Overcoming the curse of dimensionality in the approximation of high-dimensional nonlinear partial differential equations via deep neural networks in the L^p-sense

Arnulf Jentzen[1,2], Benno Kuckuck[3], and Joshua Lee Padgett[4,5]

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

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

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

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

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

January 26, 2022

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 these methods have generated in the scientific literature 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 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 curse of dimensionality 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 curse of dimensionality 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 spatially in the L^2-sense on $[a, b]^d$ and temporally at time of maturity $t = T$ without the curse of dimensionality 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 DNNs without the curse of dimensionality. It is the key contribution of this article to generalize this result by proving this statement in the L^p-sense with $p \in (0, \infty)$.

Questions and Comments

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; cf., e.g., [1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 19, 24, 25, 26, 29, 32, 35, 36, 37, 38, 39, 40, 42, 44]. 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 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 spite of the numerous highly encouraging numerical simulations for deep learning-based approximation methods for high-dimensional PDEs indicating this potential, there is thus far no mathematical result which proves that one of these methods is indeed capable of overcoming the curse of dimensionality in the numerical approximation of PDEs. However, in the last three years, a number of rigorous mathematical results have appeared in the scientific literature which show that deep neural networks (DNNs) have the expressive power to approximate solutions of high-dimensional PDEs without the curse of dimensionality 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 $\varepsilon \in (0,\infty)$; cf., e.g., [9,14,17,18,20,21,23,28,30,33,34,41]. While the articles [9,14,17,18,20,21,23,28,33,34,41] prove such DNN approximation results for linear PDEs, the article [30] establishes a DNN approximation result for certain nonlinear PDEs. More specifically, Hutzenthaler et al. [30] proves that 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 DNNs spatially in the L^2-sense on $[a,b]^d$ and temporally at time of maturity $t = T$ without the curse of dimensionality 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 DNNs without the curse of dimensionality.

It is the key contribution of this article to generalize this result by proving this statement in the L^p-sense with $p \in (0,\infty)$. 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.

Theorem 1.1. *Let $\gamma \in \mathbb{N}$, $\nu \in \{1,2,3,4\}$, $T,\kappa,p,\alpha \in (0,\infty)$, $\mathfrak{a}_1,\mathfrak{a}_2,\mathfrak{a}_3,\mathfrak{a}_4 \in C(\mathbb{R},\mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{a}_1(x) = (\max\{x,0\})^\gamma$, $\mathfrak{a}_2(x) = x^\gamma$, $\mathfrak{a}_3(x) = \max\{x,\alpha x\}$, and $\mathfrak{a}_4(x) = \ln(1+\exp(x))$, let $\mathbf{A}\colon \cup_{d\in\mathbb{N}}\mathbb{R}^d \to \cup_{d\in\mathbb{N}}\mathbb{R}^d$ satisfy for all $d \in \mathbb{N}$, $x = (x_1,\dots,x_d) \in \mathbb{R}^d$ that $\mathbf{A}(x) = (\mathfrak{a}_\nu(x_1),\dots,\mathfrak{a}_\nu(x_d))$, let $\mathbf{N} = \cup_{L\in\mathbb{N}}\cup_{l_0,l_1,\dots,l_L\in\mathbb{N}}(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$, let $\mathcal{R}\colon \mathbf{N} \to (\cup_{k,l\in\mathbb{N}}C(\mathbb{R}^k,\mathbb{R}^l))$ and $\mathcal{P}\colon \mathbf{N}\to\mathbb{N}$ satisfy for all $L \in \mathbb{N}$, $l_0,l_1,\dots,l_L \in \mathbb{N}$, $\Phi = ((W_1,B_1),(W_2,B_2),\dots,(W_L,B_L)) \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}$, $x_1 \in \mathbb{R}^{l_1}$, $\dots$, $x_L \in \mathbb{R}^{l_L}$ with $\forall\, k \in \{1,2,\dots,L\}\colon x_k = \mathbf{A}(W_k x_{k-1}+B_k)$ that*

$$\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0},\mathbb{R}^{l_L}), \quad (\mathcal{R}(\Phi))(x_0) = W_L x_{L-1} + B_L, \quad \textit{and} \quad \mathcal{P}(\Phi) = \textstyle\sum_{k=1}^{L} l_k(l_{k-1}+1), \tag{1.1}$$

let $(\mathfrak{g}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$, *let* $f\colon \mathbb{R}\to\mathbb{R}$ *be Lipschitz continuous, let* $u_d \in C^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R})$, $d\in\mathbb{N}$, *and assume for all* $d\in\mathbb{N}$, $x=(x_1,\ldots,x_d)\in\mathbb{R}^d$, $\varepsilon\in(0,1]$, $t\in(0,T]$ *that* $\mathcal{R}(\mathfrak{g}_{d,\varepsilon})\in C(\mathbb{R}^d,\mathbb{R})$, $\varepsilon|u_d(t,x)|+|u_d(0,x)-(\mathcal{R}(\mathfrak{g}_{d,\varepsilon}))(x)| \leqslant \varepsilon\kappa d^\kappa(1+\sum_{k=1}^d|x_k|)^\kappa$, $\mathcal{P}(\mathfrak{g}_{d,\varepsilon})\leqslant \kappa d^\kappa\varepsilon^{-\kappa}$, *and*

$$(\tfrac{\partial}{\partial t}u_d)(t,x) = (\Delta_x u_d)(t,x) + f(u_d(t,x)). \tag{1.2}$$

Then there exist $c\in\mathbb{R}$ *and* $(\mathfrak{u}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]}\subseteq\mathbf{N}$ *such that for all* $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ *it holds that* $\mathcal{R}(\mathfrak{u}_{d,\varepsilon})\in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{P}(\mathfrak{u}_{d,\varepsilon})\leqslant cd^c\varepsilon^{-c}$, *and*

$$\left[\int_{[0,1]^d}\left|u_d(T,x)-(\mathcal{R}(\mathfrak{u}_{d,\varepsilon}))(x)\right|^p \mathrm{d}x\right]^{1/p} \leqslant \varepsilon. \tag{1.3}$$

Note #1: We will need to update the descriptions below based on our changes in Theorem 1.1 above. . .

Theorem 1.1 is an immediate consequence of Corollary 4.15 in Section 4.3 below. Corollary 4.15, in turn, follows from Theorem 4.1 which is the main result of this article (see Section 4 below for details). In the following we provide some explanatory comments concerning the mathematical objects appearing in Theorem 1.1 above. The function $\mathbf{A}\colon \cup_{d\in\mathbb{N}}\mathbb{R}^d\to\cup_{d\in\mathbb{N}}\mathbb{R}^d$ in Theorem 1.1 above describes the multidimensional rectifier functions which we employ as activation functions in the approximating DNNs in Theorem 1.1 above. The function $\|\cdot\|\colon \cup_{d\in\mathbb{N}}\mathbb{R}^d\to[0,\infty)$ describes the standard norms on $\mathbb{R}^d$, $d\in\mathbb{N}$, in the sense that for all $d\in\mathbb{N}$ we have that $\|\cdot\|\colon\cup_{d\in\mathbb{N}}\mathbb{R}^d\to[0,\infty)$ restricted to $\mathbb{R}^d$ is nothing but the standard norm on $\mathbb{R}^d$. The set $\mathbf{N}=\cup_{L\in\mathbb{N}}\cup_{l_0,l_1,\ldots,l_L\in\mathbb{N}}(\times_{k=1}^L(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$ in Theorem 1.1 above represents the set of all neural networks which we employ to approximate the solutions of the PDEs under consideration. The function $\mathcal{R}\colon\mathbf{N}\to\cup_{k,l\in\mathbb{N}}C(\mathbb{R}^k,\mathbb{R}^l)$ in Theorem 1.1 above assigns to each neural network its realization function. More specifically, we observe that for every neural network $\Phi\in\mathbf{N}$ we have that $\mathcal{R}(\Phi)\in\cup_{k,l\in\mathbb{N}}C(\mathbb{R}^k,\mathbb{R}^l)$ is the realization function of the neural network Φ with the activation functions being multidimensional versions of the rectifier function provided by $\mathbf{A}\colon\cup_{d\in\mathbb{N}}\mathbb{R}^d\to\cup_{d\in\mathbb{N}}\mathbb{R}^d$. The function $\mathcal{P}\colon\mathbf{N}\to\mathbb{N}$ counts for every neural network $\Phi\in\mathbf{N}$ the number of real parameters employed in Φ. More formally, we note that for every neural network $\Phi\in\mathbf{N}$ we have that $\mathcal{P}(\Phi)\in\mathbb{N}$ is the number of real numbers used to describe the neural network Φ. Furthermore, we observe that $\mathcal{P}(\Phi)$ corresponds to the amount of memory that is needed on a computer to store the neural network $\Phi\in\mathbf{N}$. The real number $T\in(0,\infty)$ in Theorem 1.1 above specifies the time horizon of the PDEs (see (1.2)) whose solutions we intend to approximate by DNNs in (1.3) in Theorem 1.1 above. The real number $\kappa\in(0,\infty)$ in Theorem 1.1 above 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 above is used to specify the way we measure the error between the exact solutions of the PDEs under consideration and their DNN approximations, that is, we measure the error between the exact solutions of the PDEs under consideration and their DNN approximations in the L^p-sense (see (1.3) above for details). In Theorem 1.1 we assume that the initial conditions of the PDEs (see (1.2)) whose solutions we intend to approximate by DNNs without the curse of dimensionality can be approximated by DNNs without the curse of dimensionality. The neural networks $\mathfrak{g}_{d,\varepsilon}\in\mathbf{N}$, $d\in\mathbb{N}$, $\varepsilon\in(0,1]$, serve as such approximating DNNs for the initial conditions of the PDEs (see (1.2)) whose solutions we intend to approximate. In particular, we note that the hypothesis that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $t \in (0,T]$ we have that $\varepsilon|u_d(t,x)| + |u_d(0,x) - (\mathcal{R}(\mathfrak{g}_{d,\varepsilon}))(x)| \leqslant \varepsilon\kappa d^\kappa(1+\|x\|^\kappa)$ in Theorem 1.1 above ensures that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ we have that $(\mathcal{R}(\mathfrak{g}_{d,\varepsilon}))(x)$ c onverges to $u_d(0,x)$ as ε converges to 0. The function $f\colon \mathbb{R} \to \mathbb{R}$ in Theorem 1.1 above specifies the nonlinearity in the PDEs (see (1.2)) whose solutions we intend to approximate by DNNs in Theorem 1.1. The functions $u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, in Theorem 1.1 above describe the exact solutions of the PDEs in (1.2). Observe that the hypothesis that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $t \in (0,T]$ we have that $\varepsilon|u_d(t,x)|+|u_d(0,x)-(\mathcal{R}(\mathfrak{g}_{d,\varepsilon}))(x)| \leqslant \varepsilon\kappa d^\kappa(1+\|x\|^\kappa)$ in Theorem 1.1 above also ensures that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T]$ we have that $|u_d(t,x)| \leqslant \kappa d^\kappa(1+\|x\|^\kappa)$. Note that the fact that for all for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0,T]$ we have that $|u_d(t,x)| \leqslant \kappa d^\kappa(1 + \|x\|^\kappa)$ in particular ensures that the solutions $u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, of (1.2) grow at most polynomially. This polynomial growth of the solutions is employed in order to assure that the solutions of (1.2) 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 there exist neural networks $\mathfrak{u}_{d,\varepsilon} \in \mathbf{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ we have that the L^p-distance between the exact solution $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ at time T of the PDE in (1.2) and the realization $\mathcal{R}(\mathfrak{u}_{d,\varepsilon})\colon \mathbb{R}^d \to \mathbb{R}$ of the neural network $\mathfrak{u}_{d,\varepsilon}$ with respect to the Lebesgue measure on $[0,1]^d$ is bounded by ε and such that the number of parameters of the neural networks $\mathfrak{u}_{d,\varepsilon} \in \mathbf{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, grows at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal $1/\varepsilon$ of the prescribed approximation accuracy $\varepsilon \in (0,1]$. Theorem 1.1 is restricted to measuring the L^p-distance with respect to the Lebesgue measure on $[0,1]^d$ but our more general DNN approximation results in Section 4 below (see Theorem 4.1 and Corollary 4.15 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 DNN approximation results in Section 4 below allow measuring the L^p-distance with respect to the uniform distribution on $[a,b]^d$.

The remainder of this article is organized as follows:

2 Artificial neural network (ANN) calculus

2.1 Structured description of ANNs

Comment #2 from Josh: Consider the following options...

(i) We denote by $\mathbf{N}$ the set given by

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

(ii) We denote by $\mathbf{N}$ the set given by

$$\mathbf{N} = \bigcup_{L\in\mathbb{N}} \bigcup_{l_0,l_1,\dots,l_L\in\mathbb{N}} \big(\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \big) \tag{2.2}$$

(iii) We denote by $\mathbf{N}$ the set given by

$$\mathbf{N} = \bigcup_{L\in\mathbb{N}} \bigcup_{l_0,l_1,\dots,l_L\in\mathbb{N}} \big(\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \big) \tag{2.3}$$

(iv) We denote by $\mathbf{N}$ the set given by

$$\mathbf{N} = \cup_{L\in\mathbb{N}}\cup_{l_0,l_1,\dots,l_L\in\mathbb{N}}\big(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k})\big) \tag{2.4}$$

Comment #3 from Josh: Which of the following look better?

$\Phi \in \big(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k})\big)$, $\Phi \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$, $\Phi \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}))$

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

$$\mathbf{N} = \bigcup_{L\in\mathbb{N}}\bigcup_{l_0,l_1,\dots,l_L\in\mathbb{N}}\big(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k})\big) \tag{2.5}$$

and we denote by $\mathcal{P}\colon \mathbf{N}\to\mathbb{N}$, $\mathcal{L}\colon \mathbf{N}\to\mathbb{N}$, $\mathcal{I}\colon \mathbf{N}\to\mathbb{N}$, $\mathcal{O}\colon \mathbf{N}\to\mathbb{N}$, $\mathcal{H}\colon \mathbf{N}\to\mathbb{N}_0 = \{0,1,2,\dots\}$, $\mathcal{D}\colon \mathbf{N}\to\cup_{L\in\mathbb{N}}\mathbb{N}^L$, and $\mathbb{D}_n\colon \mathbf{N}\to\mathbb{N}_0$, $n\in\mathbb{N}_0$, the functions which satisfy for all $L\in\mathbb{N}$, $l_0,l_1,\dots,l_L\in\mathbb{N}$, $\Phi\in\big(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k})\big)$, $n\in\mathbb{N}_0$ that $\mathcal{P}(\Phi)=\sum_{k=1}^{L} l_k(l_{k-1}+1)$, $\mathcal{L}(\Phi)=L$, $\mathcal{I}(\Phi)=l_0$, $\mathcal{O}(\Phi)=l_L$, $\mathcal{H}(\Phi)=L-1$, $\mathcal{D}(\Phi)=(l_0,l_1,\dots,l_L)$, and

$$\mathbb{D}_n(\Phi) = \begin{cases} l_n & : n\leqslant L \\ 0 & : n>L \end{cases}. \tag{2.6}$$

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

Definition 2.3 (Euclidean 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 $|\!|\!|x|\!|\!| = \max_{i\in\{1,2,\dots,d\}}|x_i|$.

Definition 2.4 (Rectifier function)**.** We denote by $\mathfrak{r}\colon\mathbb{R}\to\mathbb{R}$ the function which satisfies for all $x\in\mathbb{R}$ that $\mathfrak{r}(x)=\max\{x,0\}$.

Definition 2.5 (Multidimensional version)**.** Let $d\in\mathbb{N}$ and let $a\in C(\mathbb{R},\mathbb{R})$. 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

$$\mathfrak{M}_{a,d}(x) = \big(a(x_1),\dots,a(x_d)\big). \tag{2.7}$$

Definition 2.6 (Realization associated to a DNN)**.** 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 $\forall\, k\in\{1,2,\dots,L\}\colon x_k=\mathfrak{M}_{a,l_k}(W_kx_{k-1}+B_k)$ that

$$\mathcal{R}_a(\Phi)\in C(\mathbb{R}^{l_0},\mathbb{R}^{l_L}) \qquad\text{and}\qquad (\mathcal{R}_a(\Phi))(x_0) = W_Lx_{L-1}+B_L \tag{2.8}$$

(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

$$\Phi_1 \bullet \Phi_2 = \begin{cases} \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1 \mathscr{W}_{\mathfrak{L}}, W_1 \mathscr{B}_{\mathfrak{L}} + B_1), \\ \qquad (W_2, B_2), (W_3, B_3), \dots, (W_L, B_L)\big) & \colon L > 1 < \mathfrak{L} \\ \big((W_1 \mathscr{W}_1, W_1 \mathscr{B}_1 + B_1), (W_2, B_2), (W_3, B_3), \dots, (W_L, B_L)\big) & \colon L > 1 = \mathfrak{L} \\ \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1 \mathscr{W}_{\mathfrak{L}}, W_1 \mathscr{B}_{\mathfrak{L}} + B_1)\big) & \colon L = 1 < \mathfrak{L} \\ \big((W_1 \mathscr{W}_1, W_1 \mathscr{B}_1 + B_1)\big) & \colon L = 1 = \mathfrak{L} \end{cases} \tag{2.9}$$

(cf. Definition 2.1).

2.2.1 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

$$\Phi^{\bullet n} = \begin{cases} \big(\mathrm{I}_{\mathcal{O}(\Phi)}, (0, 0, \dots, 0)\big) \in \mathbb{R}^{\mathcal{O}(\Phi) \times \mathcal{O}(\Phi)} \times \mathbb{R}^{\mathcal{O}(\Phi)} & \colon n = 0 \\ \Phi \bullet (\Phi^{\bullet (n-1)}) & \colon n \in \mathbb{N} \end{cases} \tag{2.10}$$

(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 $\mathcal{E}_{L,\Psi} \colon \{\Phi \in \mathbf{N} \colon (\mathcal{L}(\Phi) \leqslant L \text{ and } \mathcal{O}(\Phi) = \mathcal{I}(\Psi))\} \to \mathbf{N}$ the function which satisfies for all $\Phi \in \mathbf{N}$ with $\mathcal{L}(\Phi) \leqslant L$ and $\mathcal{O}(\Phi) = \mathcal{I}(\Psi)$ that

$$\mathcal{E}_{L,\Psi}(\Phi) = (\Psi^{\bullet (L - \mathcal{L}(\Phi))}) \bullet \Phi \tag{2.11}$$

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

2.3 Parallelizations of ANNs

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

$$\mathbf{P}_n \colon \big\{(\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n \colon \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)\big\} \to \mathbf{N} \tag{2.12}$$

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

$\mathbb{R}^{l_{1,k}})\big)$, $\Phi_2 = ((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \ldots, (W_{2,L}, B_{2,L})) \in \big(\times_{k=1}^{L} (\mathbb{R}^{l_{2,k} \times l_{2,k-1}} \times \mathbb{R}^{l_{2,k}})\big), \ldots,$ $\Phi_n = ((W_{n,1}, B_{n,1}), (W_{n,2}, B_{n,2}), \ldots, (W_{n,L}, B_{n,L})) \in \big(\times_{k=1}^{L} (\mathbb{R}^{l_{n,k} \times l_{n,k-1}} \times \mathbb{R}^{l_{n,k}})\big)$ that

$$\begin{aligned} \mathbf{P}_n(\Phi_1, \Phi_2, \ldots, \Phi_n) = \Bigg(&\left(\begin{pmatrix} W_{1,1} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,1} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,1} \end{pmatrix}, \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ B_{3,1} \\ \vdots \\ B_{n,1} \end{pmatrix}\right), \\ &\left(\begin{pmatrix} W_{1,2} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,2} \end{pmatrix}, \begin{pmatrix} B_{1,2} \\ B_{2,2} \\ B_{3,2} \\ \vdots \\ B_{n,2} \end{pmatrix}\right), \ldots, \\ &\left(\begin{pmatrix} W_{1,L} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,L} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,L} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,L} \end{pmatrix}, \begin{pmatrix} B_{1,L} \\ B_{2,L} \\ B_{3,L} \\ \vdots \\ B_{n,L} \end{pmatrix}\right)\Bigg) \end{aligned} \tag{2.13}$$

(cf. Definition 2.1).

2.4 Affine linear transformations as 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 neural network given by $\mathbf{A}_{W,B} = (W, B)$ (cf. Definitions 2.1 and 2.2).

Lemma 2.13. *Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^m$. Then*

(i) it holds that $\mathcal{D}(\mathbf{A}_{W,B}) = (n, m) \in \mathbb{N}^2$,

(ii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\mathbf{A}_{W,B}) \in C(\mathbb{R}^n, \mathbb{R}^m)$, and

(iii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ that $(\mathcal{R}_a(\mathbf{A}_{W,B}))(x) = Wx + B$

(cf. Definitions 2.1, 2.6, and 2.12).

Proof of Lemma 2.13. Observe that the fact that $\mathbf{A}_{W,B} \in (\mathbb{R}^{m \times n} \times \mathbb{R}^m) \subseteq \mathbf{N}$ ensures that $\mathcal{D}(\mathbf{A}_{W,B}) = (n, m) \in \mathbb{N}^2$ (cf. Definitions 2.1 and 2.12). This establishes item (i). Next, note that the fact that $\mathbf{A}_{W,B} = (W, B) \in (\mathbb{R}^{m \times n} \times \mathbb{R}^m)$ and (2.8) assure that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathbf{A}_{W,B}) \in C(\mathbb{R}^n, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathbf{A}_{W,B}))(x) = Wx + B \tag{2.14}$$

(cf. Definition 2.6). This establishes items (ii) and (iii). The proof of Lemma 2.13 is thus complete. □

Lemma 2.14. *Let $\Phi \in \mathbf{N}$ (cf. Definition 2.1). Then*

(i) it holds for all $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathcal{O}(\Phi)}$, $B \in \mathbb{R}^m$ that

$$\mathcal{D}(\mathbf{A}_{W,B} \bullet \Phi) = \big(\mathbb{D}_0(\Phi), \mathbb{D}_1(\Phi), \dots, \mathbb{D}_{\mathcal{L}(\Phi)-1}(\Phi), m\big) \in \mathbb{N}^{\mathcal{L}(\Phi)+1}, \tag{2.15}$$

(ii) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathcal{O}(\Phi)}$, $B \in \mathbb{R}^m$ that $\mathcal{R}_a(\mathbf{A}_{W,B} \bullet \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^m)$,

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

$$(\mathcal{R}_a(\mathbf{A}_{W,B} \bullet \Phi))(x) = W(\mathcal{R}_a(\Phi))(x) + B, \tag{2.16}$$

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

$$\mathcal{D}(\Phi \bullet \mathbf{A}_{W,B}) = \big(n, \mathbb{D}_1(\Phi), \mathbb{D}_2(\Phi), \dots, \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)\big) \in \mathbb{N}^{\mathcal{L}(\Phi)+1}, \tag{2.17}$$

(v) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $W \in \mathbb{R}^{\mathcal{I}(\Phi) \times n}$, $B \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that $\mathcal{R}_a(\Phi \bullet \mathbf{A}_{W,B}) \in C(\mathbb{R}^n, \mathbb{R}^{\mathcal{O}(\Phi)})$, and

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

$$(\mathcal{R}_a(\Phi \bullet \mathbf{A}_{W,B}))(x) = (\mathcal{R}_a(\Phi))(Wx + B) \tag{2.18}$$

(cf. Definitions 2.6, 2.7, and 2.12).

Proof of Lemma 2.14. Observe that Lemma 2.13 proves that for all $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^m$, $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathbf{A}_{W,B}) \in C(\mathbb{R}^n, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathbf{A}_{W,B}))(x) = Wx + B \tag{2.19}$$

(cf. Definitions 2.6 and 2.12). Combining this and, e.g., Grohs et al. [22, Proposition 2.6] establishes items (i), (ii), (iii), (iv), (v), and (vi). The proof of Lemma 2.14 is thus complete. □

2.5 Linear combinations of ANNs

2.5.1 Summations of ANNs

Definition 2.15. 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 neural network given by $\mathfrak{S}_{m,n} = \mathbf{A}_{(\mathrm{I}_m \ \mathrm{I}_m \ \dots \ \mathrm{I}_m),0}$ (cf. Definitions 2.2, 2.8, and 2.12).

Definition 2.16 (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 2.17 (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 neural network given by $\mathfrak{T}_{m,n} = \mathbf{A}_{(\mathrm{I}_m \ \mathrm{I}_m \ \dots \ \mathrm{I}_m)^*,0}$ (cf. Definitions 2.2, 2.8, 2.12, and 2.16).

Definition 2.18 (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 $k \in \mathbb{N} \cap [u, v]$ 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 neural network given by

$$\bigoplus_{k=u}^{v} \Phi_k = \Big(\mathfrak{S}_{\mathcal{O}(\Phi_u), v-u+1} \bullet \big[\mathbf{P}_{v-u+1}(\Phi_u, \Phi_{u+1}, \ldots, \Phi_v)\big] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_u), v-u+1}\Big) \in \mathbf{N} \tag{2.20}$$

(cf. Definitions 2.2, 2.7, 2.11, 2.15, and 2.17).

Definition 2.19 (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 $k \in \mathbb{N} \cap [u, v]$ 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 neural network given by

$$\boxplus_{k=u,\Psi}^{v} \Phi_k = \bigoplus_{k=u}^{v} \mathcal{E}_{\max_{j \in \{u, u+1, \ldots, v\}} \mathcal{L}(\Phi_j), \Psi}(\Phi_k) \in \mathbf{N} \tag{2.21}$$

(cf. Definitions 2.2, 2.10, and 2.18).

2.5.2 Linear combinations of certain ANNs with the same length

Definition 2.20 (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 2.21. *Let $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, $n = v - u + 1$, $h_u, h_{u+1}, \ldots, h_v \in \mathbb{R}$, $\Phi_u, \Phi_{u+1}, \ldots,$ $\Phi_v, \Psi \in \mathbf{N}$, $B_u, B_{u+1}, \ldots, B_v \in \mathbb{R}^{\mathcal{I}(\Phi)}$ satisfy $\mathcal{D}(\Phi_u) = \mathcal{D}(\Phi_{u+1}) = \ldots = \mathcal{D}(\Phi_v)$ and*

$$\Psi = \bigoplus_{k=u}^{v} \Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big) \tag{2.22}$$

(cf. Definitions 2.1, 2.7, 2.8, 2.12, 2.18, and 2.20). Then

(i) it holds that

$$\begin{aligned} \mathcal{D}(\Psi) &= \big(\mathcal{I}(\Phi_u), \textstyle\sum_{k=u}^{v} \mathbb{D}_1(\Phi_u), \sum_{k=u}^{v} \mathbb{D}_2(\Phi_u), \ldots, \sum_{k=u}^{v} \mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big) \\ &= \big(\mathcal{I}(\Phi_u), n\mathbb{D}_1(\Phi_u), n\mathbb{D}_2(\Phi_u), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big), \end{aligned} \tag{2.23}$$

(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

$$(\mathcal{R}_a(\Psi))(x) = \sum_{k=u}^{v} h_k (\mathcal{R}_a(\Phi_k))(x + B_k) \tag{2.24}$$

(cf. Definition 2.6).

Proof of Lemma 2.21. First, note that the hypothesis that $\mathcal{D}(\Phi_u) = \mathcal{D}(\Phi_{u+1}) = \ldots = \mathcal{D}(\Phi_v)$ and item (i) in Lemma 2.13 show that for all $k \in \{u, u+1, \ldots, v\}$ it holds that

$$\mathcal{D}\big(\mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big) = \mathcal{D}\big(\mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_u)}, B_k}\big) = \big(\mathcal{I}(\Phi_u), \mathcal{I}(\Phi_u)\big) \in \mathbb{N}^2. \tag{2.25}$$

This and, e.g., Grohs et al. [22, item (i) in Proposition 2.6] demonstrate that for all $k \in \{u, u+1, \ldots, v\}$ it holds that

$$\mathcal{D}\big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big) = \big(\mathcal{I}(\Phi_u), \mathbb{D}_1(\Phi_u), \mathbb{D}_2(\Phi_u), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_u)}(\Phi_u)\big). \tag{2.26}$$

Observe that this and, e.g., Grohs et al. [23, item (i) in Lemma 3.14] yield that for all $k \in \{u, u+1, \ldots, v\}$ it holds that

$$\mathcal{D}\Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big) = \mathcal{D}\big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big). \tag{2.27}$$

Combining this, (2.26), and, e.g., Grohs et al. [23, item (ii) in Lemma 3.28] establish that

$$\begin{aligned}
\mathcal{D}(\Psi) &= \mathcal{D}\Big(\oplus_{k=u}^{v}\big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\big)\Big) \\
&= \big(\mathcal{I}(\Phi_u), \textstyle\sum_{k=u}^{v} \mathbb{D}_1(\Phi_u), \sum_{k=u}^{v} \mathbb{D}_2(\Phi_u), \ldots, \sum_{k=u}^{v} \mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big) \\
&= \big(\mathcal{I}(\Phi_u), n\mathbb{D}_1(\Phi_u), n\mathbb{D}_2(\Phi_u), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_u), \mathcal{O}(\Phi_u)\big).
\end{aligned} \tag{2.28}$$

This establishes item (i). Furthermore, note that item (vi) and item (v) in Lemma 2.14 imply 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 $\mathcal{R}_a(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$ and

$$\Big(\mathcal{R}_a\big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big)(x) = (\mathcal{R}_a(\Phi_k))(x + B_k) \tag{2.29}$$

(cf. Definition 2.6). Combining this and, e.g., Grohs et al. [23, Lemma 3.14] ensures 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

$$\mathcal{R}_a\Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\Big) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)}) \tag{2.30}$$

and

$$\Big(\mathcal{R}_a\big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}\big)\big)\Big)(x) = h_k(\mathcal{R}_a(\Phi_k))(x + B_k). \tag{2.31}$$

Moreover, observe that, e.g., Grohs et al. [23, Lemma 3.28] and (2.27) assure 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

$$\begin{aligned}
(\mathcal{R}_a(\Psi))(x) &= \Big(\mathcal{R}_a\big(\oplus_{k=u}^{v}(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}))\big)\Big)(x) \\
&= \sum_{k=u}^{v} \Big(\mathcal{R}_a\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k})\big)\Big)(x) = \sum_{k=u}^{v} h_k(\mathcal{R}_a(\Phi_k))(x + B_k).
\end{aligned} \tag{2.32}$$

This establishes items (ii) and (iii). The proof of Lemma 2.21 is thus complete. □

2.5.3 Linear combinations of certain ANNs with different lengths

Lemma 2.22. *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 $j \in \mathbb{N} \cap [u, v]$ 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*

$$\Psi = \mathop{\boxplus}_{k=u, \mathfrak{J}}^{v} \Big(h_k \circledast \big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k} \big) \Big) \tag{2.33}$$

(cf. Definitions 2.1, 2.6, 2.7, 2.8, 2.12, 2.19, and 2.20). Then

(i) it holds that

$$\begin{aligned} &\mathcal{D}(\Psi) \\ &= \Big(\mathcal{I}(\Phi_u), \sum_{k=u}^{v} \mathbb{D}_1\big(\mathcal{E}_{L,\mathfrak{J}}(\Phi_k)\big), \sum_{k=u}^{v} \mathbb{D}_2\big(\mathcal{E}_{L,\mathfrak{J}}(\Phi_k)\big), \ldots, \sum_{k=u}^{v} \mathbb{D}_{L-1}\big(\mathcal{E}_{L,\mathfrak{J}}(\Phi_k)\big), \mathcal{O}(\Phi_u) \Big), \end{aligned} \tag{2.34}$$

(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

$$(\mathcal{R}_a(\Psi))(x) = \sum_{k=u}^{v} h_k (\mathcal{R}_a(\Phi_k))(x + B_k) \tag{2.35}$$

(cf. Definition 2.10).

Proof of Lemma 2.22. Note that item (i) in Lemma 2.21 establishes item (i). In addition, observe that item (vi) and item (v) in Lemma 2.14 prove that for all $k \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ 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

$$\Big(\mathcal{R}_a\big(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k} \big) \Big)(x) = (\mathcal{R}_a(\Phi_k))(x + B_k). \tag{2.36}$$

This, e.g., Grohs et al. [23, Lemma 3.14], and, e.g., Grohs et al. [22, item (ii) in Lemma 2.14] show that for all $k \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ it holds that

$$\mathcal{R}_a\Big(\mathcal{E}_{L,\mathfrak{J}}\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}) \big) \Big) = \mathcal{R}_a\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}) \big) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)}) \tag{2.37}$$

and

$$\begin{aligned} \Big(\mathcal{R}_a\Big(\mathcal{E}_{L,\mathfrak{J}}\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}) \big) \Big) \Big)(x) &= \Big(\mathcal{R}_a\big(h_k \circledast (\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)}, B_k}) \big) \Big)(x) \\ &= h_k (\mathcal{R}_a(\Phi_k))(x + B_k) \end{aligned} \tag{2.38}$$

(cf. Definition 2.10). Combining this, e.g., Grohs et al. [23, Lemma 3.28], and (2.27) demonstrates 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

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

$$= \left(\mathcal{R}_a\left(\bigoplus_{k=u}^{v} \mathcal{E}_{L,\mathfrak{J}}\left(h_k \circledast \left(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)},B_k}\right)\right)\right)\right)(x) \tag{2.39}$$
$$= \sum_{k=u}^{v}\left(\mathcal{R}_a\left(\mathcal{E}_{L,\mathfrak{J}}\left(h_k \circledast \left(\Phi_k \bullet \mathbf{A}_{\mathrm{I}_{\mathcal{I}(\Phi_k)},B_k}\right)\right)\right)\right)(x) = \sum_{k=u}^{v} h_k(\mathcal{R}_a(\Phi_k))(x + B_k)$$

(cf. Definition 2.18). This establishes items (ii) and (iii). The proof of Lemma 2.22 is thus complete. □

3 ANN representations for MLP approximations

3.1 Activation functions as neural networks

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 neural network 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. Note that the fact that $\mathfrak{i}_n \in ((\mathbb{R}^{n\times n} \times \mathbb{R}^n) \times (\mathbb{R}^{n\times n} \times \mathbb{R}^n)) \subseteq \mathbf{N}$ yields that $\mathcal{D}(\mathfrak{i}_n) = (n, n, n) \in \mathbb{N}^3$ (cf. Definitions 2.1 and 3.1). This establishes item (i). Next, observe 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 (2.8) establish 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. Note that Lemma 3.2 implies 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. [22, 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 one-dimensional identity

Definition 3.4 (Identity network)**.** Let $\gamma \in \mathbb{N}_0$. Then we denote by $\mathbf{I}_\gamma \in \mathbf{N}$ the neural network which satisfies

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

(cf. Definitions 2.1 and 2.2).

Lemma 3.5. *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, 2.20, 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. Observe that (3.5) ensures 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). Furthermore, note that (3.5) assures 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). Moreover, observe that item (i) in Lemma 2.21 (applied with $u \curvearrowleft 1$, $v \curvearrowleft 1$, $h_u \curvearrowleft (1+\alpha)^{-1}$, $\Phi_u \curvearrowleft \mathbf{I}_1$, $B_u \curvearrowleft 0$, $\Psi \curvearrowleft \Psi$ in the notation of Lemma 2.21) establishes item (iii). Combining (3.6) and item (iii) in Lemma 2.21 (applied with $u \curvearrowleft 1$, $v \curvearrowleft 1$, $h_u \curvearrowleft (1+\alpha)^{-1}$, $\Phi_u \curvearrowleft \mathbf{I}_1$, $B_u \curvearrowleft 0$, $\Psi \curvearrowleft \Psi$, $a \curvearrowleft a$ in the notation of Lemma 2.21) therefore proves items (iv) and (v). The proof of Lemma 3.5 is thus complete. □

Lemma 3.6. *Let $\gamma \in \mathbb{N} \cap [2, \infty)$, $b_1, b_2, \ldots, b_\gamma \in \mathbb{R}$, $a \in C(\mathbb{R}, \mathbb{R})$ 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

$$\Psi = \mathbf{A}_{1,c_0} \bullet \left(\bigoplus_{i=1}^{\gamma} \left(c_i \circledast \left(\mathfrak{i}_1 \bullet \mathbf{A}_{1,b_i} \right) \right) \right), \tag{3.7}$$

(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, 2.18, 2.20, and 3.1).

Proof of Lemma 3.6. Throughout this proof let $\mathbf{B} = (\mathbf{B}_{i,j})_{i,j \in \{1,2,\ldots,\gamma+1\}} \in \mathbb{R}^{(\gamma+1)\times(\gamma+1)}$ satisfy for all $i, j \in \{1, 2, \ldots, \gamma\}$ that $\mathbf{B}_{1,i+1} = 1$, $\mathbf{B}_{i,1} = 0$, $\mathbf{B}_{\gamma+1,1} = 1$, and $\mathbf{B}_{i+1,j+1} = (b_j)^{i-1}$ and let $\mathbf{D} = (\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{\gamma+1})^* \in \mathbb{R}^{1\times(\gamma+1)}$ satisfy for all $k \in \{1, 2, \ldots, \gamma+1\}$ that $\mathbf{D}_k = \mathbb{1}_{\{\gamma\}}(k)\, \gamma^{-1}$ (cf. Definition 2.16). Note that the assumption that $b_1 < b_2 < \ldots < b_\gamma$ and, e.g., Horn and Johnson [27, Eq. (0.9.11.2)] show that

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

This demonstrates that there exists a unique $\mathbf{C} = (c_0, c_1, \ldots, c_\gamma)^* \in \mathbb{R}^{1\times(\gamma+1)}$ such that $\mathbf{BC} = \mathbf{D}$. This establishes item (i). In addition, observe that Lemma 2.21 and item (i) establish item (ii). Next, note that item (i) in Lemma 2.21 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). Furthermore, observe that item (iii) in Lemma 3.2 and item (iii) in Lemma 2.21 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, 2.18, and 2.20). This and bionomial theorem imply that for all $x \in \mathbb{R}$ it holds that

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

Combining (3.11) and item (i) hence ensures that

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

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

Lemma 3.7. *Let $\gamma \in \mathbb{N} \cap [2,\infty)$, $b_1, b_2, \ldots, b_\gamma \in \mathbb{R}$, $a \in C(\mathbb{R},\mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = (\max\{x,0\})^\gamma$ 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 $c_0, c_1, \ldots, c_\gamma \in \mathbb{R}$ which satisfy for all $k \in \{0,1,\ldots,\gamma\}$ that $\mathbb{1}_{\{\gamma\}}(k)\, c_0 + \sum_{i=1}^{\gamma} c_i (b_i)^k = \mathbb{1}_{\{\gamma-1\}}(k)\, \gamma^{-1}$,

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

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

(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, 2.18, 2.20, and 3.4).

Proof of Lemma 3.7. First, note 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). Moreover, observe that item (i) in Lemma 3.6 establishes item (ii). In addition, note that Lemma 2.21 and item (ii) establish item (iii). Next, observe that item (i) in Lemma 2.21 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). Furthermore, note that item (i) and item (iii) in Lemma 2.21 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, 2.18, and 2.20). This and bionomial theorem demonstrate that for all $x \in \mathbb{R}$ it holds that

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

Combining (3.17) and item (ii) therefore yields that

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

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

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

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

(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

Lemma 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 $U^\theta_n \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that*

$$\begin{aligned}U^\theta_n(t, x) &= \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n} \left[\sum_{k=1}^{M^n} \big(\mathcal{R}_a(\mathbf{G})\big)\big(x + W^{(\theta,0,-k)}_{T-t}\big) \right] \\ &+ \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}} \left[\sum_{k=1}^{M^{n-i}} \Big((\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,i,k)}_i) - \mathbb{1}_{\mathbb{N}}(i) (\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,-i,k)}_{i-1}) \Big) \big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big) \right],\end{aligned} \tag{3.20}$$

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

$$\mathbf{U}^\theta_{n,t} = \left[\bigoplus_{k=1}^{M^n} \Big(\tfrac{1}{M^n} \circledast \big(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, W^{(\theta,0,-k)}_{T-t}} \big) \Big) \right]$$

$$\boxplus_{\mathfrak{I}}\left[\overset{n-1}{\underset{i=0,\mathfrak{I}}{\boxplus}}\left[\left(\tfrac{(T-t)}{M^{n-i}}\right)\circledast\left(\overset{M^{n-i}}{\underset{k=1,\mathfrak{I}}{\boxplus}}\left(\left(\mathbf{F}\bullet\mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}_t^{(\theta,i,k)}}\right)\bullet\mathbf{A}_{\mathrm{I}_d,W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}}\right)\right)\right]\right] \tag{3.21}$$
$$\boxplus_{\mathfrak{I}}\left[\overset{n-1}{\underset{i=0,\mathfrak{I}}{\boxplus}}\left[\left(\tfrac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\right)\circledast\left(\overset{M^{n-i}}{\underset{k=1,\mathfrak{I}}{\boxplus}}\left(\left(\mathbf{F}\bullet\mathbf{U}^{(\theta,-i,k)}_{\max\{i-1,0\},\mathcal{U}_t^{(\theta,i,k)}}\right)\bullet\mathbf{A}_{\mathrm{I}_d,W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}}\right)\right)\right]\right]$$

(cf. Definitions 2.1, 2.6, 2.7, 2.8, 2.12, 2.18, 2.19, and 2.20). 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* $\mathcal{D}(\mathbf{U}^{\theta_1}_{n,t_1})=\mathcal{D}(\mathbf{U}^{\theta_2}_{n,t_2})$,

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

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

$$\left|\!\left|\!\left|\mathcal{D}(\mathbf{U}^{\theta}_{n,t})\right|\!\right|\!\right|\leqslant\max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|\}\left(1+\sqrt{2}\right)^nM^n, \tag{3.22}$$

(iv) 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*

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

$$\mathcal{P}(\mathbf{U}^{\theta}_{n,t})\leqslant 2\big(\mathcal{L}(\mathbf{G})+n\mathcal{H}(\mathbf{F})\big)\Big[\big(1+\sqrt{2}\big)^nM^n\max\{\mathfrak{d},|\!|\!|\mathcal{D}(\mathbf{F})|\!|\!|,|\!|\!|\mathcal{D}(\mathbf{G})|\!|\!|\}\Big]^2 \tag{3.23}$$

(cf. Definition 2.3).

Proof of Lemma 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

$$\Phi^{\theta}_{n,t}=\overset{M^n}{\underset{k=1}{\oplus}}\left(\tfrac{1}{M^n}\circledast\left(\mathbf{G}\bullet\mathbf{A}_{\mathrm{I}_d,W^{(\theta,0,-k)}_{T-t}}\right)\right), \tag{3.24}$$

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

$$\Psi^{\theta,j}_{n,i,t}=\overset{M^{n-i}}{\underset{k=1,\mathfrak{I}}{\boxplus}}\left(\left(\mathbf{F}\bullet\mathbf{U}^{(\theta,(-1)^ji,k)}_{\max\{i-j,0\},\mathcal{U}_t^{(\theta,i,k)}}\right)\bullet\mathbf{A}_{\mathrm{I}_d,W^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}-t}}\right), \tag{3.25}$$

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

$$\Xi^{\theta,j}_{n,t}=\overset{n-1}{\underset{i=0,\mathfrak{I}}{\boxplus}}\left[\left(\tfrac{(-1)^j(T-t)\,\mathbb{1}_{\mathbb{N}}(i+j)}{M^{n-i}}\right)\circledast\Psi^{\theta,j}_{n,i,t}\right], \tag{3.26}$$

let $L^{\theta,j}_{n,i,t}\in\mathbb{N}$, $\theta\in\Theta$, $j\in\{0,1\}$, $n\in\mathbb{N}$, $i\in\{0,1,\dots,n-1\}$, $t\in[0,T]$, 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

$$L^{\theta,j}_{n,i,t}=\max_{k\in\{1,2,\dots,M^{n-i}\}}\mathcal{L}\left(\mathbf{F}\bullet\mathbf{U}^{(\theta,(-1)^ji,k)}_{\max\{i-j,0\},\mathcal{U}_t^{(\theta,i,k)}}\right), \tag{3.27}$$

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

$$\mathbb{L}_n=\max\left\{\mathcal{L}(\mathbf{G}),\max_{(i,j)\in\{0,1,\dots,n-1\}\times\{0,1\}}L^{0,j}_{n,i,0}\right\}. \tag{3.28}$$

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

$$\mathcal{D}(\mathbf{U}^{\theta_1}_{0,t_1}) = (1, d, 1) = \mathcal{D}(\mathbf{U}^{\theta_2}_{0,t_2}). \tag{3.29}$$

Moreover, observe 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

$$\mathcal{L}(\mathbf{U}^\theta_{0,t}) = 1, \qquad |||\mathcal{D}(\mathbf{U}^\theta_{0,t})||| = d, \qquad \text{and} \qquad (\mathcal{R}_a(\mathbf{U}^\theta_{0,t}))(x) = U^\theta_0(t,x) \tag{3.30}$$

(cf. Definition 2.3). Combining (3.29), (3.30), and the fact that the assumption that $\mathcal{R}_{\mathfrak{r}}(\mathbf{G}) \in C(\mathbb{R}^d, \mathbb{R})$ implies that $\max\{\mathfrak{d}, |||\mathcal{D}(\mathbf{F})|||, |||\mathcal{D}(\mathbf{G})|||\} \geqslant d$ hence proves items (i), (ii), (iii), and (iv) in the 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), and (iv) hold true for all $k \in \{0, 1, \ldots, n-1\}$. Note 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 2.21 (applied for every $\theta \in \Theta$, $t \in [0,T]$ with

$$\begin{gathered} u \curvearrowleft 1, \quad v \curvearrowleft M^n, \quad (h_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft (M^{-n})_{k\in\{1,2,\ldots,M^n\}}, \quad \Psi \curvearrowleft \Phi^\theta_{n,t}, \\ (\Phi_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft (\mathbf{G})_{k\in\{1,2,\ldots,M^n\}}, \quad (B_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft (W^{\theta,0,-k}_{T-t})_{k\in\{1,2,\ldots,M^n\}} \end{gathered} \tag{3.31}$$

in the notation of Lemma 2.21) assure that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\mathcal{D}\big(\Phi^\theta_{n,t}\big) = \big(d, M^n\mathbb{D}_1(\mathbf{G}), M^n\mathbb{D}_2(\mathbf{G}), \ldots, M^n\mathbb{D}_{\mathcal{L}(\mathbf{G})-1}(\mathbf{G}), 1\big) = \mathcal{D}\big(\Phi^0_{n,0}\big) \in \mathbb{N}^{\mathcal{L}(\mathbf{G})+1} \tag{3.32}$$

and

$$\big(\mathcal{R}_a(\Phi^\theta_{n,t})\big)(x) = \frac{1}{M^n}\left[\sum_{k=1}^{M^n} \big(\mathcal{R}_a(\mathbf{G})\big)\big(x + W^{(\theta,0,-k)}_{T-t}\big)\right]. \tag{3.33}$$

In addition, observe that the induction hypothesis and, e.g., Grohs et al. [22, item (ii) in Proposition 2.6] prove that for all $\theta \in \Theta$, $j \in \{0,1\}$, $i \in \{0,1,\ldots,n-1\}$, $t \in [0,T]$ it holds that

$$\begin{aligned} L^{\theta,j}_{n,i,t} &= \max_{k\in\{1,2,\ldots,M^{n-i}\}} \mathcal{L}\Big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\Big) \\ &= \max_{k\in\{1,2,\ldots,M^{n-i}\}} \Big[\mathcal{L}(\mathbf{F}) + \mathcal{L}\Big(\mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\Big) - 1\Big] \\ &= \max_{k\in\{1,2,\ldots,M^{n-i}\}} \Big[\mathcal{L}(\mathbf{F}) + \mathcal{L}\Big(\mathbf{U}^0_{\max\{i-j,0\},0}\Big) - 1\Big] \\ &= \max_{k\in\{1,2,\ldots,M^{n-i}\}} \mathcal{L}\big(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-j,0\},0}\big) = L^{0,j}_{n,i,0}. \end{aligned} \tag{3.34}$$

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 2.22 (applied for every $j \in \{0,1\}$, $i \in \{0,1,\ldots,n-1\}$, $\theta \in \Theta$, $t \in [0,T]$ with

$$\begin{gathered} u \curvearrowleft 1, \quad v \curvearrowleft M^{n-i}, \quad \mathfrak{J} \curvearrowleft \mathfrak{J}, \quad (B_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft \big(W^{\theta,0,k}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big)_{k\in\{1,2,\ldots,M^{n-i}\}}, \\ (h_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft (1)_{k\in\{1,2,\ldots,M^{n-i}\}}, \quad L \curvearrowleft L^{0,j}_{n,i,0}, \quad \Psi \curvearrowleft \Psi^{\theta,j}_{n,i,t}, \quad a \curvearrowleft a \\ (\Phi_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft \big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\big)_{k\in\{1,2,\ldots,M^{n-i}\}} \end{gathered} \tag{3.35}$$

in the notation of Lemma 2.22), and, e.g., Grohs et al. [22, 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

$$\begin{aligned}
&\mathcal{D}\big(\Psi^{\theta,j}_{n,i,t}\big) &(3.36)\\
&= \bigg(d, \sum_{k=1}^{M^{n-i}} \mathbb{D}_1\Big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}\big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\big)\Big), \sum_{k=1}^{M^{n-i}} \mathbb{D}_2\Big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}\big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\big)\Big),\\
&\qquad \ldots, \sum_{k=1}^{M^{n-i}} \mathbb{D}_{L^{0,j}_{n,i,0}-1}\Big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}\big(\mathbf{F} \bullet \mathbf{U}^{(\theta,(-1)^j i,k)}_{\max\{i-j,0\},\mathcal{U}^{(\theta,i,k)}_t}\big)\Big), 1\bigg) &(3.37)\\
&= \bigg(d, M^{n-i}\mathbb{D}_1\Big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}\big(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-j,0\},0}\big)\Big), M^{n-i}\mathbb{D}_2\Big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}\big(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-j,0\},0}\big)\Big),\\
&\qquad \ldots, M^{n-i}\mathbb{D}_{L^{0,j}_{n,i,0}-1}\Big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}\big(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-j,0\},0}\big)\Big), 1\bigg) = \mathcal{D}(\Psi^{0,j}_{n,i,0}) \in \mathbb{N}^{L^{0,j}_{n,i,0}+1}
\end{aligned}$$

and

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

Combining (3.36), (3.38), and Lemma 2.22 (applied for every $j \in \{0,1\}$, $\theta \in \Theta$, $t \in [0,T]$ with

$$\begin{aligned}
&u \curvearrowleft 1, \quad v \curvearrowleft M^{n-i}, \quad \mathfrak{J} \curvearrowleft \mathfrak{J}, \quad (\Phi_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft \big(\Psi^{\theta,j}_{n,i,t}\big)_{i\in\{0,1,\ldots,n-1\}}, \quad L \curvearrowleft L^{0,j}_{n,i,0}, \quad (3.39)\\
&(h_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft \Big(\tfrac{(-1)^j (T-t)\, \mathbb{1}_{\mathbb{N}}(i+j)}{M^{n-i}}\Big)_{i\in\{0,1,\ldots,n-1\}}, \quad (B_k)_{k\in\{u,u+1,\ldots,v\}} \curvearrowleft (\mathrm{I}_d)_{k\in\{0,1,\ldots,n-1\}}
\end{aligned}$$

in the notation of Lemma 2.22) demonstrates that for all $j \in \{0,1\}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\mathcal{D}(\Xi^{\theta,j}_{n,t}) &= \bigg(d, \sum_{i=0}^{n-1} \mathbb{D}_1\big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}(\Psi^{\theta,j}_{n,i,t})\big), \sum_{i=0}^{n-1} \mathbb{D}_2\big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}(\Psi^{\theta,j}_{n,i,t})\big),\\
&\qquad \ldots, \sum_{i=0}^{n-1} \mathbb{D}_{L^{0,j}_{n,i,0}-1}\big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}(\Psi^{\theta,j}_{n,i,t})\big), 1\bigg)\\
&= \bigg(d, \sum_{i=0}^{n-1} \mathbb{D}_1\big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}(\Psi^{0,j}_{n,i,0})\big), \sum_{i=0}^{n-1} \mathbb{D}_2\big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}(\Psi^{0,j}_{n,i,0})\big), \qquad (3.40)\\
&\qquad \ldots, \sum_{i=0}^{n-1} \mathbb{D}_{L^{0,j}_{n,i,0}-1}\big(\mathcal{E}_{L^{0,j}_{n,i,0},\mathfrak{J}}(\Psi^{0,j}_{n,i,0})\big), 1\bigg) = \mathcal{D}(\Xi^{0,j}_{n,0}) \in \mathbb{N}^{L^{0,j}_{n,i,0}+1}
\end{aligned}$$

and

$$\big(\mathcal{R}_a(\Xi^{\theta,j}_{n,t})\big)(x) = \sum_{i=0}^{n-1} \Big(\tfrac{(-1)^j (T-t)\, \mathbb{1}_{\mathbb{N}}(i+j)}{M^{n-i}}\Big)\big(\mathcal{R}_a(\Psi^{\theta,j}_{n,i,t})\big)(x) \qquad (3.41)$$

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

Next, note that (3.32), (3.36), (3.40), item (i) in Lemma 2.22 (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 2.22), and, e.g., Grohs et al. [22, item (i) in Proposition 2.6] yield 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) \\
&= \mathcal{D}\big(\Phi^0_{n,0} \boxplus_{\mathfrak{J}} \Xi^{0,0}_{n,0} \boxplus_{\mathfrak{J}} \Xi^{0,1}_{n,0}\big) = \mathcal{D}(\mathbf{U}^0_{n,0}) \in \mathbb{N}^{\mathbb{L}_n+1}.
\end{aligned} \tag{3.42}$$

Next, observe that (3.24), (3.25), (3.26), (3.33), (3.38), (3.41), and item (iii) in Lemma 2.22 (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 2.22) establish that for all $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\big(\mathcal{R}_a(\mathbf{U}^\theta_{n,t})\big)(x) &= \big(\mathcal{R}_a(\Phi^\theta_{n,t} \boxplus_{\mathfrak{J}} \Xi^{\theta,0}_{n,t} \boxplus_{\mathfrak{J}} \Xi^{\theta,1}_{n,t})\big)(x) \\
&= \big(\mathcal{R}_a(\Phi^\theta_{n,t})\big)(x) + \big(\mathcal{R}_a(\Xi^{\theta,0}_{n,t})\big)(x) + \big(\mathcal{R}_a(\Xi^{\theta,1}_{n,t})\big)(x) \\
&= \tfrac{1}{M^n} \Bigg[\sum_{k=1}^{M^n} \big(\mathcal{R}_a(\mathbf{G})\big)\big(x + W^{(\theta,0,-k)}_{T-t}\big) \Bigg] \\
&\quad + \sum_{i=0}^{n-1} \tfrac{(T-t)}{M^{n-i}} \Bigg[\sum_{k=1}^{M^{n-i}} \big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,i,k)}_i\big)\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big) \Bigg] \\
&\quad + \sum_{i=0}^{n-1} \tfrac{(t-T)\, \mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}} \Bigg[\sum_{k=1}^{M^{n-i}} \big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,-i,k)}_{\max\{i-1,0\}}\big)\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big) \Bigg] \\
&= \tfrac{1}{M^n} \Bigg[\sum_{k=1}^{M^n} \big(\mathcal{R}_a(\mathbf{G})\big)\big(x + W^{(\theta,0,-k)}_{T-t}\big) \Bigg] \\
&\quad + \sum_{i=0}^{n-1} \tfrac{(T-t)}{M^{n-i}} \Bigg[\sum_{k=1}^{M^{n-i}} \big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,i,k)}_i\big)\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big) \Bigg] \\
&\quad + \sum_{i=0}^{n-1} \tfrac{(t-T)\, \mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}} \Bigg[\sum_{k=1}^{M^{n-i}} \big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,-i,k)}_{i-1}\big)\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big) \Bigg] \\
&= U^\theta_n(t, x).
\end{aligned} \tag{3.43}$$

Furthermore, note that (3.42) and Jensen's inequality imply that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that

$$
\begin{aligned}
&\left\|\!\left|\mathcal{D}(\mathbf{U}^{\theta}_{n,t})\right|\!\right\| = \left\|\!\left|\mathcal{D}(\mathbf{U}^{0}_{n,0})\right|\!\right\| \\
&= \max\Big\{\mathfrak{d}, \max_{k\in\{1,2,\dots,\mathbb{L}_n-1\}}\Big(\mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big) + \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big) + \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big)\Big)\Big\} \\
&\leqslant \max_{k\in\{1,2,\dots,\mathbb{L}_n-1\}}\Big[\max\big\{\mathfrak{d}, \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big)\big\} + \max\big\{\mathfrak{d}, \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big)\big\} \\
&\qquad + \max\big\{\mathfrak{d}, \mathbb{D}_k\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big)\big\}\Big] \\
&\leqslant \left\|\!\left|\mathcal{D}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big)\right|\!\right\| + \left\|\!\left|\mathcal{D}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big)\right|\!\right\| + \left\|\!\left|\mathcal{D}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big)\right|\!\right\|.
\end{aligned}
\tag{3.44}
$$

This, (3.32), (3.36), (3.40), (3.42), the induction hypothesis, Jensen's inequality, and, e.g., Grohs et al. [22, item (i) in Proposition 2.6] ensure that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that

$$
\begin{aligned}
&\left\|\!\left|\mathcal{D}(\mathbf{U}^{\theta}_{n,t})\right|\!\right\| = \left\|\!\left|\mathcal{D}(\mathbf{U}^{0}_{n,0})\right|\!\right\| \leqslant \left\|\!\left|\mathcal{D}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Phi^0_{n,0})\big)\right|\!\right\| + \left\|\!\left|\mathcal{D}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,0}_{n,0})\big)\right|\!\right\| + \left\|\!\left|\mathcal{D}\big(\mathcal{E}_{\mathbb{L}_n,\mathfrak{J}}(\Xi^{0,1}_{n,0})\big)\right|\!\right\| \\
&\leqslant \max\{\mathfrak{d}, M^n \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\} + \max\Big\{\mathfrak{d}, \sum_{i=0}^{n-1}\left\|\!\left|\mathcal{D}\big(\Psi^{0,0}_{n,i,0}\big)\right|\!\right\|\Big\} + \max\Big\{\mathfrak{d}, \sum_{i=0}^{n-1}\left\|\!\left|\mathcal{D}\big(\Psi^{0,1}_{n,i,0}\big)\right|\!\right\|\Big\} \\
&\leqslant \max\{\mathfrak{d}, M^n \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\} + \max\Big\{\mathfrak{d}, \sum_{i=0}^{n-1} M^{n-i}\left\|\!\left|\mathcal{D}\big(\mathbf{F}\bullet\mathbf{U}^0_{i,0}\big)\right|\!\right\|\Big\} \\
&\quad + \max\Big\{\mathfrak{d}, \sum_{i=1}^{n-1} M^{n-i}\left\|\!\left|\mathcal{D}\big(\mathbf{F}\bullet\mathbf{U}^0_{\max\{i-1,0\},0}\big)\right|\!\right\|\Big\} \\
&\leqslant \max\{\mathfrak{d}, M^n \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\} \\
&\quad + \sum_{i=0}^{n-1} M^{n-i}\Big[\max\big\{\mathfrak{d}, \left\|\!\left|\mathcal{D}\big(\mathbf{F}\bullet\mathbf{U}^0_{i,0}\big)\right|\!\right\|\big\} + \max\big\{\mathfrak{d}, \left\|\!\left|\mathcal{D}\big(\mathbf{F}\bullet\mathbf{U}^0_{\max\{i-1,0\},0}\big)\right|\!\right\|\big\}\Big] \\
&\leqslant \max\{\mathfrak{d}, M^n \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\} \\
&\quad + \sum_{i=0}^{n-1} M^{n-i}\Big[\max\big\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{F})\right|\!\right\|, \left\|\!\left|\mathcal{D}\big(\mathbf{U}^0_{i,0}\big)\right|\!\right\|\big\} + \max\big\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{F})\right|\!\right\|, \left\|\!\left|\mathcal{D}\big(\mathbf{U}^0_{\max\{i-1,0\},0}\big)\right|\!\right\|\big\}\Big] \\
&\leqslant \big(\max\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\} + n\max\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{F})\right|\!\right\|\}\big) M^n \\
&\quad + \sum_{i=0}^{n-1} M^{n-i}\Big[\left\|\!\left|\mathcal{D}\big(\mathbf{U}^0_{i,0}\big)\right|\!\right\| + \left\|\!\left|\mathcal{D}\big(\mathbf{U}^0_{\max\{i-1,0\},0}\big)\right|\!\right\|\Big].
\end{aligned}
\tag{3.45}
$$

Combining this and, e.g., Hutzenthaler et al. [31, Corollary 4.3] (applied with $\gamma \hookleftarrow 0$, $\beta \hookleftarrow M$, $\alpha_0 \hookleftarrow \max\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\}$, $\alpha_1 \hookleftarrow \max\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{F})\right|\!\right\|\}$, $(x_i)_{i\in\{0,1,\dots,k\}} \hookleftarrow (\left\|\!\left|\mathcal{D}(\mathbf{U}^0_{i,0})\right|\!\right\|)_{i\in\{0,1,\dots,n\}}$ in the notation of Hutzenthaler et al. [31, Corollary 4.3]) assures that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that

$$
\begin{aligned}
\left\|\!\left|\mathcal{D}(\mathbf{U}^{\theta}_{n,t})\right|\!\right\| &\leqslant \tfrac{1}{2}\big(\max\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\} + \max\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{F})\right|\!\right\|\}\big)\big(1+\sqrt{2}\big)^n M^n \\
&\leqslant \max\{\mathfrak{d}, \left\|\!\left|\mathcal{D}(\mathbf{F})\right|\!\right\|, \left\|\!\left|\mathcal{D}(\mathbf{G})\right|\!\right\|\}\big(1+\sqrt{2}\big)^n M^n.
\end{aligned}
\tag{3.46}
$$

Moreover, observe that the induction hypothesis, (3.28), (3.42), Jensen's inequality, and, e.g., Grohs et al. [22, item (ii) in Proposition 2.6] prove that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that

$$
\mathcal{L}(\mathbf{U}^{\theta}_{n,t}) = \mathcal{L}(\mathbf{U}^{0}_{n,0}) = \mathbb{L}_n = \max\Big\{\mathcal{L}(\mathbf{G}), \max_{(i,j)\in\{0,1,\dots,n-1\}\times\{0,1\}} \mathcal{L}\big(\mathbf{F}\bullet\mathbf{U}^0_{\max\{i-j,0\},0}\big)\Big\}
$$

$$
\begin{aligned}
&= \max\Big\{\mathcal{L}(\mathbf{G}), \max_{(i,j)\in\{0,1,\ldots,n-1\}\times\{0,1\}}\Big(\mathcal{L}(\mathbf{F}) + \mathcal{L}\big(\mathbf{U}^0_{\max\{i-j,0\},0}\big) - 1\Big)\Big\}\\
&= \max\Big\{\mathcal{L}(\mathbf{G}), \mathcal{H}(\mathbf{F}) + \max_{(i,j)\in\{0,1,\ldots,n-1\}\times\{0,1\}} \mathcal{L}\big(\mathbf{U}^0_{\max\{i-j,0\},0}\big)\Big\}\\
&\leqslant \max\Big\{\mathcal{L}(\mathbf{G}), \mathcal{H}(\mathbf{F}) + \max_{(i,j)\in\{0,1,\ldots,n-1\}\times\{0,1\}}\Big(\max\{\mathfrak{d}, \mathcal{L}(\mathbf{G})\} + \max\{i-j,0\}\mathcal{H}(\mathbf{F})\Big)\Big\}\\
&= \max\Big\{\mathcal{L}(\mathbf{G}), \mathcal{H}(\mathbf{F}) + \big[\max\{\mathfrak{d}, \mathcal{L}(\mathbf{G})\} + (n-1)\mathcal{H}(\mathbf{F})\big]\Big\} \leqslant \max\{\mathfrak{d}, \mathcal{L}(\mathbf{G})\} + n\mathcal{H}(\mathbf{F}).
\end{aligned}
\tag{3.47}
$$

Combining (3.42), (3.43), (3.46), and (3.47) completes the induction step. Induction hence establishes items (i), (ii), (iii), and (iv). In addition, note that item (ii) and item (iii) show that for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$ it holds that

$$
\begin{aligned}
\mathcal{P}(\mathbf{U}^\theta_{n,t}) &\leqslant \sum_{k=1}^{\mathcal{L}(\mathbf{U}^\theta_{n,t})} |||\mathcal{D}(\mathbf{U}^\theta_{n,t})|||\Big[|||\mathcal{D}(\mathbf{U}^\theta_{n,t})||| + 1\Big] \leqslant 2\mathcal{L}(\mathbf{U}^\theta_{n,t})|||\mathcal{D}(\mathbf{U}^\theta_{n,t})|||^2\\
&\leqslant 2\big(\mathcal{L}(\mathbf{G}) + n\mathcal{H}(\mathbf{F})\big)\Big(\max\big\{\mathfrak{d}, |||\mathcal{D}(\mathbf{F})|||, |||\mathcal{D}(\mathbf{G})|||\big\}\Big)^2\big(1+\sqrt{2}\big)^{2n} M^{2n}.
\end{aligned}
\tag{3.48}
$$

This establishes item (v). The proof of Lemma 3.9 is thus complete. □

4 ANN approximations for PDEs

4.1 ANN approximation results with general activation functions

Theorem 4.1. *Let $p, q, r, L, C, \alpha_0, \alpha_1, \beta_0, \beta_1, T \in [0,\infty)$, $\mathfrak{q} \in [2,\infty)$, $a \in C(\mathbb{R},\mathbb{R})$, $\mathfrak{I} \in \mathbf{N}$, $(\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}_0\times(0,1]} \subseteq \mathbf{N}$, 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, and assume for all $d \in \mathbb{N}_0$, $v, w \in \mathbb{R}$, $x \in \mathbb{R}^{\max\{d,1\}}$, $\varepsilon \in (0,1]$ that $(\int_{\mathbb{R}^d}\|y\|^{pq\mathfrak{q}}\,\nu_d(dy))^{1/(pq\mathfrak{q})} \leqslant Cd^r$, $\mathcal{H}(\mathfrak{I}) = 1$, $\mathcal{R}_a(\mathfrak{I}) = \mathrm{id}_\mathbb{R}$, $\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}) \in C(\mathbb{R}^{\max\{d,1\}},\mathbb{R})$, $\max\{|f_0(v)-f_0(w)|, |(\mathcal{R}_a(\mathbf{F}_{0,\varepsilon}))(v) - (\mathcal{R}_a(\mathbf{F}_{0,\varepsilon}))(w)|\} \leqslant L|v-w|$, $\varepsilon^{\alpha_{\min\{d,1\}}}\mathcal{L}(\mathbf{F}_{d,\varepsilon}) + \varepsilon^{\beta_{\min\{d,1\}}}|||\mathcal{D}(\mathbf{F}_{d,\varepsilon})||| \leqslant C(\max\{d,1\})^p$, and*

$$
\varepsilon|(\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}))(x)| + |f_d(x) - (\mathcal{R}_a(\mathbf{F}_{d,\varepsilon}))(x)| \leqslant \varepsilon C(\max\{d,1\})^p(1+\|x\|)^{pq} \tag{4.1}
$$

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

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

$$
(\tfrac{\partial}{\partial t}u_d)(t,x) + \mathfrak{c}(\Delta_x u_d)(t,x) + f\big(u_d(t,x)\big) = 0 \tag{4.2}
$$

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

(ii) there exist $(\mathbf{U}_{d,t,\varepsilon})_{(d,t,\varepsilon)\in\mathbb{N}\times[0,T]\times(0,1]} \subseteq \mathbf{N}$ and $\eta = (\eta_\delta)_{\delta\in(0,\infty)}\colon \mathbb{R}\to\mathbb{R}$ such that for all $d \in \mathbb{N}$, $t \in [0,T]$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ it holds that $\mathcal{R}_a(\mathbf{U}_{d,t,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{P}(\mathbf{U}_{d,t,\varepsilon}) \leqslant \eta_\delta d^{p(7+4q+(2+q)\delta)}\varepsilon^{-(4+2\delta+\max\{\alpha_0,\alpha_1\}+2\max\{\beta_0,\beta_1\})}$ and

$$
\left(\int_{\mathbb{R}^d}\big|u_d(t,x) - (\mathcal{R}_a(\mathbf{U}_{d,t,\varepsilon}))(x)\big|^{\mathfrak{q}}\,\nu_d(dx)\right)^{1/\mathfrak{q}} \leqslant \varepsilon. \tag{4.3}
$$

Note #4: Update the proof of Theorem 4.1...

Proof of Theorem 4.1. Throughout this proof let $\mathfrak{B} \in [0,\infty)$, $(m_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ 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 $m_{k+1} \leqslant \mathfrak{B} m_k$, 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, 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 $v_{d,\varepsilon} \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$v_{d,\varepsilon}(t,x) = \mathbb{E}\Big[\big(\mathcal{R}_a(\mathbf{G}_{d,\varepsilon})\big)(x + W^{d,0}_{T-t})\Big] + \int_t^T \mathbb{E}\Big[\big((\mathcal{R}_a(\mathbf{F}_\varepsilon)\big)(v_{d,\varepsilon}(s, x + W^{d,0}_{s-t}))\Big]\,\mathrm{d}s, \tag{4.4}$$

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

$$\begin{aligned} U^{d.\theta}_{n,j,\varepsilon}(t,x) &= \frac{\mathbb{1}_{\mathbb{N}}(n)}{(m_j)^n}\Bigg[\sum_{k=1}^{(m_j)^n} \big(\mathcal{R}_a(\mathbf{G}_{d,\varepsilon})\big)\big(x + W^{d,(\theta,0,-k)}_{T-t}\big)\Bigg] \\ &\quad + \sum_{i=0}^{n-1} \frac{(T-t)}{(m_j)^{n-i}}\Bigg[\sum_{k=1}^{(m_j)^{n-i}} \Big[\big((\mathcal{R}_a(\mathbf{F}_\varepsilon)\big)\big(U^{d,(\theta,i,k)}_{i,j,\varepsilon}\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{d,(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big)\big) \\ &\quad - \mathbb{1}_{\mathbb{N}}(i)\big((\mathcal{R}_a(\mathbf{F}_\varepsilon)\big)\big(U^{d,(\theta,-i,k)}_{i-1,j,\varepsilon}\big(\mathcal{U}^{(\theta,i,k)}_t, x + W^{d,(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}\big)\big)\Big]\Bigg], \end{aligned} \tag{4.5}$$

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

$$\begin{aligned} \mathbf{U}^{d,\theta,\varepsilon}_{n,j,t} &= \Bigg[\bigoplus_{k=1}^{(m_j)^n} \Big(\tfrac{1}{(m_j)^n} \circledast \big(\mathbf{G}_{d,\varepsilon} \bullet \mathbf{A}_{\mathrm{I}_d, W^{d,(\theta,0,-k)}_{T-t}}\big)\Big)\Bigg] \\ &\quad \boxplus_{\mathfrak{J}} \Bigg[\mathop{\boxplus}_{i=0,\mathfrak{J}}^{n-1}\Bigg[\Big(\tfrac{(T-t)}{(m_j)^{n-i}}\Big) \circledast \Bigg(\mathop{\boxplus}_{k=1,\mathfrak{J}}^{(m_j)^{n-i}} \Big(\big(\mathbf{F}_\varepsilon \bullet \mathbf{U}^{d,(\theta,i,k),\varepsilon}_{i,j,\mathcal{U}^{(\theta,i,k)}_t}\big) \bullet \mathbf{A}_{\mathrm{I}_d, W^{d,(\theta,i,k)}_{\mathcal{U}^{d,(\theta,i,k)}_t - t}}\Big)\Bigg)\Bigg]\Bigg] \\ &\quad \boxplus_{\mathfrak{J}} \Bigg[\mathop{\boxplus}_{i=0,\mathfrak{J}}^{n-1}\Bigg[\Big(\tfrac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{(m_j)^{n-i}}\Big) \circledast \Bigg(\mathop{\boxplus}_{k=1,\mathfrak{J}}^{(m_j)^{n-i}} \Big(\big(\mathbf{F}_\varepsilon \bullet \mathbf{U}^{d,(\theta,-i,k),\varepsilon}_{\max\{i-1,0\},j,\mathcal{U}^{(\theta,i,k)}_t}\big) \bullet \mathbf{A}_{\mathrm{I}_d, W^{d,(\theta,i,k)}_{\mathcal{U}^{(\theta,i,k)}_t - t}}\Big)\Bigg)\Bigg]\Bigg] \end{aligned} \tag{4.6}$$

(cf. Lemma 3.9), let $c_d \in [1,\infty)$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$ that

$$\begin{aligned} c_d &= (C+1)d^p\big(e^{LT}(T+1)\big)^{q+1}\big((C+1)^q d^{pq} + 1\big) \\ &\quad \cdot \Bigg[1 + \bigg(\int_{\mathbb{R}^d} \|x\|^{pq\mathfrak{q}}\, \nu_d(\mathrm{d}x)\bigg)^{1/(pq\mathfrak{q})} + \Big(\mathbb{E}\big[\|W^{d,0}_T\|^{pq\mathfrak{q}}\big]\Big)^{1/(pq\mathfrak{q})}\Bigg]^{pq}, \end{aligned} \tag{4.7}$$

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

$$B_{d,\varepsilon} = \max\big\{\mathfrak{d}, |||\mathcal{D}(\mathbf{F}_\varepsilon)|||, |||\mathcal{D}(\mathbf{G}_{d,\varepsilon})|||\big\}, \tag{4.8}$$

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 $\delta_{d,\varepsilon} = \varepsilon/(c_d+1)$, and assume without loss of generality that $\max\{|f(0)|+1, \mathfrak{d}\} \leqslant C$ (cf. Definitions 2.7, 2.8, 2.12, 2.18, 2.19, and 2.20). 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 $\varepsilon|(\mathcal{R}_a(\mathbf{G}_{d,\varepsilon}))(x)| + |g_d(x) - (\mathcal{R}_a(\mathbf{G}_{d,\varepsilon}))(x)| \leqslant \varepsilon C d^p(1+\|x\|)^{pq}$ demonstrate that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ it holds that

$$|g_d(x)| \leqslant |g_d(x) - (\mathcal{R}_a(\mathbf{G}_{d,\varepsilon}))(x)| + |(\mathcal{R}_a(\mathbf{G}_{d,\varepsilon}))(x)| \leqslant \varepsilon C d^p(1+\|x\|)^{pq} + C d^p(1+\|x\|)^{pq}. \tag{4.9}$$

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

$$|g_d(x)| \leqslant (C+1)d^p(1+\|x\|)^{pq}. \tag{4.10}$$

Combining this, the assumption that for all $w, z \in \mathbb{R}$ it holds that $|f(w)-f(z)| \leqslant L|w-z|$, and Beck et al. [4, Corollary 3.10] (applied for every $d \in \mathbb{N}$ with $d \curvearrowleft d$, $m \curvearrowleft d$, $L \curvearrowleft L$, $T \curvearrowleft T$, $\mu \curvearrowleft (\mathbb{R}^d \ni x \mapsto (0,0,\dots,0) \in \mathbb{R}^d)$, $\sigma \curvearrowleft \mathrm{I}_d$, $f \curvearrowleft ([0,T] \times \mathbb{R}^d \times \mathbb{R} \ni (t,x,w) \mapsto f(w) \in \mathbb{R})$, $g \curvearrowleft g_d$, $W \curvearrowleft W^{d,0}$ in the notation of Beck et al. [4, Corollary 3.10]) establishes item (i). Next, note that the fact that for all $d \in \mathbb{N}$ the random variable $\|W^d_T/\sqrt{T}\|^2$ is a chi-squared distributed random variable with d-degrees of freedom, Jensen's inequality, and, e.g., Simon [43, Eq. (2.35)] establish that for all $d \in \mathbb{N}$ it holds that

$$\Big(\mathbb{E}\big[\|W^{d,0}_T\|^{pq\mathfrak{q}}\big]\Big)^2 \leqslant \mathbb{E}\big[\|W^{d,0}_T\|^{2pq\mathfrak{q}}\big] = (2T)^{pq\mathfrak{q}}\left[\frac{\Gamma(\frac{d}{2}+pq\mathfrak{q})}{\Gamma(\frac{d}{2})}\right] = (2T)^{pq\mathfrak{q}}\left[\prod_{k=0}^{pq\mathfrak{q}-1}\big(\tfrac{d}{2}+k\big)\right]. \tag{4.11}$$

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

$$\Big(\mathbb{E}\big[\|W^{d,0}_T\|^{pq\mathfrak{q}}\big]\Big)^{1/(pq\mathfrak{q})} \leqslant \sqrt{2T}\left[\prod_{k=0}^{pq\mathfrak{q}-1}\big(\tfrac{d}{2}+k\big)\right]^{1/(2pq\mathfrak{q})} \leqslant \sqrt{2T\big(\tfrac{d}{2}+pq\mathfrak{q}-1\big)}. \tag{4.12}$$

Combining this and the assumption that for all $d \in \mathbb{N}$ it holds that $(\int_{\mathbb{R}^d}\|x\|^{pq\mathfrak{q}}\,\nu_d(\mathrm{d}x))^{1/(pq\mathfrak{q})} \leqslant Cd^r$ ensures that there exists $\overline{C} \in [0,\infty)$ such that for all $d \in \mathbb{N}$ it holds that

$$c_d \leqslant \overline{C}\big[\tfrac{1}{3}(1+d^r+\sqrt{d})\big]^{pq\mathfrak{q}} \leqslant \overline{C}d^{(rpq+1)\mathfrak{q}}. \tag{4.13}$$

Furthermore, observe that the triangle inequality assures that for all $n \in \mathbb{N}_0$, $d \in \mathbb{N}$, $\delta \in (0,1]$, $t \in [0,T]$ it holds that

$$\begin{aligned}&\bigg(\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|u_d(t,x)-U^{d,0}_{n,j,\delta}(t,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\bigg)^{1/\mathfrak{q}} \\ &\leqslant \bigg(\int_{\mathbb{R}^d}\big|u_d(t,x)-v_{d,\delta}(t,x)\big|^{\mathfrak{q}}\,\nu_d(\mathrm{d}x)\bigg)^{1/\mathfrak{q}} + \bigg(\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|v_{d,\delta}(t,x)-U^{d,0}_{n,j,\delta}(t,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\bigg)^{1/\mathfrak{q}}.\end{aligned} \tag{4.14}$$

Moreover, note that the assumption that for all $w \in \mathbb{R}$, $\varepsilon \in (0,1]$ it holds that $|(\mathcal{R}_a(\mathbf{F}_\varepsilon))(w) - f(w)| \leqslant C\varepsilon\max\{1,|w|^q\}$ proves that for all $\varepsilon \in (0,1]$ it holds that

$$|(\mathcal{R}_a(\mathbf{F}_\varepsilon))(0)| \leqslant |(\mathcal{R}_a(\mathbf{F}_\varepsilon))(0) - f(0)| + |f(0)| \leqslant \varepsilon + |f(0)| \leqslant C. \tag{4.15}$$

In addition, 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 $|f(w) - (\mathcal{R}_a(\mathbf{F}_\delta))(w)| \leqslant C\delta \max\{1, |w|^q\}$ and $|g_d(x) - (\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)| \leqslant \delta C d^p (1 + \|x\|)^{pq}$ shows that for all $d \in \mathbb{N}$, $w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\begin{aligned} &\max\bigl\{|f(w) - (\mathcal{R}_a(\mathbf{F}_\delta))(w)|, |g_d(x) - (\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)|\bigr\} \\ &\quad \leqslant \max\bigl\{C\delta(1 + |w|^q), \delta C d^p (1 + \|x\|)^{pq}\bigr\} \leqslant \delta(C+1)d^p\bigl((1 + \|x\|)^{pq} + |w|^q\bigr). \end{aligned} \tag{4.16}$$

Combining this, (4.10), (4.15), 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 $\max\{|f(w) - f(z)|, |(\mathcal{R}_a(\mathbf{F}_\delta))(w) - (\mathcal{R}_a(\mathbf{F}_\delta))(z)|\} \leqslant L|w - z|$ and $\delta|(\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)| + |g_d(x) - (\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)| \leqslant \delta C d^p (1 + \|x\|)^{pq}$, and Hutzenthaler et al. [30, Lemma 2.3] (applied for every $d \in \mathbb{N}$, $\delta \in (0,1]$ with $f_1 \leftharpoondown f$, $f_2 \leftharpoondown \mathcal{R}_a(\mathbf{F}_\delta)$, $g_1 \leftharpoondown g_d$, $g_2 \leftharpoondown \mathcal{R}_a(\mathbf{G}_{d,\delta})$, $T \leftharpoondown T$, $L \leftharpoondown L$, $B \leftharpoondown (C+1)d^p$, $\delta \leftharpoondown \delta(C+1)d^p$, $\mathbf{W} \leftharpoondown W^{d,0}$, $u_1 \leftharpoondown u_d$, $u_2 \leftharpoondown v_{d,\delta}$, $p \leftharpoondown p$, $q \leftharpoondown q$ in the notation of Hutzenthaler et al. [30, Lemma 2.3]) demonstrate that for all $d \in \mathbb{N}$, $\delta \in (0,1]$, $t \in [0,T]$ it holds that

$$\begin{aligned} \left(\int_{\mathbb{R}^d} |u_d(t,x) - v_{d,\delta}(t,x)|^{\mathfrak{q}} \, \nu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} &\leqslant \delta(C+1)d^p \bigl(e^{LT}(T+1)\bigr)^{q+1}\bigl((C+1)^q d^{pq} + 1\bigr) \\ &\quad \cdot \left[\int_{\mathbb{R}^d} \Bigl(1 + \|x\| + \bigl(\mathbb{E}\bigl[\|W_T^{d,0}\|^{pq}\bigr]\bigr)^{1/(pq)}\Bigr)^{pq\mathfrak{q}} \nu_d(\mathrm{d}x)\right]^{1/\mathfrak{q}}. \end{aligned} \tag{4.17}$$

Combining (4.17) and the triangle inequality hence yields that for all $d \in \mathbb{N}$, $\delta \in (0,1]$, $t \in [0,T]$ it holds that

$$\left(\int_{\mathbb{R}^d} |u_d(t,x) - v_{d,\delta}(t,x)|^{\mathfrak{q}} \, \nu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} \leqslant c_d \delta. \tag{4.18}$$

Next, note that (4.15) and the assumption that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that $\delta|(\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)| + |g_d(x) - (\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)| \leqslant \delta C d^p (1 + \|x\|)^{pq}$ establish that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\max\bigl\{|(\mathcal{R}_a(\mathbf{F}_\delta))(0)|, |(\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)|\bigr\} \leqslant \max\bigl\{C, C d^p (1 + \|x\|)^{pq}\bigr\} = C d^p (1 + \|x\|)^{pq}. \tag{4.19}$$

This and the fact that for all $n \in \mathbb{N}$, $w_1, w_2, \ldots, w_n \in [0, \infty)$ it holds that $\bigl[\sum_{k=1}^n w_k\bigr]^{1/2} \leqslant \sum_{k=1}^n (w_k)^{1/2}$ imply that for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, $\delta \in (0,1]$ it holds that

$$\begin{aligned} \max\bigl\{|(\mathcal{R}_a(\mathbf{F}_\delta))(0)|, |(\mathcal{R}_a(\mathbf{G}_{d,\delta}))(x)|\bigr\} &\leqslant C d^p \bigl(1 + \bigl[\textstyle\sum_{k=1}^d |x_k|^2\bigr]^{1/2}\bigr)^{pq} \\ &\leqslant C d^p \bigl(1 + \textstyle\sum_{k=1}^d |x_k|\bigr)^{pq}. \end{aligned} \tag{4.20}$$

Combining this, the assumption that for all $d \in \mathbb{N}$, $w, z \in \mathbb{R}$, $\delta \in (0,1]$ it holds that $\max\{|f(w) - f(z)|, |(\mathcal{R}_a(\mathbf{F}_\delta))(w) - (\mathcal{R}_a(\mathbf{F}_\delta))(z)|\} \leqslant L|w - z|$, and Hutzenthaler et al. [31, Proposition 4.4] (applied for every $\delta \in (0,1]$ with $T \leftharpoondown T$, $\mathfrak{B} \leftharpoondown \mathfrak{B}$, $L \leftharpoondown \max\{L, C\}$, $p \leftharpoondown p$, $q \leftharpoondown pq$, $\mathfrak{q} \leftharpoondown \mathfrak{q}$, $(m_j)_{j\in\mathbb{N}} \leftharpoondown (m_j)_{j\in\mathbb{N}}$, $f \leftharpoondown \mathcal{R}_a(\mathbf{F}_\delta)$, $(g_d)_{d\in\mathbb{N}} \leftharpoondown (\mathcal{R}_a(\mathbf{G}_{d,\delta}))_{d\in\mathbb{N}}$, $\Theta \leftharpoondown \Theta$, $(\mathfrak{u}^\theta)_{\theta\in\Theta} \leftharpoondown (\mathfrak{u}^\theta)_{\theta\in\Theta}$, $(\mathcal{U}^\theta)_{\theta\in\Theta} \leftharpoondown (\mathcal{U}^\theta)_{\theta\in\Theta}$, $(W^{d,\theta})_{(d,\theta)\in\mathbb{N}\times\Theta} \leftharpoondown (W^{d,\theta})_{(d,\theta)\in\mathbb{N}\times\Theta}$, $(u_d)_{d\in\mathbb{N}} \leftharpoondown (v_{d,\delta})_{d\in\mathbb{N}}$, $(U^{d,\theta}_{n,j})_{(n,j,d,\theta)\in\mathbb{Z}\times\mathbb{Z}\times\mathbb{N}\times\Theta} \leftharpoondown (U^{d,\theta}_{n,j,\delta})_{(n,j,d,\theta)\in\mathbb{Z}\times\mathbb{Z}\times\mathbb{N}\times\Theta}$ in the notation of Hutzenthaler et al. [31, Proposition 4.4]) ensures that there exists $\mathcal{n}\colon \mathbb{N}\times\mathbb{R}\to\mathbb{N}$ such that for all $d\in\mathbb{N}$, $t\in[0,T]$, $\delta\in(0,1]$ it holds that

$$\left(\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|v_{d,\delta}(t,x)-U^{d,0}_{\mathcal{n}(d,\delta),\mathcal{n}(d,\delta),\delta}(t,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} \leqslant \left(\int_{\mathbb{R}^d}\delta^{\mathfrak{q}}\,\nu_d(dx)\right)^{1/\mathfrak{q}} = \delta. \tag{4.21}$$

Combining (4.14), (4.18), and (4.21) therefore assures that there exists $\mathcal{n}\colon \mathbb{N}\times\mathbb{R}\to\mathbb{N}$ such that for all $d\in\mathbb{N}$, $t\in[0,T]$, $\delta\in(0,1]$ it holds that

$$\left(\int_{\mathbb{R}^d}\mathbb{E}\Big[\big|u_d(t,x)-U^{d,0}_{\mathcal{n}(d,\delta),\mathcal{n}(d,\delta),\delta}(t,x)\big|^{\mathfrak{q}}\Big]\,\nu_d(\mathrm{d}x)\right)^{1/\mathfrak{q}} \leqslant c_d\delta+\delta. \tag{4.22}$$

This and Fubini's theorem prove that there exists $\mathcal{n}\colon \mathbb{N}\times\mathbb{R}\to\mathbb{N}$ such that for all $d\in\mathbb{N}$, $t\in[0,T]$, $\varepsilon\in(0,1]$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\int_{\mathbb{R}^d}\Big|u_d(t,x)-U^{d,0}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),\delta_{d,\varepsilon}}(t,x)\Big|^{\mathfrak{q}}\,\nu_d(\mathrm{d}x)\right]\\
&=\int_{\mathbb{R}^d}\mathbb{E}\left[\Big|u_d(t,x)-U^{d,0}_{\mathcal{n}(d,\delta_{d,\varepsilon}),\mathcal{n}(d,\delta_{d,\varepsilon}),\delta_{d,\varepsilon}}(t,x)\Big|^{\mathfrak{q}}\right]\nu_d(\mathrm{d}x)\leqslant \big(c_d\delta_{d,\varepsilon}+\delta_{d,\varepsilon}\big)^{\mathfrak{q}}=\varepsilon^{\mathfrak{q}}.
\end{aligned} \tag{4.23}$$

(4.23) hence shows that there exists $\mathcal{n}\colon \mathbb{N}\times\mathbb{R}\to\mathbb{N}$ such that for all $d\in\mathbb{N}$, $t\in[0,T]$, $\varepsilon\in(0,1]$ it holds that there exists $\omega_{d,\varepsilon}\in\Omega$ such that

$$\int_{\mathbb{R}^d}\big|u_d(t,x)-U^{d,0}_{\mathcal{n}(d,\varepsilon),\mathcal{n}(d,\varepsilon),\delta_{d,\varepsilon}}(t,x,\omega_{d,\varepsilon})\big|^{\mathfrak{q}}\,\nu_d(\mathrm{d}x)\leqslant\varepsilon^{\mathfrak{q}}. \tag{4.24}$$

Furthermore, observe that (4.6) and item (iv) in Lemma 3.9 (applied for every $d,j\in\mathbb{N}$, $\varepsilon\in(0,1]$ with $\Theta\curvearrowleft\Theta$, $d\curvearrowleft d$, $M\curvearrowleft m_j$, $\mathbf{F}\curvearrowleft\mathbf{F}_\varepsilon$, $\mathbf{G}\curvearrowleft\mathbf{G}_{d,\varepsilon}$, $(\mathcal{U}^\theta)_{\theta\in\Theta}\curvearrowleft(\mathcal{U}^\theta)_{\theta\in\Theta}$, $(W^\theta)_{\theta\in\Theta}\curvearrowleft(W^\theta)_{\theta\in\Theta}$, $(U^\theta_n)_{(n,\theta)\in\mathbb{Z}\times\Theta}\curvearrowleft(U^{d,\theta}_{n,j,\varepsilon})_{(n,\theta)\in\mathbb{Z}\times\Theta}$, $(\mathbf{U}^\theta_{n,t})_{(n,t,\theta)\in\mathbb{Z}\times[0,T]\times\Theta}\curvearrowleft(\mathbf{U}^{d,\theta,\varepsilon}_{n,j,t})_{(n,t,\theta)\in\mathbb{Z}\times[0,T]\times\Theta}$ in the notation of Lemma 3.9) demonstrate that for all $d,n,j\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$, $\varepsilon\in(0,1]$ it holds that $(\mathcal{R}_a(\mathbf{U}^{d,0,\varepsilon}_{n,j,t}))(x)=U^{d,0}_{n,j,\varepsilon}(t,x)$. Combining this and (4.24) establishes (4.3). Moreover, note that item (v) in Lemma 3.9 implies there exists $\mathcal{n}\colon \mathbb{N}\times\mathbb{R}\to\mathbb{N}$ such that for all $d\in\mathbb{N}$, $t\in[0,T]$, $\varepsilon\in(0,1]$ it holds that

$$\begin{aligned}
&\mathcal{P}\big(\mathbf{U}^{d,0,\varepsilon}_{\mathcal{n}(d,\varepsilon),\mathcal{n}(d,\varepsilon),t}\big)\\
&\leqslant 2\big(\mathcal{L}(\mathbf{G}_{d,\varepsilon})+\mathcal{n}(d,\varepsilon)\mathcal{H}(\mathbf{F}_\varepsilon)\big)\Big[\big(1+\sqrt{2}\big)^{\mathcal{n}(d,\varepsilon)}\big(m_{\mathcal{n}(d,\varepsilon)}\big)^{\mathcal{n}(d,\varepsilon)}\max\{\mathfrak{d},|||\mathcal{D}(\mathbf{F}_\varepsilon)|||,|||\mathcal{D}(\mathbf{G}_{d,\varepsilon})|||\}\Big]^2\\
&\leqslant 2\big(\mathcal{L}(\mathbf{G}_{d,\varepsilon})+\mathcal{H}(\mathbf{F}_\varepsilon)\big)\Big(\max\{2,|||\mathcal{D}(\mathbf{F}_\varepsilon)|||,|||\mathcal{D}(\mathbf{G}_{d,\varepsilon})|||\}\Big)^2\\
&\quad\cdot\Big[\big(\mathcal{n}(d,\varepsilon)\big)^{1/2}(1+\sqrt{2})^{\mathcal{n}(d,\varepsilon)}\big(m_{\mathcal{n}(d,\varepsilon)}\big)^{\mathcal{n}(d,\varepsilon)}\Big]^2.
\end{aligned} \tag{4.25}$$

In addition, observe that the assumption that for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ it holds that $\mathcal{L}(\mathbf{F}_\varepsilon)\leqslant C\varepsilon^{-\lambda}$, $|||\mathcal{D}(\mathbf{F}_\varepsilon)|||\leqslant C\varepsilon^{-\gamma}$, $\mathcal{L}(\mathbf{G}_{d,\varepsilon})\leqslant Cd^{\mathfrak{p}}\varepsilon^{-\beta}$, and $|||\mathcal{D}(\mathbf{G}_{d,\varepsilon})|||\leqslant Cd^{\mathfrak{p}}\varepsilon^{-\alpha}$ ensures that for all

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

$$\begin{aligned}
&\big(\mathcal{L}(\mathbf{G}_{d,\varepsilon}) + \mathcal{H}(\mathbf{F}_\varepsilon)\big)\Big(\max\{\mathfrak{d}, \|\!|\mathcal{D}(\mathbf{F}_\varepsilon)|\!\|, \|\!|\mathcal{D}(\mathbf{G}_{d,\varepsilon})|\!\|\}\Big)^2 \\
&\leqslant \big(Cd^p\varepsilon^{-\beta} + C\varepsilon^{-\lambda}\big)\Big(\max\{C, C\varepsilon^{-\gamma}, Cd^p\varepsilon^{-\alpha}\}\Big)^2 \\
&\leqslant 2Cd^p \max\{\varepsilon^{-\beta}, \varepsilon^{-\lambda}\}\Big(\max\{C, C\varepsilon^{-\gamma}, Cd^p\varepsilon^{-\alpha}\}\Big)^2 \\
&\leqslant 2C^3d^{3p}\varepsilon^{-\max\{\beta,\lambda\}}\big(\max\{\varepsilon^{-\gamma}, \varepsilon^{-\alpha}\}\big)^2 \leqslant 2C^3d^{3p}\varepsilon^{-(\max\{\beta,\lambda\}+2\max\{\gamma,\alpha\})}.
\end{aligned} \tag{4.26}$$

This, (4.25), and Hutzenthaler et al. [31, Proposition 4.4] (applied with $p \hookleftarrow p$, $q \hookleftarrow pq$, $\beta \hookleftarrow 1/2$, $d \hookleftarrow d$, $\alpha \hookleftarrow 1$, $\mathfrak{d} \hookleftarrow 0$, $(m_j)_{j\in\mathbb{N}} \hookleftarrow (m_j)_{j\in\mathbb{N}}$, $\mathcal{n} \hookleftarrow \mathcal{n}$ in the notation of Hutzenthaler et al. [31, Proposition 4.4]) assure that there exists $\mathcal{n}\colon \mathbb{N}\times\mathbb{R} \to \mathbb{N}$ and $\mathfrak{c} = (\mathfrak{c}_\delta)_{\delta\in(0,\infty)}\colon \mathbb{R} \to \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\delta \in (0,\infty)$ it holds that

$$\begin{aligned}
\mathcal{P}\big(\mathbf{U}^{d,0,\varepsilon}_{\mathcal{n}(d,\varepsilon),\mathcal{n}(d,\varepsilon),t}\big) &\leqslant 2C^3d^{3p}\varepsilon^{-(\max\{\beta,\lambda\}+2\max\{\gamma,\alpha\})}\Big[\big(\mathcal{n}(d,\varepsilon)\big)^{1/2}(1+\sqrt{2})^{\mathcal{n}(d,\varepsilon)}\big(m_{\mathcal{n}(d,\varepsilon)}\big)^{\mathcal{n}(d,\varepsilon)}\Big]^2 \\
&\leqslant 2C^3d^{3p}\varepsilon^{-(\max\{\beta,\lambda\}+2\max\{\gamma,\alpha\})}\big[\mathfrak{c}_\delta d^{(p+pq)(2+\delta)}\varepsilon^{-(2+\delta)}\big]^2 \\
&\leqslant (\mathfrak{c}_\delta)^2 2C^3 d^{p(7+4q+(2+q)\delta)}\varepsilon^{-(4+2\delta+\max\{\beta,\lambda\}+2\max\{\gamma,\alpha\})}.
\end{aligned} \tag{4.27}$$

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

4.2 One-dimensional neural network 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 $f\colon \mathbb{R} \to \mathbb{R}$ 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

$$w_f(h) = \sup\Big(\big\{|f(x)-f(y)| \in [0,\infty)\colon \big(x, y \in A \text{ with } |x-y| \leqslant h\big)\big\} \cup \{0\}\Big) \tag{4.28}$$

and we call w_f the modulus of continuity of f.

Lemma 4.3. *Let $a \in [-\infty,\infty]$, $b \in [a,\infty]$ and let $f\colon ([a,b]\cap\mathbb{R}) \to \mathbb{R}$ 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 $x, y \in [a,b]\cap\mathbb{R}$ that $|f(x)-f(y)| \leqslant w_f(|x-y|)$, and

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

(cf. Definition 4.2).

Proof of Lemma 4.3. First, note that (4.28) implies items (i), (ii), (iii), and (iv). Next, observe that (4.28) and the triangle inequality ensure that for all $h, \mathfrak{h} \in [0, \infty]$ it holds that

$$\begin{aligned}
&w_f(h + \mathfrak{h}) \\
&= \sup\Big(\{|f(x) - f(y)| \in [0, \infty) \colon (x, y \in [a, b] \cap \mathbb{R} \colon |x - y| \leqslant (h + \mathfrak{h}))\} \cup \{0\}\Big) \\
&\leqslant \sup\Big(\Big\{\big|f(x) - f\big(x - h\tfrac{x-y}{|x-y|}\big)\big| \in [0, \infty) \colon \big(x, y \in [a, b] \cap \mathbb{R} \colon |x - y| \leqslant (h + \mathfrak{h})\big)\Big\} \cup \{0\}\Big) \\
&\quad + \sup\Big(\Big\{\big|f\big(x - h\tfrac{x-y}{|x-y|}\big) - f(y)\big| \in [0, \infty) \colon \big(x, y \in [a, b] \cap \mathbb{R} \colon |x - y| \leqslant (h + \mathfrak{h})\big)\Big\} \cup \{0\}\Big) \\
&\leqslant \sup\Big(\Big\{\big|f(x) - f\big(x - h\tfrac{x-y}{|x-y|}\big)\big| \in [0, \infty) \colon \big(x, y \in [a, b] \cap \mathbb{R} \colon |x - y| \leqslant h\big)\Big\} \cup \{0\}\Big) \\
&\quad + \sup\Big(\Big\{\big|f\big(x - h\tfrac{x-y}{|x-y|}\big) - f(y)\big| \in [0, \infty) \colon \big(x, y \in [a, b] \cap \mathbb{R} \colon |x - y| \leqslant \mathfrak{h}\big)\Big\} \cup \{0\}\Big) \\
&= w_f(h) + w_f(\mathfrak{h})
\end{aligned} \tag{4.29}$$

(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 $|f(x) - f(y)| \leqslant L|x - y|$. Then it holds for all $h \in [0, \infty)$ that $w_f(h) \leqslant Lh$ (cf. Definition 4.2).*

Proof of Lemma 4.4. Note that the assumption that for all $x, y \in A$ it holds that $|f(x) - f(y)| \leqslant L|x - y|$ and (4.28) assure that for all $h \in [0, \infty)$ it holds that

$$\begin{aligned}
w_f(h) &= \sup\Big(\big\{|f(x) - f(y)| \in [0, \infty) \colon \big(x, y \in A \text{ with } |x - y| \leqslant h\big)\big\} \cup \{0\}\Big) \\
&\leqslant \sup\Big(\big\{L|x - y| \in [0, \infty) \colon \big(x, y \in A \text{ with } |x - y| \leqslant h\big)\big\} \cup \{0\}\Big) \leqslant \sup(\{Lh, 0\}) = Lh
\end{aligned} \tag{4.30}$$

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

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(y) = f_{k-1} + \big(\tfrac{y - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(f_k - f_{k-1}). \tag{4.31}$$

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

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

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

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(x) = f_{k-1} + \big(\tfrac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(f_k - f_{k-1}), \tag{4.32}$$

and

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

$$(\mathscr{L}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K}^{f_0,f_1,\ldots,f_K})(x) = \big(\tfrac{\mathfrak{x}_k - x}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big) f_{k-1} + \big(\tfrac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big) f_k \tag{4.33}$$

(cf. Definition 4.5).

Proof of Lemma 4.6. Observe that (4.31) proves items (i) and (ii). Furthermore, note that item (ii) shows that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ it holds that

$$(\mathscr{L}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K}^{f_0,f_1,\ldots,f_K})(x) = \Big[1 - \big(\tfrac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)\Big] f_{k-1} + \big(\tfrac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big) f_k = \big(\tfrac{\mathfrak{x}_k - x}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big) f_{k-1} + \big(\tfrac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big) f_k \tag{4.34}$$

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

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

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

$$\begin{aligned} &\big|(\mathscr{L}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K}^{f(\mathfrak{x}_0),f(\mathfrak{x}_1),\ldots,f(\mathfrak{x}_K)})(x) - (\mathscr{L}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K}^{f(\mathfrak{x}_0),f(\mathfrak{x}_1),\ldots,f(\mathfrak{x}_K)})(y)\big| \\ &\leqslant \bigg[\max_{k \in \{1,2,\ldots,K\}} \bigg(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|}\bigg)\bigg] |x - y| \end{aligned} \tag{4.35}$$

and

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

(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 $l(x) = (\mathscr{L}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K}^{f(\mathfrak{x}_0),f(\mathfrak{x}_1),\ldots,f(\mathfrak{x}_K)})(x)$ and let $L \in [0, \infty]$ satisfy

$$L = \max_{k \in \{1,2,\ldots,K\}} \bigg(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|}\bigg) \tag{4.36}$$

(cf. Definitions 4.2 and 4.5). Observe that item (iv) in Lemma 4.3, (4.36), and item (ii) in Lemma 4.6 demonstrate that for all $k \in \{1, 2, \ldots, K\}$, $x, y \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ with $x \neq y$ it holds that

$$\begin{aligned} |l(x) - l(y)| &= \Big|\big(\tfrac{x - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)\big(f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})\big) - \big(\tfrac{y - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)\big(f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})\big)\Big| \\ &= \bigg|\bigg(\frac{f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\bigg)(x - y)\bigg| \leqslant \bigg(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|}\bigg)|x - y| \leqslant L|x - y|. \end{aligned} \tag{4.37}$$

This, item (iv) in Lemma 4.3, Lemma 4.6, (4.36), and the triangle inequality yield that for all $j, k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{x}_{j-1}, \mathfrak{x}_j]$, $y \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ with $j < k$ and $x \neq y$ it holds that

$$\begin{aligned} |l(x) - l(y)| &\leqslant |l(x) - l(\mathfrak{x}_j)| + |l(\mathfrak{x}_j) - l(\mathfrak{x}_{k-1})| + |l(\mathfrak{x}_{k-1}) - l(y)| \\ &= |l(x) - l(\mathfrak{x}_j)| + |f(\mathfrak{x}_j) - f(\mathfrak{x}_{k-1})| + |l(\mathfrak{x}_{k-1}) - l(y)| \end{aligned}$$

$$\leqslant |l(x) - l(\mathfrak{r}_j)| + \left[\sum_{i=j+1}^{k-1} |f(\mathfrak{r}_i) - f(\mathfrak{r}_{i-1})|\right] + |l(\mathfrak{r}_{k-1}) - l(y)| \tag{4.38}$$

$$\leqslant |l(x) - l(\mathfrak{r}_j)| + \left[\sum_{i=j+1}^{k-1} w_f\big(|\mathfrak{r}_i - \mathfrak{r}_{i-1}|\big)\right] + |l(\mathfrak{r}_{k-1}) - l(y)|$$

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

Combining this and (4.37) establishes that for all $x, y \in [\mathfrak{r}_0, \mathfrak{r}_K]$ with $x \neq y$ it holds that $|l(x) - l(y)| \leqslant L|x - y|$. This, the fact that for all $x, y \in (-\infty, \mathfrak{r}_0]$ with $x \neq y$ it holds that $|l(x) - l(y)| = 0 \leqslant L|x - y|$, the fact that for all $x, y \in [\mathfrak{r}_K, \infty)$ with $x \neq y$ it holds that $|l(x) - l(y)| = 0 \leqslant L|x - y|$, and the triangle inequality imply that for all $x, y \in \mathbb{R}$ with $x \neq y$ it holds that $|l(x) - l(y)| \leqslant L|x - y|$. This proves item (i). Moreover, note that (4.28), Lemma 4.3, item (iii) in Lemma 4.6, and the triangle inequality ensure that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ it holds that

$$\begin{aligned}
|l(x) - f(x)| &= \left| \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) f(\mathfrak{r}_k) + \left(\frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) f(\mathfrak{r}_{k-1}) - f(x) \right| \\
&= \left| \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) (f(\mathfrak{r}_k) - f(x)) + \left(\frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) (f(\mathfrak{r}_{k-1}) - f(x)) \right| \\
&\leqslant \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) |f(\mathfrak{r}_k) - f(x)| + \left(\frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) |f(\mathfrak{r}_{k-1}) - f(x)| \qquad (4.39) \\
&\leqslant w_f\big(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|\big) \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} - \frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) \\
&= w_f\big(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|\big) \leqslant w_f\big(\max\nolimits_{j \in \{1,2,\ldots,K\}} |\mathfrak{r}_j - \mathfrak{r}_{j-1}|\big).
\end{aligned}$$

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

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

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

$$\big|(\mathscr{L}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K}^{f(\mathfrak{r}_0),f(\mathfrak{r}_1),\ldots,f(\mathfrak{r}_K)})(x) - (\mathscr{L}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K}^{f(\mathfrak{r}_0),f(\mathfrak{r}_1),\ldots,f(\mathfrak{r}_K)})(y)\big| \leqslant L|x - y| \tag{4.40}$$

and

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

(cf. Definition 4.5).

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

$$\begin{aligned}
&\big|(\mathscr{L}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K}^{f(\mathfrak{r}_0),f(\mathfrak{r}_1),\ldots,f(\mathfrak{r}_K)})(x) - (\mathscr{L}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K}^{f(\mathfrak{r}_0),f(\mathfrak{r}_1),\ldots,f(\mathfrak{r}_K)})(y)\big| \\
&\leqslant \left[\max_{k \in \{1,2,\ldots,K\}} \left(\tfrac{L|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\right)\right] |x - y| = L|x - y|
\end{aligned} \tag{4.41}$$

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

$$\sup_{x \in [\mathfrak{x}_0, \mathfrak{x}_K]} \big|(\mathscr{L}^{f(\mathfrak{x}_0), f(\mathfrak{x}_1), \dots, f(\mathfrak{x}_K)}_{\mathfrak{x}_0, \mathfrak{x}_1, \dots, \mathfrak{x}_K})(x) - f(x)\big| \leqslant L\bigg(\max_{k \in \{1,2,\dots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}|\bigg). \tag{4.42}$$

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

4.2.3 Linear interpolation with ANNs

Note #5: Need to update everything below this point. . .

Lemma 4.9. *Let $\alpha, \beta, h \in \mathbb{R}$, $\mathbf{H} \in \mathbf{N}$ satisfy $\mathbf{H} = h \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta})$ (cf. Definitions 2.1, 2.7, 2.12, 2.20, and 3.1). Then*

(i) it holds that $\mathbf{H} = ((\alpha, \beta), (h, 0))$,

(ii) it holds that $\mathcal{D}(\mathbf{H}) = (1, 1, 1) \in \mathbb{N}^3$,

(iii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{H}) \in C(\mathbb{R}, \mathbb{R})$, and

(iv) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbf{H}))(x) = h \max\{\alpha x + \beta, 0\}$

(cf. Definitions 2.4 and 2.6).

Proof of Lemma 4.9. Observe that Lemma 2.13 shows that for all $x \in \mathbb{R}$ it holds that $\mathbf{A}_{\alpha,\beta} = (\alpha, \beta)$, $\mathcal{D}(\mathbf{A}_{\alpha,\beta}) = (1, 1) \in \mathbb{N}^2$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta}) \in C(\mathbb{R}, \mathbb{R})$, and $(\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta}))(x) = \alpha x + \beta$ (cf. Definitions 2.4 and 2.6). Lemma 3.2, Lemma 3.3, (2.8), and (2.9) therefore demonstrate that for all $x \in \mathbb{R}$ it holds that $\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta} = ((\alpha, \beta), (1, 0))$, $\mathcal{D}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) = (1, 1, 1) \in \mathbb{N}^3$, $\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) \in C(\mathbb{R}, \mathbb{R})$, and

$$(\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}))(x) = \mathfrak{r}(\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta})(x)) = \max\{\alpha x + \beta, 0\}. \tag{4.43}$$

This, e.g., Grohs et al. [23, Lemma 3.14], and Definition 2.20 yield that for all $x \in \mathbb{R}$ it holds that $h \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) = ((\alpha, \beta), (h, 0))$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{H}) \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{D}(\mathbf{H}) = (1, 1, 1)$, and

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{H}))(x) = h((\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}))(x)) = h \max\{\alpha x + \beta, 0\}. \tag{4.44}$$

This establishes items (i), (ii), (iii), and (iv). The proof of Lemma 4.9 is thus complete. □

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

$$\mathbf{F} = \mathbf{A}_{1,f_0} \bullet \Bigg(\bigoplus_{k=0}^{K} \bigg(\Big(\tfrac{(f_{\min\{k+1,K\}} - f_k)}{(\mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_{\min\{k,K-1\}})} - \tfrac{(f_k - f_{\max\{k-1,0\}})}{(\mathfrak{x}_{\max\{k,1\}} - \mathfrak{x}_{\max\{k-1,0\}})}\Big) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{x}_k})\bigg)\Bigg) \tag{4.45}$$

(cf. Definitions 2.1, 2.7, 2.12, 2.18, 2.20, and 3.1). Then

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

(ii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$*,*

(iii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) = \mathscr{L}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K}^{f(\mathfrak{x}_0),f(\mathfrak{x}_1),\ldots,f(\mathfrak{x}_K)}$*, and*

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

(cf. Definitions 2.4, 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

$$c_k = \frac{(f_{\min\{k+1,K\}} - f_k)}{(\mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_{\min\{k,K-1\}})} - \frac{(f_k - f_{\max\{k-1,0\}})}{(\mathfrak{x}_{\max\{k,1\}} - \mathfrak{x}_{\max\{k-1,0\}})} \tag{4.46}$$

and let $\Phi_0, \Phi_1, \ldots, \Phi_K \in ((\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \times (\mathbb{R}^{1\times 1} \times \mathbb{R}^1)) \subseteq \mathbf{N}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that $\Phi_k = c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{x}_k})$. Observe that Lemma 4.9 assures that for all $k \in \{0, 1, \ldots, K\}$ it holds that $\mathcal{R}_{\mathfrak{r}}(\Phi_k) \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{D}(\Phi_k) = (1, 1, 1) \in \mathbb{N}^3$, and $\forall\, x \in \mathbb{R}\colon (\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = c_k \max\{x - \mathfrak{x}_k, 0\}$ (cf. Definitions 2.4 and 2.6). This, Lemma 2.14, (4.45), and e.g., Grohs et al. [23, Lemma 3.28] assure that $\mathcal{D}(\mathbf{F}) = (1, K+1, 1) \in \mathbb{N}^3$ and $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$. This establishes items (i) and (ii). Moreover, note that item (i) and (2.6) imply that

$$\mathcal{P}(\mathbf{F}) = 2(K + 1) + (K + 2) = 3K + 4. \tag{4.47}$$

This proves item (iv). Next observe that (4.46), Lemma 2.14, and e.g., Grohs et al. [23, Lemma 3.28] ensure that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = f_0 + \sum_{k=0}^{K} (\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = f_0 + \sum_{k=0}^{K} c_k \max\{x - \mathfrak{x}_k, 0\}. \tag{4.48}$$

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

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = f_0 + 0 = f_0. \tag{4.49}$$

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

$$\sum_{n=0}^{k-1} c_n = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}. \tag{4.50}$$

We now prove (4.50) by induction on $k \in \{1, 2, \ldots, K\}$. For the base case $k = 1$ observe that (4.46) assures that $\sum_{n=0}^{0} c_n = c_0 = \frac{f_1 - f_0}{\mathfrak{x}_1 - \mathfrak{x}_0}$. This proves (4.50) 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 (4.46) 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

$$\sum_{n=0}^{k-1} c_n = c_{k-1} + \sum_{n=0}^{k-2} c_n = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} - \tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}} + \tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}} = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}. \tag{4.51}$$

Induction thus proves (4.50). In addition, observe that (4.48), (4.50), 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

$$\begin{aligned}(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_{k-1}) &= \sum_{n=0}^{K} c_n (\max\{x - \mathfrak{x}_n, 0\} - \max\{\mathfrak{x}_{k-1} - \mathfrak{x}_n, 0\}) \\ = \sum_{n=0}^{k-1} c_n [(x - \mathfrak{x}_n) - (\mathfrak{x}_{k-1} - \mathfrak{x}_n)] &= \sum_{n=0}^{k-1} c_n (x - \mathfrak{x}_{k-1}) = \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}).\end{aligned} \tag{4.52}$$

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

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = f_{k-1} + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\big)(x - \mathfrak{r}_{k-1}). \tag{4.53}$$

We now prove (4.53) by induction on $k \in \{1, 2, \ldots, K\}$. For the base case $k = 1$ observe that (4.49) and (4.52) demonstrate that for all $x \in [\mathfrak{r}_0, \mathfrak{r}_1]$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{r}_0) + (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{r}_0) = f_0 + \big(\tfrac{f_1 - f_0}{\mathfrak{r}_1 - \mathfrak{r}_0}\big)(x - \mathfrak{r}_0). \tag{4.54}$$

This proves (4.53) 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 (4.52) implies that for all $k \in \{2, 3, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ with $\forall\, y \in [\mathfrak{r}_{k-2}, \mathfrak{r}_{k-1}] \colon (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y) = f_{k-2} + (\frac{f_{k-1} - f_{k-2}}{\mathfrak{r}_{k-1} - \mathfrak{r}_{k-2}})(y - \mathfrak{r}_{k-2})$ it holds that

$$\begin{aligned}(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) &= (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{r}_{k-1}) + (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{r}_{k-1}) \\ &= f_{k-2} + \big(\tfrac{f_{k-1} - f_{k-2}}{\mathfrak{r}_{k-1} - \mathfrak{r}_{k-2}}\big)(\mathfrak{r}_{k-1} - \mathfrak{r}_{k-2}) + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\big)(x - \mathfrak{r}_{k-1}) = f_{k-1} + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\big)(x - \mathfrak{r}_{k-1}).\end{aligned} \tag{4.55}$$

Induction thus proves (4.53). Furthermore, observe that (4.46) and (4.50) ensure that

$$\sum_{n=0}^{K} c_n = c_K + \sum_{n=0}^{K-1} c_n = -\tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}} + \tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}} = 0. \tag{4.56}$$

The fact that $\forall\, k \in \{0, 1, \ldots, K\} \colon \mathfrak{r}_k \leqslant \mathfrak{r}_K$ and (4.48) hence imply that for all $x \in [\mathfrak{r}_K, \infty)$ it holds that

$$\begin{aligned}(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{r}_K) &= \left[\sum_{n=0}^{K} c_n(\max\{x - \mathfrak{r}_n, 0\} - \max\{\mathfrak{r}_K - \mathfrak{r}_n, 0\})\right] \\ &= \sum_{n=0}^{K} c_n[(x - \mathfrak{r}_n) - (\mathfrak{r}_K - \mathfrak{r}_n)] = \sum_{n=0}^{K} c_n(x - \mathfrak{r}_K) = 0.\end{aligned} \tag{4.57}$$

This and (4.53) show that for all $x \in [\mathfrak{r}_K, \infty)$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{r}_K) = f_{K-1} + \big(\tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}}\big)(\mathfrak{r}_K - \mathfrak{r}_{K-1}) = f_K. \tag{4.58}$$

Combining this, (4.49), (4.53), and (4.31) establishes item (iii). The proof of Lemma 4.10 is thus complete. □

4.2.4 ANN approximations of one-dimensional functions

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

$$\mathbf{F} = \mathbf{A}_{1, f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1, K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1, 0\}}))}{(b-a)} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1, -\mathfrak{r}_k}) \right) \right) \tag{4.59}$$

(cf. Definitions 2.1, 2.7, 2.12, 2.18, 2.20, and 3.1). Then

(i) it holds that $\mathcal{D}(\mathbf{F}) = (1, K + 1, 1)$,

(ii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$*,*

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

(iv) it holds that $\sup_{x \in [a,b]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leqslant L(b - a)K^{-1}$*, and*

(v) it holds that $\mathcal{P}(\mathbf{F}) = 3K + 4$

(cf. Definitions 2.4 and 2.6).

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

$$\frac{(f(\mathfrak{r}_{\min\{k+1,K\}}) - f(\mathfrak{r}_k))}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \frac{(f(\mathfrak{r}_k) - f(\mathfrak{r}_{\max\{k-1,0\}}))}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} = \frac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)}. \tag{4.60}$$

This and items (i), (ii), and (iv) in Lemma 4.10 prove items (i), (ii), and (v). In addition, observe that (4.60) and item (iii) in Lemma 4.10 demonstrate that

$$\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) = \mathscr{L}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K}. \tag{4.61}$$

Combining this with the assumption that $\forall\, x, y \in [a, b] \colon |f(x) - f(y)| \leqslant L|x - y|$ and item (i) in Lemma 4.8 establishes item (iii). Moreover, note that (4.61), the assumption that $\forall\, x, y \in [a, b] \colon |f(x) - f(y)| \leqslant L|x - y|$, item (ii) in Lemma 4.8, and the fact that $\forall\, k \in \{1, 2, \ldots, K\} \colon \mathfrak{r}_k - \mathfrak{r}_{k-1} = (b - a)K^{-1}$ demonstrate that for all $x \in [a, b]$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leqslant L\left(\max_{k \in \{1,2,\ldots,K\}} |\mathfrak{r}_k - \mathfrak{r}_{k-1}|\right) = L(b - a)K^{-1}. \tag{4.62}$$

This establishes item (iv). The proof of Lemma 4.11 is thus complete. □

Lemma 4.12. *Let* $L, a \in \mathbb{R}$*,* $b \in [a, \infty)$*,* $\xi \in [a, b]$*, let* $f \colon [a, b] \to \mathbb{R}$ *satisfy for all* $x, y \in [a, b]$ *that* $|f(x) - f(y)| \leqslant L|x - y|$*, and let* $\mathbf{F} \in \mathbf{N}$ *satisfy* $\mathbf{F} = \mathbf{A}_{1,f(\xi)} \bullet (0 \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\xi}))$ *(cf. Definitions 2.1, 2.7, 2.12, 2.20, and 3.1). Then*

(i) it holds that $\mathcal{D}(\mathbf{F}) = (1, 1, 1)$*,*

(ii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$*,*

(iii) it holds for all $x \in \mathbb{R}$ *that* $(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = f(\xi)$*,*

(iv) it holds that $\sup_{x \in [a,b]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leqslant L \max\{\xi - a, b - \xi\}$*, and*

(v) it holds that $\mathcal{P}(\mathbf{F}) = 4$

(cf. Definitions 2.4 and 2.6).

Proof of Lemma 4.12. Note that items (i) and (ii) in Lemma 2.14, and items (ii) and (iii) in Lemma 4.9 establish items (i) and (ii). In addition, observe that item (iii) in Lemma 2.14 and, e.g., Grohs et al. [23, item (iii) in Lemma 3.14] assure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) &= (\mathcal{R}_{\mathfrak{r}}(0 \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\xi})))(x) + f(\xi) \\ &= 0\big((\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\xi}))(x)\big) + f(\xi) = f(\xi) \end{aligned} \tag{4.63}$$

(cf. Definitions 2.4 and 2.6). This establishes item (iii). Next note that (4.63), the fact that $\xi \in [a,b]$, and the fact that for all $x, y \in [a,b]$ it holds that $|f(x) - f(y)| \leqslant L|x-y|$ assure that for all $x \in [a,b]$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| = |f(\xi) - f(x)| \leqslant L|x - \xi| \leqslant L \max\{\xi - a, b - \xi\}. \tag{4.64}$$

This establishes item (iv). Moreover, observe that (2.6) and item (i) assure that

$$\mathcal{P}(\mathbf{F}) = 1(1+1) + 1(1+1) = 4. \tag{4.65}$$

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

Corollary 4.13. *Let $\varepsilon \in (0,\infty)$, $L, a \in \mathbb{R}$, $b \in (a,\infty)$, $K \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that $\mathfrak{r}_k = a + \frac{k(b-a)}{\max\{K,1\}}$, let $f\colon [a,b] \to \mathbb{R}$ satisfy for all $x, y \in [a,b]$ that $|f(x) - f(y)| \leqslant L|x-y|$, and let $\mathbf{F} \in \mathbf{N}$ satisfy*

$$\mathbf{F} = \mathbf{A}_{1, f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)} \right) \circledast (\mathbf{i}_1 \bullet \mathbf{A}_{1, -\mathfrak{r}_k}) \right) \right) \tag{4.66}$$

(cf. Definitions 2.1, 2.7, 2.12, 2.18, 2.20, and 3.1). Then

(i) it holds that $\mathcal{D}(\mathbf{F}) = (1, K+1, 1)$,

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

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

(iv) it holds that $\sup_{x \in [a,b]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leqslant \frac{L(b-a)}{\max\{K,1\}} \leqslant \varepsilon$, and

(v) it holds that $\mathcal{P}(\mathbf{F}) = 3K + 4 \leqslant 3L(b-a)\varepsilon^{-1} + 7$

(cf. Definitions 2.4 and 2.6).

Proof of Corollary 4.13. Note that the fact that $K \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$ implies that $\frac{L(b-a)}{\max\{K,1\}} \leqslant \varepsilon$. This, items (i), (ii), (iii), and (iv) in Lemma 4.11, and items (i), (ii), (iii), and (iv) in Lemma 4.12 establish items (i), (ii), (iii), and (iv). Moreover, note that the fact that $K \leqslant 1 + \frac{L(b-a)}{\varepsilon}$, item (v) in Lemma 4.11, and item (v) in Lemma 4.12 assure that

$$\mathcal{P}(\mathbf{F}) = 3K + 4 \leqslant \frac{3L(b-a)}{\varepsilon} + 7. \tag{4.67}$$

This establishes item (v). The proof of Corollary 4.13 is thus complete. □

Corollary 4.14. *Let $\varepsilon \in (0,1]$, $L \in [0,\infty)$, $q \in (1,\infty)$ and let $f\colon \mathbb{R} \to \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leqslant L|x-y|$. Then there exists $\mathbf{F} \in \mathbf{N}$ such that*

(i) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$,

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

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

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

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

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

(cf. Definitions 2.1, 2.4, and 2.6).

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

$$c_k = \frac{K(g(\mathfrak{r}_{\min\{k+1,K\}}) - 2g(\mathfrak{r}_k) + g(\mathfrak{r}_{\max\{k-1,0\}}))}{2a}, \tag{4.68}$$

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

$$\mathbf{F} = \mathbf{A}_{1,g(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k})) \right) \tag{4.69}$$

(cf. Definitions 2.1, 2.7, 2.12, 2.18, 2.20, and 3.1). Observe that Corollary 4.13 implies that

(I) it holds that $\mathcal{D}(\mathbf{F}) = (1, K + 1, 1)$,

(II) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) \in C(\mathbb{R}, \mathbb{R})$,

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

(IV) it holds that $\sup_{x\in[-a,a]}|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - g(x)| \leqslant \frac{2La}{\max\{K,1\}} \leqslant \varepsilon$, and

(V) it holds that $\mathcal{P}(\mathbf{F}) = 3K + 4$

(cf. Definitions 2.4 and 2.6). This and the fact that for all $x \in [-a, a]$ it holds that $g(x) = f(x)$ establish items (i), (ii), and (iv). Next note that the triangle inequality, item (iv), the fact that $f(-a) = g(-a) = (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(-a)$, the fact that $f(a) = g(a) = (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(a)$, and the fact that for all $x, y \in \mathbb{R}$ it holds that $|f(x) - f(y)| \leqslant L|x - y|$ ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| &\leqslant |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(a)| + |f(a) - f(0)| + |f(0) - f(x)| \\ &= |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - g(a)| + |f(a) - f(0)| + |f(0) - f(x)| \\ &\leqslant |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(a)| + |f(a) - f(0)| + |f(0) - f(x)| \\ &\leqslant L|x - a| + L|a| + L|x| = L(|x - a| + a + |x|) \end{aligned} \tag{4.70}$$

and

$$\begin{aligned} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| &\leqslant |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(-a)| + |f(-a) - f(0)| + |f(0) - f(x)| \\ &= |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - g(-a)| + |f(-a) - f(0)| + |f(0) - f(x)| \\ &\leqslant |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(-a)| + |f(-a) - f(0)| + |f(0) - f(x)| \end{aligned} \tag{4.71}$$

$$\leqslant L|x+a| + L|a| + L|x| = L(|x+a| + a + |x|).$$

(4.70) hence assures that for all $x \in (a, \infty)$ it holds that

$$\begin{aligned}\frac{|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)|}{\max\{1, |x|^q\}} &\leqslant \frac{L(|x-a| + a + |x|)}{\max\{1, |x|^q\}} = \frac{L(x-a+a+x)}{\max\{1, |x|^q\}} \\ &= \frac{2L|x|}{\max\{1, |x|^q\}} \leqslant \frac{\max\{1, 2L\}}{\max\{1, |x|^q\}} \leqslant \frac{\max\{1, 2L\}}{a^{q-1}} = \varepsilon.\end{aligned} \tag{4.72}$$

Moreover, (4.71) demonstrates that for all $x \in (-\infty, -a)$ it holds that

$$\begin{aligned}\frac{|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)|}{\max\{1, |x|^q\}} &\leqslant \frac{L(|x+a| + a + |x|)}{\max\{1, |x|^q\}} = \frac{L(-(x+a)+a-x)}{\max\{1, |x|^q\}} \\ &= \frac{2L|x|}{\max\{1, |x|^q\}} \leqslant \frac{\max\{1, 2L\}}{\max\{1, |x|^q\}} \leqslant \frac{\max\{1, 2L\}}{a^{q-1}} = \varepsilon.\end{aligned} \tag{4.73}$$

Combining this, (4.72), item (IV), and the fact that for all $x \in [-a, a]$ it holds that $f(x) = g(x)$ shows that for all $x \in \mathbb{R}$ it holds $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leqslant \varepsilon \max\{1, |x|^q\}$. This establishes item (v). In addition, observe that item (I), the fact that $\max\{1, 2L\} = \varepsilon a^{q-1}$, and the fact that $K \leqslant 1 + \frac{2La}{\varepsilon}$ prove that

$$K \leqslant 1 + \frac{2La}{\varepsilon} \leqslant 1 + \frac{\max\{1, 2L\}a}{\varepsilon} = 1 + a^q \leqslant 2a^q \leqslant 2\left(\frac{\max\{1, 2L\}}{\varepsilon}\right)^{q/(q-1)}. \tag{4.74}$$

This establishes item (iii). Next observe that item (V) implies that

$$\mathcal{P}(\mathbf{F}) = 3K + 4 = 3(K+1) + 1 = 3(\mathbb{D}_1(\mathbf{F})) + 1. \tag{4.75}$$

This and item (iii) guarantee that

$$\begin{aligned}3(\mathbb{D}_1(\mathbf{F})) + 1 \leqslant 4(\mathbb{D}_1(\mathbf{F})) &\leqslant 4\big[2(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)} + 1\big] \\ &\leqslant 8(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)} + 4 \leqslant 12(\max\{1, 2L\})^{q/(q-1)} \varepsilon^{-q/(q-1)}.\end{aligned} \tag{4.76}$$

Combining (4.75) and (4.76) therefore establishes item (vi). The proof of Corollary 4.14 is thus complete. □

4.3 ANN approximation results with general polynomial convergence rates

Corollary 4.15. *Let $T, \kappa, \mathfrak{q} \in (0, \infty)$, $f \in C(\mathbb{R}, \mathbb{R})$, let $\|\cdot\| \colon \bigcup_{d \in \mathbb{N}} \mathbb{R}^d \to [0, \infty)$ satisfy for all $d \in \mathbb{N}$, $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $\|x\| = [\sum_{k=1}^d |x_k|^2]^{1/2}$, let $\mathbf{G}_{d,\varepsilon} \in \mathbf{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, let $u_d \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, and assume for all $d \in \mathbb{N}$, $v, w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $t \in [0,T)$ that $|f(v) - f(w)| \leqslant \kappa |v - w|$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\varepsilon |(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| + \varepsilon |u_d(t,x)| + |u_d(T,x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| \leqslant \varepsilon \kappa d^\kappa (1 + \|x\|^\kappa)$, $\mathcal{P}(\mathbf{G}_{d,\varepsilon}) \leqslant \kappa d^\kappa \varepsilon^{-\kappa}$, and*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) + \tfrac{1}{2}(\Delta_x u_d)(t,x) + f\big(u_d(t,x)\big) = 0 \tag{4.77}$$

(cf. Definitions 2.1, 2.3, 2.4, and 2.6). Then there exist $\mathbf{U}_{d,t,\varepsilon} \in \mathbf{N}$, $d \in \mathbb{N}$, $t \in [0,T]$, $\varepsilon \in (0,1]$, and $c \in (0,\infty)$ which satisfy for all $d \in \mathbb{N}$, $t \in [0,T]$, $\varepsilon \in (0,1]$ that $\mathcal{R}_{\mathfrak{r}}(\mathbf{U}_{d,t,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{P}(\mathbf{U}_{d,\varepsilon}) \leqslant c d^c \varepsilon^{-c}$, and

$$\left(\int_{[0,1]^d} \left|u_d(t,x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{U}_{d,t,\varepsilon}))(x)\right|^{\mathfrak{q}} \mathrm{d}x\right)^{1/\mathfrak{q}} \leqslant \varepsilon. \tag{4.78}$$

Proof of Corollary 4.15. The proof of Corollary 4.15 is thus complete. □

Acknowledgments

The first author acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Müenster: Dynamics-Geometry-Structure. The third author acknowledges funding by the National Science Foundation (NSF 1903450).

References

[1] 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.

[2] 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.

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

[4] 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.

[5] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *J. Mach. Learn. Res. 20* (2019), Paper No. 74, 25.

[6] Becker, S., Cheridito, P., and Jentzen, A. Pricing and hedging American-style options with deep learning. *Journal of Risk and Financial Management 13*, 7 (Jul 2020), 158.

[7] Becker, S., Cheridito, P., Jentzen, A., and Welti, T. Solving high-dimensional optimal stopping problems using deep learning. *European J. Appl. Math. 32*, 3 (2021), 470–514.

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

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

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

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

[12] 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.

[13] 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.

[14] 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.

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

[16] 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).

[17] Gonon, L., Grohs, P., Jentzen, A., Kofler, D., and Šiška, D. Uniform error estimates for artificial neural network approximations for heat equations. *arXiv:1911.09647* (2019), 70 pages.

[18] Gonon, L., and Schwab, C. Deep ReLU network expression rates for option prices in high-dimensional, exponential Lévy models. Tech. Rep. 2020-52, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2020.

[19] Goudenège, L., Molent, A., and Zanette, A. Machine learning for pricing American options in high-dimensional Markovian and non-Markovian models. *Quantitative Finance 20*, 4 (2020), 573–591.

[20] Grohs, P., and Herrmann, L. Deep neural network approximation for high-dimensional elliptic PDEs with boundary conditions. *arXiv:2007.05384* (2020), 22 pages.

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

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

[23] Grohs, P., Jentzen, A., and Salimova, D. Deep neural network approximations for Monte Carlo algorithms. *arXiv:1908.10828* (2019), 45 pages.

[24] 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.

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

[26] 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.

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

[28] 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.

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

[30] Hutzenthaler, M., Jentzen, A., Kruse, T., and Nguyen, T. A. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *SN Part. Diff. Equ. Appl. 1*, 2 (2020), 1–34.

[31] 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.

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

[33] 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.

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

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

[36] Lye, K. O., Mishra, S., and Ray, D. Deep learning observables in computational fluid dynamics. *J. Comput. Phys. 410* (2020), 109339, 26.

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

[38] 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), 1–48.

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

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

[41] 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.

[42] Salvador, B., Oosterlee, C. W., and van der Meer, R. Financial option valuation by unsupervised learning with artificial neural networks. *Mathematics 9*, 1 (2021), 46.

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

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