \\ &+ \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}} \Bigg[\sum_{k=1}^{M^{n-i}} \big((\mathcal{R}_a(\mathbf{F}) \circ U_i^{(\theta,i,k)}) - \mathbb{1}_{\mathbb{N}}(i)(\mathcal{R}_a(\mathbf{F}) \circ U_{\max\{i-1,0\}}^{(\theta,-i,k)})\big)\big(\mathcal{U}_t^{(\theta,i,k)}, x + W_{\mathcal{U}_t^{(\theta,i,k)}-t}^{(\theta,i,k)}\big)\Bigg], \end{aligned} \tag{3.20}$$

and let $\mathbf{U}_{n,t}^\theta \in \mathbf{N}$, $t \in [0, T]$, $n \in \mathbb{Z}$, $\theta \in \Theta$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}$, $t \in [0, T]$ that $\mathbf{U}_{0,t}^\theta = ((0\ 0\ \ldots\ 0), 0) \in \mathbb{R}^{1 \times d} \times \mathbb{R}^1$ and

$$\begin{aligned} \mathbf{U}_{n,t}^\theta = &\Bigg[\bigoplus_{k=1}^{M^n} \Big(\tfrac{1}{M^n} \circledast \big(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, W_{T-t}^{(\theta,0,-k)}}\big)\Big)\Bigg] \\ &\boxplus_{\mathfrak{J}} \Bigg[\mathop{\boxplus}_{i=0,\mathfrak{J}}^{n-1} \Bigg[\Big(\tfrac{(T-t)}{M^{n-i}}\Big) \circledast \Bigg(\mathop{\boxplus}_{k=1,\mathfrak{J}}^{M^{n-i}} \Big(\big(\mathbf{F} \bullet \mathbf{U}_{i,\mathcal{U}_t^{(\theta,i,k)}}^{(\theta,i,k)}\big) \bullet \mathbf{A}_{\mathrm{I}_d, W_{\mathcal{U}_t^{(\theta,i,k)}-t}^{(\theta,i,k)}}\Big)\Bigg)\Bigg]\Bigg] \\ &\boxplus_{\mathfrak{J}} \Bigg[\mathop{\boxplus}_{i=0,\mathfrak{J}}^{n-1} \Bigg[\Big(\tfrac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\Big) \circledast \Bigg(\mathop{\boxplus}_{k=1,\mathfrak{J}}^{M^{n-i}} \Big(\big(\mathbf{F} \bullet \mathbf{U}_{\max\{i-1,0\},\mathcal{U}_t^{(\theta,i,k)}}^{(\theta,-i,k)}\big) \bullet \mathbf{A}_{\mathrm{I}_d, W_{\mathcal{U}_t^{(\theta,i,k)}-t}^{(\theta,i,k)}}\Big)\Bigg)\Bigg]\Bigg] \end{aligned} \tag{3.21}$$

(cf. Definitions 2.1, 2.6, 2.7, 2.8, 2.12, 2.16, 2.17, and 2.18). Then

(i) it holds for all $\theta_1, \theta_2 \in \Theta$, $n \in \mathbb{N}_0$, $t_1, t_2 \in [0, T]$ that $\mathcal{D}(\mathbf{U}_{n,t_1}^{\theta_1}) = \mathcal{D}(\mathbf{U}_{n,t_2}^{\theta_2})$,

(ii) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0, T]$ that $\mathcal{R}_a(\mathbf{U}_{n,t}^\theta) \in C(\mathbb{R}^d, \mathbb{R})$,

(iii) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$ that $\mathcal{L}(\mathbf{U}^\theta_{n,t}) \leq \max\{\mathfrak{d}, \mathcal{L}(\mathbf{G})\} + n\mathcal{H}(\mathbf{F})$,

(iv) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$ that

$$\left|\!\left|\!\left|\mathcal{D}(\mathbf{U}^\theta_{n,t})\right|\!\right|\!\right| \leq \max\bigl\{\mathfrak{d}, |||\mathcal{D}(\mathbf{F})|||, |||\mathcal{D}(\mathbf{G})|||\bigr\}(3M)^n, \tag{3.22}$$

(v) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that $U^\theta_n(t,x) = ((\mathcal{R}_a(\mathbf{U}^\theta_{n,t}))(x)$, and

(vi) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$ that

$$\mathcal{P}(\mathbf{U}^\theta_{n,t}) \leq 2\bigl(\max\{\mathfrak{d}, \mathcal{L}(\mathbf{G})\} + n\mathcal{H}(\mathbf{F})\bigr)\Bigl(\max\bigl\{\mathfrak{d}, |||\mathcal{D}(\mathbf{F})|||, |||\mathcal{D}(\mathbf{G})|||\bigr\}\Bigr)^2(3M)^{2n} \tag{3.23}$$

(cf. Definition 2.3).

Proof of Proposition 3.9. Throughout this proof let $\Phi^\theta_{n,t} \in \mathbf{N}$, $\theta \in \Theta$, $n \in \mathbb{N}$, $t \in [0,T]$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}$, $t \in [0,T]$ that

$$\Phi^\theta_{n,t} = \bigoplus_{k=1}^{M^n}\Bigl(\tfrac{1}{M^n} \circledast \bigl(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, W^{(\theta,0,-k)}_{T-t}}\bigr)\Bigr), \tag{3.24}$$

let $\Psi^{\theta,j}_{n,i,t} \in \mathbf{N}$, $\theta \in \Theta$, $j \in \{0,1\}$, $n \in \mathbb{N}$, $i \in \{0,1,\dots,n-1\}$, $t \in [0,T]$, satisfy for all $\theta \in \Theta$, $j \in \{0,1\}$, $n \in \mathbb{N}$, $i \in \{0,1,\dots,n-1\}$, $t \in [0,T]$ that

$$\Psi^{\theta,j}_{n,i,t} = \boxplus_{k=1,\mathfrak{J}}^{M^{n-i}}\Bigl(\bigl(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\bigr) \bullet \mathbf{A}_{\mathrm{I}_d, W^{(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}}\Bigr), \tag{3.25}$$

let $\Xi^{\theta,j}_{n,t} \in \mathbf{N}$, $\theta \in \Theta$, $j \in \{0,1\}$, $n \in \mathbb{N}$, $t \in [0,T]$, satisfy for all $\theta \in \Theta$, $j \in \{0,1\}$, $n \in \mathbb{N}$, $t \in [0,T]$ that

$$\Xi^{\theta,j}_{n,t} = \boxplus_{i=0,\mathfrak{J}}^{n-1}\Bigl[\Bigl(\tfrac{(-1)^j (T-t)\, \mathbb{1}_{\mathbb{N}}(i-j+1)}{M^{n-i}}\Bigr) \circledast \Psi^{\theta,j}_{n,i,t}\Bigr], \tag{3.26}$$

let $L_i \in \mathbb{N}$, $i \in \mathbb{Z}$, satisfy for all $i \in \mathbb{Z}$ that

$$L_i = \mathcal{L}\Bigl(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i,0\},0}\Bigr), \tag{3.27}$$

let $\mathfrak{L}_n \in \mathbb{N}$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$ that

$$\mathfrak{L}_n = \max_{i \in \{-1,0,\dots,n-1\}} L_i, \tag{3.28}$$

and let $\mathbb{L}_n \in \mathbb{N}$, $n \in \mathbb{N}$ satisfy for all $n \in \mathbb{N}$ that

$$\mathbb{L}_n = \max\{\mathcal{L}(\mathbf{G}), \mathfrak{L}_n, \mathfrak{L}_{n-1}\}. \tag{3.29}$$

We prove items (i), (ii), (iii), (iv), and (v) by induction on $n \in \mathbb{N}_0$. For the base case $n = 0$ observe that the fact for all $\theta \in \Theta$, $t \in [0,T]$ it holds that $\mathbf{U}^\theta_{0,t} = ((0\ 0\ \dots\ 0), 0) \in \mathbb{R}^{1\times d} \times \mathbb{R}^1$ implies that for all $\theta_1, \theta_2 \in \Theta$, $t_1, t_2 \in [0,T]$ it holds that

$$\mathcal{D}(\mathbf{U}^{\theta_1}_{0,t_1}) = (d,1) = \mathcal{D}(\mathbf{U}^{\theta_2}_{0,t_2}) \tag{3.30}$$

and that $\mathcal{R}_a(\mathbf{U}^{\theta_1}_{0,t_1}) \in C(\mathbb{R}^d,\mathbb{R})$. Furthermore, note that the fact that (3.20) implies that for all $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ it holds that $U^\theta_0(t,x)=0$ and the fact for all $\theta\in\Theta$, $t\in[0,T]$ it holds that $\mathbf{U}^\theta_{0,t}=((0\ 0\ \ldots\ 0),0)\in\mathbb{R}^{1\times d}\times\mathbb{R}^1$ ensure that for all $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ it holds that

$$\mathcal{L}(\mathbf{U}^\theta_{0,t})=1, \qquad \left|\!\left|\!\left|\mathcal{D}(\mathbf{U}^\theta_{0,t})\right|\!\right|\!\right|=d, \qquad \text{and} \qquad (\mathcal{R}_a(\mathbf{U}^\theta_{0,t}))(x)=U^\theta_0(t,x) \tag{3.31}$$

(cf. Definition 2.3). Combining (3.30), (3.31), and the fact that the assumption that $\mathcal{R}_a(\mathbf{G})\in C(\mathbb{R}^d,\mathbb{R})$ implies that $\max\{\mathfrak{d},|||\mathcal{D}(\mathbf{F})|||,|||\mathcal{D}(\mathbf{G})|||\}\geq d$ hence proves items (i), (ii), (iii), (iv), and (v) in the base case $n=0$. For the induction step $\mathbb{N}_0\ni(n-1)\dashrightarrow n\in\mathbb{N}$ let $n\in\mathbb{N}$ and assume that items (i), (ii), (iii), (iv), and (v) hold true for all $k\in\{0,1,\ldots,n-1\}$. Observe that the hypothesis that for every $\theta\in\Theta$, $t\in[0,T]$ it holds that $W^\theta_t\in\mathbb{R}^d$ and Lemma 2.19 (applied for every $\theta\in\Theta$, $t\in[0,T]$ with

$$\begin{split}&u \curvearrowleft 1, \quad v\curvearrowleft M^n, \quad (h_k)_{k\in\{u,u+1,\ldots,v\}}\curvearrowleft(M^{-n})_{k\in\{1,2,\ldots,M^n\}}, \quad (t_k)_{k\in\{u,u+1,\ldots,v\}}\curvearrowleft(1)_{k\in\{1,2,\ldots,M^n\}},\\ &\Psi\curvearrowleft\Phi^\theta_{n,t}, \quad (\Phi_k)_{k\in\{u,u+1,\ldots,v\}}\curvearrowleft(\mathbf{G})_{k\in\{1,2,\ldots,M^n\}}, \quad (B_k)_{k\in\{u,u+1,\ldots,v\}}\curvearrowleft(W^{\theta,0,-k}_{T-t})_{k\in\{1,2,\ldots,M^n\}}\end{split} \tag{3.32}$$

in the notation of Lemma 2.19) assure that for all $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ it holds that

$$\mathcal{D}\big(\Phi^\theta_{n,t}\big)=\big(d,M^n\mathbb{D}_1(\mathbf{G}),M^n\mathbb{D}_2(\mathbf{G}),\ldots,M^n\mathbb{D}_{\mathcal{L}(\mathbf{G})-1}(\mathbf{G}),1\big)\in\mathbb{N}^{\mathcal{L}(\mathbf{G})+1} \tag{3.33}$$

and

$$\big(\mathcal{R}_a(\Phi^\theta_{n,t})\big)(x)=\frac{1}{M^n}\left[\sum_{k=1}^{M^n}\big(\mathcal{R}_a(\mathbf{G})\big)\big(x+W^{(\theta,0,-k)}_{T-t}\big)\right]. \tag{3.34}$$

Moreover, note that the induction hypothesis and, e.g., Grohs et al. [35, item (ii) in Proposition 2.6] prove that for all $\theta\in\Theta$, $j\in\{0,1\}$, $i\in\{0,1,\ldots,n-1\}$, $k\in\{1,2,\ldots,M^{n-i}\}$, $t\in[0,T]$ it holds that

$$\begin{split}\mathcal{L}\Big(\mathbf{F}\bullet\mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\Big)&=\mathcal{L}(\mathbf{F})+\mathcal{L}\Big(\mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\Big)-1\\ &=\mathcal{L}(\mathbf{F})+\mathcal{L}\Big(\mathbf{U}^0_{\max\{i-j,0\},0}\Big)-1\\ &=\mathcal{L}\Big(\mathbf{F}\bullet\mathbf{U}^0_{\max\{i-j,0\},0}\Big)=L_{i-j}.\end{split} \tag{3.35}$$

This, the induction hypothesis, the hypothesis that for all $\theta\in\Theta$, $t\in[0,T]$ it holds that $W^\theta_t\in\mathbb{R}^d$, the hypothesis that for all $\theta\in\Theta$, $t\in[0,T]$ it holds that $\mathcal{U}^\theta_t\in[0,T]$, Lemma 2.20 (applied for every $j\in\{0,1\}$, $i\in\{0,1,\ldots,n-1\}$, $\theta\in\Theta$, $t\in[0,T]$ with

$$\begin{split}&u\curvearrowleft 1, \quad v\curvearrowleft M^{n-i}, \quad \mathfrak{J}\curvearrowleft\mathfrak{J}, \quad (B_k)_{k\in\{u,u+1,\ldots,v\}}\curvearrowleft\big(W^{\theta,i,k}_{\mathcal{U}^{(\theta,i,k)}_t-t}\big)_{k\in\{1,2,\ldots,M^{n-i}\}},\\ &(h_k)_{k\in\{u,u+1,\ldots,v\}}\curvearrowleft(1)_{k\in\{1,2,\ldots,M^{n-i}\}}, \quad L\curvearrowleft L_{i-j}, \quad \Psi\curvearrowleft\Psi^{\theta,j}_{n,i,t}, \quad a\curvearrowleft a\\ &(\Phi_k)_{k\in\{u,u+1,\ldots,v\}}\curvearrowleft\big(\mathbf{F}\bullet\mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\big)_{k\in\{1,2,\ldots,M^{n-i}\}}\end{split} \tag{3.36}$$

in the notation of Lemma 2.20), and, e.g., Grohs et al. [35, Proposition 2.6] show that for all $\theta\in\Theta$, $j\in\{0,1\}$, $i\in\{0,1,\ldots,n-1\}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ it holds that

$$\mathcal{D}\big(\Psi^{\theta,j}_{n,i,t}\big)$$

$$
\begin{aligned}
&= \Bigg(d, \sum_{k=1}^{M^{n-i}} \mathbb{D}_1\Big(\mathcal{E}_{L_{i-j},\mathfrak{I}}\big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}_t^{(\theta,i,k)}}\big)\Big), \sum_{k=1}^{M^{n-i}} \mathbb{D}_2\Big(\mathcal{E}_{L_{i-j},\mathfrak{I}}\big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}_t^{(\theta,i,k)}}\big)\Big), \\
&\qquad \ldots, \sum_{k=1}^{M^{n-i}} \mathbb{D}_{L_{i-j}-1}\Big(\mathcal{E}_{L_{i-j},\mathfrak{I}}\big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}_t^{(\theta,i,k)}}\big)\Big), 1\Bigg) \\
&= \Bigg(d, M^{n-i}\mathbb{D}_1\big(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-j,0\},0}\big), M^{n-i}\mathbb{D}_2\big(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-j,0\},0}\big), \\
&\qquad \ldots, M^{n-i}\mathbb{D}_{L_{i-j}-1}\big(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-j,0\},0}\big), 1\Bigg) \in \mathbb{N}^{L_{i-j}+1}
\end{aligned} \tag{3.37}
$$

and

$$
\begin{aligned}
\big(\mathcal{R}_a(\Psi^{\theta,j}_{n,i,t})\big)(x) &= \sum_{k=1}^{M^{n-i}} \Big(\mathcal{R}_a\big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}_t^{(\theta,i,k)}}\big)\Big)\big(x + W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}\big) \\
&= \sum_{k=1}^{M^{n-i}} \Big(\mathcal{R}_a(\mathbf{F}) \circ \mathcal{R}_a\big(\mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}_t^{(\theta,i,k)}}\big)\Big)\big(x + W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}\big) \\
&= \sum_{k=1}^{M^{n-i}} \Big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\}}\Big)\big(\mathcal{U}_t^{(\theta,i,k)}, x + W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}\big).
\end{aligned} \tag{3.38}
$$

This and (3.28) demonstrate that for all $\theta \in \Theta$, $n \in \mathbb{N}$, $j \in \{0,1\}$, $t \in [0,T]$ it holds that

$$
\begin{aligned}
\max_{i \in \{0,1,\ldots,n-1\}} \mathcal{L}\Big(\Big(\tfrac{(-1)^j (T-t)\, \mathbb{1}_{\mathbb{N}}(i-j+1)}{M^{n-i}}\Big) \circledast \Psi^{\theta,j}_{n,i,t}\Big) &= \max_{i \in \{0,1,\ldots,n-1\}} \mathcal{L}\Big(\Psi^{\theta,j}_{n,i,t}\Big) \\
&= \max_{i \in \{0,1,\ldots,n-1\}} L_{i-j} = \mathfrak{L}_{n-j}.
\end{aligned} \tag{3.39}
$$

Combining (3.37), (3.38), (3.39), e.g., Grohs et al. [35, item (i) in Proposition 2.6], and Lemma 2.20 (applied for every $j \in \{0,1\}$, $\theta \in \Theta$, $t \in [0,T]$ with

$$
\begin{aligned}
&u \curvearrowleft 0, \quad v \curvearrowleft n-1, \quad \mathfrak{J} \curvearrowleft \mathfrak{I}, \quad (\Phi_k)_{k \in \{u,u+1,\ldots,v\}} \curvearrowleft \big(\Psi^{\theta,j}_{n,i,t}\big)_{i \in \{0,1,\ldots,n-1\}}, \quad L \curvearrowleft \mathfrak{L}_{n-j}, \\
&(h_k)_{k \in \{u,u+1,\ldots,v\}} \curvearrowleft \Big(\tfrac{(-1)^j (T-t)\, \mathbb{1}_{\mathbb{N}}(i-j+1)}{M^{n-i}}\Big)_{i \in \{0,1,\ldots,n-1\}}, \quad (B_k)_{k \in \{u,u+1,\ldots,v\}} \curvearrowleft ((0,0,\ldots,0))_{k \in \{0,1,\ldots,n-1\}}
\end{aligned} \tag{3.40}
$$

in the notation of Lemma 2.20) yields that for all $j \in \{0,1\}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
\mathcal{D}\big(\Xi^{\theta,j}_{n,t}\big) &= \Bigg(d, \sum_{i=0}^{n-1} \mathbb{D}_1\big(\mathcal{E}_{\mathfrak{L}_{n-j},\mathfrak{I}}\big(\Psi^{\theta,j}_{n,i,t}\big)\big), \sum_{i=0}^{n-1} \mathbb{D}_2\big(\mathcal{E}_{\mathfrak{L}_{n-j},\mathfrak{I}}\big(\Psi^{\theta,j}_{n,i,t}\big)\big), \\
&\qquad \ldots, \sum_{i=0}^{n-1} \mathbb{D}_{\mathfrak{L}_{n-j}-1}\big(\mathcal{E}_{\mathfrak{L}_{n-j},\mathfrak{I}}\big(\Psi^{\theta,j}_{n,i,t}\big)\big), 1\Bigg) \\
&= \Bigg(d, \sum_{i=0}^{n-1} \mathbb{D}_1\big(\mathcal{E}_{\mathfrak{L}_{n-j},\mathfrak{I}}\big(\Psi^{0,j}_{n,i,0}\big)\big), \sum_{i=0}^{n-1} \mathbb{D}_2\big(\mathcal{E}_{\mathfrak{L}_{n-j},\mathfrak{I}}\big(\Psi^{0,j}_{n,i,0}\big)\big), \\
&\qquad \ldots, \sum_{i=0}^{n-1} \mathbb{D}_{\mathfrak{L}_{n-j}-1}\big(\mathcal{E}_{\mathfrak{L}_{n-j},\mathfrak{I}}\big(\Psi^{0,j}_{n,i,0}\big)\big), 1\Bigg) \in \mathbb{N}^{\mathfrak{L}_{n-j}+1}
\end{aligned} \tag{3.41}
$$

and

$$
\big(\mathcal{R}_a(\Xi^{\theta,j}_{n,t})\big)(x) = \sum_{i=0}^{n-1} \Big(\tfrac{(-1)^j (T-t)\, \mathbb{1}_{\mathbb{N}}(i-j+1)}{M^{n-i}}\Big) \big(\mathcal{R}_a(\Psi^{\theta,j}_{n,i,t})\big)(x) \tag{3.42}
$$

$$= \sum_{i=0}^{n-1} \Big(\tfrac{(-1)^j (T-t)\, \mathbb{1}_{\mathbb{N}}(i-j+1)}{M^{n-i}} \Big) \Bigg[\sum_{k=1}^{M^{n-i}} \big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\}} \big) \big(\mathcal{U}_t^{(\theta,i,k)}, x + W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t} \big) \Bigg].$$

In addition, observe that (3.33), (3.41), item (i) in Lemma 2.20 (applied for every $\theta \in \Theta$, $t \in [0,T]$ with $u \curvearrowleft 1$, $v \curvearrowleft 3$, $L \curvearrowleft \mathbb{L}_n$, $\Phi_1 \curvearrowleft \Phi^\theta_{n,t}$, $\Phi_2 \curvearrowleft \Xi^{\theta,0}_{n,t}$, $\Phi_3 \curvearrowleft \Xi^{\theta,1}_{n,t}$, $\mathfrak{J} \curvearrowleft \mathfrak{J}$, $h_1 \curvearrowleft 1$, $h_2 \curvearrowleft 1$, $h_3 \curvearrowleft 1$, $B_1 \curvearrowleft 0$, $B_2 \curvearrowleft 0$, $B_3 \curvearrowleft 0$ in the notation of Lemma 2.20), and, e.g., Grohs et al. [35, item (i) in Proposition 2.6] establish that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that

$$\begin{aligned}
\mathcal{D}(\mathbf{U}^\theta_{n,t}) &= \mathcal{D}\big(\Phi^\theta_{n,t} \boxplus_{\mathfrak{J}} \Xi^{\theta,0}_{n,t} \boxplus_{\mathfrak{J}} \Xi^{\theta,1}_{n,t}\big) \\
&= \Big(d, \mathbb{D}_1\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^\theta_{n,t})\big) + \mathbb{D}_1\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{\theta,0}_{n,t})\big) + \mathbb{D}_1\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{\theta,1}_{n,t})\big), \\
&\qquad \mathbb{D}_2\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^\theta_{n,t})\big) + \mathbb{D}_2\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{\theta,0}_{n,t})\big) + \mathbb{D}_2\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{\theta,1}_{n,t})\big), \\
&\qquad \ldots, \mathbb{D}_{\mathbb{L}_n-1}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^\theta_{n,t})\big) + \mathbb{D}_{\mathbb{L}_n-1}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{\theta,0}_{n,t})\big) + \mathbb{D}_{\mathbb{L}_n-1}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{\theta,1}_{n,t})\big), 1 \Big) \\
&= \Big(d, \mathbb{D}_1\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big) + \mathbb{D}_1\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big) + \mathbb{D}_1\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big), \\
&\qquad \mathbb{D}_2\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big) + \mathbb{D}_2\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big) + \mathbb{D}_2\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big), \\
&\qquad \ldots, \mathbb{D}_{\mathbb{L}_n-1}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big) + \mathbb{D}_{\mathbb{L}_n-1}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big) + \mathbb{D}_{\mathbb{L}_n-1}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big), 1 \Big) \in \mathbb{N}^{\mathbb{L}_n+1}.
\end{aligned} \tag{3.43}$$

Moreover, it then holds for all $\theta \in \Theta$, $t \in [0,T]$ that $\mathcal{R}_a(\mathbf{U}^\theta_{n,t}) \in C(\mathbb{R}^d, \mathbb{R})$. Next, note that (3.24), (3.25), (3.26), (3.34), (3.42), and item (iii) in Lemma 2.20 (applied for every $\theta \in \Theta$, $t \in [0,T]$ with $u \curvearrowleft 1$, $v \curvearrowleft 3$, $L \curvearrowleft \mathbb{L}_n$, $\Phi_1 \curvearrowleft \Phi^\theta_{n,t}$, $\Phi_2 \curvearrowleft \Xi^{\theta,0}_{n,t}$, $\Phi_3 \curvearrowleft \Xi^{\theta,1}_{n,t}$, $\mathfrak{J} \curvearrowleft \mathfrak{J}$, $h_1 \curvearrowleft 1$, $h_2 \curvearrowleft 1$, $h_3 \curvearrowleft 1$, $B_1 \curvearrowleft 0$, $B_2 \curvearrowleft 0$, $B_3 \curvearrowleft 0$ in the notation of Lemma 2.20) imply that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\big(\mathcal{R}_a(\mathbf{U}^\theta_{n,t})\big)(x) &= \big(\mathcal{R}_a(\Phi^\theta_{n,t} \boxplus_{\mathfrak{J}} \Xi^{\theta,0}_{n,t} \boxplus_{\mathfrak{J}} \Xi^{\theta,1}_{n,t})\big)(x) \\
&= \big(\mathcal{R}_a(\Phi^\theta_{n,t})\big)(x) + \big(\mathcal{R}_a(\Xi^{\theta,0}_{n,t})\big)(x) + \big(\mathcal{R}_a(\Xi^{\theta,1}_{n,t})\big)(x) \\
&= \tfrac{1}{M^n} \Bigg[\sum_{k=1}^{M^n} \big(\mathcal{R}_a(\mathbf{G})\big)\big(x + W^{(\theta,0,-k)}_{T-t}\big) \Bigg] \\
&\quad + \sum_{i=0}^{n-1} \tfrac{(T-t)}{M^{n-i}} \Bigg[\sum_{k=1}^{M^{n-i}} \big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,i,k)}_i\big)\big(\mathcal{U}_t^{(\theta,i,k)}, x + W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}\big) \Bigg] \\
&\quad + \sum_{i=0}^{n-1} \tfrac{(t-T)\, \mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}} \Bigg[\sum_{k=1}^{M^{n-i}} \big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,-i,k)}_{\max\{i-1,0\}}\big)\big(\mathcal{U}_t^{(\theta,i,k)}, x + W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}\big) \Bigg] \\
&= U^\theta_n(t,x).
\end{aligned} \tag{3.44}$$

Furthermore, observe that (3.43) and Jensen's inequality ensure that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that

$$\begin{aligned}
&\big|\!\big|\!\big|\mathcal{D}(\mathbf{U}^\theta_{n,t})\big|\!\big|\!\big| = \big|\!\big|\!\big|\mathcal{D}(\mathbf{U}^0_{n,0})\big|\!\big|\!\big| \\
&= \max\Big\{ d, \max_{k \in \{1,2,\ldots,\mathbb{L}_n-1\}} \Big(\mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big) + \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big) + \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big) \Big) \Big\} \\
&\le \max_{k \in \{1,2,\ldots,\mathbb{L}_n-1\}} \Big[\max\big\{ d, \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big) \big\} + \max\big\{ d, \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big) \big\}
\end{aligned} \tag{3.45}$$

$$+\max\bigl\{d,\mathbb{D}_k\bigl(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\bigr)\bigr\}\Bigr]$$
$$\leq \big|\!\big|\!\big|\mathcal{D}\bigl(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^{0}_{n,0})\bigr)\big|\!\big|\!\big| + \big|\!\big|\!\big|\mathcal{D}\bigl(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\bigr)\big|\!\big|\!\big| + \big|\!\big|\!\big|\mathcal{D}\bigl(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\bigr)\big|\!\big|\!\big|.$$

This, (3.33), (3.37), (3.41), Jensen's inequality, and, e.g., Grohs et al. [35, item (i) in Proposition 2.6] assure that for all $\theta\in\Theta$, $t\in[0,T]$ it holds that

$$\begin{aligned}
&\big|\!\big|\!\big|\mathcal{D}(\mathbf{U}^{\theta}_{n,t})\big|\!\big|\!\big|\\
&\leq \max\{\mathfrak{d}, M^n|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|\} + \sum_{i=0}^{n-1}\max\bigl\{\mathfrak{d},\big|\!\big|\!\big|\mathcal{D}\bigl(\Psi^{0,0}_{n,i,0}\bigr)\big|\!\big|\!\big|\bigr\} + \sum_{i=0}^{n-1}\max\bigl\{\mathfrak{d},\big|\!\big|\!\big|\mathcal{D}\bigl(\Psi^{0,1}_{n,i,0}\bigr)\big|\!\big|\!\big|\bigr\}\\
&\leq \max\{\mathfrak{d}, M^n|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|\}\\
&\quad+\sum_{i=0}^{n-1}M^{n-i}\Bigl[\max\bigl\{\mathfrak{d},\big|\!\big|\!\big|\mathcal{D}\bigl(\mathbf{F}\bullet\mathbf{U}^{0}_{i,0}\bigr)\big|\!\big|\!\big|\bigr\}+\max\bigl\{\mathfrak{d},\big|\!\big|\!\big|\mathcal{D}\bigl(\mathbf{F}\bullet\mathbf{U}^{0}_{\max\{i-1,0\},0}\bigr)\big|\!\big|\!\big|\bigr\}\Bigr] \qquad (3.46)\\
&\leq \max\{\mathfrak{d}, M^n|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|\}\\
&\quad+\sum_{i=0}^{n-1}M^{n-i}\Bigl[\max\bigl\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|,\big|\!\big|\!\big|\mathcal{D}\bigl(\mathbf{U}^{0}_{i,0}\bigr)\big|\!\big|\!\big|\bigr\}+\max\bigl\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|,\big|\!\big|\!\big|\mathcal{D}\bigl(\mathbf{U}^{0}_{\max\{i-1,0\},0}\bigr)\big|\!\big|\!\big|\bigr\}\Bigr].
\end{aligned}$$

The induction hypothesis hence proves that for all $\theta\in\Theta$, $t\in[0,T]$ it holds that

$$\begin{aligned}
\big|\!\big|\!\big|\mathcal{D}(\mathbf{U}^{\theta}_{n,t})\big|\!\big|\!\big| &\leq \max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|\}M^n\\
&\qquad+\sum_{i=0}^{n-1}M^{n-i}\max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|\}\Bigl[(3M)^i+(3M)^{\max\{i-1,0\}}\Bigr]\\
&=\max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|\}M^n\Biggl[1+\Biggl(\sum_{i=0}^{n-1}3^i\Biggr)+1+\tfrac{1}{M}\sum_{i=1}^{n-1}3^{i-1}\Biggr] \qquad (3.47)\\
&\leq \max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|\}M^n\Bigl[2+\tfrac{3^n-1}{2}+\tfrac{3^{n-1}-1}{2}\Bigr]\\
&=\max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|\}(3M)^n\Bigl[\tfrac{1}{3^n}+\tfrac{2}{3}\Bigr]\\
&\leq \max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|\}(3M)^n.
\end{aligned}$$

Moreover, note that the induction hypothesis, (3.29), (3.43), Jensen's inequality, and, e.g., Grohs et al. [35, item (ii) in Proposition 2.6] show that for all $\theta\in\Theta$, $t\in[0,T]$ it holds that

$$\begin{aligned}
\mathcal{L}(\mathbf{U}^{\theta}_{n,t}) = \mathcal{L}(\mathbf{U}^{0}_{n,0}) = \mathbb{L}_n &= \max\Bigl\{\mathcal{L}(\mathbf{G}),\max_{i\in\{-1,0,\ldots,n-1\}}\mathcal{L}\bigl(\mathbf{F}\bullet\mathbf{U}^{0}_{\max\{i,0\},0}\bigr)\Bigr\}\\
&=\max\Bigl\{\mathcal{L}(\mathbf{G}),\max_{i\in\{-1,0,\ldots,n-1\}}\Bigl(\mathcal{L}(\mathbf{F})+\mathcal{L}\bigl(\mathbf{U}^{0}_{\max\{i,0\},0}\bigr)-1\Bigr)\Bigr\}\\
&=\max\Bigl\{\mathcal{L}(\mathbf{G}),\mathcal{H}(\mathbf{F})+\max_{i\in\{-1,0,\ldots,n-1\}}\mathcal{L}\bigl(\mathbf{U}^{0}_{\max\{i,0\},0}\bigr)\Bigr\} \qquad (3.48)\\
&\leq\max\Bigl\{\mathcal{L}(\mathbf{G}),\mathcal{H}(\mathbf{F})+\max_{i\in\{0,1,\ldots,n-1\}}\Bigl(\max\{\mathfrak{d},\mathcal{L}(\mathbf{G})\}+\max\{i,0\}\mathcal{H}(\mathbf{F})\Bigr)\Bigr\}\\
&=\max\Bigl\{\mathcal{L}(\mathbf{G}),\mathcal{H}(\mathbf{F})+\bigl[\max\{\mathfrak{d},\mathcal{L}(\mathbf{G})\}+(n-1)\mathcal{H}(\mathbf{F})\bigr]\Bigr\}\leq\max\{\mathfrak{d},\mathcal{L}(\mathbf{G})\}+n\mathcal{H}(\mathbf{F}).
\end{aligned}$$

Combining (3.43), (3.44), (3.47), and (3.48) completes the induction step. Induction hence establishes items (i), (ii), (iii), (iv), and (v). In addition, observe that item (iii) and item (iv) demonstrate that for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$ it holds that

$$\begin{aligned}\mathcal{P}(\mathbf{U}^\theta_{n,t}) &\leq \sum_{k=1}^{\mathcal{L}(\mathbf{U}^\theta_{n,t})} |||\mathcal{D}(\mathbf{U}^\theta_{n,t})||| \Big[|||\mathcal{D}(\mathbf{U}^\theta_{n,t})||| + 1\Big] \leq 2\mathcal{L}(\mathbf{U}^\theta_{n,t})|||\mathcal{D}(\mathbf{U}^\theta_{n,t})|||^2 \\ &\leq 2\big(\max\{\mathfrak{d}, \mathcal{L}(\mathbf{G})\} + n\mathcal{H}(\mathbf{F})\big)\Big(\max\big\{\mathfrak{d}, |||\mathcal{D}(\mathbf{F})|||, |||\mathcal{D}(\mathbf{G})|||\big\}\Big)^2 (3M)^{2n}.\end{aligned} \tag{3.49}$$

This establishes item (vi). The proof of Proposition 3.9 is thus complete. □

4 ANN approximations for PDEs

In this section we use the ANN representations for MLP approximations from Section 3 to state and prove in Theorem 4.1 in Section 4.1 below the main ANN approximation result of this work. In our proof of Theorem 4.1 we employ the error estimates for suitable MLP approximations from Hutzenthaler et al. [48] while the arguments in our proof of Theorem 4.1 are inspired by the arguments in Hutzenthaler et al. [46].

In the elementary results in Section 4.2 we show that the Lipschitz continuous nonlinearity of the PDE in (4.5) in Theorem 4.1 can be approximated with suitable convergence rates by ANNs with ReLU, leaky ReLU, or softplus activation functions. In the situation of the ReLU activation function we use a linear interpolation technique similar as, for example, in Hutzenthaler et al. [46, Section 3.4].

In Section 4.3 below we combine the ANN approximation result in Theorem 4.1 from Section 4.1 and the ANN approximation results for the Lipschitz continuous PDE nonlinearities from Section 4.2 with the ANN presentation results for the one-dimensional identity function $\mathbb{R} \ni x \mapsto x \in \mathbb{R}$ from Section 3.2 to establish the ANN approximation results in Corollary 4.15 and Corollary 4.16. Theorem 1.1 in the introduction, in turn, is an immediate consequence of Corollary 4.16 from Section 4.3.

4.1 ANN approximation results with general activation functions

Theorem 4.1. *Let $L, \kappa, \alpha_0, \alpha_1, \beta_0, \beta_1, T \in (0,\infty)$, $p, r \in \mathbb{N}$, $\mathfrak{q} \in [2,\infty)$, $a \in C(\mathbb{R},\mathbb{R})$, for every $d \in \mathbb{N}_0$ let $f_d \in C(\mathbb{R}^{\max\{d,1\}}, \mathbb{R})$, for every $d \in \mathbb{N}$ let $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure with*

$$\textstyle\int_{\mathbb{R}^d} \|y\|^{p^2\mathfrak{q}}\, \nu_d(dy) \leq \kappa d^{rp^2\mathfrak{q}}, \tag{4.1}$$

let $\mathfrak{J} \in \mathbf{N}$ satisfy $\mathcal{H}(\mathfrak{J}) = 1$ and $\mathcal{R}_a(\mathfrak{J}) = \mathrm{id}_\mathbb{R}$, let $(\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}_0\times(0,1]} \subseteq \mathbf{N}$ satisfy for all $d \in \mathbb{N}_0$, $x \in \mathbb{R}^{\max\{d,1\}}$, $\varepsilon \in (0,1]$ that

$$\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}) \in C(\mathbb{R}^{\max\{d,1\}}, \mathbb{R}), \quad \varepsilon^{\alpha_{\min\{d,1\}}}\mathcal{L}(\mathbf{F}_{d,\varepsilon}) + \varepsilon^{\beta_{\min\{d,1\}}}|||\mathcal{D}(\mathbf{F}_{d,\varepsilon})||| \leq \kappa(\max\{d,1\})^p, \tag{4.2}$$

$$\text{and} \quad \varepsilon|(\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}))(x)| + |f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}))(x)| \leq \varepsilon\kappa(\max\{d,1\})^p(1 + \|x\|)^p, \tag{4.3}$$

and assume for all $v, w \in \mathbb{R}$, $\varepsilon \in (0,1]$ that

$$\max\{|f_0(v) - f_0(w)|, |(\mathcal{R}_a(\mathbf{F}_{0,\varepsilon}))(v) - (\mathcal{R}_a(\mathbf{F}_{0,\varepsilon}))(w)|\} \leq L|v - w| \tag{4.4}$$

(cf. Definitions 2.1, 2.3, and 2.6). Then

(i) *for every $d \in \mathbb{N}$ there exists a unique at most polynomially growing viscosity solution $u_d \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ of*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) + \tfrac{1}{2}(\Delta_x u_d)(t,x) + f_0\big(u_d(t,x)\big) = 0 \tag{4.5}$$

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

(ii) *there exist $(\mathbf{U}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ and $\eta\colon (0,\infty) \to \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ it holds that*

$$\begin{aligned}\mathcal{P}(\mathbf{U}_{d,\varepsilon}) \le \eta(\delta) d^{3p+2(2+\delta)(2rp+p+2)p+(rp+p+1)p(\max\{\alpha_0,\alpha_1\}+2\max\{\beta_0,\beta_1\})} \\ \cdot \varepsilon^{-(2(2+\delta)+\max\{\alpha_0,\alpha_1\}+2\max\{\beta_0,\beta_1\})},\end{aligned} \tag{4.6}$$

$$\mathcal{R}_a(\mathbf{U}_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R}), \quad \text{and} \quad \left[\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}_a(\mathbf{U}_{d,\varepsilon}))(x)|^{\mathfrak{q}} \nu_d(dx)\right]^{1/\mathfrak{q}} \le \varepsilon. \tag{4.7}$$

Proof of Theorem 4.1. Throughout this proof let $\mathfrak{B} \in [1,\infty)$, $(m_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ satisfy for all $k \in \mathbb{N}$ that $\liminf_{j\to\infty} m_j = \infty$, $\limsup_{j\to\infty} {}^{(m_j)^{\mathfrak{q}/2}}/_{j} < \infty$, and $m_{k+1} \le \mathfrak{B} m_k$, let $\mathfrak{d} \in \mathbb{N}$ satisfy $\mathcal{D}(\mathfrak{J}) = (1, \mathfrak{d}, 1)$, let $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathfrak{u}^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, be i.i.d. random variables, assume for all $t \in (0,1)$ that $\mathbb{P}(\mathfrak{u}^0 \le t) = t$, let $\mathcal{U}^\theta\colon [0,T]\times\Omega \to [0,T]$, $\theta \in \Theta$, satisfy for all $t \in [0,T]$, $\theta \in \Theta$ that $\mathcal{U}^\theta_t = t + (T-t)\mathfrak{u}^\theta$, let $W^{d,\theta}\colon [0,T]\times\Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, $\theta \in \Theta$, be independent standard Brownian motions, assume for every $d \in \mathbb{N}$ that $(\mathcal{U}^\theta)_{\theta\in\Theta}$ and $(W^{d,\theta})_{\theta\in\Theta}$ are independent, let $U^{d,\theta}_{n,j,\varepsilon}\colon [0,T]\times\mathbb{R}^d\times\Omega \to \mathbb{R}$, $d, j, n \in \mathbb{Z}$, $\theta \in \Theta$, $\varepsilon \in (0,1]$, satisfy for all $\varepsilon \in (0,1]$, $n \in \mathbb{N}_0$, $d, j \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} U^{d,\theta}_{n,j,\varepsilon}(t,x) = {} & \frac{\mathbb{1}_{\mathbb{N}}(n)}{(m_j)^n}\left[\sum_{k=1}^{(m_j)^n} \big(\mathcal{R}_a(\mathbf{F}_{d,\varepsilon})\big)\big(x + W^{d,(\theta,0,-k)}_{T-t}\big)\right] \\ & + \sum_{i=0}^{n-1} \frac{(T-t)}{(m_j)^{n-i}}\left[\sum_{k=1}^{(m_j)^{n-i}} \Big[\big((\mathcal{R}_a(\mathbf{F}_{0,\varepsilon})\big)\big(U^{d,(\theta,i,k)}_{i,j,\varepsilon}\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{d,(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big)\big) \right. \\ & \left. - \mathbb{1}_{\mathbb{N}}(i)\big((\mathcal{R}_a(\mathbf{F}_{0,\varepsilon})\big)\big(U^{d,(\theta,-i,k)}_{\max\{i-1,0\},j,\varepsilon}\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{d,(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big)\big)\Big]\right], \end{aligned} \tag{4.8}$$

let $\mathbf{U}^{d,\theta,\varepsilon}_{n,j,t}\colon \Omega \to \mathbf{N}$, $d, j, n \in \mathbb{Z}$, $\theta \in \Theta$, $t \in [0,T]$, $\varepsilon \in (0,1]$, satisfy for all $\varepsilon \in (0,1]$, $\theta \in \Theta$, $d, j, n \in \mathbb{N}$, $t \in [0,T]$, $\omega \in \Omega$ that $\mathbf{U}^{d,\theta,\varepsilon}_{0,j,t}(\omega) = ((0\ 0\ \ldots\ 0), 0) \in \mathbb{R}^{1\times d} \times \mathbb{R}^1$ and

$$\begin{aligned} \mathbf{U}^{d,\theta,\varepsilon}_{n,j,t} = {} & \left[\bigoplus_{k=1}^{(m_j)^n} \Big(\tfrac{1}{(m_j)^n} \circledast \big(\mathbf{F}_{d,\varepsilon} \bullet \mathbf{A}_{\mathrm{I}_d, W^{d,(\theta,0,-k)}_{T-t}}\big)\Big)\right] \\ & \boxplus_{\mathfrak{J}} \left[\mathop{\boxplus}_{i=0,\mathfrak{J}}^{n-1} \left[\Big(\tfrac{(T-t)}{(m_j)^{n-i}}\Big) \circledast \left(\mathop{\boxplus}_{k=1,\mathfrak{J}}^{(m_j)^{n-i}} \Big(\big(\mathbf{F}_{0,\varepsilon} \bullet \mathbf{U}^{d,(\theta,i,k),\varepsilon}_{i,j,\mathcal{U}^{(\theta,i,k)}_t}\big) \bullet \mathbf{A}_{\mathrm{I}_d, W^{d,(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}}\Big)\right)\right]\right] \\ & \boxplus_{\mathfrak{J}} \left[\mathop{\boxplus}_{i=0,\mathfrak{J}}^{n-1} \left[\Big(\tfrac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{(m_j)^{n-i}}\Big) \circledast \left(\mathop{\boxplus}_{k=1,\mathfrak{J}}^{(m_j)^{n-i}} \Big(\big(\mathbf{F}_{0,\varepsilon} \bullet \mathbf{U}^{d,(\theta,-i,k),\varepsilon}_{\max\{i-1,0\},j,\mathcal{U}^{(\theta,i,k)}_t}\big) \bullet \mathbf{A}_{\mathrm{I}_d, W^{d,(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}}\Big)\right)\right]\right] \end{aligned} \tag{4.9}$$

(cf. Proposition 3.9), assume without loss of generality that $\max\{|f_0(0)|, \mathfrak{d}, 1\} \leq \kappa$, let $c_d, \mathfrak{c}_d \in [1,\infty)$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$ that

$$c_d = \kappa 2^{p-1} d^p \big(e^{LT}(T+1)\big)^{p+1}((2\kappa d^p)^p + 1) \cdot \left(1 + \left(\int_{\mathbb{R}^d} \|x\|^{p^2\mathfrak{q}}\, \nu_d(\mathrm{d}x)\right)^{1/p^2\mathfrak{q}} + \big(\mathbb{E}\big[\|W_T^{d,0}\|^{p^2}\big]\big)^{1/p^2}\right)^{p^2} \tag{4.10}$$

and

$$\mathfrak{c}_d = \kappa 2^{2(p+1)} d^p e^{LT}(T+1)(\sqrt{\mathfrak{q}} - 1)\left(1 + \left(\int_{\mathbb{R}^d} \|x\|^{p\mathfrak{q}}\, \nu_d(\mathrm{d}x)\right)^{\frac{1}{\mathfrak{q}}} + \left(\sup_{s\in[0,T]} \mathbb{E}\big[\|W_s^{d,0}\|^{p\mathfrak{q}}\big]\right)^{\frac{1}{\mathfrak{q}}}\right), \tag{4.11}$$

let $\mathcal{n}\colon \mathbb{N}\times(0,1] \to [1,\infty]$ satisfy for all $d \in \mathbb{N}$, $\delta \in (0,1]$ that

$$\mathcal{n}(d,\delta) = \inf\left(\left\{ n \in \mathbb{N}\colon \mathfrak{c}_d \left(\frac{(1+2LT)\exp\left(\frac{(m_n)^{\frac{\mathfrak{q}}{2}}}{n}\right)}{(m_n)^{\frac{1}{2}}}\right)^n \leq \delta \right\} \cup \{\infty\}\right), \tag{4.12}$$

let κ_δ, $\delta \in (0,\infty)$, satisfy for all $\delta \in (0,\infty)$ that $\kappa_\delta = 2(2+\delta) + \max\{\alpha_0, \alpha_1\} + 2\max\{\beta_0, \beta_1\}$, 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 $\delta_{d,\varepsilon} = \varepsilon/(c_d+1)$ (cf. Definitions 2.7, 2.8, 2.12, 2.16, 2.17, and 2.18). Note that the assumption that for all $w, z \in \mathbb{R}$, $\varepsilon \in (0,1]$ it holds that $|(\mathcal{R}_a(\mathbf{F}_{0,\varepsilon}))(z) - (\mathcal{R}_a(\mathbf{F}_{0,\varepsilon}))(w)| \leq L|z-w|$, the assumption that for all $d \in \mathbb{N}_0$, $x \in \mathbb{R}^{\max\{d,1\}}$, $\varepsilon \in (0,1]$ it holds that $|(\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}))(x)| \leq \kappa(\max\{d,1\})^p(1+\|x\|)^p$, and Beck et al. [6, Corollary 3.10] (applied for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ with $d \curvearrowleft d$, $m \curvearrowleft d$, $L \curvearrowleft L$, $T \curvearrowleft T$, $\mu \curvearrowleft (\mathbb{R}^d \ni x \mapsto (0,0,\ldots,0) \in \mathbb{R}^d)$, $\sigma \curvearrowleft \mathrm{I}_d$, $f \curvearrowleft ([0,T]\times\mathbb{R}^d\times\mathbb{R} \ni (t,x,w) \mapsto (\mathcal{R}_a(\mathbf{F}_{0,\varepsilon}))(w) \in \mathbb{R})$, $g \curvearrowleft \mathcal{R}_a(\mathbf{F}_{d,\varepsilon})$, $W \curvearrowleft W^{d,0}$ in the notation of Beck et al. [6, Corollary 3.10]) yield that for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists a unique at most polynomially growing $v_{d,\varepsilon} \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ such that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$v_{d,\varepsilon}(t,x) = \mathbb{E}\Big[\big(\mathcal{R}_a(\mathbf{F}_{d,\varepsilon})\big)(x + W_{T-t}^{d,0})\Big] + \int_t^T \mathbb{E}\Big[\big(\mathcal{R}_a(\mathbf{F}_{0,\varepsilon})\big)(v_{d,\varepsilon}(s, x + W_{s-t}^{d,0}))\Big]\,\mathrm{d}s. \tag{4.13}$$

Next, observe that the triangle inequality and the assumption that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ it holds that $\varepsilon|(\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}))(x)| + |f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}))(x)| \leq \varepsilon\kappa d^p(1+\|x\|)^p$ establish that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that

$$|f_d(x)| \leq |f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,1}))(x)| + |(\mathcal{R}_a(\mathbf{F}_{d,1}))(x)| \leq \kappa d^p(1+\|x\|)^p. \tag{4.14}$$

Combining this, the assumption that for all $w, z \in \mathbb{R}$ it holds that $|f_0(w) - f_0(z)| \leq L|w-z|$, and Beck et al. [9, Theorem 1.1] (applied for every $d \in \mathbb{N}$ with $d \curvearrowleft d$, $m \curvearrowleft d$, $L \curvearrowleft L$, $T \curvearrowleft T$, $\mu \curvearrowleft (\mathbb{R}^d \ni x \mapsto (0,0,\ldots,0) \in \mathbb{R}^d)$, $\sigma \curvearrowleft \mathrm{I}_d$, $f \curvearrowleft (\mathbb{R}^d\times\mathbb{R} \ni (x,w) \mapsto f_0(w) \in \mathbb{R})$, $g \curvearrowleft f_d$, $W \curvearrowleft W^{d,0}$ in the notation of Beck et al. [9, Theorem 1.1]) establishes item (i). Furthermore, note that the fact that for all $d \in \mathbb{N}$, $s \in (0,T]$ the random variable $\|W_s^{d,0}/\sqrt{s}\|^2$

is a chi-squared distributed random variable with d-degrees of freedom, Jensen's inequality, and, e.g., Simon [62, Eq. (2.35)] imply that for all $\gamma, d \in \mathbb{N}$, $s \in [0,T]$ it holds that

$$\Big(\mathbb{E}\big[\|W_s^{d,0}\|^{\gamma}\big]\Big)^2 \leq \mathbb{E}\big[\|W_s^{d,0}\|^{2\gamma}\big] = (2s)^{\gamma}\left[\frac{\Gamma(\frac{d}{2}+\gamma)}{\Gamma(\frac{d}{2})}\right] = (2s)^{\gamma}\left[\prod_{k=0}^{\gamma-1}\big(\tfrac{d}{2}+k\big)\right]. \tag{4.15}$$

This ensures that for all $d \in \mathbb{N}$ it holds that

$$\Big(\mathbb{E}\big[\|W_T^{d,0}\|^{p^2}\big]\Big)^{1/(p^2)} \leq \sqrt{2T}\left[\prod_{k=0}^{p^2-1}\big(\tfrac{d}{2}+k\big)\right]^{1/(2p^2)} \leq \sqrt{2T\big(\tfrac{d}{2}+p^2-1\big)}. \tag{4.16}$$

Combining this and the assumption that for all $d \in \mathbb{N}$ it holds that $\int_{\mathbb{R}^d}\|x\|^{p^2\mathfrak{q}}\,\nu_d(\mathrm{d}x) \leq \kappa d^{rp^2\mathfrak{q}}$ assures that there exists $\overline{C} \in [1,\infty)$ such that for all $d \in \mathbb{N}$ it holds that

$$c_d \leq \overline{C} d^{p+(r+1)p^2}. \tag{4.17}$$

Moreover, observe that the triangle inequality proves that for all $n \in \mathbb{N}_0$, $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
&\bigg(\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|u_d(0,x) - U_{n,j,\delta}^{d,0}(0,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\bigg)^{1/\mathfrak{q}} \\
&\leq \bigg(\int_{\mathbb{R}^d}\big|u_d(0,x) - v_{d,\delta}(0,x)\big|^{\mathfrak{q}}\,\nu_d(\mathrm{d}x)\bigg)^{1/\mathfrak{q}} + \bigg(\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|v_{d,\delta}(0,x) - U_{n,j,\delta}^{d,0}(0,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\bigg)^{1/\mathfrak{q}}.
\end{aligned} \tag{4.18}$$

In addition, note that the assumption that for all $x \in \mathbb{R}$, $\delta \in (0,1]$ it holds that $|f_0(x) - (\mathcal{R}_a(\mathbf{F}_{0,\delta}))(x)| \leq \delta\kappa(1+|x|)^p$ shows that for all $\delta \in (0,1]$ it holds that

$$|(\mathcal{R}_a(\mathbf{F}_{0,\delta}))(0)| \leq |(\mathcal{R}_a(\mathbf{F}_{0,\delta}))(0) - f_0(0)| + |f_0(0)| \leq \delta\kappa + |f_0(0)| \leq \kappa + |f_0(0)| \leq 2\kappa. \tag{4.19}$$

Next, observe that the assumption that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $w \in \mathbb{R}$, $\delta \in (0,1]$ it holds that $|f_0(w) - (\mathcal{R}_a(\mathbf{F}_{0,\delta}))(w)| \leq \delta\kappa(1+|w|)^p$ and $|f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,\delta}))(x)| \leq \delta\kappa d^p(1+\|x\|)^p$ demonstrates that for all $d \in \mathbb{N}$, $w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}
&\max\big\{|f_0(w) - (\mathcal{R}_a(\mathbf{F}_{0,\delta}))(w)|, |f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,\delta}))(x)|\big\} \\
&\leq \max\big\{\delta\kappa(1+|w|)^p, \delta\kappa d^p(1+\|x\|)^p\big\} \leq \delta\kappa d^p \max\big\{(1+\|x\|+|w|)^p, (1+\|x\|)^p\big\} \\
&\leq \delta\kappa d^p(1+\|x\|+|w|)^p \leq \delta\kappa d^p 2^{p-1}((1+\|x\|)^p + |w|^p) \leq \delta\kappa d^p 2^{p-1}((1+\|x\|)^{p^2} + |w|^p).
\end{aligned} \tag{4.20}$$

Combining this, (4.14), (4.19), the assumption that for all $d \in \mathbb{N}$, $w, z \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that $\max\{|f_0(w) - f_0(z)|, |(\mathcal{R}_a(\mathbf{F}_{0,\delta}))(w) - (\mathcal{R}_a(\mathbf{F}_{0,\delta}))(z)|\} \leq L|w-z|$ and $\delta|(\mathcal{R}_a(\mathbf{F}_{d,\delta}))(x)| + |f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,\delta}))(x)| \leq \delta\kappa d^p(1+\|x\|)^p$, and Hutzenthaler et al. [46, Lemma 2.3] (applied for every $d \in \mathbb{N}$, $\delta \in (0,1]$ with $f_1 \curvearrowleft f_0$, $f_2 \curvearrowleft \mathcal{R}_a(\mathbf{F}_{0,\delta})$, $g_1 \curvearrowleft f_d$, $g_2 \curvearrowleft \mathcal{R}_a(\mathbf{F}_{d,\delta})$, $T \curvearrowleft T$, $L \curvearrowleft L$, $B \curvearrowleft 2\kappa d^p$, $\delta \curvearrowleft \delta\kappa 2^{p-1}d^p$, $\mathbf{W} \curvearrowleft W^{d,0}$, $u_1 \curvearrowleft u_d$, $u_2 \curvearrowleft v_{d,\delta}$, $p \curvearrowleft p$, $q \curvearrowleft p$ in the notation of Hutzenthaler et al. [46, Lemma 2.3]) yield that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\bigg(\int_{\mathbb{R}^d}|u_d(0,x) - v_{d,\delta}(0,x)|^{\mathfrak{q}}\,\nu_d(\mathrm{d}x)\bigg)^{1/\mathfrak{q}} \leq \delta\kappa 2^{p-1}d^p\big(e^{LT}(T+1)\big)^{p+1}((2\kappa d^p)^p + 1) \tag{4.21}$$

$$\cdot \left(\int_{\mathbb{R}^d} \left(1 + \|x\| + \left(\mathbb{E}\big[\|W_T^{d,0}\|^{p^2} \big] \right)^{1/p^2} \right)^{p^2 \mathfrak{q}} \nu_d(\mathrm{d}x) \right)^{1/\mathfrak{q}}.$$

Combining (4.21) and the triangle inequality therefore establishes that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\left(\int_{\mathbb{R}^d} |u_d(0,x) - v_{d,\delta}(0,x)|^{\mathfrak{q}} \, \nu_d(\mathrm{d}x) \right)^{1/\mathfrak{q}} \leq c_d \delta. \tag{4.22}$$

Furthermore, note that (4.19) and the assumption that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that $\delta |(\mathcal{R}_a(\mathbf{F}_{d,\delta}))(x)| + |f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,\delta}))(x)| \leq \delta \kappa d^p (1 + \|x\|)^p$ imply that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\max\big\{ |(\mathcal{R}_a(\mathbf{F}_{0,\delta}))(0)|, |(\mathcal{R}_a(\mathbf{F}_{d,\delta}))(x)| \big\} \leq \max\big\{ 2\kappa, \kappa d^p (1 + \|x\|)^p \big\} \leq 2\kappa d^p (1 + \|x\|)^p. \tag{4.23}$$

This, the assumption that for all $v, w \in \mathbb{R}$, $\delta \in (0,1]$ it holds that $|(\mathcal{R}_a(\mathbf{F}_{0,\delta}))(v) - (\mathcal{R}_a(\mathbf{F}_{0,\delta}))(w)| \leq L|v - w|$, and Hutzenthaler et al. [48, Corollary 3.15] (applied for every $\delta \in (0,1]$, $d, j \in \mathbb{N}$ with $T \curvearrowleft T$, $L \curvearrowleft L$, $\mathfrak{L} \curvearrowleft 2^p \kappa d^p$, $p \curvearrowleft p$, $\mathfrak{p} \curvearrowleft \mathfrak{q}$, $m \curvearrowleft m_j$, $f \curvearrowleft ([0,T] \times \mathbb{R}^d \times \mathbb{R} \ni (t,x,v) \mapsto (\mathcal{R}_a(\mathbf{F}_{0,\delta}))(v) \in \mathbb{R})$, $g \curvearrowleft \mathcal{R}_a(\mathbf{F}_{d,\delta})$, $\Theta \curvearrowleft \Theta$, $(\mathfrak{u}^\theta)_{\theta \in \Theta} \curvearrowleft (\mathfrak{u}^\theta)_{\theta \in \Theta}$, $(\mathcal{U}^\theta)_{\theta \in \Theta} \curvearrowleft (\mathcal{U}^\theta)_{\theta \in \Theta}$, $(W^\theta)_{\theta \in \Theta} \curvearrowleft (W^{d,\theta})_{\theta \in \Theta}$, $u \curvearrowleft v_{d,\delta}$, $(U_n^\theta)_{(n,\theta) \in \mathbb{Z} \times \Theta} \curvearrowleft (U_{n,j,\delta}^{d,\theta})_{(n,\theta) \in \mathbb{Z} \times \Theta}$ in the notation of Hutzenthaler et al. [48, Corollary 3.15]) assure that for all $\delta \in (0,1]$, $d, j \in \mathbb{N}$, $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} &\mathbb{E}\Big[\big| v_{d,\delta}(0,x) - U_{n,j,\delta}^{d,0}(0,x) \big|^{\mathfrak{q}} \Big] \\ &\leq \left(\frac{2(\sqrt{\mathfrak{q}-1}) 2^p \kappa d^p (T+1) \exp(LT)(1+2LT)^n}{(m_j)^{n/2} \exp(-(m_j)^{\mathfrak{q}/2}/\mathfrak{q})} \sup_{s \in [0,T]} \big(\mathbb{E}\big[(1 + \|x + W_s^{d,0}\|^p)^{\mathfrak{q}} \big] \big)^{1/\mathfrak{q}} \right)^{\mathfrak{q}}. \end{aligned} \tag{4.24}$$

It follows from (4.24) for all $\delta \in (0,1]$, $d, j \in \mathbb{N}$, $n \in \mathbb{N}_0$ that

$$\begin{aligned} &\left(\int_{\mathbb{R}^d} \mathbb{E}\Big[\big| v_{d,\delta}(0,x) - U_{n,j,\delta}^{d,0}(0,x) \big|^{\mathfrak{q}} \Big] \, \nu_d(\mathrm{d}x) \right)^{1/\mathfrak{q}} \\ &\leq \frac{(\sqrt{\mathfrak{q}-1}) 2^{p+1} \kappa d^p (T+1) \exp(LT)(1+2LT)^n}{(m_j)^{n/2} \exp(-(m_j)^{\mathfrak{q}/2}/\mathfrak{q})} \left(\int_{\mathbb{R}^d} \sup_{s \in [0,T]} \mathbb{E}\big[(1 + \|x + W_s^{d,0}\|^p)^{\mathfrak{q}} \big] \, \nu_d(\mathrm{d}x) \right)^{1/\mathfrak{q}}. \end{aligned} \tag{4.25}$$

Moreover, note that Jensen's inequality and the triangle inequality ensure for all $x \in \mathbb{R}$, $s \in [0,T]$, $d \in \mathbb{N}$ that $(1 + \|x + W_s^{d,0}\|^p)^{\mathfrak{q}} \leq 2^{\mathfrak{q}-2+p\mathfrak{q}} \big(1 + \|x\|^{p\mathfrak{q}} + \|W_s^{d,0}\|^{p\mathfrak{q}} \big)$. This and the assumption that for all $d \in \mathbb{N}$ it holds that ν_d is a probability measure imply for all $d \in \mathbb{N}$ that

$$\int_{\mathbb{R}^d} \sup_{s \in [0,T]} \mathbb{E}\big[(1 + \|x + W_s^{d,0}\|^p)^{\mathfrak{q}} \big] \, \nu_d(\mathrm{d}x) \leq 2^{(p+1)\mathfrak{q}} \left(1 + \int_{\mathbb{R}^d} \|x\|^{p\mathfrak{q}} \, \nu_d(\mathrm{d}x) + \sup_{s \in [0,T]} \mathbb{E}\big[\|W_s^{d,0}\|^{p\mathfrak{q}} \big] \right). \tag{4.26}$$

This, (4.25), and (4.11) ensure for all $\delta \in (0,1]$, $d, j \in \mathbb{N}$, $n \in \mathbb{N}_0$ that

$$\left(\int_{\mathbb{R}^d} \mathbb{E}\Big[\big|v_{d,\delta}(0,x) - U^{d,0}_{n,j,\delta}(0,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} \le \mathfrak{c}_d \left(\frac{(1+2LT)\exp\left(\frac{(m_j)^{\frac{\mathfrak{q}}{2}}}{n}\right)}{(m_j)^{1/2}}\right)^n. \tag{4.27}$$

Moreover, note that (4.15) assures for all $s \in [0,T]$, $d \in \mathbb{N}$ that

$$\mathbb{E}\big[\|W^{d,0}_s\|^{p\mathfrak{q}}\big] \le 1 + (2T+1)^{\frac{p(\mathfrak{q}+1)}{2}} \left(\frac{d}{2} + p(\mathfrak{q}+1) - 1\right)^{\frac{p(\mathfrak{q}+1)}{2}}. \tag{4.28}$$

It follows for all $d \in \mathbb{N}$ that

$$\left(\sup_{s\in[0,T]} \mathbb{E}\big[\|W^{d,0}_s\|^{p\mathfrak{q}}\big]\right)^{\frac{1}{\mathfrak{q}}} \le 1 + (2T+1)^p \left(\frac{d}{2} + p(\mathfrak{q}+1) - 1\right)^p. \tag{4.29}$$

Furthermore, the assumptions that for all $d \in \mathbb{N}$ it holds that $(\int_{\mathbb{R}^d} \|y\|^{p^2\mathfrak{q}}\,\nu_d(\mathrm{d}y)) \le \kappa d^{rp^2\mathfrak{q}}$ and that ν_d is a probability measure imply for all $d \in \mathbb{N}$ that

$$\left(\int_{\mathbb{R}^d} \|x\|^{p\mathfrak{q}}\,\nu_d(\mathrm{d}x)\right)^{\frac{1}{\mathfrak{q}}} \le \left(1 + \int_{\mathbb{R}^d} \|x\|^{p^2\mathfrak{q}}\,\nu_d(\mathrm{d}x)\right)^{\frac{1}{\mathfrak{q}}} \le 1 + \kappa d^{rp^2}. \tag{4.30}$$

Combining this, (4.29), and (4.11) proves that there exists $\overline{\mathfrak{C}} \in [1,\infty)$ such that for all $d \in \mathbb{N}$ it holds that

$$\mathfrak{c}_d \le \overline{\mathfrak{C}} d^{rp^2+p}. \tag{4.31}$$

The fact that $\limsup_{n\to\infty} {(m_n)^{\mathfrak{q}/2}}/{n} < \infty$ and the fact that $\liminf_{n\to\infty} m_n = \infty$ imply that

$$\limsup_{n\to\infty} \left(\frac{(1+2LT)\exp\left(\frac{(m_n)^{\frac{\mathfrak{q}}{2}}}{n}\right)}{(m_n)^{1/2}}\right)^n = 0. \tag{4.32}$$

This, (4.31), and (4.12) show that it holds for all $d \in \mathbb{N}$, $\delta \in (0,1]$ that $\mathcal{n}(d,\delta) < \infty$. It thus follows from (4.12) and (4.27) for all $\delta \in (0,1]$, $d \in \mathbb{N}$ that

$$\left(\int_{\mathbb{R}^d} \mathbb{E}\Big[\big|v_{d,\delta}(0,x) - U^{d,0}_{\mathcal{n}(d,\delta),\mathcal{n}(d,\delta),\delta}(0,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} \le \mathfrak{c}_d \left(\frac{(1+2LT)\exp\left(\frac{(m_{\mathcal{n}(d,\delta)})^{\frac{\mathfrak{q}}{2}}}{\mathcal{n}(d,\delta)}\right)}{(m_{\mathcal{n}(d,\delta)})^{1/2}}\right)^{\mathcal{n}(d,\delta)} \le \delta. \tag{4.33}$$

Combining this, (4.18), and (4.22) hence ensures that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\left(\int_{\mathbb{R}^d} \mathbb{E}\Big[\big|u_d(0,x) - U^{d,0}_{\mathcal{n}(d,\delta),\mathcal{n}(d,\delta),\delta}(0,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} \le c_d\delta + \delta. \tag{4.34}$$

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

$$\begin{aligned}&\mathbb{E}\bigg[\int_{\mathbb{R}^d}\Big|u_d(0,x) - U^{d,0}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),\delta_{d,\varepsilon}}(0,x)\Big|^{\mathfrak{q}}\,\nu_d(\mathrm{d}x)\bigg]\\&= \int_{\mathbb{R}^d}\mathbb{E}\bigg[\Big|u_d(0,x) - U^{d,0}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),\delta_{d,\varepsilon}}(0,x)\Big|^{\mathfrak{q}}\bigg]\,\nu_d(\mathrm{d}x) \leq \big(\mathfrak{c}_d\delta_{d,\varepsilon} + \delta_{d,\varepsilon}\big)^{\mathfrak{q}} = \varepsilon^{\mathfrak{q}}.\end{aligned}\tag{4.35}$$

Hence, there exists $(\omega_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \Omega$, which is assumed to be fixed for the remainder of this proof, such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\int_{\mathbb{R}^d}\Big|u_d(0,x) - U^{d,0}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),\delta_{d,\varepsilon}}(0,x,\omega_{d,\varepsilon})\Big|^{\mathfrak{q}}\,\nu_d(\mathrm{d}x) \leq \varepsilon^{\mathfrak{q}}.\tag{4.36}$$

Furthermore, note that (4.9) and items (ii) and (v) in Proposition 3.9 (applied for every $d, j \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\omega \in \Omega$ with $\Theta \curvearrowleft \Theta$, $d \curvearrowleft d$, $M \curvearrowleft m_j$, $\mathbf{F} \curvearrowleft \mathbf{F}_{0,\varepsilon}$, $\mathbf{G} \curvearrowleft \mathbf{F}_{d,\varepsilon}$, $(\mathcal{U}^\theta)_{\theta\in\Theta} \curvearrowleft (\mathcal{U}^\theta(\omega))_{\theta\in\Theta}$, $(W^\theta)_{\theta\in\Theta} \curvearrowleft (W^\theta(\omega))_{\theta\in\Theta}$, $(U^\theta_n)_{(n,\theta)\in\mathbb{Z}\times\Theta} \curvearrowleft (U^{d,\theta}_{n,j,\varepsilon}(\omega))_{(n,\theta)\in\mathbb{Z}\times\Theta}$, $(\mathbf{U}^\theta_{n,t})_{(n,t,\theta)\in\mathbb{Z}\times[0,T]\times\Theta} \curvearrowleft (\mathbf{U}^{d,\theta,\varepsilon}_{n,j,t}(\omega))_{(n,t,\theta)\in\mathbb{Z}\times[0,T]\times\Theta}$ in the notation of Proposition 3.9) prove that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x \in \mathbb{R}^d$ it holds that $(\mathcal{R}_a(\mathbf{U}^{d,0,\delta_{d,\varepsilon}}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),0}(\omega_{d,\varepsilon}))) \in C(\mathbb{R}^d,\mathbb{R})$ and

$$\Big(\mathcal{R}_a\Big(\mathbf{U}^{d,0,\delta_{d,\varepsilon}}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),0}(\omega_{d,\varepsilon})\Big)\Big)(x) = U^{d,0}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),\delta_{d,\varepsilon}}(0,x,\omega_{d,\varepsilon}).\tag{4.37}$$

Combining this and (4.36) establishes (4.7). Moreover, observe that item (vi) in Proposition 3.9 implies that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}&\mathcal{P}\big(\mathbf{U}^{d,0,\delta_{d,\varepsilon}}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),0}(\omega_{d,\varepsilon})\big)\\&\leq 2\big(\max\{\mathfrak{d},\mathcal{L}(\mathbf{F}_{d,\delta_{d,\varepsilon}})\} + \mathcal{n}(d,\delta_{d,\varepsilon})\mathcal{H}(\mathbf{F}_{0,\delta_{d,\varepsilon}})\big)\big(\max\{\mathfrak{d},|||\mathcal{D}(\mathbf{F}_{0,\delta_{d,\varepsilon}})|||,|||\mathcal{D}(\mathbf{F}_{d,\delta_{d,\varepsilon}})|||\}\big)^2\\&\quad\cdot(3m_{\mathcal{n}(d,\delta_{d,\varepsilon})})^{2\mathcal{n}(d,\delta_{d,\varepsilon})}\\&\leq 2\big(\max\{\mathfrak{d},\mathcal{L}(\mathbf{F}_{d,\delta_{d,\varepsilon}})\} + \mathcal{H}(\mathbf{F}_{0,\delta_{d,\varepsilon}})\big)\big(\max\{\mathfrak{d},|||\mathcal{D}(\mathbf{F}_{0,\delta_{d,\varepsilon}})|||,|||\mathcal{D}(\mathbf{F}_{d,\delta_{d,\varepsilon}})|||\}\big)^2\\&\quad\cdot\Big[\big(\mathcal{n}(d,\delta_{d,\varepsilon})\big)^{1/2}\big(3m_{\mathcal{n}(d,\delta_{d,\varepsilon})}\big)^{\mathcal{n}(d,\delta_{d,\varepsilon})}\Big]^2.\end{aligned}\tag{4.38}$$

In addition, note that the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{L}(\mathbf{F}_{0,\varepsilon}) \leq \kappa\varepsilon^{-\alpha_0}$, $|||\mathcal{D}(\mathbf{F}_{0,\varepsilon})||| \leq \kappa\varepsilon^{-\beta_0}$, $\mathcal{L}(\mathbf{F}_{d,\varepsilon}) \leq \kappa d^p\varepsilon^{-\alpha_1}$, and $|||\mathcal{D}(\mathbf{F}_{d,\varepsilon})||| \leq \kappa d^p\varepsilon^{-\beta_1}$ ensures that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}&\big(\max\{\mathfrak{d},\mathcal{L}(\mathbf{F}_{d,\delta_{d,\varepsilon}})\} + \mathcal{H}(\mathbf{F}_{0,\delta_{d,\varepsilon}})\big)\big(\max\{\mathfrak{d},|||\mathcal{D}(\mathbf{F}_{0,\delta_{d,\varepsilon}})|||,|||\mathcal{D}(\mathbf{F}_{d,\delta_{d,\varepsilon}})|||\}\big)^2\\&\leq \big(\kappa d^p\delta_{d,\varepsilon}^{-\alpha_1} + \kappa\delta_{d,\varepsilon}^{-\alpha_0}\big)\Big(\max\{\kappa,\kappa\delta_{d,\varepsilon}^{-\beta_0},\kappa d^p\delta_{d,\varepsilon}^{-\beta_1}\}\Big)^2\\&\leq 2\kappa^3 d^{3p}\max\{\delta_{d,\varepsilon}^{-\alpha_1},\delta_{d,\varepsilon}^{-\alpha_0}\}\Big(\max\{\delta_{d,\varepsilon}^{-\beta_0},\delta_{d,\varepsilon}^{-\beta_1}\}\Big)^2\\&\leq 2\kappa^3 d^{3p}\delta_{d,\varepsilon}^{-(\max\{\alpha_0,\alpha_1\}+2\max\{\beta_0,\beta_1\})}.\end{aligned}\tag{4.39}$$

It follows from (4.12) that it holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ with $\mathcal{n}(d,\delta_{d,\varepsilon}) \in \mathbb{N}\cap[2,\infty)$ that

$$\mathfrak{c}_d\left(\frac{(1+2LT)\exp\Big(\frac{(m_{\mathcal{n}(d,\delta_{d,\varepsilon})-1})^{\frac{\mathfrak{q}}{2}}}{\mathcal{n}(d,\delta_{d,\varepsilon})-1}\Big)}{(m_{\mathcal{n}(d,\delta_{d,\varepsilon})-1})^{1/2}}\right)^{(\mathcal{n}(d,\delta_{d,\varepsilon})-1)} > \delta_{d,\varepsilon}.\tag{4.40}$$

This implies for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ with $\mathscr{n}(d,\delta_{d,\varepsilon}) \in \mathbb{N} \cap [2,\infty)$ that

$$\begin{aligned}
&\big(\mathscr{n}(d,\delta_{d,\varepsilon})\big)^{1/2}\big(3m_{\mathscr{n}(d,\delta_{d,\varepsilon})}\big)^{\mathscr{n}(d,\delta_{d,\varepsilon})} \\
&\le \big(\mathscr{n}(d,\delta_{d,\varepsilon})\big)^{1/2}\big(3m_{\mathscr{n}(d,\delta_{d,\varepsilon})}\big)^{\mathscr{n}(d,\delta_{d,\varepsilon})}\left(\mathfrak{c}_d\delta_{d,\varepsilon}^{-1}\left(\frac{(1+2LT)\exp\Big(\frac{(m_{\mathscr{n}(d,\delta_{d,\varepsilon})-1})^{\frac{\mathfrak{q}}{2}}}{\mathscr{n}(d,\delta_{d,\varepsilon})-1}\Big)}{(m_{\mathscr{n}(d,\delta_{d,\varepsilon})-1})^{1/2}}\right)^{(\mathscr{n}(d,\delta_{d,\varepsilon})-1)}\right)^{2+\delta} \\
&\le \frac{\mathfrak{c}_d^{2+\delta}}{\delta_{d,\varepsilon}^{2+\delta}}\sup_{n\in\mathbb{N}}\left(\frac{m_{n+1}^{n+1}}{m_n^{n(1+\frac{\delta}{2})}}(n+1)^{\frac{1}{2}}\left(3(1+2LT)\exp\left(\frac{m_n^{\frac{\mathfrak{q}}{2}}}{n}\right)\right)^{n(2+\delta)}\right).
\end{aligned} \tag{4.41}$$

Observe that the fact that it holds for all $n \in \mathbb{N}$ that $m_{n+1} \le \mathfrak{B}m_n$ ensures for all $n \in \mathbb{N}$ that $({m_{n+1}}/{m_n})^n \le \mathfrak{B}^n$ and $m_{n+1} \le \mathfrak{B}^n m_1$. Therefore, it holds for all $n \in \mathbb{N}$ that ${m_{n+1}^{n+1}}/{m_n^{n(1+\frac{\delta}{2})}} \le {\mathfrak{B}^{2n}m_1}/{m_n^{n\delta/2}}$. This and (4.41) yield for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ with $\mathscr{n}(d,\delta_{d,\varepsilon}) \in \mathbb{N} \cap [2,\infty)$ that

$$\big(\mathscr{n}(d,\delta_{d,\varepsilon})\big)^{1/2}\big(3m_{\mathscr{n}(d,\delta_{d,\varepsilon})}\big)^{\mathscr{n}(d,\delta_{d,\varepsilon})} \le \frac{\mathfrak{c}_d^{2+\delta}m_1}{\delta_{d,\varepsilon}^{2+\delta}}\sup_{n\in\mathbb{N}}\left(\frac{(n+1)^{\frac{1}{2n}}\left(3(1+2LT)\mathfrak{B}\exp\left(\frac{m_n^{\frac{\mathfrak{q}}{2}}}{n}\right)\right)^{2+\delta}}{m_n^{\frac{\delta}{2}}}\right)^n. \tag{4.42}$$

Observe that the fact that $m_1 \in \mathbb{N}$, the fact that $\mathfrak{B} \in [1,\infty)$, the fact that $\mathfrak{q} \in [2,\infty)$, the fact that for all $d \in \mathbb{N}$ it holds that $\mathfrak{c}_d \in [1,\infty)$, and the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\delta_{d,\varepsilon} \in (0,1]$ show for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that

$$\begin{aligned}
&\frac{\mathfrak{c}_d^{2+\delta}m_1}{\delta_{d,\varepsilon}^{2+\delta}}\sup_{n\in\mathbb{N}}\left(\frac{(n+1)^{\frac{1}{2n}}\left(3(1+2LT)\mathfrak{B}\exp\left(\frac{m_n^{\frac{\mathfrak{q}}{2}}}{n}\right)\right)^{2+\delta}}{m_n^{\frac{\delta}{2}}}\right)^n \\
&\ge m_1\cdot\frac{2^{\frac{1}{2}}\left(3(1+2LT)\mathfrak{B}\exp\left(m_1^{\frac{\mathfrak{q}}{2}}\right)\right)^{2+\delta}}{m_1^{\frac{\delta}{2}}} \ge 3m_1m_1^{\mathfrak{q}+\frac{\delta}{2}\mathfrak{q}-\frac{\delta}{2}} \ge 3m_1.
\end{aligned} \tag{4.43}$$

Combining this and the fact that it holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that $\mathscr{n}(d,\delta_{d,\varepsilon}) < \infty$ assures that (4.42) holds for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$. This, (4.38), and (4.39) ensure for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that

$$\begin{aligned}
\mathcal{P}\big(\mathbf{U}^{d,0,\delta_{d,\varepsilon}}_{\mathscr{n}(d,\delta_{d,\varepsilon}),\mathscr{n}(d,\delta_{d,\varepsilon}),0}(\omega_{d,\varepsilon})\big) &\le 4\kappa^3 d^{3p}\delta_{d,\varepsilon}^{-(2(2+\delta)+\max\{\alpha_0,\alpha_1\}+2\max\{\beta_0,\beta_1\})}\mathfrak{c}_d^{2(2+\delta)}m_1^2 \\
&\quad\cdot\sup_{n\in\mathbb{N}}\left(\frac{(n+1)^{\frac{1}{2n}}\left(3(1+2LT)\mathfrak{B}\exp\left(\frac{m_n^{\frac{\mathfrak{q}}{2}}}{n}\right)\right)^{2+\delta}}{m_n^{\frac{\delta}{2}}}\right)^{2n}.
\end{aligned} \tag{4.44}$$

Note that (4.17) establishes for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ that

$$\delta_{d,\varepsilon}^{-(2(2+\delta)+\max\{\alpha_0,\alpha_1\}+2\max\{\beta_0,\beta_1\})} = \left(\frac{c_d+1}{\varepsilon}\right)^{\kappa_\delta} \leq (2\overline{C})^{\kappa_\delta}\varepsilon^{-\kappa_\delta} d^{(p+(r+1)p^2)\kappa_\delta}. \tag{4.45}$$

Moreover, (4.31) proves for all $d \in \mathbb{N}$, $\delta \in (0,\infty)$ that

$$\mathfrak{c}_d^{2(2+\delta)} \leq \overline{\mathfrak{C}}^{2(2+\delta)} d^{2(2+\delta)(rp^2+p)}. \tag{4.46}$$

Furthermore, the fact that $\limsup_{n\to\infty} {}^{(m_n)^{\mathfrak{q}/2}}\!/_{n} < \infty$, the fact that $\liminf_{n\to\infty} m_n = \infty$, and the fact that $\lim_{n\to\infty}(n+1)^{\frac{1}{2n}} = 1$ yield for all $\delta \in (0,\infty)$ that

$$\sup_{n\in\mathbb{N}} \left(\frac{(n+1)^{\frac{1}{2n}}\left(3(1+2LT)\mathfrak{B}\exp\left(\frac{m_n^{\frac{\mathfrak{q}}{2}}}{n}\right)\right)^{2+\delta}}{m_n^{\frac{\delta}{2}}} \right)^{2n} < \infty. \tag{4.47}$$

Combining (4.44), (4.45), (4.46), and (4.47) shows that there exists $\eta\colon (0,\infty) \to \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ it holds that

$$\mathcal{P}\big(\mathbf{U}^{d,0,\delta_{d,\varepsilon}}_{n(d,\delta_{d,\varepsilon}),n(d,\delta_{d,\varepsilon}),0}(\omega_{d,\varepsilon})\big) \leq \eta(\delta) d^{3p+2(2+\delta)(rp^2+p)+(p+(r+1)p^2)\kappa_\delta}\varepsilon^{-\kappa_\delta}. \tag{4.48}$$

This establishes item (ii). The proof of Theorem 4.1 is thus complete. □

4.2 One-dimensional ANN approximation results

4.2.1 The modulus of continuity

Definition 4.2 (Modulus of continuity)**.** Let $A \subseteq \mathbb{R}$ be a set and let $f\colon A \to \mathbb{R}$ be a function. Then we denote by $w_f\colon [0,\infty] \to [0,\infty]$ the function which satisfies for all $h \in [0,\infty]$ that

$$w_f(h) = \sup\Big(\big\{|f(x)-f(y)| \in [0,\infty)\colon \big(x,y \in A \text{ with } |x-y| \leq h\big)\big\} \cup \{0\}\Big) \tag{4.49}$$

and we call w_f the modulus of continuity of f.

Lemma 4.3. *Let $b_1 \in [-\infty,\infty]$, $b_2 \in [b_1,\infty]$ and let $f\colon ([b_1,b_2]\cap\mathbb{R}) \to \mathbb{R}$ be a function. Then*

(i) it holds that w_f is non-decreasing,

(ii) it holds that f is uniformly continuous if and only if $\lim_{h\searrow 0} w_f(h) = 0$,

(iii) it holds that f is globally bounded if and only if $w_f(\infty) < \infty$,

(iv) it holds for all $x,y \in [b_1,b_2]\cap\mathbb{R}$ that $|f(x)-f(y)| \leq w_f(|x-y|)$, and

(v) it holds for all $h,\mathfrak{h} \in [0,\infty]$ that $w_f(h+\mathfrak{h}) \leq w_f(h) + w_f(\mathfrak{h})$

(cf. Definition 4.2).

Proof of Lemma 4.3. First, observe that (4.49) implies items (i), (ii), (iii), and (iv). Next, note that (4.49) ensures that for all $h, \mathfrak{h} \in [0, \infty]$ and for all $x, y \in [b_1, b_2] \cap \mathbb{R}$ that satisfy $|x - y| \leq \max\{h, \mathfrak{h}\}$ it holds that $|f(x) - f(y)| \leq w_f(h) + w_f(\mathfrak{h})$. Furthermore, observe that (4.49) assures that for all $h, \mathfrak{h} \in [0, \infty]$ and for all $x, y \in [b_1, b_2] \cap \mathbb{R}$ that satisfy $\max\{h, \mathfrak{h}\} < |x-y| \leq (h+\mathfrak{h})$ it holds that $x - h\frac{x-y}{|x-y|} \in [b_1, b_2] \cap \mathbb{R}$, that $|x - \big(x - h\frac{x-y}{|x-y|}\big)| = h$, that $|\big(x - h\frac{x-y}{|x-y|}\big) - y| = |x - y| - h \leq \mathfrak{h}$, and thus

$$\begin{aligned} |f(x) - f(y)| &\leq \big|f(x) - f\big(x - h\tfrac{x-y}{|x-y|}\big)\big| + \big|f\big(x - h\tfrac{x-y}{|x-y|}\big) - f(y)\big| \\ &\leq w_f(h) + w_f(\mathfrak{h}). \end{aligned} \tag{4.50}$$

Combining both cases, (4.49) proves that for all $h, \mathfrak{h} \in [0, \infty]$ it holds that $w_f(h + \mathfrak{h}) \leq w_f(h) + w_f(\mathfrak{h})$ (cf. Definition 4.2). This establishes item (v). The proof of Lemma 4.3 is thus complete. ◻

Lemma 4.4. *Let $A \subseteq \mathbb{R}$, $L \in [0, \infty)$, and let $f \colon A \to \mathbb{R}$ satisfy for all $x, y \in A$ that $|f(x) - f(y)| \leq L|x - y|$. Then it holds for all $h \in [0, \infty)$ that $w_f(h) \leq Lh$ (cf. Definition 4.2).*

Proof of Lemma 4.4. Note that the assumption that for all $x, y \in A$ it holds that $|f(x) - f(y)| \leq L|x - y|$ and (4.49) show that for all $h \in [0, \infty)$ it holds that

$$\begin{aligned} w_f(h) &= \sup\Big(\big\{|f(x) - f(y)| \in [0, \infty) \colon \big(x, y \in A \text{ with } |x - y| \leq h\big)\big\} \cup \{0\}\Big) \\ &\leq \sup\Big(\big\{L|x - y| \in [0, \infty) \colon \big(x, y \in A \text{ with } |x - y| \leq h\big)\big\} \cup \{0\}\Big) \leq \sup(\{Lh, 0\}) = Lh \end{aligned} \tag{4.51}$$

(cf. Definition 4.2). The proof of Lemma 4.4 is thus complete. ◻

4.2.2 Linear interpolation of one-dimensional functions

Definition 4.5 (Linear interpolation function)**.** Let $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K, f_0, f_1, \ldots, f_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \cdots < \mathfrak{r}_K$. Then we denote by $\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K} \colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $k \in \{1, 2, \ldots, K\}$, $x \in (-\infty, \mathfrak{r}_0)$, $y \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k)$, $z \in [\mathfrak{r}_K, \infty)$ that $(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) = f_0$, $(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(z) = f_K$, and

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y) = f_{k-1} + \big(\tfrac{y - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\big)(f_k - f_{k-1}). \tag{4.52}$$

Lemma 4.6. *Let $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K, f_0, f_1, \ldots, f_K \in \mathbb{R}$ satisfy that $\mathfrak{r}_0 < \mathfrak{r}_1 < \cdots < \mathfrak{r}_K$. Then*

(i) it holds for all $k \in \{0, 1, \ldots, K\}$ that $(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(\mathfrak{r}_k) = f_k$,

(ii) it holds for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ that

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) = f_{k-1} + \big(\tfrac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\big)(f_k - f_{k-1}), \tag{4.53}$$

and

(iii) it holds for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ that

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) = \big(\tfrac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\big) f_{k-1} + \big(\tfrac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\big) f_k \tag{4.54}$$

(cf. Definition 4.5).

Proof of Lemma 4.6. Observe that (4.52) demonstrates items (i) and (ii). Moreover, note that item (ii) yields that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ it holds that

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) = \left[1 - \left(\tfrac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)\right] f_{k-1} + \left(\tfrac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) f_k = \left(\tfrac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) f_{k-1} + \left(\tfrac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) f_k \tag{4.55}$$

(cf. Definition 4.5). This proves item (iii). The proof of Lemma 4.6 is thus complete. □

Lemma 4.7. *Let $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \cdots < \mathfrak{r}_K$ and let $f \colon [\mathfrak{r}_0, \mathfrak{r}_K] \to \mathbb{R}$ be a function. Then*

(i) it holds for all $x, y \in \mathbb{R}$ that

$$\begin{aligned} &\big|(\mathscr{L}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - (\mathscr{L}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y)\big| \\ &\leq \left[\max_{k \in \{1, 2, \ldots, K\}} \left(\tfrac{w_f(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|)}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\right)\right] |x - y| \end{aligned} \tag{4.56}$$

and

(ii) it holds that $\sup_{x \in [\mathfrak{r}_0, \mathfrak{r}_K]} \big|(\mathscr{L}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - f(x)\big| \leq w_f(\max_{k \in \{1, 2, \ldots, K\}} |\mathfrak{r}_k - \mathfrak{r}_{k-1}|)$

(cf. Definitions 4.2 and 4.5).

Proof of Lemma 4.7. Throughout this proof let $l \colon \mathbb{R} \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}$ that $l(x) = (\mathscr{L}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x)$ and let $L \in [0, \infty]$ satisfy

$$L = \max_{k \in \{1, 2, \ldots, K\}} \left(\tfrac{w_f(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|)}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\right) \tag{4.57}$$

(cf. Definitions 4.2 and 4.5). Observe that item (iv) in Lemma 4.3, (4.57), and item (ii) in Lemma 4.6 establish that for all $k \in \{1, 2, \ldots, K\}$, $x, y \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ it holds that

$$\begin{aligned} |l(x) - l(y)| &= \left|\left(\tfrac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)\big(f(\mathfrak{r}_k) - f(\mathfrak{r}_{k-1})\big) - \left(\tfrac{y - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)\big(f(\mathfrak{r}_k) - f(\mathfrak{r}_{k-1})\big)\right| \\ &= \left|\left(\frac{f(\mathfrak{r}_k) - f(\mathfrak{r}_{k-1})}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(x - y)\right| \leq \left(\frac{w_f(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|)}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\right)|x - y| \leq L|x - y|. \end{aligned} \tag{4.58}$$

This, item (iv) in Lemma 4.3, Lemma 4.6, (4.57), and the triangle inequality imply that for all $j, k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{j-1}, \mathfrak{r}_j]$, $y \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ with $j < k$ it holds that

$$\begin{aligned} |l(x) - l(y)| &\leq |l(x) - l(\mathfrak{r}_j)| + |l(\mathfrak{r}_j) - l(\mathfrak{r}_{k-1})| + |l(\mathfrak{r}_{k-1}) - l(y)| \\ &= |l(x) - l(\mathfrak{r}_j)| + |f(\mathfrak{r}_j) - f(\mathfrak{r}_{k-1})| + |l(\mathfrak{r}_{k-1}) - l(y)| \\ &\leq |l(x) - l(\mathfrak{r}_j)| + \left[\sum_{i=j+1}^{k-1} |f(\mathfrak{r}_i) - f(\mathfrak{r}_{i-1})|\right] + |l(\mathfrak{r}_{k-1}) - l(y)| \\ &\leq |l(x) - l(\mathfrak{r}_j)| + \left[\sum_{i=j+1}^{k-1} w_f\big(|\mathfrak{r}_i - \mathfrak{r}_{i-1}|\big)\right] + |l(\mathfrak{r}_{k-1}) - l(y)| \end{aligned} \tag{4.59}$$

$$\leq L\left((\mathfrak{x}_j - x) + \left[\sum_{i=j+1}^{k-1} (\mathfrak{x}_i - \mathfrak{x}_{i-1})\right] + (y - \mathfrak{x}_{k-1})\right) = L|x-y|.$$

Combining this and (4.58) ensures that for all $x, y \in [\mathfrak{x}_0, \mathfrak{x}_K]$ it holds that $|l(x) - l(y)| \leq L|x-y|$. This, the fact that for all $x, y \in (-\infty, \mathfrak{x}_0]$ it holds that $|l(x) - l(y)| = 0 \leq L|x-y|$, the fact that for all $x, y \in [\mathfrak{x}_K, \infty)$ it holds that $|l(x) - l(y)| = 0 \leq L|x-y|$, and the triangle inequality assure that for all $x, y \in \mathbb{R}$ it holds that $|l(x) - l(y)| \leq L|x-y|$. This proves item (i). In addition, note that (4.49), Lemma 4.3, item (iii) in Lemma 4.6, and the triangle inequality prove that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ it holds that

$$\begin{aligned}
|l(x) - f(x)| &= \left|\left(\frac{\mathfrak{x}_k - x}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right) f(\mathfrak{x}_{k-1}) + \left(\frac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right) f(\mathfrak{x}_k) - f(x)\right| \\
&= \left|\left(\frac{\mathfrak{x}_k - x}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right) (f(\mathfrak{x}_{k-1}) - f(x)) + \left(\frac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right) (f(\mathfrak{x}_k) - f(x))\right| \\
&\leq \left(\frac{\mathfrak{x}_k - x}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right) |f(\mathfrak{x}_{k-1}) - f(x)| + \left(\frac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right) |f(\mathfrak{x}_k) - f(x)| \qquad (4.60) \\
&\leq w_f\big(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|\big) \left(\frac{\mathfrak{x}_k - x}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} + \frac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right) \\
&= w_f\big(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|\big) \leq w_f\big(\max\nolimits_{j \in \{1,2,\ldots,K\}} |\mathfrak{x}_j - \mathfrak{x}_{j-1}|\big).
\end{aligned}$$

This establishes item (ii). The proof of Lemma 4.7 is thus complete. ∎

Lemma 4.8. *Let $K \in \mathbb{N}$, $L, \mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K \in \mathbb{R}$ satisfy $\mathfrak{x}_0 < \mathfrak{x}_1 < \cdots < \mathfrak{x}_K$ and let $f \colon [\mathfrak{x}_0, \mathfrak{x}_K] \to \mathbb{R}$ satisfy for all $x, y \in [\mathfrak{x}_0, \mathfrak{x}_K]$ that $|f(x) - f(y)| \leq L|x-y|$. Then*

(i) it holds for all $x, y \in \mathbb{R}$ that

$$\big|(\mathscr{L}^{f(\mathfrak{x}_0), f(\mathfrak{x}_1), \ldots, f(\mathfrak{x}_K)}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(x) - (\mathscr{L}^{f(\mathfrak{x}_0), f(\mathfrak{x}_1), \ldots, f(\mathfrak{x}_K)}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(y)\big| \leq L|x-y| \tag{4.61}$$

and

(ii) it holds that $\sup_{x \in [\mathfrak{x}_0, \mathfrak{x}_K]} \big|(\mathscr{L}^{f(\mathfrak{x}_0), f(\mathfrak{x}_1), \ldots, f(\mathfrak{x}_K)}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(x) - f(x)\big| \leq L(\max_{k \in \{1,2,\ldots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}|)$

(cf. Definition 4.5).

Proof of Lemma 4.8. First, observe that the assumption that for all $x, y \in [\mathfrak{x}_0, \mathfrak{x}_K]$ it holds that $|f(x) - f(y)| \leq L|x-y|$, Lemma 4.4, and item (i) in Lemma 4.7 show that for all $x, y \in \mathbb{R}$ it holds that

$$\begin{aligned}
&\big|(\mathscr{L}^{f(\mathfrak{x}_0), f(\mathfrak{x}_1), \ldots, f(\mathfrak{x}_K)}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(x) - (\mathscr{L}^{f(\mathfrak{x}_0), f(\mathfrak{x}_1), \ldots, f(\mathfrak{x}_K)}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(y)\big| \\
&\leq \left[\max_{k \in \{1,2,\ldots,K\}} \left(\tfrac{L|\mathfrak{x}_k - \mathfrak{x}_{k-1}|}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|}\right)\right] |x-y| = L|x-y|
\end{aligned} \tag{4.62}$$

(cf. Definition 4.5). This proves item (i). Next, note that the assumption that for all $x, y \in [\mathfrak{x}_0, \mathfrak{x}_K]$ it holds that $|f(x) - f(y)| \leq L|x-y|$, Lemma 4.4, and item (ii) in Lemma 4.7 demonstrate that

$$\sup_{x \in [\mathfrak{x}_0, \mathfrak{x}_K]} \big|(\mathscr{L}^{f(\mathfrak{x}_0), f(\mathfrak{x}_1), \ldots, f(\mathfrak{x}_K)}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(x) - f(x)\big| \leq L\left(\max_{k \in \{1,2,\ldots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}|\right). \tag{4.63}$$

This establishes item (ii). The proof of Lemma 4.8 is thus complete. ∎

4.2.3 Linear interpolation with ANNs

Lemma 4.9. *Let $K \in \mathbb{N}_0$, $f_0, c_0, c_1, \ldots, c_K, \beta_0, \beta_1, \ldots, \beta_K, \alpha_0, \alpha_1, \ldots, \alpha_K \in \mathbb{R}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\mathbf{F} \in \mathbf{N}$ satisfy*

$$\mathbf{F} = \mathbf{A}_{1,f_0} \bullet \left(\overset{K}{\underset{k=0}{\oplus}} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha_k,\beta_k})) \right) \tag{4.64}$$

(cf. Definitions 2.1, 2.7, 2.12, 2.16, 2.18, and 3.1). Then

(i) it holds that $\mathcal{D}(\mathbf{F}) = (1, K+1, 1) \in \mathbb{N}^3$,

(ii) it holds that $\mathcal{R}_a(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$,

(iii) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_a(\mathbf{F}))(x) = f_0 + \sum_{k=0}^{K} c_k(a(\alpha_k x + \beta_k))$, and

(iv) it holds that $\mathcal{P}(\mathbf{F}) = 3K + 4 = 3(\mathbb{D}_1(\mathbf{F})) + 1$

(cf. Definition 2.6).

Proof of Lemma 4.9. Note that, e.g., Grohs et al. [35, item (i) in Proposition 2.6], item (i) in Lemma 2.19, and item (i) in Lemma 3.2 prove item (i). In addition, e.g., Grohs et al. [35, item (v) in Proposition 2.6], item (ii) in Lemma 2.19, and item (ii) in Lemma 3.2 establish item (ii). It follows from, e.g., Grohs et al. [35, item (v) in Proposition 2.6], item (iii) in Lemma 2.19, and item (iii) in Lemma 3.2 that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\mathbf{F}))(x) &= f_0 + \left(\mathcal{R}_a\left(\overset{K}{\underset{k=0}{\oplus}} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha_k,\beta_k}))\right)\right)(x) \\ &= f_0 + \sum_{k=0}^{K} c_k((\mathcal{R}_a(\mathfrak{i}_1))(\alpha_k x + \beta_k)) = f_0 + \sum_{k=0}^{K} c_k(a(\alpha_k x + \beta_k)).\end{aligned} \tag{4.65}$$

This shows item (iii). Moreover, note that item (i) implies that

$$\mathcal{P}(\mathbf{F}) = (K+1)(1+1) + (K+1+1) = 3K+4 = 3(K+1)+1 = 3(\mathbb{D}_1(\mathbf{F})) + 1. \tag{4.66}$$

This proves item (iv). The proof of Lemma 4.9 is thus complete. ☐

Lemma 4.10. *Let $K \in \mathbb{N}$, $f_0, f_1, \ldots, f_K, \mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \cdots < \mathfrak{r}_K$, let $\mathbf{F} \in \mathbf{N}$ satisfy*

$$\mathbf{F} = \mathbf{A}_{1,f_0} \bullet \left(\overset{K}{\underset{k=0}{\oplus}} \left(\left(\tfrac{(f_{\min\{k+1,K\}} - f_k)}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \tfrac{(f_k - f_{\max\{k-1,0\}})}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right), \tag{4.67}$$

and let $\mathfrak{r} \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{r}(x) = \max\{x, 0\}$ (cf. Definitions 2.1, 2.7, 2.12, 2.16, 2.18, and 3.1). Then it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) = \mathscr{L}^{f_0,f_1,\ldots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K}$ (cf. Definitions 2.6 and 4.5).

Proof of Lemma 4.10. Throughout this proof let $c_0, c_1, \ldots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that

$$c_k = \frac{(f_{\min\{k+1,K\}} - f_k)}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \frac{(f_k - f_{\max\{k-1,0\}})}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})}. \tag{4.68}$$

Observe that item (iii) in Lemma 4.9 ensures that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = f_0 + \sum_{k=0}^{K} c_k \max\{x - \mathfrak{x}_k, 0\}. \tag{4.69}$$

This and the fact that $\forall\, k \in \{0, 1, \ldots, K\} \colon \mathfrak{x}_0 \le \mathfrak{x}_k$ assure that for all $x \in (-\infty, \mathfrak{x}_0]$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = f_0 + 0 = f_0. \tag{4.70}$$

Next we claim that for all $k \in \{1, 2, \ldots, K\}$ it holds that

$$\sum_{n=0}^{k-1} c_n = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}. \tag{4.71}$$

We now prove (4.71) by induction on $k \in \{1, 2, \ldots, K\}$. For the base case $k = 1$ observe that (4.68) assures that $\sum_{n=0}^{0} c_n = c_0 = \frac{f_1 - f_0}{\mathfrak{x}_1 - \mathfrak{x}_0}$. This proves (4.71) in the base case $k = 1$. For the induction step from $\{1, 2, \ldots, K-1\} \ni (k-1) \dashrightarrow k \in \{2, 3, \ldots, K\}$ note that (4.68) ensures that for all $k \in \{2, 3, \ldots, K\}$ with $\sum_{n=0}^{k-2} c_n = \frac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}}$ it holds that

$$\sum_{n=0}^{k-1} c_n = c_{k-1} + \sum_{n=0}^{k-2} c_n = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} - \tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}} + \tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}} = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}. \tag{4.72}$$

Induction thus proves (4.71). In addition, observe that (4.69), (4.71), and the fact that $\forall\, k \in \{1, 2, \ldots, K\} \colon \mathfrak{x}_{k-1} < \mathfrak{x}_k$ show that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ it holds that

$$\begin{split}(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_{k-1}) &= \sum_{n=0}^{K} c_n (\max\{x - \mathfrak{x}_n, 0\} - \max\{\mathfrak{x}_{k-1} - \mathfrak{x}_n, 0\}) \\ &= \sum_{n=0}^{k-1} c_n [(x - \mathfrak{x}_n) - (\mathfrak{x}_{k-1} - \mathfrak{x}_n)] = \sum_{n=0}^{k-1} c_n (x - \mathfrak{x}_{k-1}) = \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}).\end{split} \tag{4.73}$$

Next we claim that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = f_{k-1} + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}). \tag{4.74}$$

We now prove (4.74) by induction on $k \in \{1, 2, \ldots, K\}$. For the base case $k = 1$ observe that (4.70) and (4.73) demonstrate that for all $x \in [\mathfrak{x}_0, \mathfrak{x}_1]$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_0) + (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_0) = f_0 + \big(\tfrac{f_1 - f_0}{\mathfrak{x}_1 - \mathfrak{x}_0}\big)(x - \mathfrak{x}_0). \tag{4.75}$$

This proves (4.74) in the base case $k = 1$. For the induction step from $\{1, 2, \ldots, K-1\} \ni (k-1) \dashrightarrow k \in \{2, 3, \ldots, K\}$ note that (4.73) implies that for all $k \in \{2, 3, \ldots, K\}$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ with $\forall\, y \in [\mathfrak{x}_{k-2}, \mathfrak{x}_{k-1}] \colon (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y) = f_{k-2} + (\frac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}})(y - \mathfrak{x}_{k-2})$ it holds that

$$\begin{split}&(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_{k-1}) + (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_{k-1}) \\ &= f_{k-2} + \big(\tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}}\big)(\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}) + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}) = f_{k-1} + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}).\end{split} \tag{4.76}$$

Induction thus proves (4.74). Furthermore, observe that (4.68) and (4.71) ensure that

$$\sum_{n=0}^{K} c_n = c_K + \sum_{n=0}^{K-1} c_n = -\tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}} + \tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}} = 0. \tag{4.77}$$

The fact that $\forall\, k \in \{0, 1, \ldots, K\} \colon \mathfrak{r}_k \le \mathfrak{r}_K$ and (4.69) hence imply that for all $x \in [\mathfrak{r}_K, \infty)$ it holds that

$$\begin{aligned}(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{r}_K) &= \left[\sum_{n=0}^{K} c_n(\max\{x - \mathfrak{r}_n, 0\} - \max\{\mathfrak{r}_K - \mathfrak{r}_n, 0\})\right] \\ &= \sum_{n=0}^{K} c_n[(x - \mathfrak{r}_n) - (\mathfrak{r}_K - \mathfrak{r}_n)] = \sum_{n=0}^{K} c_n(x - \mathfrak{r}_K) = 0.\end{aligned} \tag{4.78}$$

This and (4.74) show that for all $x \in [\mathfrak{r}_K, \infty)$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{r}_K) = f_{K-1} + \left(\tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}}\right)(\mathfrak{r}_K - \mathfrak{r}_{K-1}) = f_K. \tag{4.79}$$

Combining this, (4.70), (4.74), and (4.52) establishes that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) = \mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K}$. The proof of Lemma 4.10 is thus complete. ☐

4.2.4 ANN approximations of one-dimensional functions

Lemma 4.11. *Let $K \in \mathbb{N}$, $\beta \in (0, \infty)$, $L, b_1, \mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K \in \mathbb{R}$, $b_2 \in (b_1, \infty)$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that $\mathfrak{r}_k = b_1 + \frac{k(b_2 - b_1)}{K}$, let $\mathfrak{r}, a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{r}(x) = \max\{x, 0\}$ and $a(x) = \frac{1}{\beta}\ln(1+\exp(\beta x))$, let $f \colon [b_1, b_2] \to \mathbb{R}$ satisfy for all $x, y \in [b_1, b_2]$ that $|f(x) - f(y)| \le L|x - y|$, and let $\mathbf{F} \in \mathbf{N}$ satisfy*

$$\mathbf{F} = \mathbf{A}_{1, f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b_1 - b_2)} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1, -\mathfrak{r}_k}) \right) \right) \tag{4.80}$$

(cf. Definitions 2.1, 2.7, 2.12, 2.16, 2.18, and 3.1). Then

(i) it holds for all $x, y \in \mathbb{R}$ that $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \le L|x - y|$,

(ii) it holds that $\sup_{x \in [b_1, b_2]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \le L(b_2 - b_1)K^{-1}$, and

(iii) it holds for all $x, y \in \mathbb{R}$ that $|(\mathcal{R}_a(\mathbf{F}))(x) - (\mathcal{R}_a(\mathbf{F}))(y)| \le L|x - y|$

(cf. Definition 2.6).

Proof of Lemma 4.11. Note that the fact that $\forall\, k \in \{0, 1, \ldots, K\} \colon \mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}} = \mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}} = (b_2 - b_1)K^{-1}$ assures that for all $k \in \{0, 1, \ldots, K\}$ it holds that

$$\tfrac{(f(\mathfrak{r}_{\min\{k+1,K\}}) - f(\mathfrak{r}_k))}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \tfrac{(f(\mathfrak{r}_k) - f(\mathfrak{r}_{\max\{k-1,0\}}))}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} = \tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b_2 - b_1)}. \tag{4.81}$$

This and Lemma 4.10 demonstrate that

$$\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) = \mathscr{L}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K}. \tag{4.82}$$

Combining this with the assumption that $\forall\, x, y \in [b_1, b_2]\colon |f(x) - f(y)| \leq L|x - y|$ and item (i) in Lemma 4.8 establishes item (i). Moreover, note that (4.82), the assumption that $\forall\, x, y \in [b_1, b_2]\colon |f(x) - f(y)| \leq L|x - y|$, item (ii) in Lemma 4.8, and the fact that $\forall\, k \in \{1, 2, \ldots, K\}\colon \mathfrak{x}_k - \mathfrak{x}_{k-1} = (b_2 - b_1)K^{-1}$ demonstrate that for all $x \in [b_1, b_2]$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq L\bigg(\max_{k\in\{1,2,\ldots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}|\bigg) = L(b_2 - b_1)K^{-1}. \tag{4.83}$$

This establishes item (ii). Next, observe that item (iii) in Lemma 4.9 and (4.81) imply for all $x \in \mathbb{R}$ that

$$\begin{aligned}
(\mathcal{R}_a(\mathbf{F}))(x) &= f(\mathfrak{x}_0) + \sum_{k=0}^{K}\Big(\tfrac{(f(\mathfrak{x}_{\min\{k+1,K\}})-f(\mathfrak{x}_k))}{(\mathfrak{x}_{\min\{k+1,K\}}-\mathfrak{x}_{\min\{k,K-1\}})} - \tfrac{(f(\mathfrak{x}_k)-f(\mathfrak{x}_{\max\{k-1,0\}}))}{(\mathfrak{x}_{\max\{k,1\}}-\mathfrak{x}_{\max\{k-1,0\}})}\Big)a(x - \mathfrak{x}_k)\\
&= f(\mathfrak{x}_0) + \sum_{k=0}^{K-1} \frac{f(\mathfrak{x}_{k+1}) - f(\mathfrak{x}_k)}{\mathfrak{x}_{k+1} - \mathfrak{x}_k} a(x - \mathfrak{x}_k) - \sum_{k=1}^{K} \frac{f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} a(x - \mathfrak{x}_k)\\
&= f(\mathfrak{x}_0) + \sum_{k=1}^{K} \frac{f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} (a(x - \mathfrak{x}_{k-1}) - a(x - \mathfrak{x}_k)).
\end{aligned} \tag{4.84}$$

Note that a is differentiable and it holds for all $x \in \mathbb{R}$ that $\frac{da(x)}{dx} = \frac{e^{\beta x}}{1+e^{\beta x}}$. It thus follows from (4.84) that $\mathcal{R}_a(\mathbf{F})$ is differentiable and it holds for all $x \in \mathbb{R}$ that

$$\frac{d(\mathcal{R}_a(\mathbf{F}))(x)}{dx} = \sum_{k=1}^{K} \frac{f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} \bigg(\frac{e^{\beta(x-\mathfrak{x}_{k-1})}}{1 + e^{\beta(x-\mathfrak{x}_{k-1})}} - \frac{e^{\beta(x-\mathfrak{x}_k)}}{1 + e^{\beta(x-\mathfrak{x}_k)}}\bigg). \tag{4.85}$$

This, the triangle inequality, the assumption that $\forall\, x, y \in [b_1, b_2]\colon |f(x) - f(y)| \leq L|x - y|$, and the fact that for all $k \in \{1, \ldots, K\}$, $x \in \mathbb{R}$ it holds that

$$\frac{e^{\beta(x-\mathfrak{x}_{k-1})}}{1 + e^{\beta(x-\mathfrak{x}_{k-1})}} - \frac{e^{\beta(x-\mathfrak{x}_k)}}{1 + e^{\beta(x-\mathfrak{x}_k)}} = \frac{e^{\beta(x-\mathfrak{x}_{k-1})} - e^{\beta(x-\mathfrak{x}_k)}}{(1 + e^{\beta(x-\mathfrak{x}_{k-1})})(1 + e^{\beta(x-\mathfrak{x}_k)})} \geq 0 \tag{4.86}$$

yield for all $x \in \mathbb{R}$ that

$$\begin{aligned}
\left|\frac{d(\mathcal{R}_a(\mathbf{F}))(x)}{dx}\right| &\leq \sum_{k=1}^{K} \frac{|f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})|}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \bigg(\frac{e^{\beta(x-\mathfrak{x}_{k-1})}}{1 + e^{\beta(x-\mathfrak{x}_{k-1})}} - \frac{e^{\beta(x-\mathfrak{x}_k)}}{1 + e^{\beta(x-\mathfrak{x}_k)}}\bigg)\\
&\leq L\bigg(\sum_{k=1}^{K} \frac{e^{\beta(x-\mathfrak{x}_{k-1})}}{1 + e^{\beta(x-\mathfrak{x}_{k-1})}} - \sum_{k=1}^{K} \frac{e^{\beta(x-\mathfrak{x}_k)}}{1 + e^{\beta(x-\mathfrak{x}_k)}}\bigg)\\
&= L\bigg(\frac{e^{\beta(x-\mathfrak{x}_0)}}{1 + e^{\beta(x-\mathfrak{x}_0)}} - \frac{e^{\beta(x-\mathfrak{x}_K)}}{1 + e^{\beta(x-\mathfrak{x}_K)}}\bigg) \leq \frac{Le^{\beta(x-\mathfrak{x}_0)}}{1 + e^{\beta(x-\mathfrak{x}_0)}} \leq L.
\end{aligned} \tag{4.87}$$

Hence, it holds for all $x, y \in \mathbb{R}$ that $|(\mathcal{R}_a(\mathbf{F}))(x) - (\mathcal{R}_a(\mathbf{F}))(y)| \leq L|x - y|$. This establishes item (iii). The proof of Lemma 4.11 is thus complete. ∎

Lemma 4.12. *Let $\varepsilon \in (0, 1]$, $L \in [0, \infty)$, $q \in (1, \infty)$, let $b \in [1, \infty)$ satisfy* $\max\{1, 2L\} = \varepsilon b^{q-1}$, *let $K \in \mathbb{N} \cap [\frac{2Lb}{\varepsilon}, \frac{2Lb}{\varepsilon} + 1]$, let $f\colon \mathbb{R} \to \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leq$*

$L|x-y|$, let $\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K, c_0, c_1, \ldots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that $\mathfrak{x}_k = -b + \frac{2kb}{K}$ and

$$c_k = \frac{K(f(\mathfrak{x}_{\min\{k+1,K\}}) - 2f(\mathfrak{x}_k) + f(\mathfrak{x}_{\max\{k-1,0\}}))}{2b}, \tag{4.88}$$

let $\mathbf{F} \in \mathbf{N}$ satisfy that

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{x}_0)} \bullet \left(\bigoplus_{k=0}^{K} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{x}_k})) \right), \tag{4.89}$$

and let $\mathfrak{r} \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{r}(x) = \max\{x, 0\}$ (cf. Definitions 2.1, 2.7, 2.12, 2.16, 2.18, and 3.1). Then

(i) it holds for all $x, y \in \mathbb{R}$ that $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y)| \leq L|x-y|$,

(ii) it holds that $\sup_{x \in [-b,b]} |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)| \leq \frac{2Lb}{K} \leq \varepsilon$,

(iii) it holds for all $x \in \mathbb{R}$ that $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)| \leq \varepsilon \max\{1, |x|^q\}$,

(iv) it holds that $\mathbb{D}_1(\mathbf{F}) \leq 2(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)} + 1$, and

(v) it holds that $\mathcal{P}(\mathbf{F}) = 3(\mathbb{D}_1(\mathbf{F})) + 1 \leq 12(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)}$

(cf. Definition 2.6).

Proof of Lemma 4.12. Note that item (i) in Lemma 4.11 yields item (i). Next, observe that the fact that $K \in \mathbb{N} \cap [\frac{2Lb}{\varepsilon}, \frac{2Lb}{\varepsilon} + 1]$ implies that $\frac{2Lb}{K} \leq \varepsilon$. This and item (ii) in Lemma 4.11 establish item (ii). The triangle inequality, item (i), the fact that $f(-b) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(-b)$, the fact that $f(b) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(b)$, and the fact that for all $x, y \in \mathbb{R}$ it holds that $|f(x) - f(y)| \leq L|x-y|$ ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)| &\leq |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(b)| + |f(b) - f(0)| + |f(0) - f(x)| \\ &= |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(b)| + |f(b) - f(0)| + |f(0) - f(x)| \\ &\leq L|x-b| + L|b| + L|x| = L(|x-b| + b + |x|) \end{aligned} \tag{4.90}$$

and

$$\begin{aligned} |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)| &\leq |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(-b)| + |f(-b) - f(0)| + |f(0) - f(x)| \\ &= |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(-b)| + |f(-b) - f(0)| + |f(0) - f(x)| \\ &\leq L|x+b| + L|b| + L|x| = L(|x+b| + b + |x|). \end{aligned} \tag{4.91}$$

It follows from (4.90) that for all $x \in (b, \infty)$ it holds that

$$\begin{aligned} \frac{|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)|}{\max\{1, |x|^q\}} &\leq \frac{L(|x-b| + b + |x|)}{\max\{1, |x|^q\}} = \frac{L(x - b + b + x)}{\max\{1, |x|^q\}} \\ &= \frac{2L|x|}{\max\{1, |x|^q\}} \leq \frac{\max\{1, 2L\}}{|x|^{q-1}} \leq \frac{\max\{1, 2L\}}{b^{q-1}} = \varepsilon. \end{aligned} \tag{4.92}$$

Moreover, (4.91) demonstrates that for all $x \in (-\infty, -b)$ it holds that

$$\begin{aligned}\frac{|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)|}{\max\{1, |x|^q\}} &\leq \frac{L(|x+b| + b + |x|)}{\max\{1, |x|^q\}} = \frac{L(-(x+b) + b - x)}{\max\{1, |x|^q\}} \\ &= \frac{2L|x|}{\max\{1, |x|^q\}} \leq \frac{\max\{1, 2L\}}{|x|^{q-1}} \leq \frac{\max\{1, 2L\}}{b^{q-1}} = \varepsilon.\end{aligned} \tag{4.93}$$

Combining this, (4.92), and item (ii) shows that for all $x \in \mathbb{R}$ it holds that $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq \varepsilon \max\{1, |x|^q\}$. This establishes item (iii). In addition, observe that the fact that $\max\{1, 2L\} = \varepsilon b^{q-1}$ and the fact that $K \leq 1 + \frac{2Lb}{\varepsilon}$ prove that

$$K \leq 1 + \frac{2Lb}{\varepsilon} \leq 1 + \frac{\max\{1, 2L\}b}{\varepsilon} = 1 + b^q \leq 2b^q = 2\left(\frac{\max\{1, 2L\}}{\varepsilon}\right)^{q/(q-1)}. \tag{4.94}$$

This and the fact that $\mathbb{D}_1(\mathbf{F}) = K + 1$ (cf. item (i) in Lemma 4.9) establish item (iv). Moreover, observe that item (iv) in Lemma 4.9 and item (iv) ensure that

$$\begin{aligned}\mathcal{P}(\mathbf{F}) = 3(\mathbb{D}_1(\mathbf{F})) + 1 \leq 4(\mathbb{D}_1(\mathbf{F})) &\leq 8(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)} + 4 \\ &\leq 12(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)}.\end{aligned} \tag{4.95}$$

This establishes item (v). The proof of Lemma 4.12 is thus complete. □

Corollary 4.13. *Let $\varepsilon \in (0, 1]$, $L \in [0, \infty)$, $q \in (1, \infty)$, $\alpha \in [0, \infty)\backslash\{1\}$, let $f \colon \mathbb{R} \to \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leq L|x - y|$, and let $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, \alpha x\}$. Then there exists $\mathbf{G} \in \mathbf{N}$ such that*

(i) it holds that $\mathcal{R}_a(\mathbf{G}) \in C(\mathbb{R}, \mathbb{R})$,

(ii) it holds that $\mathcal{H}(\mathbf{G}) = 1$,

(iii) it holds for all $x, y \in \mathbb{R}$ that $|(\mathcal{R}_a(\mathbf{G}))(x) - (\mathcal{R}_a(\mathbf{G}))(y)| \leq L|x - y|$,

(iv) it holds for all $x \in \mathbb{R}$ that $|(\mathcal{R}_a(\mathbf{G}))(x) - f(x)| \leq \varepsilon \max\{1, |x|^q\}$,

(v) it holds that $\mathbb{D}_1(\mathbf{G}) \leq 4(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)} + 2$, and

(vi) it holds that $\mathcal{P}(\mathbf{G}) = 3(\mathbb{D}_1(\mathbf{G})) + 1 \leq 24(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)}$

(cf. Definitions 2.1 and 2.6).

Proof of Corollary 4.13. Throughout this proof let $b \in [1, \infty)$ satisfy $\max\{1, 2L\} = \varepsilon b^{q-1}$, let $K \in \mathbb{N} \cap [\frac{2Lb}{\varepsilon}, \frac{2Lb}{\varepsilon} + 1]$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K$, $c_0, c_1, \ldots, c_K$, $h_0, h_1, \ldots, h_{2K+1}$, $\alpha_0, \alpha_1, \ldots, \alpha_{2K+1}$, $\beta_0, \beta_1, \ldots, \beta_{2K+1} \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that $\mathfrak{r}_k = -b + \frac{2kb}{K}$,

$$c_k = \frac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{2b}, \tag{4.96}$$

$h_k = \frac{c_k|1-\alpha|\alpha}{(1-\alpha)(1-\alpha^2)}$, $h_{k+K+1} = \frac{c_k|1-\alpha|}{(1-\alpha)(1-\alpha^2)}$, $\alpha_k = \frac{-|1-\alpha|}{1-\alpha}$, $\alpha_{k+K+1} = \frac{|1-\alpha|}{1-\alpha}$, $\beta_k = \frac{|1-\alpha|\mathfrak{r}_k}{1-\alpha}$, and $\beta_{k+K+1} = \frac{-|1-\alpha|\mathfrak{r}_k}{1-\alpha}$, let $\mathbf{F} \in \mathbf{N}$ satisfy that

$$\mathbf{F} = \mathbf{A}_{1, f(\mathfrak{r}_0)} \bullet \left(\overset{K}{\underset{k=0}{\oplus}} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1, -\mathfrak{r}_k})) \right), \tag{4.97}$$

let $\mathbf{G} \in \mathbf{N}$ satisfy that

$$\mathbf{G} = \mathbf{A}_{1,f(\mathfrak{x}_0)} \bullet \left(\overset{2K+1}{\underset{k=0}{\oplus}} (h_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha_k,\beta_k})) \right), \tag{4.98}$$

and let $\mathfrak{r} \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{r}(x) = \max\{x, 0\}$ (cf. Definitions 2.1, 2.7, 2.12, 2.16, 2.18, and 3.1). Note that items (i) and (ii) in Lemma 4.9 establish items (i) and (ii). Furthermore, observe that the fact that for all $x \in \mathbb{R}$ it holds that $a(x) = \max\{x, \alpha x\} = \frac{x+\alpha x+|x-\alpha x|}{2}$ implies for all $x \in \mathbb{R}$ that

$$\begin{aligned}
&\frac{|1-\alpha|\alpha}{(1-\alpha)(1-\alpha^2)} a\left(-\frac{|1-\alpha|x}{1-\alpha}\right) + \frac{|1-\alpha|}{(1-\alpha)(1-\alpha^2)} a\left(\frac{|1-\alpha|x}{1-\alpha}\right) \\
&= \frac{|1-\alpha|\alpha}{(1-\alpha)(1-\alpha^2)} \frac{\frac{-|1-\alpha|x}{1-\alpha} + \frac{-\alpha|1-\alpha|x}{1-\alpha} + \left|\frac{-|1-\alpha|x}{1-\alpha} - \frac{-\alpha|1-\alpha|x}{1-\alpha}\right|}{2} \\
&\quad + \frac{|1-\alpha|}{(1-\alpha)(1-\alpha^2)} \frac{\frac{|1-\alpha|x}{1-\alpha} + \frac{\alpha|1-\alpha|x}{1-\alpha} + \left|\frac{|1-\alpha|x}{1-\alpha} - \frac{\alpha|1-\alpha|x}{1-\alpha}\right|}{2} \\
&= \frac{|1-\alpha|\Big(-\alpha|1-\alpha|x - \alpha^2|1-\alpha|x + |1-\alpha|x + \alpha|1-\alpha|x\Big)}{2(1-\alpha)(1-\alpha^2)(1-\alpha)} \\
&\quad + \frac{|1-\alpha|\Big(\alpha\big|-|1-\alpha|x + \alpha|1-\alpha|x\big| + \big||1-\alpha|x - \alpha|1-\alpha|x\big|\Big)}{2(1-\alpha)(1-\alpha^2)|1-\alpha|} \\
&= \frac{|1-\alpha|^2(1-\alpha^2)x}{2(1-\alpha)^2(1-\alpha^2)} + \frac{\alpha\big|(-1+\alpha)|1-\alpha|x\big| + \big|(1-\alpha)|1-\alpha|x\big|}{2(1-\alpha)(1-\alpha^2)} \\
&= \frac{x}{2} + \frac{(\alpha+1)|1-\alpha||1-\alpha||x|}{2(1-\alpha)(1-\alpha)(1+\alpha)} = \frac{x+|x|}{2} = \max\{x, 0\} = \mathfrak{r}(x).
\end{aligned} \tag{4.99}$$

Combining this and item (iii) in Lemma 4.9 shows for all $x \in \mathbb{R}$ that

$$\begin{aligned}
(\mathcal{R}_a(\mathbf{G}))(x) &= f(\mathfrak{x}_0) + \sum_{k=0}^{K} \frac{c_k|1-\alpha|\alpha}{(1-\alpha)(1-\alpha^2)} \left(a\left(-\frac{|1-\alpha|x}{1-\alpha} + \frac{|1-\alpha|\mathfrak{x}_k}{1-\alpha} \right) \right) \\
&\quad + \sum_{k=0}^{K} \frac{c_k|1-\alpha|}{(1-\alpha)(1-\alpha^2)} \left(a\left(\frac{|1-\alpha|x}{1-\alpha} - \frac{|1-\alpha|\mathfrak{x}_k}{1-\alpha} \right) \right) \\
&= f(\mathfrak{x}_0) + \sum_{k=0}^{K} c_k \mathfrak{r}(x - \mathfrak{x}_k) = (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x).
\end{aligned} \tag{4.100}$$

Therefore, items (i) and (iii) in Lemma 4.12 establish items (iii) and (iv). Note that item (i) in Lemma 4.9 shows that $\mathbb{D}_1(\mathbf{G}) = 2(K+1) = 2\mathbb{D}_1(\mathbf{F})$. Therefore, item (iv) in Lemma 4.12 implies item (v). It follows from item (iv) in Lemma 4.9 that $\mathcal{P}(\mathbf{G}) = 3(\mathbb{D}_1(\mathbf{G})) + 1 \leq 2(3(\mathbb{D}_1(\mathbf{F})) + 1) = 2\mathcal{P}(\mathbf{F})$. This and item (v) in Lemma 4.12 ensure item (vi). The proof of Corollary 4.13 is thus complete. □

Corollary 4.14. *Let $\varepsilon \in (0, 1]$, $L \in [0, \infty)$, $q \in (1, \infty)$, let $f : \mathbb{R} \to \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leq L|x - y|$, and let $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \ln(1 + \exp(x))$. Then there exists $\mathbf{G} \in \mathbf{N}$ such that* *(i) it holds that* $\mathcal{R}_a(\mathbf{G}) \in C(\mathbb{R}, \mathbb{R})$*,*

(ii) it holds that $\mathcal{H}(\mathbf{G}) = 1$*,*

(iii) it holds for all $x, y \in \mathbb{R}$ *that* $|(\mathcal{R}_a(\mathbf{G}))(x) - (\mathcal{R}_a(\mathbf{G}))(y)| \leq L|x - y|$*,*

(iv) it holds for all $x \in \mathbb{R}$ *that* $|(\mathcal{R}_a(\mathbf{G}))(x) - f(x)| \leq 2\varepsilon \max\{1, |x|^q\}$*,*

(v) it holds that $\mathbb{D}_1(\mathbf{G}) \leq 2(\max\{1, 2L\})^{q/(q-1)}\varepsilon^{-q/(q-1)} + 1$*, and*

(vi) it holds that $\mathcal{P}(\mathbf{G}) = 3(\mathbb{D}_1(\mathbf{G})) + 1 \leq 12(\max\{1, 2L\})^{q/(q-1)}\varepsilon^{-q/(q-1)}$

(cf. Definitions 2.1 and 2.6).

Proof of Corollary 4.14. Throughout this proof let $b \in [1, \infty)$ satisfy $\max\{1, 2L\} = \varepsilon b^{q-1}$, let $K \in \mathbb{N} \cap [\frac{2Lb}{\varepsilon}, \frac{2Lb}{\varepsilon} + 1]$, let $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K, c_0, c_1, \ldots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that $\mathfrak{r}_k = -b + \frac{2kb}{K}$ and

$$c_k = \frac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{2b}, \tag{4.101}$$

let $\mathbf{F} \in \mathbf{N}$ satisfy that

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\overset{K}{\underset{k=0}{\oplus}} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k})) \right), \tag{4.102}$$

let $\beta = \max\{2, 2K^2 L \ln(2)\varepsilon^{-1}\}$, let $\mathfrak{r}, \mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{r}(x) = \max\{x, 0\}$ and

$$\mathfrak{a}(x) = \frac{1}{\beta} \ln(1 + \exp(\beta x)), \tag{4.103}$$

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

$$\mathbf{G} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\overset{K}{\underset{k=0}{\oplus}} \left(\frac{c_k}{\beta} \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\beta,-\beta\mathfrak{r}_k}) \right) \right) \tag{4.104}$$

(cf. Definitions 2.1, 2.7, 2.12, 2.16, 2.18, and 3.1). Note that items (i) and (ii) in Lemma 4.9 establish items (i) and (ii). Moreover, item (iii) in Lemma 4.9 implies for all $x \in \mathbb{R}$ that

$$\begin{aligned} (\mathcal{R}_a(\mathbf{G}))(x) &= f(\mathfrak{r}_0) + \sum_{k=0}^{K} \frac{c_k}{\beta}(a(\beta x - \beta \mathfrak{r}_k)) = f(\mathfrak{r}_0) + \sum_{k=0}^{K} \frac{c_k}{\beta} \ln(1 + \exp(\beta(x - \mathfrak{r}_k))) \\ &= f(\mathfrak{r}_0) + \sum_{k=0}^{K} c_k(\mathfrak{a}(x - \mathfrak{r}_k)) = (\mathcal{R}_\mathfrak{a}(\mathbf{F}))(x). \end{aligned} \tag{4.105}$$

This and item (iii) in Lemma 4.11 prove item (iii). Observe that the triangle inequality and item (iii) in Lemma 4.12 show for all $x \in \mathbb{R}$ that

$$\begin{aligned} |(\mathcal{R}_a(\mathbf{G}))(x) - f(x)| &\leq |(\mathcal{R}_a(\mathbf{G}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x)| + |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)| \\ &\leq |(\mathcal{R}_a(\mathbf{G}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x)| + \varepsilon \max\{1, |x|^q\}. \end{aligned} \tag{4.106}$$

Furthermore, (4.105), item (iii) in Lemma 4.9, and the triangle inequality imply for all $x \in \mathbb{R}$ that

$$\begin{aligned}|(\mathcal{R}_a(\mathbf{G}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x)| &= |(\mathcal{R}_{\mathfrak{a}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x)| \\ &\leq \sum_{k=0}^{K} |c_k| \, |\mathfrak{a}(x - \mathfrak{x}_k) - \mathfrak{r}(x - \mathfrak{x}_k)|.\end{aligned} \tag{4.107}$$

Note that it holds for all $x \in [0, \infty)$ that $|\frac{1}{\beta} \ln(1 + \exp(\beta x)) - x| = |\frac{1}{\beta} \ln(\frac{1+\exp(\beta x)}{\exp(\beta x)})| \leq \frac{1}{\beta} \ln(2)$ and that it holds for all $x \in (-\infty, 0)$ that $|\frac{1}{\beta} \ln(1 + \exp(\beta x))| \leq \frac{1}{\beta} \ln(2)$. This and (4.107) yield for all $x \in \mathbb{R}$ that

$$|(\mathcal{R}_a(\mathbf{G}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x)| \leq \frac{\ln(2)}{\beta} \sum_{k=0}^{K} |c_k|. \tag{4.108}$$

Moreover, (4.101), the triangle inequality, and the assumption that $\forall x, y \in \mathbb{R} \colon |f(x) - f(y)| \leq L|x - y|$ imply for all $k \in \{0, 1, \ldots, K\}$ that

$$\begin{aligned}|c_k| &\leq \tfrac{K}{2b} \big(|f(\mathfrak{x}_{\min\{k+1,K\}}) - f(\mathfrak{x}_k)| + |f(\mathfrak{x}_{\max\{k-1,0\}}) - f(\mathfrak{x}_k)| \big) \\ &\leq \tfrac{KL}{2b} \big(|\mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_k| + |\mathfrak{x}_{\max\{k-1,0\}} - \mathfrak{x}_k| \big) \leq 2KL.\end{aligned} \tag{4.109}$$

This, (4.108), and the fact that $\beta \geq 2K^2 L \ln(2) \varepsilon^{-1}$ show for all $x \in \mathbb{R}$ that

$$|(\mathcal{R}_a(\mathbf{G}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x)| \leq \tfrac{2\ln(2) K^2 L}{\beta} \leq \varepsilon. \tag{4.110}$$

Combining this and (4.106) proves item (iv). Moreover, observe that items (i) and (iv) in Lemma 4.9 show that $\mathbb{D}_1(\mathbf{G}) = K + 1 = \mathbb{D}_1(\mathbf{F})$ and $\mathcal{P}(\mathbf{G}) = 3(\mathbb{D}_1(\mathbf{G})) + 1 = 3(\mathbb{D}_1(\mathbf{F})) + 1 = \mathcal{P}(\mathbf{F})$. Therefore, items (iv) and (v) in Lemma 4.12 establish items (v) and (vi). The proof of Corollary 4.14 is thus complete. □

4.3 ANN approximation results with specific activation functions

Corollary 4.15. *Let $\gamma, T, \kappa, \mathfrak{c} \in (0, \infty)$, $r \in \mathbb{N}$, $p \in \mathbb{N} \backslash \{1\}$, $\mathfrak{q} \in [2, \infty)$, let $f \colon \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous, for every $d \in \mathbb{N}$ let $u_d \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that*

$$(\tfrac{\partial}{\partial t} u_d)(t, x) + \mathfrak{c}(\Delta_x u_d)(t, x) + f\big(u_d(t, x)\big) = 0, \tag{4.111}$$

let $\nu \in \{0, 1\}$, $\alpha \in [0, \infty) \backslash \{1\}$, $\mathfrak{a}_0, \mathfrak{a}_1 \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{a}_0(x) = \max\{x, \alpha x\}$ and $\mathfrak{a}_1(x) = \ln(1 + \exp(x))$, for every $d \in \mathbb{N}$ let $\mu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be a probability measure with

$$\textstyle\int_{\mathbb{R}^d} \|y\|^{p^2 \mathfrak{q}} \mu_d(\mathrm{d}y) \leq \kappa d^{r p^2 \mathfrak{q}}, \tag{4.112}$$

and assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that there exists $\mathbf{G} \in \mathbf{N}$ such that for all $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{G}) \in C(\mathbb{R}^d, \mathbb{R}), \qquad \mathcal{P}(\mathbf{G}) \leq \kappa d^p \varepsilon^{-\gamma}, \qquad \textit{and} \tag{4.113}$$

$$\varepsilon |u_d(t, x)| + |u_d(T, x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{G}))(x)| \leq \varepsilon \kappa d^p (1 + \|x\|)^p \tag{4.114}$$

(cf. Definitions 2.1, 2.3, and 2.6). Then there exists $c \in \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists $\mathbf{U} \in \mathbf{N}$ such that

$$\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}) \in C(\mathbb{R}^d, \mathbb{R}), \qquad \mathcal{P}(\mathbf{U}) \le c d^c \varepsilon^{-c}, \qquad \text{and} \tag{4.115}$$

$$\sup_{q \in (0,\mathfrak{q}]} \left[\textstyle\int_{\mathbb{R}^d} |u_d(0,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}))(x)|^q \mu_d(\mathrm{d}x)\right]^{1/q} \le \varepsilon. \tag{4.116}$$

Proof of Corollary 4.15. Throughout this proof let $\mathscr{f}\colon \mathbb{R} \to \mathbb{R}$ satisfy for all $z \in \mathbb{R}$ that $\mathscr{f}(z) = (2\mathfrak{c})^{-1} f(z)$ and for every $d \in \mathbb{N}$ let $\mathscr{u}_d\colon [0, 2\mathfrak{c}T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$, $s \in [0, 2\mathfrak{c}T]$ that

$$\mathscr{u}_d(s,x) = u_d\big((2\mathfrak{c})^{-1}s, x\big). \tag{4.117}$$

Observe that (4.113) and (4.114) assure that there exist $(\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ which satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that $\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\mathbf{F}_{d,\varepsilon}) \le \kappa d^p \varepsilon^{-\gamma}$, and

$$\varepsilon|u_d(t,x)| + |u_d(T,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{d,\varepsilon}))(x)| \le \varepsilon\kappa d^p (1 + \|x\|)^p. \tag{4.118}$$

Note that (4.118) proves that for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$|u_d(t,x)| \le \kappa d^p (1 + \|x\|)^p. \tag{4.119}$$

This assures that for all $d \in \mathbb{N}$ it holds that u_d is at most polynomially growing. Therefore, we obtain that for all $d \in \mathbb{N}$ it holds that $\mathscr{u}_d$ is at most polynomially growing. Furthermore, observe that it holds for all $d \in \mathbb{N}$ that $\mathscr{u}_d \in C^{1,2}([0, 2\mathfrak{c}T] \times \mathbb{R}^d, \mathbb{R})$. In addition, (4.111) implies for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $s \in [0, 2\mathfrak{c}T]$ that

$$(\tfrac{\partial}{\partial t}\mathscr{u}_d)(s,x) + \tfrac{1}{2}(\Delta_x \mathscr{u}_d)(s,x) + \mathscr{f}\big(\mathscr{u}_d(s,x)\big) = 0. \tag{4.120}$$

Note that the assumption that f is Lipschitz continuous establishes that there exists $L \in [0,\infty)$ such that for all $w, z \in \mathbb{R}$ it holds that $|f(z) - f(w)| \le L|z - w|$. This yields for all $w, z \in \mathbb{R}$ that $|\mathscr{f}(z) - \mathscr{f}(w)| \le (2\mathfrak{c})^{-1} L|z - w|$. Next, observe that (4.118) and the triangle inequality ensure for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ that

$$\begin{aligned}
&\varepsilon|(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{d,\varepsilon}))(x)| + |\mathscr{u}_d(2\mathfrak{c}T, x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{d,\varepsilon}))(x)| \\
&= \varepsilon|(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{d,\varepsilon}))(x)| + |u_d(T,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{d,\varepsilon}))(x)| \le \varepsilon|u_d(T,x)| + 2|u_d(T,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{d,\varepsilon}))(x)| \\
&\le 2\varepsilon\kappa d^p (1 + \|x\|)^p.
\end{aligned} \tag{4.121}$$

Furthermore, note that the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{P}(\mathbf{F}_{d,\varepsilon}) \le \kappa d^p \varepsilon^{-\gamma}$ and Lemma 2.4 establish that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that $\varepsilon^\gamma \mathcal{L}(\mathbf{F}_{d,\varepsilon}) + \varepsilon^\gamma |||\mathcal{D}(\mathbf{F}_{d,\varepsilon})||| \le \kappa d^p$. Moreover, observe that it follows from Corollaries 4.13 and 4.14 that there exists $(\mathbf{F}_{0,\varepsilon})_{\varepsilon\in(0,1]} \subseteq \mathbf{N}$ such that for all $\varepsilon \in (0,1]$ the following properties hold true:

(I) it holds that $\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}) \in C(\mathbb{R}, \mathbb{R})$,

(II) it holds for all $x, y \in \mathbb{R}$ that $|(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(y)| \le \frac{L}{2\mathfrak{c}}|x - y|$,

(III) it holds for all $x \in \mathbb{R}$ that $|(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(x) - \mathscr{f}(x)| \le 2\varepsilon \max\{1, |x|^2\}$,

(IV) it holds that $\mathcal{P}(\mathbf{F}_{0,\varepsilon}) \le 24(\max\{1, \mathfrak{c}^{-1}L\})^2 \varepsilon^{-2}$.

Note that item (IV) and Lemma 2.4 prove for all $\varepsilon \in (0,1]$ that $\varepsilon^2 \mathcal{L}(F_{0,\varepsilon}) + \varepsilon^2 |||\mathcal{D}(\mathbf{F}_{0,\varepsilon})||| \leq 24(\max\{1, \mathfrak{c}^{-1}L\})^2 \varepsilon^{-2}$. Furthermore, observe that items (II) and (III) and the triangle inequality imply for all $\varepsilon \in (0,1]$, $x \in \mathbb{R}$ that

$$\begin{aligned} |(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(x)| &\leq |(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(0)| + |(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(0) - \mathscr{f}(0)| + |\mathscr{f}(0)| \\ &\leq \tfrac{L}{2\mathfrak{c}}|x| + 2\varepsilon + |\mathscr{f}(0)| \leq \big(2 + (2\mathfrak{c})^{-1}(L + |f(0)|)\big)(1+|x|)^2. \end{aligned} \tag{4.122}$$

This, item (III), the fact that $\forall x \in \mathbb{R}\colon 1 + |x|^2 \leq (1+|x|)^2$, and the assumption that $p \geq 2$ demonstrate for all $\varepsilon \in (0,1]$, $x \in \mathbb{R}$ that

$$\begin{aligned} &\varepsilon|(\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(x)| + |\mathscr{f}(x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{F}_{0,\varepsilon}))(x)| \\ &\leq \varepsilon\big(\big(2 + (2\mathfrak{c})^{-1}(L + |f(0)|)\big)(1+|x|)^2 + 2\max\{1, |x|^2\}\big) \\ &\leq \varepsilon(4 + (2\mathfrak{c})^{-1}(L + |f(0)|))(1+|x|)^p. \end{aligned} \tag{4.123}$$

Moreover, items (i), (iii), and (v) in Lemma 3.5 and item (ii) in Lemma 3.8 ensure that there exists $\mathfrak{J} \in \mathbf{N}$ such that $\mathcal{H}(\mathfrak{J}) = 1$ and $\mathcal{R}_{\mathfrak{a}_\nu}(\mathfrak{J}) = \mathrm{id}_\mathbb{R}$. Theorem 4.1 (applied with $\kappa \curvearrowleft \max\{2\kappa, 24(\max\{1, \mathfrak{c}^{-1}L\})^2, 4 + (2\mathfrak{c})^{-1}(L + |f(0)|)\}$, $(u_d)_{d\in\mathbb{N}} \curvearrowleft (\mathscr{u}_d)_{d\in\mathbb{N}}$, $L \curvearrowleft (2\mathfrak{c})^{-1}L$, $\alpha_0 \curvearrowleft 2$, $\beta_0 \curvearrowleft 2$, $\alpha_1 \curvearrowleft \gamma$, $\beta_1 \curvearrowleft \gamma$, $T \curvearrowleft 2\mathfrak{c}T$, $a \curvearrowleft \mathfrak{a}_\nu$, $f_0 \curvearrowleft \mathscr{f}$, $(f_d)_{d\in\mathbb{N}} \curvearrowleft (\mathbb{R}^d \ni x \mapsto u_d(T,x) \in \mathbb{R})_{d\in\mathbb{N}}$, $(\nu_d)_{d\in\mathbb{N}} \curvearrowleft (\mu_d)_{d\in\mathbb{N}}$ in the notation of Theorem 4.1) establishes that there exist $(\mathbf{U}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ and $c \in (0,\infty)$ which satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that $\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\mathbf{U}_{d,\varepsilon}) \leq c d^c \varepsilon^{-c}$, and

$$\left(\int_{\mathbb{R}^d} \big|\mathscr{u}_d(0,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}_{d,\varepsilon}))(x)\big|^{\mathfrak{q}} \mu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} \leq \varepsilon. \tag{4.124}$$

The fact that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $\mathscr{u}_d(0,x) = u_d(0,x)$ and Jensen's inequality hence imply (4.116). The proof of Corollary 4.15 is thus complete. ◻

Corollary 4.16. *Let $T, \kappa, \mathfrak{c}, p \in (0,\infty)$, $a \in \mathbb{R}$, $b \in (a,\infty)$, let $f\colon \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous, for every $d \in \mathbb{N}$ let $u_d \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) = \mathfrak{c}(\Delta_x u_d)(t,x) + f\big(u_d(t,x)\big), \tag{4.125}$$

let $\nu \in \{0,1\}$, $\alpha \in [0,\infty)\backslash\{1\}$, $\mathfrak{a}_0, \mathfrak{a}_1 \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{a}_0(x) = \max\{x, \alpha x\}$ and $\mathfrak{a}_1(x) = \ln(1 + \exp(x))$, and assume for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that there exists $\mathbf{G} \in \mathbf{N}$ such that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{G}) \in C(\mathbb{R}^d, \mathbb{R}), \qquad \mathcal{P}(\mathbf{G}) \leq \kappa d^\kappa \varepsilon^{-\kappa}, \qquad \text{and} \tag{4.126}$$

$$\varepsilon|u_d(t,x)| + |u_d(0,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{G}))(x)| \leq \varepsilon\kappa d^\kappa (1 + \|x\|)^\kappa \tag{4.127}$$

(cf. Definitions 2.1, 2.3, and 2.6). Then there exists $c \in \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists $\mathbf{U} \in \mathbf{N}$ such that

$$\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}) \in C(\mathbb{R}^d, \mathbb{R}), \qquad \mathcal{P}(\mathbf{U}) \leq c d^c \varepsilon^{-c}, \qquad \text{and} \tag{4.128}$$

$$\left[\tfrac{1}{(b-a)^d} \int_{[a,b]^d} |u_d(T,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}))(x)|^p \, \mathrm{d}x\right]^{1/p} \leq \varepsilon. \tag{4.129}$$

Proof of Corollary 4.16. Let $\mathfrak{p} = \inf\{k \in \mathbb{N} \colon k \geq \max\{\kappa, 2\}\}$ and $\mathfrak{q} = \max\{p, 2\}$. For every $d \in \mathbb{N}$ let $v_d \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$, $t \in [0,T]$ that $v_d(t,x) = u_d(T-t,x)$. Note that (4.125) shows that it holds for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T]$ that

$$(\tfrac{\partial}{\partial t} v_d)(t,x) = -\mathfrak{c}(\Delta_x v_d)(t,x) - f\big(v_d(t,x)\big). \tag{4.130}$$

Observe that (4.127) implies for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\varepsilon|v_d(t,x)| + |v_d(T,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{G}))(x)| \leq \varepsilon\kappa d^\kappa (1 + \|x\|)^\kappa. \tag{4.131}$$

For all $d \in \mathbb{N}$ let $\mu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be the uniform distribution on $[a,b]^d$ and note that there exists $K \in [1,\infty)$ such that for all $d \in \mathbb{N}$ it holds that $\int_{\mathbb{R}^d} \|y\|^{\mathfrak{p}^2\mathfrak{q}} \mu_d(\mathrm{d}y) \leq K^{\mathfrak{p}^2\mathfrak{q}} d^{\mathfrak{p}^2\mathfrak{q}}$. Corollary 4.15 (applied with $p \curvearrowleft \mathfrak{p}$, $\mathfrak{q} \curvearrowleft \mathfrak{q}$, $\gamma \curvearrowleft \kappa$, $r \curvearrowleft 1$, $\kappa \curvearrowleft \max\{\kappa, K^{\mathfrak{p}^2\mathfrak{q}}\}$, $(u_d)_{d\in\mathbb{N}} \curvearrowleft (v_d)_{d\in\mathbb{N}}$ in the notation of Corollary 4.15) establishes that there exists $c \in \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists $\mathbf{U} \in \mathbf{N}$ such that $\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\mathbf{U}) \leq c d^c \varepsilon^{-c}$, and

$$\sup_{q \in (0,\mathfrak{q}]} \left(\int_{\mathbb{R}^d} \big| v_d(0,x) - (\mathcal{R}_{\mathfrak{a}_\nu}(\mathbf{U}))(x) \big|^q \mu_d(\mathrm{d}x) \right)^{1/q} \leq \varepsilon. \tag{4.132}$$

This, the definition of $(\mu_d)_{d\in\mathbb{N}}$, the fact that $p \in (0,\mathfrak{q}]$, and the fact that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $v_d(0,x) = u_d(T,x)$ prove (4.129). The proof of Corollary 4.16 is thus complete. ☐

Acknowledgments

This work has been partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure. In addition, this work has been partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the frame of the priority programme SPP 2298 "Theoretical Foundations of Deep Learning" – Project no. 464123384. Furthermore, funding by the National Science Foundation (NSF 1903450) is gratefully acknowledged. Moreover, the second author gratefully acknowledges the support of the startup fund project of the Shenzhen Research Institute of Big Data under grant No. T00120220001.

References

[1] Baggenstos, J., and Salimova, D. Approximation properties of residual neural networks for Kolmogorov PDEs. *Discrete Contin. Dyn. Syst. Ser. B 28*, 5 (2023), 3193–3215.

[2] Bao, G., Ye, X., Zang, Y., and Zhou, H. Numerical solution of inverse problems by weak adversarial networks. *Inverse Problems 36*, 11 (2020), 115003, 31.

[3] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep splitting method for parabolic PDEs. *SIAM J. Sci. Comput. 43*, 5 (2021), A3135–A3154.

[4] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving the Kolmogorov PDE by means of deep learning. *J. Sci. Comput. 88*, 3 (2021), Paper No. 73, 28.

[5] Beck, C., E, W., and Jentzen, A. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *J. Nonlinear Sci. 29*, 4 (2019), 1563–1619.

[6] Beck, C., Gonon, L., Hutzenthaler, M., and Jentzen, A. On existence and uniqueness properties for solutions of stochastic fixed point equations. *Discrete Contin. Dyn. Syst. Ser. B 26*, 9 (2021), 4927–4962.

[7] Beck, C., Gonon, L., and Jentzen, A. Overcoming the curse of dimensionality in the numerical approximation of high-dimensional semilinear elliptic partial differential equations. *arXiv:2003.00596* (2020), 50 pages.

[8] Beck, C., Hornung, F., Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of Allen-Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. *J. Numer. Math. 28*, 4 (2020), 197–222.

[9] Beck, C., Hutzenthaler, M., and Jentzen, A. On nonlinear Feynman–Kac formulas for viscosity solutions of semilinear parabolic partial differential equations. *Stochastics and Dynamics 21*, 08 (2021), 2150048.

[10] Beck, C., Hutzenthaler, M., Jentzen, A., and Kuckuck, B. An overview on deep learning-based approximation methods for partial differential equations. *Discrete Contin. Dyn. Syst. Ser. B 28*, 6 (2023), 3697–3746.

[11] Becker, S., Braunwarth, R., Hutzenthaler, M., Jentzen, A., and von Wurstemberger, P. Numerical simulations for full history recursive multilevel Picard approximations for systems of high-dimensional partial differential equations. *Commun. Comput. Phys. 28*, 5 (2020), 2109–2138.

[12] Bellman, R. *Dynamic programming.* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2010. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus.

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

[14] Berner, J., Grohs, P., and Jentzen, A. Analysis of the generalization error: empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *SIAM J. Math. Data Sci. 2*, 3 (2020), 631–657.

[15] Blechschmidt, J., and Ernst, O. G. Three ways to solve partial differential equations with neural network—a review. *GAMM-Mitt. 44*, 2 (2021), Paper No. e202100006, 29.

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

[17] Chassagneux, J.-F., Chen, J., and Frikha, N. Deep Runge–Kutta schemes for BSDEs. *arXiv:2212.14372* (2022), 33 pages.

[18] Cioica-Licht, P. A., Hutzenthaler, M., and Werner, P. T. Deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear partial differential equations. *arXiv:2205.14398* (2022), 34 pages.

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

[20] E, W., Han, J., and Jentzen, A. Algorithms for solving high dimensional PDEs: from nonlinear Monte Carlo to machine learning. *Nonlinearity 35*, 1 (2022), 278–310.

[21] E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *J. Sci. Comput. 79*, 3 (2019), 1534–1571.

[22] E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. *Partial Differ. Equ. Appl. 2*, 6 (2021), Paper No. 80, 31.

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

[24] Elbrächter, D., Grohs, P., Jentzen, A., and Schwab, C. DNN expression rate analysis of high-dimensional PDEs: Application to option pricing. *Constr. Approx.* (2021), 1–69.

[25] Feng, D., Yang, Z., and Zou, S. Fractional weak adversarial networks for the stationary fractional advection dispersion equations. *arXiv:2305.19571* (2023), 19 pages.

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

[27] Germain, M., Pham, H., and Warin, X. Approximation error analysis of some deep backward schemes for nonlinear PDEs. *SIAM J. Sci. Comput. 44*, 1 (2022), A28–A56.

[28] Germain, M., Pham, H., and Warin, X. *Neural Networks–Based Algorithms for Stochastic Control and PDEs in Finance.* Cambridge University Press, 2023, pp. 426––452.

[29] Giles, M. B., Jentzen, A., and Welti, T. Generalised multilevel Picard approximations. *arXiv:1911.03188* (2019), 61 pages.

[30] Gonon, L., Grohs, P., Jentzen, A., Kofler, D., and Šiška, D. Uniform error estimates for artificial neural network approximations for heat equations. *IMA J. Numer. Anal. 42*, 3 (2022), 1991–2054.

[31] Gonon, L., and Schwab, C. Deep ReLU network expression rates for option prices in high-dimensional, exponential Lévy models. *Finance Stoch. 25*, 4 (2021), 615–657.

[32] Grohs, P., and Herrmann, L. Deep neural network approximation for high-dimensional parabolic Hamilton–Jacobi–Bellman equations. *arXiv:2103.05744* (2021), 23 pages.

[33] Grohs, P., and Herrmann, L. Deep neural network approximation for high-dimensional elliptic PDEs with boundary conditions. *IMA J. Numer. Anal. 42*, 3 (2022), 2055–2082.

[34] Grohs, P., Hornung, F., Jentzen, A., and von Wurstemberger, P. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *Mem. Amer. Math. Soc. 284*, 1410 (2023), v+93.

[35] Grohs, P., Hornung, F., Jentzen, A., and Zimmermann, P. Space-time error estimates for deep neural network approximations for differential equations. *Adv. Comput. Math. 49*, 1 (2023), Paper No. 4, 78.

[36] Grohs, P., Jentzen, A., and Salimova, D. Deep neural network approximations for solutions of PDEs based on Monte Carlo algorithms. *Partial Differ. Equ. Appl. 3*, 4 (2022), Paper No. 45, 41.

[37] Han, J., Jentzen, A., and E, W. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. USA 115*, 34 (2018), 8505–8510.

[38] Han, J., and Long, J. Convergence of the deep BSDE method for coupled FBSDEs. *Probab. Uncertain. Quant. Risk 5* (2020), Paper No. 5, 33.

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

[40] Horn, R. A., and Johnson, C. R. *Matrix analysis.* Cambridge University Press, Cambridge, 1985.

[41] Hornung, F., Jentzen, A., and Salimova, D. Space-time deep neural network approximations for high-dimensional partial differential equations. *arXiv:2006.02199* (2020), 52 pages.

[42] Hu, Z., Shukla, K., Karniadakis, G. E., and Kawaguchi, K. Tackling the curse of dimensionality with physics–informed neural networks. *arXiv:2307.12306* (2023), 37 pages.

[43] Huré, C., Pham, H., and Warin, X. Deep backward schemes for high-dimensional nonlinear PDEs. *Math. Comp. 89*, 324 (2020), 1547–1579.

[44] Hutzenthaler, M., Jentzen, A., and Kruse, T. Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. *Found. Comput. Math. 22*, 4 (2022), 905–966.

[45] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. Multilevel Picard approximations for high-dimensional semilinear second-order PDEs with Lipschitz nonlinearities. *arXiv:2009.02484* (2020), 37 pages.

[46] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *Partial Differ. Equ. Appl. 1*, 2 (2020), Paper No. 10, 34.

[47] Hutzenthaler, M., Jentzen, A., Kruse, T., Nguyen, T. A., and von Wurstemberger, P. Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *Proc. A. 476*, 2244 (2020), 20190630, 25.

[48] Hutzenthaler, M., Jentzen, A., Kuckuck, B., and Padgett, J. L. Strong L^p-error analysis of nonlinear Monte Carlo approximations for high-dimensional semilinear partial differential equations. *arXiv:2110.08297* (2021), 42 pages. Revision requested from Numerical Algorithms.

[49] Hutzenthaler, M., Jentzen, A., and von Wurstemberger, P. Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *Electron. J. Probab. 25* (2020), Paper No. 101, 73.

[50] Hutzenthaler, M., and Kruse, T. Multilevel Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *SIAM J. Numer. Anal. 58*, 2 (2020), 929–961.

[51] Hutzenthaler, M., Kruse, T., and Nguyen, T. A. Multilevel Picard approximations for McKean-Vlasov stochastic differential equations. *J. Math. Anal. Appl. 507*, 1 (2022), Paper No. 125761, 14.

[52] Hutzenthaler, M., and Nguyen, T. A. Multilevel Picard approximations of high-dimensional semilinear partial differential equations with locally monotone coefficient functions. *Appl. Numer. Math. 181* (2022), 151–175.

[53] Jacquier, A., and Oumgari, M. Deep curve-dependent PDEs for affine rough volatility. *SIAM J. Financial Math. 14*, 2 (2023), 353–382.

[54] Jentzen, A., Salimova, D., and Welti, T. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *Commun. Math. Sci. 19*, 5 (2021), 1167–1205.

[55] Kutyniok, G., Petersen, P., Raslan, M., and Schneider, R. A theoretical analysis of deep neural networks and parametric PDEs. *Constr. Approx. 55*, 1 (2022), 73–125.

[56] Neufeld, A., and Wu, S. Multilevel Picard approximation algorithm for semilinear partial integro-differential equations and its complexity analysis. *arXiv:2205.09639* (2022), 54 pages.

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

[58] Nüsken, N., and Richter, L. Solving high-dimensional Hamilton-Jacobi-Bellman PDEs using neural networks: perspectives from the theory of controlled diffusions and measures on path space. *Partial Differ. Equ. Appl. 2*, 4 (2021), Paper No. 48, 48.

[59] Pham, H., Warin, X., and Germain, M. Neural networks-based backward scheme for fully nonlinear PDEs. *Partial Differ. Equ. Appl. 2*, 1 (2021), Paper No. 16, 24.

[60] Raissi, M., Perdikaris, P., and Karniadakis, G. E. Physics–informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys. 378* (2019), 686–707.

[61] Reisinger, C., and Zhang, Y. Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. *Anal. Appl. (Singap.) 18*, 6 (2020), 951–999.

[62] Simon, M. K. *Probability distributions involving Gaussian random variables: A handbook for engineers and scientists.* Springer Science & Business Media, 2007.

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

[64] Valsecchi Oliva, P., Wu, Y., He, C., and Ni, H. Towards fast weak adversarial training to solve high dimensional parabolic partial differential equations using XNODE-WAN. *J. Comput. Phys. 463* (2022), Paper No. 111233, 17.

[65] Zang, Y., Bao, G., Ye, X., and Zhou, H. Weak adversarial networks for high-dimensional partial differential equations. *J. Comput. Phys. 411* (2020), 109409, 14.