Expressivity of deep neural networks for space-time solutions of high-dimensional nonlinear PDEs

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

[1] Faculty of Mathematics and Computer Science, University of Münster, Münster, Germany, e-mail: ajentzen@uni-muenster.de

[2] Faculty of Mathematics and Computer Science, University of Münster, Münster, Germany, e-mail: bkuckuck@uni-muenster.de

[3] Department of Mathematics and Statistics, Texas Tech University, Lubbock, Texas, USA, e-mail: joshua.padgett@ttu.edu

[4] Center for Astrophysics, Space Physics, and Engineering Research and Department of Physics, Baylor University, Waco, Texas, USA, e-mail: joshua.padgett@ttu.edu

Last compiled: September 14, 2020 4:10pm -05:00

Abstract

Recently, several variants of deep learning based approximation methods for partial differential equations (PDEs) have been proposed and a number of very encouraging numerical simulations have indicated the potential of such approximation methods to overcome the curse of dimensionality in the numerical approximation of high-dimensional PDEs. Nonetheless, there is as yet no comprehensive mathematical theory which explains why these methods seem to overcome the curse of dimensionality. However, there are now several partial results available in the scientific literature which rigorously prove that deep neural network (DNN) approximations indeed overcome the curse of dimensionality in the approximation of PDEs in the sense that the number of real parameters used to describe the approximating DNNs grows at most polynomially in both the PDE dimension $d \in \mathbb{N} = \{1, 2, 3, \dots\}$ and the reciprocal $1/\varepsilon$ of the prescribed approximation accuracy $\varepsilon \in (0, \infty)$. In the case of nonlinear PDEs these prior works study DNN approximations for solutions of PDEs only at the time of maturity $T \in (0, \infty)$ and it remained an open question whether DNN approximations can also approximate entire solutions of nonlinear PDEs on the space-time region $[0, T] \times \mathbb{R}^d$ without the curse of dimensionality. It is precisely the subject of this article to overcome this obstacle. In particular, the main result of this article shows that for all $a \in \mathbb{R}$, $b \in [a, \infty)$ it holds that DNNs can approximate solutions of PDEs with Lipschitz nonlinearities on the space-time region $[0, T] \times [a, b]^d$ without the curse of dimensionality.

Contents

1 Introduction

Recently, several variants of deep learning based approximation methods for partial differential equations (PDEs) have been proposed and a number of very encouraging numerical simulations have indicated the potential of such approximation methods to overcome the curse of dimensionality in the numerical approximation of high-dimensional PDEs (cf., e.g., [1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 13, 14, 16, 21, 22, 23, 25, 27, 30, 31, 32, 33, 34, 35, 37]). There is as yet no comprehensive mathematical theory which explains why these methods seem to overcome the curse of dimensionality. However, there are now several partial results available in the scientific literature which rigorously prove that deep neural network (DNN) approximations indeed overcome the curse of dimensionality in the approximation of PDEs in the sense that the number of real parameters used to describe the approximating DNNs grows at most polynomially in both the PDE dimension $d \in \mathbb{N} = \{1, 2, 3, \dots\}$ and the reciprocal $1/\varepsilon$ of the prescribed approximation accuracy $\varepsilon \in (0, \infty)$; cf., e.g., [?, 7, 12, 15, 17, 18, 20, 24, 26, 28, 29, 36]. The articles [7, 12, 15, 17, 18, 20, 24, 28, 29, 36] study DNN approximations for linear PDEs and the articles [?, 26] study DNN approximations for nonlinear PDEs. Except for the articles [19, 24] in the case of linear PDEs, all of the above articles study DNN approximations for solutions of PDEs at the time of maturity $T \in (0, \infty)$ but do not provide approximations for the entire PDE solution on $[0, T] \times \mathbb{R}^d$ and it remained an open question whether DNN approximations can also approximate solutions of nonlinear PDEs on the space-time region $[0, T] \times \mathbb{R}^d$ without the curse of dimensionality.

It is precisely the subject of this article to overcome this obstacle and to prove that DNNs have the power to approximate solutions of certain nonlinear PDEs on the entire space-time region $[0, T] \times \mathbb{R}^d$ without the curse of dimensionality. In particular, the main result of this article, **??** in **??** below, shows that for all $a \in \mathbb{R}$, $b \in [a, \infty)$ it holds that DNNs can approximate solutions of PDEs with Lipschitz nonlinearities on the space-time region $[0, T] \times [a, b]^d$ without the curse of dimensionality. In order to lay out the findings of this work in more detail, we present in Theorem 1.1 below a special case of **??** in **??** below.

Theorem 1.1. *Let* $\mathbf{A} \colon (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d) \to (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d)$ *and* $\|\cdot\| \colon (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d) \to [0, \infty)$ *satisfy for all* $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ *that* $\mathbf{A}(x) = (\max\{x_1, 0\}, \max\{x_2, 0\}, \dots, \max\{x_d, 0\})$ *and* $\|x\| = [\sum_{k=1}^d (x_k)^2]^{1/2}$, *let* $\mathbf{N} = \bigcup_{L \in \mathbb{N}} \bigcup_{l_0, l_1, \dots, l_L \in \mathbb{N}} (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, *let* $\mathcal{R} \colon \mathbf{N} \to (\bigcup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ *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* $\forall\, k \in \{1, 2, \dots, L-1\} \colon x_k = \mathbf{A}(W_k x_{k-1} + B_k)$ *that* $\mathcal{R}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L})$, $(\mathcal{R}(\Phi))(x_0) = W_L x_{L-1} + B_L$, *and* $\mathcal{P}(\Phi) = \sum_{k=1}^L l_k(l_{k-1} + 1)$, *let* $T, \kappa, p \in (0, \infty)$, $(\mathfrak{g}_{d,\varepsilon})_{(d,\varepsilon) \in (\mathbb{N} \times (0,1])} \subseteq \mathbf{N}$, *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* $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $t \in [0, T]$ *that* $\mathcal{R}(\mathfrak{g}_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\varepsilon \|(\nabla_x u_d)(0, x)\| + \varepsilon |u_d(t, x)| + |u_d(0, x) - (\mathcal{R}(\mathfrak{g}_{d,\varepsilon}))(x)| \leq \varepsilon \kappa d^\kappa (1 + \|x\|^\kappa)$, $\mathcal{P}(\mathfrak{g}_{d,\varepsilon}) \leq \kappa d^\kappa \varepsilon^{-\kappa}$, *and*

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

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

$$\left[\int_{[0,T] \times [0,1]^d} |u_d(y) - (\mathcal{R}(\mathfrak{u}_{d,\varepsilon}))(y)|^p \, dy \right]^{1/p} \leq \varepsilon. \tag{1.2}$$

Theorem 1.1 is an immediate consequence of **??** in **??** below. **??**, in turn, follows from **??** which is the main result of this article (see **??** below for details). In the following we provide some explanatory comments concerning the mathematical objects appearing in Theorem 1.1 above. The function $\mathbf{A} \colon (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d) \to (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d)$ 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 $\|\cdot\| \colon (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d) \to [0, \infty)$ describes the standard norms on $\mathbb{R}^d$, $d \in \mathbb{N}$, in the sense that for all $d \in \mathbb{N}$ it holds that $\|\cdot\| \colon (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d) \to [0, \infty)$ restricted to $\mathbb{R}^d$ is nothing but the standard norm on $\mathbb{R}^d$. The set $\mathbf{N} = \bigcup_{L \in \mathbb{N}} \bigcup_{l_0, l_1, \dots, l_L \in \mathbb{N}} (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$ in Theorem 1.1 above represents the set of all neural networks which we employ to approximate the solutions of the PDEs under consideration. The function $\mathcal{R} \colon \mathbf{N} \to (\bigcup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ in Theorem 1.1 above assigns to each neural network its realization function. More specifically, we observe that for every neural network $\Phi \in \mathbf{N}$ it holds that $\mathcal{R}(\Phi) \in (\bigcup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l))$ is the realization function of the neural network Φ with the activation functions being multidimensional versions of the rectifier function provided by $\mathbf{A} \colon (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d) \to (\bigcup_{d \in \mathbb{N}} \mathbb{R}^d)$. 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 $\Phi \in \mathbf{N}$ it holds 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 (1.1)) whose solutions we intend to approximate by DNNs in (1.2) 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 (1.2) above for details). In Theorem 1.1 we assume that the initial conditions of the PDEs (see (1.1)) whose solutions we intend to approximate by DNNs without the curse of dimensionality can be approximated by DNNs without the curse of dimensionality. The neural networks $(\mathfrak{g}_{d,\varepsilon})_{(d,\varepsilon)\in\mathbb{N}\times(0,1]} \subseteq \mathbf{N}$ serve as such approximating DNNs for the initial conditions of the PDEs (see (1.1)) whose solutions we intend to approximate. In particular, we note that the hypothesis that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$, $t \in [0,T]$ it holds that $\varepsilon\|(\nabla_x u_d)(0,x)\| + \varepsilon|u_d(t,x)| + |u_d(0,x) - (\mathcal{R}(\mathfrak{g}_{d,\varepsilon}))(x)| \leq \varepsilon\kappa d^\kappa(1+\|x\|^\kappa)$ in Theorem 1.1 above ensures that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $(\mathcal{R}(\mathfrak{g}_{d,\varepsilon}))(x)$ converges 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 (1.1)) 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 (1.1). Theorem 1.1 establishes that there exist neural networks $\mathfrak{u}_{d,\varepsilon} \in \mathbf{N}$, $(d,\varepsilon) \in \mathbb{N}\times(0,1]$, such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that the L^p-distance between the exact solution $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ of the PDE in (1.1) and the realization $\mathcal{R}(\mathfrak{u}_{d,\varepsilon})\colon \mathbb{R}^{d+1} \to \mathbb{R}$ of the neural network $\mathfrak{u}_{d,\varepsilon}$ with respect to the Lebesgue measure on the space-time region $[0,T]\times[0,1]^d$ is bounded by ε and such that the number of parameters of the neural networks $\mathfrak{u}_{d,\varepsilon} \in \mathbf{N}$, $(d,\varepsilon) \in \mathbb{N}\times(0,1]$, 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 $[0,T]\times[0,1]^d$ but our more general DNN approximation results in **??** below (see **??** and **??** in **??**) allow measuring the L^p-distance with respect to more general probability measures on $[0,T]\times\mathbb{R}^d$. In particular, for all $a \in \mathbb{R}$, $b \in (a,\infty)$ we have that the more general DNN approximation results in **??** below allow measuring the L^p-distance with respect to the uniform distribution on $[0,T]\times[a,b]^d$.

The rest of this article is structured in the following way:

2 Properties of solutions of partial differential equations (PDEs)

A comment from Josh: Should I refer to the equations as "stochastic fixed point equations", rather than PDEs?

2.1 An a priori bound for solutions of PDEs

A comment from Josh: This is new...

Definition 2.1 (The Euclidean norm)**.** *We denote by $\|\cdot\| : (\bigcup_{d\in\mathbb{N}} \mathbb{R}^d) \to [0,\infty)$ the function which satisfies for all $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ that $\|x\| = [\sum_{i=1}^d |x_i|^2]^{1/2}$.*

A comment from Josh: This is new...
Note: The result can be cited from the paper with Nguyen, if we want...

Lemma 2.2. *Let $d \in \mathbb{N}$, $T, L, C \in (0,\infty)$, $p, q \in [1,\infty)$, let $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R},\mathbb{R})$ and $g \in C(\mathbb{R}^d,\mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ that $|f(t,x,v)-f(t,x,w)| \leq L|v-w|$ and $|g(x)| \leq C(1+\|x\|)^p$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\mathbf{W}\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard Brownian motion, let $u \in C([0,T]\times\mathbb{R}^d,\mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that $\mathbb{E}[|g(x+\mathbf{W}_{T-t})| + \int_t^T |f(s,x+\mathbf{W}_{s-t},u(s,x+\mathbf{W}_{s-t}))|\,ds] < \infty$ and*

$$u(t,x) = \mathbb{E}[g(x+\mathbf{W}_{T-t})] + \int_t^T \mathbb{E}[f(s,x+\mathbf{W}_{s-t},u(s,x+\mathbf{W}_{s-t}))]\,ds \tag{2.1}$$

(cf. Definition 2.1). Then it holds for all $x \in \mathbb{R}^d$ that

$$\sup_{t\in[0,T]} \left(\mathbb{E}\Big[\big|u(t, x + \mathbf{W}_t)\big|^q\Big]\right)^{1/q} \leq e^{LT}(T+1)C \left[\sup_{t\in[0,T]} \left(\mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_t\|\big)^{pq}\Big]\right)^{1/q}\right]. \tag{2.2}$$

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

$$\mu_t(B) = \mathbb{P}(x + \mathbf{W}_t \in B). \tag{2.3}$$

The integral transformation theorem, (2.1), and the triangle inequality show for all $t \in [0,T]$ that

$$\begin{aligned}
&\left(\mathbb{E}\Big[|u(t, x + \mathbf{W}_t)|^q\Big]\right)^{1/q} = \left(\int_{\mathbb{R}^d} |u(t,z)|^q \, \mu_t(dz)\right)^{1/q} \\
&= \left(\int_{\mathbb{R}^d} \left|\mathbb{E}\left[g(z + \mathbf{W}_{T-t}) + \int_t^T f(s, z + \mathbf{W}_{s-t}, u(s, z + \mathbf{W}_{s-t}))\, ds\right]\right|^q \mu_t(dz)\right)^{1/q} \\
&\leq \left(\int_{\mathbb{R}^d} \big|\mathbb{E}\big[g(z + \mathbf{W}_{T-t})\big]\big|^q \, \mu_t(dz)\right)^{1/q} \\
&\quad + \int_t^T \left(\int_{\mathbb{R}^d} \big|\mathbb{E}\big[f(s, z + \mathbf{W}_{s-t}, u(s, z + \mathbf{W}_{s-t}))\big]\big|^q \, \mu_t(dz)\right)^{1/q} ds.
\end{aligned} \tag{2.4}$$

Next, Jensen's inequality, Fubini's theorem, (2.3), the fact that $\mathbf{W}$ has independent and stationary increments, and the fact that for all $x \in \mathbb{R}^d$ it holds that $|g(x)| \leq C(1 + \|x\|^p)$ demonstrate that for all $t \in [0,T]$ it holds that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \big|\mathbb{E}\big[g(z + \mathbf{W}_{T-t})\big]\big|^q \, \mu_t(dz) \leq \int_{\mathbb{R}^d} \mathbb{E}\Big[|g(z + \mathbf{W}_T - \mathbf{W}_t)|^q\Big] \, \mu_t(dz) \\
&= \mathbb{E}\Big[|g(x + \mathbf{W}_t + \mathbf{W}_T - \mathbf{W}_t)|^q\Big] = \mathbb{E}\Big[|g(x + \mathbf{W}_T)|^q\Big] \leq \mathbb{E}\Big[C^q\Big(1 + \|x + \mathbf{W}_T\|\Big)^{pq}\Big].
\end{aligned} \tag{2.5}$$

Furthermore, Jensen's inequality, Fubini's theorem, (2.3), the fact that $\mathbf{W}$ has independent and stationary increments, the triangle inequality, the fact that for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ that $|f(t,x,v) - f(t,x,w)| \leq L|v - w|$, and the fact that for all $x \in \mathbb{R}^d$ it holds that $|g(x)| \leq C(1 + \|x\|^p)$ demonstrate for all $t \in [0,T]$ that

$$\begin{aligned}
&\int_t^T \left(\int_{\mathbb{R}^d} \big|\mathbb{E}\big[f(s, z + \mathbf{W}_{s-t}, u(s, z + \mathbf{W}_{s-t}))\big]\big|^q \, \mu_t(dz)\right)^{1/q} ds \\
&\quad\leq \int_t^T \left(\mathbb{E}\Big[\big|f(s, z + \mathbf{W}_s, u(s, z + \mathbf{W}_s))\big|^q\Big]\right)^{1/q} ds \\
&\quad\leq \int_t^T \left(\mathbb{E}\Big[|f(s, x + \mathbf{W}_s, 0)|^q\Big]\right)^{1/q} ds \\
&\quad\quad + \int_t^T \left(\mathbb{E}\Big[|f(s, z + \mathbf{W}_s, u(s, z + \mathbf{W}_s)) - f(s, x + \mathbf{W}_s, 0)|^q\Big]\right)^{1/q} ds \\
&\leq T \sup_{s\in[0,T]} \left(\mathbb{E}\Big[C^q\Big(1 + \|x + \mathbf{W}_s\|\Big)^{pq}\Big]\right)^{1/q} + \int_t^T \left(\mathbb{E}\Big[L^q\, |u(s, x + \mathbf{W}_s)|^q\Big]\right)^{1/q} ds.
\end{aligned} \tag{2.6}$$

Combining this with (2.4) and (2.5) implies that for all $t \in [0,T]$ it holds that

$$\begin{aligned}&\left(\mathbb{E}\Big[|u(t,x+\mathbf{W}_t)|^q\Big]\right)^{1/q}\\&\leq (T+1)C\sup_{s\in[0,T]}\left(\mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big]\right)^{1/q}+L\int_t^T\left(\mathbb{E}\Big[|u(s,x+\mathbf{W}_s)|^q\Big]\right)^{1/q}ds.\end{aligned}\tag{2.7}$$

Next, **A comment from Josh: Add citation...** shows that

$$\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{|u(s,y)|}{(1+\|y\|)^p}\leq\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{|u(s,y)|}{1+\|y\|^p}<\infty.\tag{2.8}$$

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

$$\begin{aligned}\int_0^T\left(\mathbb{E}\Big[|u(s,x+\mathbf{W}_s)|^q\Big]\right)^{1/q}ds&\leq\left[\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{|u(s,y)|}{(1+\|y\|)^p}\right]\int_0^T\left(\mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big]\right)^{1/q}ds\\&\leq\left[\sup_{s\in[0,T]}\sup_{y\in\mathbb{R}^d}\frac{|u(s,y)|}{(1+\|y\|)^p}\right]T\left(1+\|x\|+\left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{pq}\Big]\right)^{1/(pq)}\right)^p<\infty.\end{aligned}\tag{2.9}$$

This, Gronwall's integral inequality, and (2.7) establish for all $t\in[0,T]$ that

$$\left(\mathbb{E}\Big[|u(t,x+\mathbf{W}_t)|^q\Big]\right)^{1/q}\leq e^{LT}(T+1)C\sup_{s\in[0,T]}\left(\mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big]\right)^{1/q}.\tag{2.10}$$

The proof of Lemma 2.2 is thus completed. □

2.2 Stability properties for solutions of PDEs

A comment from Josh: This is new...

Lemma 2.3. *Let $d\in\mathbb{N}$, $T,L,C,B\in(0,\infty)$, $p,q\in[1,\infty)$, let $f_1,f_2\in C([0,T]\times\mathbb{R}^d\times\mathbb{R},\mathbb{R})$ and $g_1,g_2\in C(\mathbb{R}^d,\mathbb{R})$ satisfy for all $t\in[0,T]$, $x\in\mathbb{R}^d$, $v,w\in\mathbb{R}$, $i\in\{1,2\}$ that $|f_i(t,x,v)-f_i(t,x,w)|\leq L|v-w|$, $|g_i(x)|\leq C(1+\|x\|)^p$, and $\max\{|f_1(t,x,v)-f_2(t,x,v)|,|g_1(x)-g_2(x)|\}\leq B((1+\|x\|)^{pq}+|v|^q)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\mathbf{W}\colon[0,T]\times\Omega\to\mathbb{R}^d$ be a standard Brownian motion, and let $u_1,u_2\in C([0,T]\times\mathbb{R}^d,\mathbb{R})$ satisfy for all $t\in[0,T]$, $x\in\mathbb{R}^d$, $i\in\{1,2\}$ that $\mathbb{E}[|g_i(x+\mathbf{W}_{T-t})|+\int_t^T|f_i(s,x+\mathbf{W}_{s-t},u_i(s,x+\mathbf{W}_{s-t}))|\,ds]<\infty$ and*

$$u_i(t,x)=\mathbb{E}[g_i(x+\mathbf{W}_{T-t})]+\int_t^T\mathbb{E}[f_i(s,x+\mathbf{W}_{s-t},u_i(s,x+\mathbf{W}_{s-t}))]\,ds\tag{2.11}$$

(cf. Definition 2.1). Then it holds for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\mathbb{E}\Big[\big|u_1(t,x+\mathbf{W}_t)-u_2(t,x+\mathbf{W}_t)\big|\Big]\leq B\big(e^{LT}(T+1)\big)^{q+1}\big(C^q+1\big)\left[\sup_{s\in[0,T]}\mathbb{E}\Big[\big(1+\|x+\mathbf{W}_s\|\big)^{pq}\Big]\right].\tag{2.12}$$

Proof of Lemma 2.3. First, (2.11), the triangle inequality, and the fact that $\mathbf{W}$ has stationary increments show for all $t\in[0,T]$, $x\in\mathbb{R}^d$ that

$$\begin{aligned}&|u_1(t,x)-u_2(t,x)|\\&\leq\mathbb{E}\Big[\big|g_1(x+\mathbf{W}_{T-t})-g_2(x+\mathbf{W}_{T-t})\big|\Big]\\&\quad+\int_t^T\mathbb{E}\Big[\big|f_1(s,x+\mathbf{W}_{s-t},u_1(s,x+\mathbf{W}_{s-t}))-f_1(s,x+\mathbf{W}_{s-t},u_2(s,x+\mathbf{W}_{s-t}))\big|\Big]\,ds\\&\quad+\int_t^T\mathbb{E}\Big[\big|f_1(s,x+\mathbf{W}_{s-t},u_2(s,x+\mathbf{W}_{s-t}))-f_2(s,x+\mathbf{W}_{s-t},u_2(s,x+\mathbf{W}_{s-t}))\big|\Big]\,ds.\end{aligned}\tag{2.13}$$

This, Fubini's theorem, the fact that $\mathbf{W}$ has independent increments, and the fact that for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$, $i \in \{1,2\}$ that $|f_i(t,x,v) - f_i(t,x,w)| \leq L|v-w|$ ensure that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\mathbb{E}\Big[\big|u_1(t,x+\mathbf{W}_t) - u_2(t,x+\mathbf{W}_t)\big|\Big] \\
&\leq \mathbb{E}\Big[\big|g_1(x+\mathbf{W}_T) - g_2(x+\mathbf{W}_T)\big|\Big] \\
&\quad + \int_t^T \mathbb{E}\Big[|f_1(s,x+\mathbf{W}_s,u_1(s,x+\mathbf{W}_s)) - f_1(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s))|\Big]\,ds \\
&\quad + \int_t^T \mathbb{E}\Big[|f_1(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s)) - f_2(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s))|\Big]\,ds \\
&\leq \mathbb{E}\Big[\big|g_1(x+\mathbf{W}_T) - g_2(x+\mathbf{W}_T)\big|\Big] \\
&\quad + L\int_t^T \mathbb{E}\Big[|u_1(s,x+\mathbf{W}_s) - u_2(s,x+\mathbf{W}_s)|\Big]\,ds \\
&\quad + T\sup_{s\in[0,T]} \mathbb{E}\Big[|f_1(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s)) - f_2(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s))|\Big].
\end{aligned} \tag{2.14}$$

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

$$\begin{aligned}
&\sup_{s\in[0,T]} \mathbb{E}\Big[\big|u_1(s,x+\mathbf{W}_s) - u_2(s,x+\mathbf{W}_s)\big|\Big] \\
&\leq e^{LT}\Bigg(\mathbb{E}\Big[\big|g_1(x+\mathbf{W}_T) - g_2(x+\mathbf{W}_T)\big|\Big] \\
&\qquad + T\sup_{s\in[0,T]} \mathbb{E}\Big[|f_1(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s)) - f_2(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s))|\Big]\Bigg).
\end{aligned} \tag{2.15}$$

Furthermore, the fact that for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $v \in \mathbb{R}$, $i \in \{1,2\}$ it holds that $\max\{|f_1(t,x,v) - f_2(t,x,v)|, |g_1(x) - g_2(x)|\} \leq B((1+\|x\|)^{pq} + |v|^q)$, the triangle inequality, and Lemma 2.2 imply for all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\mathbb{E}\Big[\big|g_1(x+\mathbf{W}_T) - g_2(x+\mathbf{W}_T)\big|\Big] \\
&\qquad + T\sup_{s\in[0,T]} \mathbb{E}\Big[|f_1(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s)) - f_2(s,x+\mathbf{W}_s,u_2(s,x+\mathbf{W}_s))|\Big] \\
&\leq B\sup_{s\in[0,T]} \mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big] + BT\sup_{s\in[0,T]} \mathbb{E}\Big[|u_2(x+\mathbf{W}_s)|^q\Big]. \\
&\leq B\sup_{s\in[0,T]} \mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big] + BT(e^{LT}(T+1)C)^q \sup_{s\in[0,T]} \mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big] \\
&\leq B(T+1)\left(e^{LT}(T+1)\right)^q (C^q+1) \sup_{s\in[0,T]} \mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big].
\end{aligned} \tag{2.16}$$

This, (2.15), and the triangle inequality yield that

$$\begin{aligned}
&\sup_{s\in[0,T]} \mathbb{E}\Big[\big|u_1(s,x+\mathbf{W}_s) - u_2(s,x+\mathbf{W}_s)\big|\Big] \\
&\leq B\left(e^{LT}(T+1)\right)^{q+1} (C^q+1) \sup_{s\in[0,T]} \mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_s\|\Big)^{pq}\Big].
\end{aligned} \tag{2.17}$$

The proof of Lemma 2.3 is thus completed. □

A comment from Josh: This is new...

Corollary 2.4. *Let $d \in \mathbb{N}$, $T, L, C, B \in (0,\infty)$, $p, q \in [1,\infty)$, let $f_1, f_2 \in C([0,T] \times \mathbb{R}^d \times \mathbb{R}, \mathbb{R})$ and $g_1, g_2 \in C(\mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$, $i \in \{1,2\}$ that $|f_i(t,x,v) - f_i(t,x,w)| \le L|v-w|$, $|g_i(x)| \le C(1+\|x\|)^p$, and $\max\{|f_1(t,x,v) - f_2(t,x,v)|, |g_1(x) - g_2(x)|\} \le B((1+\|x\|)^{pq} + |v|^q)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathbf{W}\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion, and let $u_1, u_2 \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $i \in \{1,2\}$ that $\mathbb{E}[|g_i(x + \mathbf{W}_{T-t})| + \int_t^T |f_i(s, x + \mathbf{W}_{s-t}, u_i(s, x + \mathbf{W}_{s-t}))|\, ds] < \infty$ and*

$$u_i(t,x) = \mathbb{E}[g_i(x + \mathbf{W}_{T-t})] + \int_t^T \mathbb{E}[f_i(s, x + \mathbf{W}_{s-t}, u_i(s, x + \mathbf{W}_{s-t}))]\, ds \tag{2.18}$$

(cf. Definition 2.1). Then it holds for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\big|u_1(t,x) - u_2(t,x)\big| \le B\big(e^{LT}(T+1)\big)^{q+1}(C^q+1)\left[\sup_{s\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_s\|\big)^{pq}\Big]\right]. \tag{2.19}$$

Note: Finish updating this proof...

Proof of Corollary 2.4. Throughout this proof let $V_{i,\mathfrak{t}}\colon [0, T-\mathfrak{t}] \times \mathbb{R}^d \to \mathbb{R}$, $\mathfrak{t} \in [0,T]$, and $F_{i,\mathfrak{t}}\colon [0, T-\mathfrak{t}] \times \mathbb{R}^d \times \mathbb{R}, \mathbb{R})$, $\mathfrak{t} \in [0,T]$, be the functions which satisfy for all $t \in [0, T-\mathfrak{t}]$, $x \in \mathbb{R}^d$, $v \in \mathbb{R}$, $i \in \{1,2\}$ that $V_{i,\mathfrak{t}}(t,x) = u_i(t+\mathfrak{t}, x)$ and $F_{i,\mathfrak{t}}(t,x,v) = f_i(t+\mathfrak{t}, x, v)$. Note that (2.18) and **A comment from Josh: Add reference/reason...** ensure for all $t \in [0, T-\mathfrak{t}]$, $x \in \mathbb{R}^d$, $i \in \{1,2\}$ it holds that

$$\begin{aligned} V_{i,\mathfrak{t}}(t,x) &= u_i(t+\mathfrak{t}, x) \\ &= \mathbb{E}[g_i(x + \mathbf{W}_{T-(t+\mathfrak{t})})] + \int_{(t+\mathfrak{t})}^T \mathbb{E}[f_i(s, x + \mathbf{W}_{s-(t+\mathfrak{t})}, u_i(s, x + \mathbf{W}_{s-(t+\mathfrak{t})}))]\, ds \\ &= \mathbb{E}[g_i(x + \mathbf{W}_{(T-\mathfrak{t})-t})] + \int_t^{(T-\mathfrak{t})} \mathbb{E}[f_i(s+\mathfrak{t}, x + \mathbf{W}_{s-t}, u_i(s+\mathfrak{t}, x + \mathbf{W}_{s-t}))]\, ds \\ &= \mathbb{E}[g_i(x + \mathbf{W}_{(T-\mathfrak{t})-t})] + \int_t^{(T-\mathfrak{t})} \mathbb{E}[F_{i,\mathfrak{t}}(s, x + \mathbf{W}_{s-t}, V_{i,\mathfrak{t}}(s, x + \mathbf{W}_{s-t}))]\, ds \end{aligned} \tag{2.20}$$

and

$$\begin{aligned} &\mathbb{E}\left[|g_i(x + \mathbf{W}_{(T-\mathfrak{t})-t})| + \int_t^{(T-\mathfrak{t})} |F_{i,\mathfrak{t}}(s, x + \mathbf{W}_{s-t}, V_{i,\mathfrak{t}}(s, x + \mathbf{W}_{s-t}))|\, ds\right] \\ &= \mathbb{E}\left[|g_i(x + \mathbf{W}_{T-(t+\mathfrak{t})})| + \int_{(t+\mathfrak{t})}^T |f_i(s, x + \mathbf{W}_{s-(t+\mathfrak{t})}, u_i(s, x + \mathbf{W}_{s-(t+\mathfrak{t})}))|\, ds\right] < \infty. \end{aligned} \tag{2.21}$$

Further note that for all $t \in [0, T-\mathfrak{t}]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ $i \in \{1,2\}$ it holds that

$$|F_{i,\mathfrak{t}}(t,x,v) - F_{i,\mathfrak{t}}(t,x,w)| = |f_i(t+\mathfrak{t}, x, v) - f_i(t+\mathfrak{t}, x, w)| \le L|v-w| \tag{2.22}$$

and

$$|F_{1,\mathfrak{t}}(t,x,v) - F_{2,\mathfrak{t}}(t,x,v)| = |f_1(t+\mathfrak{t}, x, v) - f_(t+\mathfrak{t}, x, v)| \le B(1+|v|^q)^{pq}). \tag{2.23}$$

In addition, note that the hypothesis that $u_1, u_2 \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ ensures that for all $\mathfrak{t} \in [0,T]$ it holds that $V_{1,\mathfrak{t}}, V_{2,\mathfrak{t}} \in C([0, T-\mathfrak{t}] \times \mathbb{R}^d, \mathbb{R})$. Combining this, (2.20), (2.21), (2.22), and (2.23) with Lemma 2.3 (with $u_1 = V_{1,\mathfrak{t}}$, $u_2 = V_{2,\mathfrak{t}}$, $f_1 = F_{1,\mathfrak{t}}$, $f_2 = F_{2,\mathfrak{t}}$, $g_1 = g_1$, $g_2 = g_2$, $T = T - \mathfrak{t}$ in the notation of Lemma 2.3) demonstrates for every $\mathfrak{t} \in [0,T]$, $t \in [0, T-\mathfrak{t}]$, $x \in \mathbb{R}^d$ that

$$\mathbb{E}\Big[\big|V_{1,\mathfrak{t}}(t, x + \mathbf{W}_t) - V_{2,\mathfrak{t}}(t, x + \mathbf{W}_t)\big|\Big] \le B\big(e^{LT}(T+1)\big)^{q+1}(C^q+1)\left[\sup_{s\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_s\|\big)^{pq}\Big]\right]. \tag{2.24}$$

This implies for all $\mathfrak{t} \in [0,T]$ that

$$\begin{aligned} \mathbb{E}\Big[\big|V_{1,\mathfrak{t}}(0, x + \mathbf{W}_0) - V_{2,\mathfrak{t}}(0, x + \mathbf{W}_0)\big|\Big] &= \mathbb{E}\Big[\big|u_1(\mathfrak{t}, x) - u_2(\mathfrak{t}, x)\big|\Big] = |u_1(\mathfrak{t}, x) - u_2(\mathfrak{t}, x)| \\ &\leq B\big(e^{LT}(T+1)\big)^{q+1}(C^q + 1)\left[\sup_{s\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_s\|\big)^{pq}\Big]\right]. \end{aligned} \tag{2.25}$$

The proof of Corollary 2.4 is thus completed. □

2.3 Temporal regularity properties for solutions of PDEs

A comment from Josh: This is new...

Lemma 2.5. *Let $d \in \mathbb{N}$, $C \in (0,\infty)$, $p \in [1,\infty)$, let $g \in C^1(\mathbb{R}^d, \mathbb{R})$ satisfy for all $x \in \mathbb{R}^d$ that $\|(\nabla g)(x)\| \leq C(1 + \|x\|)^p$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathbf{W}\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion (cf. Definition 2.1). Then it holds for all $t, \mathfrak{t} \in [0,T]$, $x \in \mathbb{R}^d$ that*

$$\mathbb{E}\Big[\big|g(x + \mathbf{W}_{\mathfrak{t}}) - g(x + \mathbf{W}_t)\big|\Big] \leq C|\mathfrak{t} - t|^{1/2}\,(d+2)^{1/2}\left[\sup_{s\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_s\|\big)^p\Big]\right]. \tag{2.26}$$

A comment from Josh: Double-check this proof...

Proof of Lemma 2.5. Note that the fact that for all $x \in \mathbb{R}^d$ it holds that $\|(\nabla g)(x)\| \leq C(1 + \|x\|)^p$, the fundamental theorem of calculus, the Cauchy-Schwarz inequality, and the fact that $\mathbf{W}$ has independent increments assure for all $t, \mathfrak{t} \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned} \mathbb{E}\Big[\big|g(x + \mathbf{W}_{\mathfrak{t}}) - g(x + \mathbf{W}_t)\big|\Big] &\leq \mathbb{E}\left[\left(\sup_{s\in[0,T]} \|(\nabla g)(x + \mathbf{W}_s)\|\right)\big\|\mathbf{W}_{|\mathfrak{t}-t|}\big\|\right] \\ &\leq \mathbb{E}\left[\sup_{s\in[0,T]} C\,(1 + \|x + \mathbf{W}_s\|)^p\right]\mathbb{E}\big[\big\|\mathbf{W}_{|\mathfrak{t}-t|}\big\|\big] \\ &\leq C\left[\sup_{s\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_s\|\big)^p\Big]\right]\mathbb{E}\big[\big\|\mathbf{W}_{|\mathfrak{t}-t|}\big\|\big]. \end{aligned} \tag{2.27}$$

Note that the fact that for all $t, \mathfrak{t} \in [0,T]$ the random variable $\|\mathbf{W}_{|\mathfrak{t}-t|}/\sqrt{|\mathfrak{t}-t|}\|^2$ is chi-squared distributed with d degrees of freedom and Jenson's inequality imply that for all $t, \mathfrak{t} \in [0,T]$ it holds that

$$\big|\mathbb{E}\big[\|\mathbf{W}_{|\mathfrak{t}-t|}\|\big]\big|^2 \leq \mathbb{E}\big[\|\mathbf{W}_{|\mathfrak{t}-t|}\|^2\big] = 2|t-\mathfrak{t}|\left[\frac{\Gamma\left(\frac{d}{2}+2\right)}{\Gamma\left(\frac{d}{2}\right)}\right]^{1/2} = 2|t-\mathfrak{t}|\left[\prod_{j=0}^{1}\left(\frac{d}{2}+j\right)\right]^{1/2} \leq |t-\mathfrak{t}|(d+2). \tag{2.28}$$

Combining (2.27) and (2.28) then yields the desired result. The proof of Lemma 2.5 is thus completed. □

A comment from Josh: This is new...
Note: I had to add an additional regularity assumption to f...

Lemma 2.6. *Let $d \in \mathbb{N}$, $T, L, C \in (0,\infty)$, $p \in [1,\infty)$, let $f \in C([0,T] \times \mathbb{R}^d \times \mathbb{R}, \mathbb{R})$ and $g \in C^1(\mathbb{R}^d, \mathbb{R})$ satisfy for all $s, t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ that $|f(s,x,v) - f(t,x,w)| \leq L(|s-t| + |v-w|)$ and $\max\{|f(t,x,0)|, \|(\nabla g)(x)\|\} \leq C(1 + \|x\|)^p$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathbf{W}\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion, and let $u \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that $\mathbb{E}[|g(x + \mathbf{W}_{T-t})| + \int_t^T |f(s, x + \mathbf{W}_{s-t}, u(s, x + \mathbf{W}_{s-t}))|\,ds] < \infty$ and*

$$u(t,x) = \mathbb{E}[g(x + \mathbf{W}_{T-t})] + \int_t^T \mathbb{E}[f(s, x + \mathbf{W}_{s-t}, u(s, x + \mathbf{W}_{s-t}))]\,ds \tag{2.29}$$

(cf. Definition 2.1). Then it holds for all $t, \mathfrak{t} \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\mathbb{E}\Big[|u(\mathfrak{t}, x + \mathbf{W}_t) - u(t, x + \mathbf{W}_t)|\Big] \leq \sqrt{|\mathfrak{t} - t|}\left(e^{LT}\left[C\Big(1 + (d+2)^{1/2}\Big) + L\big(T + Ce^{LT}\big)(T+1)\right]\left[\sup_{s \in [0,T]} \mathbb{E}\Big[(1 + \|x + \mathbf{W}_s\|)^p\Big]\right]\right). \tag{2.30}$$

A comment from Josh: I should probably clean up this bound (above)...

Note: Double-check proof...

Proof of Lemma 2.6. Throughout this proof let $\delta = \mathfrak{t} - t$. Without loss of generality, assume that $t \leq \mathfrak{t}$. First, (2.29), the triangle inequality, and the fact that $\mathbf{W}$ has stationary increments show for all $t, \mathfrak{t} \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\mathbb{E}\Big[\big|u(\mathfrak{t}, x) - u(t,x)\big|\Big] = \mathbb{E}\Big[\big|u(t+\delta, x) - u(t,x)\big|\Big] \\
&= \mathbb{E}\Bigg[\bigg|\mathbb{E}\big[g\big(x + \mathbf{W}_{T-(t+\delta)}\big) - g(x + \mathbf{W}_{T-t})\big] \\
&\quad + \int_{t+\delta}^T \mathbb{E}\big[f(s, x + \mathbf{W}_{s-(t+\delta)}, u(s, x + \mathbf{W}_{s-(t+\delta)}))\big]\, ds - \int_t^T \mathbb{E}\big[f(s, x + \mathbf{W}_{s-t}, u(s, x + \mathbf{W}_{s-t})\big]\, ds\bigg|\Bigg] \\
&= \mathbb{E}\Bigg[\bigg|\mathbb{E}\big[g\big(x + \mathbf{W}_{T-(t+\delta)}\big) - g(x + \mathbf{W}_{T-t})\big] \\
&\quad + \int_t^{T-\delta} \mathbb{E}\big[f(s+\delta, x + \mathbf{W}_{s-t}, u(s+\delta, x + \mathbf{W}_{s-t})\big]\, ds - \int_t^T \mathbb{E}\big[f(s, x + \mathbf{W}_{s-t}, u(s, x + \mathbf{W}_{s-t})\big]\, ds\bigg|\Bigg] \\
&\leq \mathbb{E}\Big[\big|g\big(x + \mathbf{W}_{T-(t+\delta)}\big) - g(x + \mathbf{W}_{T-t})\big|\Big] \\
&\quad + \int_t^{T-\delta} \mathbb{E}\Big[\big|f(s, x + \mathbf{W}_{s-t}, u(s+\delta, x + \mathbf{W}_{s-t}) - f(s, x + \mathbf{W}_{s-t}, u(s, x + \mathbf{W}_{s-t})\big|\Big]\, ds \\
&\quad + \int_t^{T-\delta} \mathbb{E}\Big[\big|f(s+\delta, x + \mathbf{W}_{s-t}, u(s+\delta, x + \mathbf{W}_{s-t}) - f(s, x + \mathbf{W}_{s-t}, u(s+\delta, x + \mathbf{W}_{s-t})\big|\Big]\, ds \\
&\quad + \int_{T-\delta}^T \mathbb{E}\Big[\big|f(s, x + \mathbf{W}_{s-t}, u(s, x + \mathbf{W}_{s-t})\big|\Big]\, ds.
\end{aligned} \tag{2.31}$$

This, Fubini's theorem, the fact that $\mathbf{W}$ has independent increments, and the fact that for all $s, t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ that $|f(s,x,v) - f(t,x,w)| \leq L(|s-t| + |v-w|)$ ensure that for all $t, \mathfrak{t} \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\mathbb{E}\Big[\big|u(t+\delta, x + \mathbf{W}_t) - u(t, x + \mathbf{W}_t)\big|\Big] \\
&\leq \mathbb{E}\Big[\big|g(x + \mathbf{W}_{T-\delta}) - g(x + \mathbf{W}_T)\big|\Big] \\
&\quad + \int_t^{T-\delta} \mathbb{E}\Big[\big|f(s, x + \mathbf{W}_s, u(s+\delta, x + \mathbf{W}_s) - f(s, x + \mathbf{W}_s, u(s, x + \mathbf{W}_s)\big|\Big]\, ds \\
&\quad + \int_t^{T-\delta} \mathbb{E}\Big[\big|f(s+\delta, x + \mathbf{W}_s, u(s+\delta, x + \mathbf{W}_s) - f(s, x + \mathbf{W}_s, u(s, x + \mathbf{W}_s)\big|\Big]\, ds \\
&\quad + \int_{T-\delta}^T \mathbb{E}\Big[\big|f(s, x + \mathbf{W}_s, u(s, x + \mathbf{W}_s)\big|\Big]\, ds \\
&\leq \mathbb{E}\Big[\big|g(x + \mathbf{W}_{T-\delta}) - g(x + \mathbf{W}_T)\big|\Big] + L\int_t^{T-\delta} \mathbb{E}\Big[\big|u(s+\delta, x + \mathbf{W}_s) - u(s, x + \mathbf{W}_s)\big|\Big]\, ds \\
&\quad + \delta LT + \delta \sup_{s \in [0,T]} \mathbb{E}\Big[\big|f(s, x + \mathbf{W}_s, u(s, x + \mathbf{W}_s)\big|\Big].
\end{aligned} \tag{2.32}$$

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

$$\begin{aligned}
&\sup_{t\in[0,T-\delta]} \mathbb{E}\Big[\big|u(t+\delta, x+\mathbf{W}_t) - u(t, x+\mathbf{W}_t)\big|\Big] \\
&\le e^{LT}\Bigg(\mathbb{E}\Big[\,\big|g(x+\mathbf{W}_{T-\delta}) - g(x+\mathbf{W}_T)\big|\Big] + \delta LT \\
&\quad + \delta \sup_{s\in[0,T]} \mathbb{E}\Big[\big|f(s, x+\mathbf{W}_s, u(s, x+\mathbf{W}_s)\big|\Big]\Bigg).
\end{aligned} \tag{2.33}$$

Note that Lemma 2.2, the fact that for all $s,t \in [0,T]$, $x \in \mathbb{R}^d$, $v,w \in \mathbb{R}$ that $|f(s,x,v) - f(t,x,w)| \le L(|s-t| + |v-w|)$, and the fact that for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that $|f(t,x,0)| \le C(1+\|x\|)^p$, and the triangle inequality assure that for all $s \in [0,T]$ it holds that

$$\begin{aligned}
&\mathbb{E}\Big[\big|f(s, x+\mathbf{W}_s, u(s, x+\mathbf{W}_s)\big|\Big] \\
&\le \mathbb{E}\Big[\big|f(s, x+\mathbf{W}_s, u(s, x+\mathbf{W}_s) - f(s, x+\mathbf{W}_s, 0)\big|\Big] + \mathbb{E}\Big[\big|f(s, x+\mathbf{W}_s, 0)\big|\Big] \\
&\le L\left(\mathbb{E}\Big[|u(s, x+\mathbf{W}_s)|\Big]\right) + C\left(\mathbb{E}\Big[(1+\|x+\mathbf{W}_s\|)^p\Big]\right) \\
&\le L\left(e^{LT}(T+1)C\left[\sup_{t\in[0,T]} \mathbb{E}\Big[(1+\|x+\mathbf{W}_t\|)^p\Big]\right]\right) + C\left(\mathbb{E}\Big[(1+\|x+\mathbf{W}_s\|)^p\Big]\right) \\
&\le (L+1)Ce^{LT}(T+1)\left[\sup_{t\in[0,T]} \mathbb{E}\Big[(1+\|x+\mathbf{W}_t\|)^p\Big]\right].
\end{aligned} \tag{2.34}$$

This, (2.33), Lemma 2.5, and the triangle inequality then imply for all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\sup_{s\in[0,T-\delta]} \mathbb{E}\Big[\big|u(s+\delta, x+\mathbf{W}_s) - u(s, x+\mathbf{W}_s)\big|^q\Big] \\
&\le e^{LT}\Bigg(C\,(\delta(d+2))^{1/2}\left(\sup_{s\in[0,T]} \mathbb{E}\Big[(1+\|x+\mathbf{W}_s)\|)^p\Big]\right) + \delta LT \\
&\quad + \sqrt{\delta}(L+1)Ce^{LT}(T+1)\left[\sup_{s\in[0,T]} \mathbb{E}\Big[(1+\|x+\mathbf{W}_s\|)^p\Big]\right]\Bigg) \\
&\le \sqrt{\delta}\left(e^{LT}\left[C\left(1+(d+2)^{1/2}\right) + L\left(T+Ce^{LT}\right)(T+1)\right]\left[\sup_{s\in[0,T]} \mathbb{E}\Big[(1+\|x+\mathbf{W}_s\|)^{pq}\Big]\right]\right).
\end{aligned} \tag{2.35}$$

The proof of Lemma 2.6 is thus completed. □

A comment from Josh: This is new...

Corollary 2.7. *Let $d \in \mathbb{N}$, $T, L, C \in (0,\infty)$, $p, q \in [1,\infty)$, let $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R}, \mathbb{R})$ and $g \in C^1(\mathbb{R}^d, \mathbb{R})$ satisfy for all $s,t \in [0,T]$, $x \in \mathbb{R}^d$, $v,w \in \mathbb{R}$ that $|f(s,x,v) - f(t,x,w)| \le L(|s-t| + |v-w|)$ and $\max\{|f(t,x,0)|, \|(\nabla g)(x)\|\} \le C(1+\|x\|)^p$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathbf{W}\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard Brownian motion, and let $u \in C([0,T]\times\mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that $\mathbb{E}[|g(x+\mathbf{W}_{T-t})| + \int_t^T |f(s, x+\mathbf{W}_{s-t}, u(s, x+\mathbf{W}_{s-t}))|\,ds] < \infty$ and*

$$u(t,x) = \mathbb{E}[g(x+\mathbf{W}_{T-t})] + \int_t^T \mathbb{E}[f(s, x+\mathbf{W}_{s-t}, u(s, x+\mathbf{W}_{s-t}))]\,ds \tag{2.36}$$

(cf. Definition 2.1). Then it holds for all $t, \mathfrak{t} \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&|u(\mathfrak{t}, x) - u(t,x)| \\
&\le \sqrt{|\mathfrak{t}-t|}\left(e^{LT}\left[C\left(1+(d+2)^{1/2}\right) + L\left(T+Ce^{LT}\right)(T+1)\right]\left[\sup_{s\in[0,T]} \mathbb{E}\Big[(1+\|x+\mathbf{W}_s\|)^p\Big]\right]\right).
\end{aligned} \tag{2.37}$$

Proof of Corollary 2.7. **A comment from Josh: Add proof...** □

2.4 Full history recursive multilevel Picard (MLP) approximations of solutions of PDEs

A comment from Josh: This is new...

Lemma 2.8. *Let $d, M \in \mathbb{N}$, $T, L, C \in (0,\infty)$, $p, \mathfrak{q} \in [1,\infty)$, let $f \in C([0,T]\times\mathbb{R}^d\times\mathbb{R},\mathbb{R})$ and $g \in C(\mathbb{R}^d,\mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$, $v, w \in \mathbb{R}$ that $|f(t,x,v)-f(t,x,w)| \leq L|v-w|$ and $|g(x)| \leq C(1+\|x\|)^p$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\mathbf{W}\colon [0,T]\times\Omega \to \mathbb{R}^d$ be a standard Brownian motion, let $u \in C([0,T]\times\mathbb{R}^d,\mathbb{R})$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that $\mathbb{E}[|g(x+\mathbf{W}_{T-t})| + \int_t^T |f(s,x+\mathbf{W}_{s-t},u(s,x+\mathbf{W}_{s-t}))|\,ds] < \infty$ and*

$$u(t,x) = \mathbb{E}[g(x+\mathbf{W}_{T-t})] + \int_t^T \mathbb{E}[f(s,x+\mathbf{W}_{s-t},u(s,x+\mathbf{W}_{s-t}))]\,ds, \tag{2.38}$$

let $\Theta = (\bigcup_{n\in\mathbb{N}}\mathbb{Z}^n)$, let $\mathfrak{u}^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, be independent uniformly distributed 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 $W^\theta\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be independent standard Brownian motions, assume for all $\theta \in \Theta$ that $\mathcal{U}^\theta$ and W^θ are independent, for every $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$ let $X^\theta_{t,s,x}\colon \Omega \to \mathbb{R}^d$ satisfy $Y^\theta_{t,s} = W^\theta_s - W^\theta_t$, and let $U^\theta_n\colon [0,T]\times\mathbb{R}^d\times\Omega \to \mathbb{R}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, satisfy for all $n \in \mathbb{N}_0$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} U^\theta_n(t,x) &= \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n}\left[\sum_{k=1}^{M^n} g\Big(x+Y^{(\theta,0,-k)}_{t,T}\Big)\right] \\ &\quad + \sum_{i=0}^{n-1}\frac{(T-t)}{M^{n-i}}\left[\sum_{k=1}^{M^{n-i}}\Big(f\Big(\mathcal{U}^{(\theta,i,k)}_t, x+Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}, U^{(\theta,i,k)}_i\Big(\mathcal{U}^{(\theta,i,k)}_t, x+Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}\Big)\Big)\right. \\ &\quad \left. -\mathbb{1}_{\mathbb{N}}(i)\, f\Big(\mathcal{U}^{(\theta,i,k)}_t, x+Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}, U^{(\theta,-i,k)}_{\max\{i-1,0\}}\Big(\mathcal{U}^{(\theta,i,k)}_t, x+Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}\Big)\Big)\Big)\right] \end{aligned} \tag{2.39}$$

(cf. Definition 2.1). Then it holds for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\Big(\mathbb{E}\Big[|U^\theta_n(t,x)-u(t,x)|^{\mathfrak{q}}\Big]\Big)^{1/\mathfrak{q}} \leq e^{LT}(T+1)C\left[\sup_{s\in[0,T]}\mathbb{E}\Big[(1+\|x+\mathbf{W}_s\|)^{p\mathfrak{q}}\Big]^{1/\mathfrak{q}}\right]\left(\frac{e^{M/2}(1+2LT)^n}{M^{n/2}}\right). \tag{2.40}$$

Proof of Lemma 2.8. **A comment from Josh: Add proof...** □

3 Artificial neural network (ANN) calculus

3.1 ANNs

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

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

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

$$\mathbb{D}_n(\Phi) = \begin{cases} l_n & : n \leq L \\ 0 & : n > L \end{cases} \tag{3.2}$$

Definition 3.2 (Neural network). *We say that Φ is a neural network if and only if it holds that $\Phi \in \mathbf{N}$ (cf. Definition 3.1).*

Definition 3.3 (Maximum norm). *We denote by $|||\cdot||| : (\bigcup_{d\in\mathbb{N}} \mathbb{R}^d) \to [0,\infty)$ the function which satisfies for all $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ that*

$$|||x||| = \max_{i\in\{1,2,\dots,d\}} |x_i|. \tag{3.3}$$

3.2 Realizations of DNNs

Definition 3.4 (Rectifier function). *We denote by $\mathfrak{r}\colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that*

$$\mathfrak{r}(x) = \max\{x, 0\}. \tag{3.4}$$

Definition 3.5 (Multidimensional version). *Let $d \in \mathbb{N}$ and let $a \in C(\mathbb{R}, \mathbb{R})$ be a function. Then we denote by $\mathfrak{M}_{a,d}\colon \mathbb{R}^d \to \mathbb{R}^d$ the function which satisfies for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ that*

$$\mathfrak{M}_{a,d}(x) = (a(x_1), a(x_2), \dots, a(x_d)). \tag{3.5}$$

Definition 3.6 (Realization associated to a DNN). *Let $a \in C(\mathbb{R}, \mathbb{R})$. Then we denote by $\mathcal{R}_a\colon \mathbf{N} \to \big(\bigcup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)\big)$ 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 $\forall\, k \in \mathbb{N} \cap (0, L)\colon x_k = \mathfrak{M}_{a,l_k}(W_k x_{k-1} + B_k)$ that*

$$\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}) \qquad \text{and} \qquad (\mathcal{R}_a(\Phi))(x_0) = W_L x_{L-1} + B_L \tag{3.6}$$

(cf. Definitions 3.1 and 3.5).

A comment from Josh: Do I need this result?

Lemma 3.7. *Let $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

(i) it holds that $\mathcal{D}(\Phi) \in \mathbb{N}^{\mathcal{L}(\Phi)+1}$ and

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

(cf. Definition 3.6).

Proof of Lemma 3.7. Note that the assumption that $\Phi \in \mathbf{N} = \bigcup_{L\in\mathbb{N}} \bigcup_{(l_0,l_1,\dots,l_L)\in\mathbb{N}^{L+1}} (\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$ ensures that there exist $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$ such that

$$\Phi \in \big(\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big). \tag{3.7}$$

Observe that (3.7) assures that

$$\mathcal{L}(\Phi) = L, \qquad \mathcal{I}(\Phi) = l_0, \qquad \mathcal{O}(\Phi) = l_L, \tag{3.8}$$

$$\text{and} \qquad \mathcal{D}(\Phi) = (l_0, l_1, \dots, l_L) \in \mathbb{N}^{L+1} = \mathbb{N}^{\mathcal{L}(\Phi)+1}. \tag{3.9}$$

This establishes item (i). Moreover, note that (3.8) and (3.6) show that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$. This establishes item (ii). The proof of Lemma 3.7 is thus completed. □

3.3 Compositions of ANNs

3.3.1 Standard compositions of ANNs

Definition 3.8 (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* $L, \mathfrak{L} \in \mathbb{N}$, $l_0, l_1, \dots, l_L, \mathfrak{l}_0, \mathfrak{l}_1, \dots, \mathfrak{l}_{\mathfrak{L}} \in \mathbb{N}$, $\Phi_1 = ((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L)) \in \big(\bigtimes_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\big)$, $\Phi_2 = ((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}}, \mathscr{B}_{\mathfrak{L}})) \in \big(\bigtimes_{k=1}^{\mathfrak{L}} (\mathbb{R}^{\mathfrak{l}_k \times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k})\big)$ *with* $l_0 = \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2) = \mathfrak{l}_{\mathfrak{L}}$ *that*

$$\Phi_1 \bullet \Phi_2 = \begin{cases} \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1 \mathscr{W}_{\mathfrak{L}}, W_1 \mathscr{B}_{\mathfrak{L}} + B_1), \\ \qquad (W_2, B_2), (W_3, B_3), \dots, (W_L, B_L)\big) & : L > 1 < \mathfrak{L} \\ \big((W_1 \mathscr{W}_1, W_1 \mathscr{B}_1 + B_1), (W_2, B_2), (W_3, B_3), \dots, (W_L, B_L)\big) & : L > 1 = \mathfrak{L} \\ \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \dots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1 \mathscr{W}_{\mathfrak{L}}, W_1 \mathscr{B}_{\mathfrak{L}} + B_1)\big) & : L = 1 < \mathfrak{L} \\ \big((W_1 \mathscr{W}_1, W_1 \mathscr{B}_1 + B_1)\big) & : L = 1 = \mathfrak{L} \end{cases} \tag{3.10}$$

(cf. Definition 3.1).

3.3.2 Elementary properties of standard compositions of ANNs

A comment from Josh: I think I need this result...

Proposition 3.9. *Let* $\Phi_1, \Phi_2 \in \mathbf{N}$ *satisfy that* $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ *(cf. Definition 3.1). Then*

(i) it holds that

$$\mathcal{D}(\Phi_1 \bullet \Phi_2) = (\mathbb{D}_0(\Phi_2), \mathbb{D}_1(\Phi_2), \dots, \mathbb{D}_{\mathcal{L}(\Phi_2)-1}(\Phi_2), \mathbb{D}_1(\Phi_1), \mathbb{D}_2(\Phi_1), \dots, \mathbb{D}_{\mathcal{L}(\Phi_1)}(\Phi_1)), \tag{3.11}$$

(ii) it holds that

$$[\mathcal{L}(\Phi_1 \bullet \Phi_2) - 1] = [\mathcal{L}(\Phi_1) - 1] + [\mathcal{L}(\Phi_2) - 1], \tag{3.12}$$

(iii) it holds that

$$\mathcal{H}(\Phi_1 \bullet \Phi_2) = \mathcal{H}(\Phi_1) + \mathcal{H}(\Phi_2), \tag{3.13}$$

(iv) it holds that

$$\begin{aligned} \mathcal{P}(\Phi_1 \bullet \Phi_2) &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + \mathbb{D}_1(\Phi_1)(\mathbb{D}_{\mathcal{L}(\Phi_2)-1}(\Phi_2) + 1) \\ &\quad - \mathbb{D}_1(\Phi_1)(\mathbb{D}_0(\Phi_1) + 1) - \mathbb{D}_{\mathcal{L}(\Phi_2)}(\Phi_2)(\mathbb{D}_{\mathcal{L}(\Phi_2)-1}(\Phi_2) + 1) \\ &\leq \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + \mathbb{D}_1(\Phi_1)\mathbb{D}_{\mathcal{L}(\Phi_2)-1}(\Phi_2), \end{aligned} \tag{3.14}$$

and

(v) it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathcal{R}_a(\Phi_1 \bullet \Phi_2) \in C(\mathbb{R}^{\mathcal{I}(\Phi_2)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$ *and*

$$\mathcal{R}_a(\Phi_1 \bullet \Phi_2) = [\mathcal{R}_a(\Phi_1)] \circ [\mathcal{R}_a(\Phi_2)] \tag{3.15}$$

(cf. Definitions 3.6 and 3.8).

Proof of Proposition 3.9. Throughout this proof let $a \in C(\mathbb{R}, \mathbb{R})$, let $L_k \in \mathbb{N}$, $k \in \{1, 2\}$, satisfy for all $k \in \{1, 2\}$ that $L_k = \mathcal{L}(\Phi_k)$, let $l_{1,0}, l_{1,1}, \dots, l_{1,\mathcal{L}(\Phi_1)}, l_{2,0}, l_{2,1}, \dots, l_{2,\mathcal{L}(\Phi_2)} \in \mathbb{N}$, $\big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \dots, (W_{k,L_k}, B_{k,L_k})\big) \in (\times_{j=1}^{L_k} (\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}}))$, $k \in \{1, 2\}$, satisfy for all $k \in \{1, 2\}$ that

$$\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \dots, (W_{k,L_k}, B_{k,L_k})\big), \tag{3.16}$$

let $L_3 \in \mathbb{N}$, $l_{3,0}, l_{3,1}, \ldots, l_{3,L_3} \in \mathbb{N}$, $\Phi_3 = \big((W_{3,1}, B_{3,1}), \ldots, (W_{3,L_3}, B_{3,L_3})\big) \in (\times_{j=1}^{L_3}(\mathbb{R}^{l_{3,j}\times l_{3,j-1}} \times \mathbb{R}^{l_{3,j}}))$ satisfy that $\Phi_3 = \Phi_1 \bullet \Phi_2$, let $x_0 \in \mathbb{R}^{l_{2,0}}, x_1 \in \mathbb{R}^{l_{2,1}}, \ldots, x_{L_2-1} \in \mathbb{R}^{l_{2,L_2-1}}$ satisfy that

$$\forall\, j \in \mathbb{N} \cap (0, L_2)\colon\ x_j = \mathfrak{M}_{a,l_{2,j}}(W_{2,j}x_{j-1} + B_{2,j}) \tag{3.17}$$

(cf. Definition 3.5), let $y_0 \in \mathbb{R}^{l_{1,0}}, y_1 \in \mathbb{R}^{l_{1,1}}, \ldots, y_{L_1-1} \in \mathbb{R}^{l_{1,L_1-1}}$ satisfy that $y_0 = W_{2,L_2}x_{L_2-1} + B_{2,L_2}$ and

$$\forall\, j \in \mathbb{N} \cap (0, L_1)\colon\ y_j = \mathfrak{M}_{a,l_{1,j}}(W_{1,j}y_{j-1} + B_{1,j}), \tag{3.18}$$

and let $z_0 \in \mathbb{R}^{l_{3,0}}, z_1 \in \mathbb{R}^{l_{3,1}}, \ldots, z_{L_3-1} \in \mathbb{R}^{l_{3,L_3-1}}$ satisfy that $z_0 = x_0$ and

$$\forall\, j \in \mathbb{N} \cap (0, L_3)\colon\ z_j = \mathfrak{M}_{a,l_{3,j}}(W_{3,j}z_{j-1} + B_{3,j}). \tag{3.19}$$

Note that (3.10) ensures that

$$\Phi_3 = \Phi_1 \bullet \Phi_2 = \begin{cases} \begin{array}{r}\big((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \ldots, (W_{2,L_2-1}, B_{2,L_2-1}),\\ (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), (W_{1,2}, B_{1,2}),\\ (W_{1,3}, B_{1,3}), \ldots, (W_{1,L_1}, B_{1,L_1})\big)\end{array} & : L_1 > 1 < L_2 \\ \begin{array}{r}\big((W_{1,1}W_{2,1}, W_{1,1}B_{2,1} + B_{1,1}), (W_{1,2}, B_{1,2}),\\ (W_{1,3}, B_{1,3}), \ldots, (W_{1,L_1}, B_{1,L_1})\big)\end{array} & : L_1 > 1 = L_2 \\ \begin{array}{r}\big((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \ldots, (W_{2,L_2-1}, B_{2,L_2-1}),\\ (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1})\big)\end{array} & : L_1 = 1 < L_2 \\ (W_{1,1}W_{2,1}, W_{1,1}B_{2,1} + B_{1,1}) & : L_1 = 1 = L_2 \end{cases}. \tag{3.20}$$

Hence, we obtain that

$$\begin{aligned}[\mathcal{L}(\Phi_1 \bullet \Phi_2) - 1] &= [(L_2 - 1) + 1 + (L_1 - 1)] - 1\\ &= [L_1 - 1] + [L_2 - 1] = [\mathcal{L}(\Phi_1) - 1] + [\mathcal{L}(\Phi_2) - 1]\end{aligned} \tag{3.21}$$

$$\text{and} \qquad \mathcal{D}(\Phi_1 \bullet \Phi_2) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1}, l_{1,2}, \ldots, l_{1,L_1}). \tag{3.22}$$

This establishes items (i), (ii), and (iii). In addition, observe that (3.22) demonstrates that

$$\begin{aligned}
&\mathcal{P}(\Phi_1 \bullet \Phi_2) = \sum_{j=1}^{L_3} l_{3,j}(l_{3,j-1} + 1)\\
&= \left[\sum_{j=1}^{L_2-1} l_{3,j}(l_{3,j-1} + 1)\right] + l_{3,L_2}(l_{3,L_2-1} + 1) + \left[\sum_{j=L_2+1}^{L_3} l_{3,j}(l_{3,j-1} + 1)\right]\\
&= \left[\sum_{j=1}^{L_2-1} l_{2,j}(l_{2,j-1} + 1)\right] + l_{1,1}(l_{2,L_2-1} + 1) + \left[\sum_{j=L_2+1}^{L_3} l_{1,j-L_2+1}(l_{1,j-L_2} + 1)\right]\\
&= \left[\sum_{j=1}^{L_2-1} l_{2,j}(l_{2,j-1} + 1)\right] + \left[\sum_{j=2}^{L_1} l_{1,j}(l_{1,j-1} + 1)\right] + l_{1,1}\big(l_{2,L_2-1} + 1\big)\\
&= \left[\sum_{j=1}^{L_2} l_{2,j}(l_{2,j-1} + 1)\right] + \left[\sum_{j=1}^{L_1} l_{1,j}(l_{1,j-1} + 1)\right] + l_{1,1}(l_{2,L_2-1} + 1)\\
&\quad - l_{2,L_2}(l_{2,L_2-1} + 1) - l_{1,1}(l_{1,0} + 1)\\
&= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1}(l_{2,L_2-1} + 1) - l_{2,L_2}(l_{2,L_2-1} + 1)\\
&\quad - l_{1,1}(l_{1,0} + 1)\\
&\le \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1}l_{2,L_2-1}.
\end{aligned} \tag{3.23}$$

This establishes item (iv). Moreover, observe that (3.20) and the fact that $a \in C(\mathbb{R},\mathbb{R})$ ensure that

$$\mathcal{R}_a(\Phi_1 \bullet \Phi_2) \in C(\mathbb{R}^{l_{2,0}}, \mathbb{R}^{l_{1,L_1}}) = C(\mathbb{R}^{\mathcal{I}(\Phi_2)}, \mathbb{R}^{\mathcal{O}(\Phi_1)}). \tag{3.24}$$

Next note that (3.21) implies that $L_3 = L_1 + L_2 - 1$. This, (3.20), and (3.22) ensure that

$$(l_{3,0}, l_{3,1}, \ldots, l_{3,L_1+L_2-1}) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1}, l_{1,2}, \ldots, l_{1,L_1}), \tag{3.25}$$

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_2) \colon\ (W_{3,j}, B_{3,j}) = (W_{2,j}, B_{2,j})\big], \tag{3.26}$$

$$(W_{3,L_2}, B_{3,L_2}) = (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), \tag{3.27}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_2, L_1 + L_2) \colon\ (W_{3,j}, B_{3,j}) = (W_{1,j+1-L_2}, B_{1,j+1-L_2})\big]. \tag{3.28}$$

This, (3.17), (3.19), and induction imply that for all $j \in \mathbb{N}_0 \cap [0, L_2)$ it holds that $z_j = x_j$. Combining this with (3.27) and the fact that $y_0 = W_{2,L_2}x_{L_2-1} + B_{2,L_2}$ ensures that

$$\begin{aligned} W_{3,L_2}z_{L_2-1} + B_{3,L_2} &= W_{3,L_2}x_{L_2-1} + B_{3,L_2} \\ &= W_{1,1}W_{2,L_2}x_{L_2-1} + W_{1,1}B_{2,L_2} + B_{1,1} \\ &= W_{1,1}(W_{2,L_2}x_{L_2-1} + B_{2,L_2}) + B_{1,1} = W_{1,1}y_0 + B_{1,1}. \end{aligned} \tag{3.29}$$

Next we claim that for all $j \in \mathbb{N} \cap [L_2, L_1 + L_2)$ it holds that

$$W_{3,j}z_{j-1} + B_{3,j} = W_{1,j+1-L_2}y_{j-L_2} + B_{1,j+1-L_2}. \tag{3.30}$$

We prove (3.30) by induction on $j \in \mathbb{N} \cap [L_2, L_1 + L_2)$. Note that (3.29) establishes (3.30) in the base case $j = L_2$. For the induction step note that the fact that $L_3 = L_1 + L_2 - 1$, (3.18), (3.19), (3.25), and (3.28) imply that for all $j \in \mathbb{N} \cap [L_2, \infty) \cap (0, L_1 + L_2 - 1)$ with

$$W_{3,j}z_{j-1} + B_{3,j} = W_{1,j+1-L_2}y_{j-L_2} + B_{1,j+1-L_2} \tag{3.31}$$

it holds that

$$\begin{aligned} &W_{3,j+1}z_j + B_{3,j+1} = W_{3,j+1}\mathfrak{M}_{a,l_{3,j}}(W_{3,j}z_{j-1} + B_{3,j}) + B_{3,j+1} \\ &= W_{1,j+2-L_2}\mathfrak{M}_{a,l_{1,j+1-L_2}}(W_{1,j+1-L_2}y_{j-L_2} + B_{1,j+1-L_2}) + B_{1,j+2-L_2} \\ &= W_{1,j+2-L_2}y_{j+1-L_2} + B_{1,j+2-L_2}. \end{aligned} \tag{3.32}$$

Induction hence proves (3.30). Next observe that (3.30) and the fact that $L_3 = L_1 + L_2 - 1$ assure that

$$W_{3,L_3}z_{L_3-1} + B_{3,L_3} = W_{3,L_1+L_2-1}z_{L_1+L_2-2} + B_{3,L_1+L_2-1} = W_{1,L_1}y_{L_1-1} + B_{1,L_1}. \tag{3.33}$$

The fact that $\Phi_3 = \Phi_1 \bullet \Phi_2$, (3.17), (3.18), and (3.19) therefore prove that

$$\begin{aligned} [\mathcal{R}_a(\Phi_1 \bullet \Phi_2)](x_0) &= [\mathcal{R}_a(\Phi_3)](x_0) = [\mathcal{R}_a(\Phi_3)](z_0) = W_{3,L_3}z_{L_3-1} + B_{3,L_3} \\ &= W_{1,L_1}y_{L_1-1} + B_{1,L_1} = [\mathcal{R}_a(\Phi_1)](y_0) \\ &= [\mathcal{R}_a(\Phi_1)]\big(W_{2,L_2}x_{L_2-1} + B_{2,L_2}\big) \\ &= [\mathcal{R}_a(\Phi_1)]\big([\mathcal{R}_a(\Phi_2)](x_0)\big) = [(\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))](x_0). \end{aligned} \tag{3.34}$$

Combining this with (3.24) establishes item (v). The proof of Proposition 3.9 is thus completed. □

A comment from Josh: Do I need this result?

Corollary 3.10. *Let $L_1, L_2, L_3 \in \mathbb{N}$, $l_{1,0}, l_{1,1}, \ldots, l_{1,L_1}, l_{2,0}, l_{2,1}, \ldots, l_{2,L_2}, l_{3,0}, l_{3,1}, \ldots, l_{3,L_3} \in \mathbb{N}$ satisfy that $l_{1,0} = l_{2,L_2}$ and let $\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big) \in (\times_{j=1}^{L_k}(\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}}))$, $k \in \{1, 2, 3\}$, satisfy that $\Phi_3 = \Phi_1 \bullet \Phi_2$ (cf. Definitions 3.1 and 3.8). Then* *(i)* *it holds that*

$$L_3 = \mathcal{L}(\Phi_3) = \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2) - 1 = L_1 + L_2 - 1 \geq \max\{L_1, L_2\}, \tag{3.35}$$

(ii) *it holds for all* $j \in \mathbb{N} \cap (0, L_2)$ *that*

$$(W_{3,j}, B_{3,j}) = (W_{2,j}, B_{2,j}), \tag{3.36}$$

(iii) *it holds that*

$$(W_{3,L_2}, B_{3,L_2}) = (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), \tag{3.37}$$

and

(iv) *it holds for all* $j \in \mathbb{N} \cap (L_2, L_1 + L_2) = \mathbb{N} \cap (L_2, \infty) \cap [1, L_3]$ *that*

$$(W_{3,j}, B_{3,j}) = (W_{1,j-L_2+1}, B_{1,j-L_2+1}). \tag{3.38}$$

Proof of Corollary 3.10. Observe that item (ii) in Proposition 3.9 proves item (i). Moreover, note that (3.10) establishes items (ii), (iii), and (iv). The proof of Corollary 3.10 is thus completed. □

3.3.3 Associativity of standard compositions of ANNs

A comment from Josh: I think I need this result...

Lemma 3.11. *Let* $\Phi_1, \Phi_2, \Phi_3 \in \mathbf{N}$ *satisfy that* $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ *and* $\mathcal{I}(\Phi_2) = \mathcal{O}(\Phi_3)$ *(cf. Definition 3.1). Then it holds that*

$$(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3) \tag{3.39}$$

(cf. Definition 3.8).

Proof of Lemma 3.11. Throughout this proof let $\Phi_4, \Phi_5, \Phi_6, \Phi_7 \in \mathbf{N}$ satisfy that $\Phi_4 = \Phi_1 \bullet \Phi_2$, $\Phi_5 = \Phi_2 \bullet \Phi_3$, $\Phi_6 = \Phi_4 \bullet \Phi_3$, and $\Phi_7 = \Phi_1 \bullet \Phi_5$, let $L_k \in \mathbb{N}$, $k \in \{1, 2, \ldots, 7\}$, satisfy for all $k \in \{1, 2, \ldots, 7\}$ that $L_k = \mathcal{L}(\Phi_k)$, let $l_{k,0}, l_{k,1}, \ldots, l_{k,L_k} \in \mathbb{N}$, $k \in \{1, 2, \ldots, 7\}$, and let $\big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big) \in (\times_{j=1}^{L_k} (\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}}))$, $k \in \{1, 2, \ldots, 7\}$, satisfy for all $k \in \{1, 2, \ldots, 7\}$ that

$$\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big). \tag{3.40}$$

Proposition 3.9 and the fact that for all $k \in \{1, 2, 3\}$ it holds that $\mathcal{L}(\Phi_k) = L_k$ proves that

$$\begin{aligned} \mathcal{L}(\Phi_6) &= \mathcal{L}((\Phi_1 \bullet \Phi_2) \bullet \Phi_3) = \mathcal{L}(\Phi_1 \bullet \Phi_2) + \mathcal{L}(\Phi_3) - 1 \\ &= \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2) + \mathcal{L}(\Phi_3) - 2 = L_1 + L_2 + L_3 - 2 \\ &= \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2 \bullet \Phi_3) - 1 = \mathcal{L}(\Phi_1 \bullet (\Phi_2 \bullet \Phi_3)) = \mathcal{L}(\Phi_7). \end{aligned} \tag{3.41}$$

Next note that Corollary 3.10, (3.40), and the fact that $\Phi_4 = \Phi_1 \bullet \Phi_2$ imply that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_2) \colon\ (W_{4,j}, B_{4,j}) = (W_{2,j}, B_{2,j})\big], \tag{3.42}$$

$$(W_{4,L_2}, B_{4,L_2}) = (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), \tag{3.43}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_2, L_1 + L_2) \colon\ (W_{4,j}, B_{4,j}) = (W_{1,j+1-L_2}, B_{1,j+1-L_2})\big]. \tag{3.44}$$

Hence, we obtain that

$$\big[\forall\, j \in \mathbb{N} \cap (L_3 - 1, L_2 + L_3 - 1) \colon (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3})\big], \tag{3.45}$$

$$(W_{4,L_2}, B_{4,L_2}) = (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), \tag{3.46}$$

and

$$\big[\forall\, j \in \mathbb{N} \cap (L_2 + L_3 - 1, L_1 + L_2 + L_3 - 1)\colon \\ (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{1,j+2-L_2-L_3}, B_{1,j+2-L_2-L_3})\big]. \tag{3.47}$$

In addition, observe that Corollary 3.10, (3.40), and the fact that $\Phi_5 = \Phi_2 \bullet \Phi_3$ demonstrate that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_3)\colon\ (W_{5,j}, B_{5,j}) = (W_{3,j}, B_{3,j})\big], \tag{3.48}$$

$$(W_{5,L_3}, B_{5,L_3}) = (W_{2,1} W_{3,L_3}, W_{2,1} B_{3,L_3} + B_{2,1}), \tag{3.49}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_3, L_2 + L_3)\colon\ (W_{5,j}, B_{5,j}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3})\big]. \tag{3.50}$$

Moreover, note that Corollary 3.10, (3.40), and the fact that $\Phi_6 = \Phi_4 \bullet \Phi_3$ ensure that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_3)\colon\ (W_{6,j}, B_{6,j}) = (W_{3,j}, B_{3,j})\big], \tag{3.51}$$

$$(W_{6,L_3}, B_{6,L_3}) = (W_{4,1} W_{3,L_3}, W_{4,1} B_{3,L_3} + B_{4,1}), \tag{3.52}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_3, L_4 + L_3)\colon\ (W_{6,j}, B_{6,j}) = (W_{4,j+1-L_3}, B_{4,j+1-L_3})\big]. \tag{3.53}$$

Furthermore, observe that Corollary 3.10, (3.40), and the fact that $\Phi_7 = \Phi_1 \bullet \Phi_5$ show that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_5)\colon\ (W_{7,j}, B_{7,j}) = (W_{5,j}, B_{5,j})\big], \tag{3.54}$$

$$(W_{7,L_5}, B_{7,L_5}) = (W_{1,1} W_{5,L_5}, W_{1,1} B_{5,L_5} + B_{1,1}), \tag{3.55}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_5, L_1 + L_5)\colon\ (W_{7,j}, B_{7,j}) = (W_{1,j+1-L_5}, B_{1,j+1-L_5})\big]. \tag{3.56}$$

This, the fact that $L_3 \le L_2 + L_3 - 1 = L_5$, (3.48), and (3.51) imply that for all $j \in \mathbb{N} \cap (0, L_3)$ it holds that

$$(W_{6,j}, B_{6,j}) = (W_{3,j}, B_{3,j}) = (W_{5,j}, B_{5,j}) = (W_{7,j}, B_{7,j}). \tag{3.57}$$

In addition, observe that (3.42), (3.43), (3.48), (3.49), (3.52), (3.54), (3.55), and the fact that $L_5 = L_2 + L_3 - 1$ demonstrate that

$$\begin{aligned}
(W_{6,L_3}, B_{6,L_3}) &= (W_{4,1} W_{3,L_3}, W_{4,1} B_{3,L_3} + B_{4,1}) \\
&= \begin{cases} (W_{2,1} W_{3,L_3}, W_{2,1} B_{3,L_3} + B_{2,1}) & \colon L_2 > 1 \\ (W_{1,1} W_{2,1} W_{3,L_3}, W_{1,1} W_{2,1} B_{3,L_3} + W_{1,1} B_{2,1} + B_{1,1}) & \colon L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{2,1} W_{3,L_3}, W_{2,1} B_{3,L_3} + B_{2,1}) & \colon L_2 > 1 \\ (W_{1,1}(W_{2,1} W_{3,L_3}), W_{1,1}(W_{2,1} B_{3,L_3} + B_{2,1}) + B_{1,1}) & \colon L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{5,L_3}, B_{5,L_3}) & \colon L_2 > 1 \\ (W_{1,1} W_{5,L_3}, W_{1,1} B_{5,L_3} + B_{1,1}) & \colon L_2 = 1 \end{cases} \\
&= (W_{7,L_3}, B_{7,L_3}).
\end{aligned} \tag{3.58}$$

Next note that the fact that $L_5 = L_2 + L_3 - 1 < L_1 + L_2 + L_3 - 1 = L_3 + L_4$, (3.53), (3.45), (3.50), and (3.54) ensure that for all $j \in \mathbb{N}$ with $L_3 < j < L_5$ it holds that

$$\begin{aligned}
(W_{6,j}, B_{6,j}) &= (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3}) \\
&= (W_{5,j}, B_{5,j}) = (W_{7,j}, B_{7,j}).
\end{aligned} \tag{3.59}$$

Moreover, observe that the fact that $L_5 = L_2 + L_3 - 1 < L_1 + L_2 + L_3 - 1 = L_3 + L_4$, (3.53), (3.58), (3.43), (3.50), and (3.55) prove that

$$\begin{aligned}
(W_{6,L_5}, B_{6,L_5}) &= \begin{cases} (W_{4,L_5+1-L_3}, B_{4,L_5+1-L_3}) & : L_2 > 1 \\ (W_{6,L_3}, B_{6,L_3}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{4,L_2}, B_{4,L_2}) & : L_2 > 1 \\ (W_{7,L_3}, B_{7,L_3}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}) & : L_2 > 1 \\ (W_{7,L_5}, B_{7,L_5}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{1,1}W_{5,L_5}, W_{1,1}B_{5,L_5} + B_{1,1}) & : L_2 > 1 \\ (W_{7,L_5}, B_{7,L_5}) & : L_2 = 1 \end{cases} \\
&= (W_{7,L_5}, B_{7,L_5}).
\end{aligned} \tag{3.60}$$

Furthermore, note that (3.53), (3.47), (3.56), and the fact that $L_5 = L_2 + L_3 - 1 \geq L_3$ assure that for all $j \in \mathbb{N}$ with $L_5 < j \leq L_6$ it holds that

$$\begin{aligned}
(W_{6,j}, B_{6,j}) &= (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{1,j+2-L_2-L_3}, B_{1,j+2-L_2-L_3}) \\
&= (W_{1,j+1-L_5}, B_{1,j+1-L_5}) = (W_{7,j}, B_{7,j}).
\end{aligned} \tag{3.61}$$

Combining this with (3.41), (3.57), (3.58), (3.59), and (3.60) establishes that

$$(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_4 \bullet \Phi_3 = \Phi_6 = \Phi_7 = \Phi_1 \bullet \Phi_5 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3). \tag{3.62}$$

The proof of Lemma 3.11 is thus completed. □

3.3.4 Compositions of ANNs and affine linear transformations

A comment from Josh: I think I need this result...

Corollary 3.12. *Let $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

(i) it holds for all $\mathbb{A} \in \mathbf{N}$ with $\mathcal{L}(\mathbb{A}) = 1$ and $\mathcal{I}(\mathbb{A}) = \mathcal{O}(\Phi)$ that

$$\mathcal{P}(\mathbb{A} \bullet \Phi) \leq \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A})}{\mathcal{O}(\Phi)}\right\}\right] \mathcal{P}(\Phi) \tag{3.63}$$

and

(ii) it holds for all $\mathbb{A} \in \mathbf{N}$ with $\mathcal{L}(\mathbb{A}) = 1$ and $\mathcal{I}(\Phi) = \mathcal{O}(\mathbb{A})$ that

$$\mathcal{P}(\Phi \bullet \mathbb{A}) \leq \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A})+1}{\mathcal{I}(\Phi)+1}\right\}\right] \mathcal{P}(\Phi) \tag{3.64}$$

(cf. Definition 3.8).

Proof of Corollary 3.12. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\mathbb{A}_1, \mathbb{A}_2 \in \mathbf{N}$ satisfy that $\mathcal{L}(\mathbb{A}_1) = \mathcal{L}(\mathbb{A}_2) = 1$, $\mathcal{I}(\mathbb{A}_1) = \mathcal{O}(\Phi)$, $\mathcal{I}(\Phi) = \mathcal{O}(\mathbb{A}_2)$, and $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$. Observe that item (iv) in Proposition 3.9, the fact that $\mathcal{O}(\Phi) = l_L$, the fact that $\mathcal{I}(\Phi) = l_0$, and the fact that for all $k \in \{1, 2\}$ it holds that $\mathcal{D}(\mathbb{A}_k) = (\mathcal{I}(\mathbb{A}_k), \mathcal{O}(\mathbb{A}_k))$ ensure that

$$\begin{aligned}
\mathcal{P}(\mathbb{A}_1 \bullet \Phi) &= \left[\sum_{m=1}^{L-1} l_m(l_{m-1} + 1)\right] + \left[\mathcal{O}(\mathbb{A}_1)\right](l_{L-1} + 1) \\
&= \left[\sum_{m=1}^{L-1} l_m(l_{m-1} + 1)\right] + \left[\tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right] l_L(l_{L-1} + 1) \\
&\leq \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right\}\right] \left[\sum_{m=1}^{L-1} l_m(l_{m-1} + 1)\right] + \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right\}\right] l_L(l_{L-1} + 1) \\
&= \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right\}\right] \left[\sum_{m=1}^{L} l_m(l_{m-1} + 1)\right] = \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{\mathcal{O}(\Phi)}\right\}\right] \mathcal{P}(\Phi)
\end{aligned} \tag{3.65}$$

and

$$\begin{aligned}\mathcal{P}(\Phi \bullet \mathbb{A}_2) &= \left[\sum_{m=2}^{L} l_m(l_{m-1}+1)\right] + l_1\big[\mathcal{I}(\mathbb{A}_2)+1\big] \\ &= \left[\sum_{m=2}^{L} l_m(l_{m-1}+1)\right] + \left[\tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right] l_1(l_0+1) \\ &\leq \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right\}\right]\left[\sum_{m=2}^{L} l_m(l_{m-1}+1)\right] + \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right\}\right] l_1(l_0+1) \\ &= \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right\}\right]\left[\sum_{m=1}^{L} l_m(l_{m-1}+1)\right] = \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{\mathcal{I}(\Phi)+1}\right\}\right] \mathcal{P}(\Phi).\end{aligned} \tag{3.66}$$

This establishes items (i) and (ii). The proof of Corollary 3.12 is thus completed. □

3.3.5 Powers and extensions of ANNs

Definition 3.13 (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 3.14 (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*

$$\Phi^{\bullet n} = \begin{cases} \big(\mathrm{I}_{\mathcal{O}(\Phi)}, (0,0,\ldots,0)\big) \in \mathbb{R}^{\mathcal{O}(\Phi)\times\mathcal{O}(\Phi)} \times \mathbb{R}^{\mathcal{O}(\Phi)} & : n = 0 \\ \Phi \bullet (\Phi^{\bullet(n-1)}) & : n \in \mathbb{N} \end{cases} \tag{3.67}$$

(cf. Definitions 3.1, 3.8, and 3.13).

Definition 3.15 (Extension of ANNs). *Let $L \in \mathbb{N}$, $\Psi \in \mathbf{N}$ satisfy that $\mathcal{I}(\Psi) = \mathcal{O}(\Psi)$. Then we denote by $\mathcal{E}_{L,\Psi}\colon \{\Phi \in \mathbf{N}\colon (\mathcal{L}(\Phi) \leq L \text{ and } \mathcal{O}(\Phi) = \mathcal{I}(\Psi))\} \to \mathbf{N}$ the function which satisfies for all $\Phi \in \mathbf{N}$ with $\mathcal{L}(\Phi) \leq L$ and $\mathcal{O}(\Phi) = \mathcal{I}(\Psi)$ that*

$$\mathcal{E}_{L,\Psi}(\Phi) = (\Psi^{\bullet(L-\mathcal{L}(\Phi))}) \bullet \Phi \tag{3.68}$$

(cf. Definitions 3.1, 3.8, and 3.14).

A comment from Josh: Do I need this result?

Lemma 3.16. *Let $d, \mathfrak{i} \in \mathbb{N}$, $\Psi \in \mathbf{N}$ satisfy that $\mathcal{D}(\Psi) = (d, \mathfrak{i}, d)$ (cf. Definition 3.1). Then*

(i) it holds for all $n \in \mathbb{N}_0$ that $\mathcal{L}(\Psi^{\bullet n}) = n+1$, $\mathcal{D}(\Psi^{\bullet n}) \in \mathbb{N}^{n+2}$, and

$$\mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d,d) & : n = 0 \\ (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & : n \in \mathbb{N} \end{cases} \tag{3.69}$$

and

(ii) it holds for all $\Phi \in \mathbf{N}$, $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$ with $\mathcal{O}(\Phi) = d$ that $\mathcal{L}\big(\mathcal{E}_{L,\Psi}(\Phi)\big) = L$ and

$$\begin{aligned}&\mathcal{P}(\mathcal{E}_{L,\Psi}(\Phi)) \\ &\leq \begin{cases} \mathcal{P}(\Phi) & : \mathcal{L}(\Phi) = L \\ \big[\big(\max\{1, \tfrac{\mathfrak{i}}{d}\}\big)\mathcal{P}(\Phi) + \big((L - \mathcal{L}(\Phi) - 1)\,\mathfrak{i} + d\big)(\mathfrak{i}+1)\big] & : \mathcal{L}(\Phi) < L \end{cases}\end{aligned} \tag{3.70}$$

(cf. Definitions 3.14, 3.15, and 3.15).

Proof of Lemma 3.16. Throughout this proof let $\Phi \in \mathbf{N}$, $l_0, l_1, \ldots, l_{\mathcal{L}(\Phi)} \in \mathbb{N}$ satisfy that $\mathcal{O}(\Phi) = d$ and $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_{\mathcal{L}(\Phi)}) \in \mathbb{N}^{\mathcal{L}(\Phi)+1}$ and let $a_{L,k} \in \mathbb{N}$, $k \in \mathbb{N}_0 \cap [0, L]$, $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$, satisfy for all $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$, $k \in \mathbb{N}_0 \cap [0, L]$ that

$$a_{L,k} = \begin{cases} l_k & : k < \mathcal{L}(\Phi) \\ \mathfrak{i} & : \mathcal{L}(\Phi) \leq k < L \\ d & : k = L \end{cases}. \tag{3.71}$$

We claim that for all $n \in \mathbb{N}_0$ it holds that

$$\mathcal{L}(\Psi^{\bullet n}) = n + 1 \qquad \text{and} \qquad \mathbb{N}^{n+2} \ni \mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d, d) & : n = 0 \\ (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & : n \in \mathbb{N} \end{cases}. \tag{3.72}$$

We now prove (3.72) by induction on $n \in \mathbb{N}_0$. Note that the fact that $\Psi^{\bullet 0} = (\mathrm{I}_d, 0) \in \mathbb{R}^{d \times d} \times \mathbb{R}^d$ (cf. Definition 3.13) establishes (3.69) in the base case $n = 0$. For the induction step $\mathbb{N}_0 \ni n \to n + 1 \in \mathbb{N}$ assume that there exists $n \in \mathbb{N}_0$ such that

$$\mathcal{L}(\Psi^{\bullet n}) = n + 1 \qquad \text{and} \qquad \mathbb{N}^{n+2} \ni \mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d, d) & : n = 0 \\ (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & : n \in \mathbb{N} \end{cases}. \tag{3.73}$$

Observe that Lemma 3.7, (3.67), items (i) and (ii) in Proposition 3.9, (3.73), and the hypothesis that $\mathcal{D}(\Psi) = (d, \mathfrak{i}, d)$ imply that

$$\begin{aligned} \mathcal{L}(\Psi^{\bullet(n+1)}) = \mathcal{L}(\Psi \bullet (\Psi^{\bullet n})) = \mathcal{L}(\Psi) + \mathcal{L}(\Psi^{\bullet n}) - 1 = 2 + (n+1) - 1 = (n+1) + 1 \\ \text{and} \qquad \mathcal{D}(\Psi^{\bullet(n+1)}) = \mathcal{D}(\Psi \bullet (\Psi^{\bullet n})) = (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) \in \mathbb{N}^{n+3}. \end{aligned} \tag{3.74}$$

Induction thus proves (3.72). Next note that (3.72) establishes item (i). In addition, observe that items (i) and (ii) in Proposition 3.9, item (i), (3.68), and (3.71) ensure that for all $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$ it holds that

$$\begin{aligned} \mathcal{L}\big(\mathcal{E}_{L,\Psi}(\Phi)\big) &= \mathcal{L}\big((\Psi^{\bullet(L - \mathcal{L}(\Phi))}) \bullet \Phi\big) = \mathcal{L}\big(\Psi^{\bullet(L - \mathcal{L}(\Phi))}\big) + \mathcal{L}(\Phi) - 1 \\ &= (L - \mathcal{L}(\Phi) + 1) + \mathcal{L}(\Phi) - 1 = L \end{aligned} \tag{3.75}$$

and

$$\mathcal{D}\big(\mathcal{E}_{L,\Psi}(\Phi)\big) = \mathcal{D}\big((\Psi^{\bullet(L - \mathcal{L}(\Phi))}) \bullet \Phi\big) = (a_{L,0}, a_{L,1}, \ldots, a_{L,L}). \tag{3.76}$$

Combining this with (3.71) demonstrates that

$$\mathcal{L}\big(\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi)\big) = \mathcal{L}(\Phi) \tag{3.77}$$

and

$$\begin{aligned} \mathcal{D}\big(\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi)\big) &= (a_{\mathcal{L}(\Phi),0}, a_{\mathcal{L}(\Phi),1}, \ldots, a_{\mathcal{L}(\Phi),\mathcal{L}(\Phi)}) \\ &= (l_0, l_1, \ldots, l_{\mathcal{L}(\Phi)}) = \mathcal{D}(\Phi). \end{aligned} \tag{3.78}$$

Hence, we obtain that

$$\mathcal{P}\big(\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi)\big) = \mathcal{P}(\Phi). \tag{3.79}$$

Next note that (3.71), (3.76), and the fact that $l_{\mathcal{L}(\Phi)} = \mathcal{O}(\Phi) = d$ imply that for all $L \in \mathbb{N} \cap (\mathcal{L}(\Phi), \infty)$ it holds that

$$\begin{aligned}
\mathcal{P}\big(\mathcal{E}_{L,\Psi}(\Phi)\big) &= \sum_{k=1}^{L} a_{L,k}(a_{L,k-1}+1) \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} a_{L,k}(a_{L,k-1}+1)\right] + \left[\sum_{k=\mathcal{L}(\Phi)}^{L} a_{L,k}(a_{L,k-1}+1)\right] \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + \left[\sum_{k=\mathcal{L}(\Phi)}^{\mathcal{L}(\Phi)} a_{L,k}(a_{L,k-1}+1)\right] \\
&\quad + \left[\sum_{k=\mathcal{L}(\Phi)+1}^{L} a_{L,k}(a_{L,k-1}+1)\right] \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + a_{L,\mathcal{L}(\Phi)}(a_{L,\mathcal{L}(\Phi)-1}+1) \\
&\quad + \left[\sum_{k=\mathcal{L}(\Phi)+1}^{L-1} a_{L,k}(a_{L,k-1}+1)\right] + \left[\sum_{k=L}^{L} a_{L,k}(a_{L,k-1}+1)\right] \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + \mathfrak{i}(l_{\mathcal{L}(\Phi)-1}+1) \\
&\quad + \big(L-1-(\mathcal{L}(\Phi)+1)+1\big)\mathfrak{i}(\mathfrak{i}+1) + a_{L,L}(a_{L,L-1}+1) \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + \tfrac{\mathfrak{i}}{d}\big[l_{\mathcal{L}(\Phi)}(l_{\mathcal{L}(\Phi)-1}+1)\big] \\
&\quad + \big(L-\mathcal{L}(\Phi)-1\big)\mathfrak{i}(\mathfrak{i}+1) + d(\mathfrak{i}+1) \\
&\le \left[\max\{1,\tfrac{\mathfrak{i}}{d}\}\right]\left[\sum_{k=1}^{\mathcal{L}(\Phi)} l_k(l_{k-1}+1)\right] + \big(L-\mathcal{L}(\Phi)-1\big)\mathfrak{i}(\mathfrak{i}+1) + d(\mathfrak{i}+1) \\
&= \left[\max\{1,\tfrac{\mathfrak{i}}{d}\}\right]\mathcal{P}(\Phi) + \big(L-\mathcal{L}(\Phi)-1\big)\mathfrak{i}(\mathfrak{i}+1) + d(\mathfrak{i}+1).
\end{aligned} \tag{3.80}$$

Combining this with (3.79) establishes (3.70). The proof of Lemma 3.16 is thus completed. □

A comment from Josh: This result is needed for properties of generalized parallelizations...

Lemma 3.17. *Let $a \in C(\mathbb{R},\mathbb{R})$, $\mathbb{I} \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}^{\mathcal{I}(\mathbb{I})}$ that $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\mathbb{I})$ and $(\mathcal{R}_a(\mathbb{I}))(x) = x$ (cf. Definitions 3.1 and 3.6). Then*

(i) it holds for all $n \in \mathbb{N}_0$, $x \in \mathbb{R}^{\mathcal{I}(\mathbb{I})}$ that

$$\mathcal{R}_a(\mathbb{I}^{\bullet n}) \in C(\mathbb{R}^{\mathcal{I}(\mathbb{I})},\mathbb{R}^{\mathcal{I}(\mathbb{I})}) \qquad \textit{and} \qquad (\mathcal{R}_a(\mathbb{I}^{\bullet n}))(x) = x \tag{3.81}$$

and

(ii) it holds for all $\Phi \in \mathbf{N}$, $L \in \mathbb{N} \cap [\mathcal{L}(\Phi),\infty)$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ with $\mathcal{O}(\Phi) = \mathcal{I}(\mathbb{I})$ that

$$\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi)) \in C(\mathbb{R}^{\mathcal{I}(\Phi)},\mathbb{R}^{\mathcal{O}(\Phi)}) \quad \textit{and} \quad \big(\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi))\big)(x) = \big(\mathcal{R}_a(\Phi)\big)(x) \tag{3.82}$$

(cf. Definitions 3.14 and 3.15).

Proof of Lemma 3.17. Throughout this proof let $\Phi \in \mathbf{N}$, $L, d \in \mathbb{N}$ satisfy that $\mathcal{L}(\Phi) \le L$ and $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\Phi) = d$. We claim that for all $n \in \mathbb{N}_0$ it holds that

$$\mathcal{R}_a(\mathbb{I}^{\bullet n}) \in C(\mathbb{R}^d,\mathbb{R}^d) \qquad \text{and} \qquad \forall\, x \in \mathbb{R}^d\colon\ (\mathcal{R}_a(\mathbb{I}^{\bullet n}))(x) = x. \tag{3.83}$$

We now prove (3.83) by induction on $n \in \mathbb{N}_0$. Note that (3.67) and the fact that $\mathcal{O}(\mathbb{I}) = d$ demonstrate that $\mathcal{R}_a(\mathbb{I}^{\bullet 0}) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and $\forall\, x \in \mathbb{R}^d\colon (\mathcal{R}_a(\mathbb{I}^{\bullet 0}))(x) = x$. This establishes (3.83) in the base case $n = 0$. For the induction step observe that for all $n \in \mathbb{N}_0$ with $\mathcal{R}_a(\mathbb{I}^{\bullet n}) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and $\forall\, x \in \mathbb{R}^d\colon (\mathcal{R}_a(\mathbb{I}^{\bullet n}))(x) = x$ it holds that

$$\mathcal{R}_a(\mathbb{I}^{\bullet (n+1)}) = \mathcal{R}_a(\mathbb{I} \bullet (\mathbb{I}^{\bullet n})) = (\mathcal{R}_a(\mathbb{I})) \circ (\mathcal{R}_a(\mathbb{I}^{\bullet n})) \in C(\mathbb{R}^d, \mathbb{R}^d) \tag{3.84}$$

and

$$\begin{split}\forall\, x \in \mathbb{R}^d\colon \big(\mathcal{R}_a(\mathbb{I}^{\bullet (n+1)})\big)(x) &= \big([\mathcal{R}_a(\mathbb{I})] \circ [\mathcal{R}_a(\mathbb{I}^{\bullet n})]\big)(x) \\ &= (\mathcal{R}_a(\mathbb{I}))\big(\big(\mathcal{R}_a(\mathbb{I}^{\bullet n})\big)(x)\big) = (\mathcal{R}_a(\mathbb{I}))(x) = x.\end{split} \tag{3.85}$$

Induction thus proves (3.83). Next observe that (3.83) establishes item (i). Moreover, note that (3.68), item (v) in Proposition 3.9, item (i), and the fact that $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\Phi)$ ensure that

$$\begin{split}&\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi)) = \mathcal{R}_a((\mathbb{I}^{\bullet (L - \mathcal{L}(\Phi))}) \bullet \Phi) \\ &\in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\mathbb{I})}) = C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{I}(\mathbb{I})}) = C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})\end{split} \tag{3.86}$$

and

$$\begin{split}\forall\, x \in \mathbb{R}^{\mathcal{I}(\Phi)}\colon \big(\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi))\big)(x) &= \big(\mathcal{R}_a(\mathbb{I}^{\bullet (L - \mathcal{L}(\Phi))})\big)\big((\mathcal{R}_a(\Phi))(x)\big) \\ &= (\mathcal{R}_a(\Phi))(x).\end{split} \tag{3.87}$$

This establishes item (ii). The proof of Lemma 3.17 is thus completed. □

Lemma 3.18. *Let $d, \mathfrak{i}, L, \mathfrak{L} \in \mathbb{N}$, $l_0, l_1, \ldots, l_{L-1} \in \mathbb{N}$, $\Phi, \Psi \in \mathbf{N}$ satisfy $\mathfrak{L} \geq L$, $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_{L-1}, d)$ and $\mathcal{D}(\Psi) = (d, \mathfrak{i}, d)$ (cf. Definition 3.1). Then it holds that $\mathcal{D}(\mathcal{E}_{\mathfrak{L},\Psi}(\Phi)) \in \mathbb{N}^{\mathfrak{L}+1}$ and*

$$\mathcal{D}(\mathcal{E}_{\mathfrak{L},\Psi}(\Phi)) = \begin{cases} (l_0, l_1, \ldots, l_{L-1}, d) & \colon \mathfrak{L} = L \\ (l_0, l_1, \ldots, l_{L-1}, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & \colon \mathfrak{L} > L \end{cases} \tag{3.88}$$

(cf. Definition 3.15).

Proof of Lemma 3.18. Observe that item (i) in Lemma 3.16 ensures that $\mathcal{L}(\Psi^{\bullet (\mathfrak{L} - L)}) = \mathfrak{L} - L + 1$, $\mathcal{D}(\Psi^{\bullet (\mathfrak{L} - L)}) \in \mathbb{N}^{\mathfrak{L} - L + 2}$, and

$$\mathcal{D}(\Psi^{\bullet (\mathfrak{L} - L)}) = \begin{cases} (d, d) & \colon \mathfrak{L} = L \\ (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & \colon \mathfrak{L} > L \end{cases} \tag{3.89}$$

(cf. Definition 3.14). Combining this with Proposition 3.9 shows that $\mathcal{L}((\Psi^{\bullet (\mathfrak{L} - L)}) \bullet \Phi) = \mathcal{L}(\Psi^{\bullet (\mathfrak{L} - L)}) + \mathcal{L}(\Phi) - 1 = \mathfrak{L}$, $\mathcal{D}((\Psi^{\bullet (\mathfrak{L} - L)}) \bullet \Phi) \in \mathbb{N}^{\mathfrak{L}+1}$, and

$$\mathcal{D}((\Psi^{\bullet (\mathfrak{L} - L)}) \bullet \Phi) = \begin{cases} (l_0, l_1, \ldots, l_{L-1}, d) & \colon \mathfrak{L} = L \\ (l_0, l_1, \ldots, l_{L-1}, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & \colon \mathfrak{L} > L. \end{cases} \tag{3.90}$$

This and (3.68) establish (3.88). The proof of Lemma 3.18 is thus completed. □

3.4 Parallelizations of ANNs

3.4.1 Parallelizations of ANNs with the same length

Definition 3.19 (Parallelization of ANNs)**.** *Let $n \in \mathbb{N}$. Then we denote by*

$$\mathbf{P}_n \colon \big\{(\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n \colon \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)\big\} \to \mathbf{N} \tag{3.91}$$

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

$(W_{2,2}, B_{2,2}), \ldots, (W_{2,L}, B_{2,L})) \in \left(\bigtimes_{k=1}^{L}(\mathbb{R}^{l_{2,k}\times l_{2,k-1}} \times \mathbb{R}^{l_{2,k}})\right)$, $\ldots$, $\Phi_n = ((W_{n,1}, B_{n,1}), (W_{n,2}, B_{n,2}), \ldots, (W_{n,L}, B_{n,L})) \in \left(\bigtimes_{k=1}^{L}(\mathbb{R}^{l_{n,k}\times l_{n,k-1}} \times \mathbb{R}^{l_{n,k}})\right)$ *that*

$$\begin{split}\mathbf{P}_n(\Phi_1, \Phi_2, \ldots, \Phi_n) = \Bigg(\left(\begin{pmatrix} W_{1,1} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,1} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,1} \end{pmatrix}, \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ B_{3,1} \\ \vdots \\ B_{n,1} \end{pmatrix}\right), \\ \left(\begin{pmatrix} W_{1,2} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,2} \end{pmatrix}, \begin{pmatrix} B_{1,2} \\ B_{2,2} \\ B_{3,2} \\ \vdots \\ B_{n,2} \end{pmatrix}\right), \ldots, \\ \left(\begin{pmatrix} W_{1,L} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,L} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,L} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,L} \end{pmatrix}, \begin{pmatrix} B_{1,L} \\ B_{2,L} \\ B_{3,L} \\ \vdots \\ B_{n,L} \end{pmatrix}\right)\Bigg)\end{split} \tag{3.92}$$

(cf. Definition 3.1).

A comment from Josh: I think I need this result...

Lemma 3.20. *Let $n, L \in \mathbb{N}$, $\Phi_1, \Phi_2, \ldots, \Phi_n \in \mathbf{N}$ satisfy that $L = \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)$ (cf. Definition 3.1). Then it holds that*

$$\mathbf{P}_n(\Phi_1, \Phi_2, \ldots, \Phi_n) \in \left(\times_{k=1}^{L}\left(\mathbb{R}^{(\sum_{j=1}^{n}\mathbb{D}_j(\Phi_k))\times(\sum_{j=1}^{n}\mathbb{D}_j(\Phi_{k-1}))} \times \mathbb{R}^{(\sum_{j=1}^{n}\mathbb{D}_j(\Phi_k))}\right)\right) \tag{3.93}$$

(cf. Definition 3.19).

Proof of Lemma 3.20. Note that (3.92) establishes (3.93). The proof of Lemma 3.20 is thus completed. ☐

A comment from Josh: I think I need this result...

Proposition 3.21. *Let $a \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n$ satisfy that $\mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)$ (cf. Definition 3.1). Then*

(i) it holds that

$$\mathcal{R}_a(\mathbf{P}_n(\Phi)) \in C\left(\mathbb{R}^{[\sum_{j=1}^{n}\mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^{n}\mathcal{O}(\Phi_j)]}\right) \tag{3.94}$$

and

(ii) it holds for all $x_1 \in \mathbb{R}^{\mathcal{I}(\Phi_1)}, x_2 \in \mathbb{R}^{\mathcal{I}(\Phi_2)}, \ldots, x_n \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ that

$$\begin{split}&\big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(x_1, x_2, \ldots, x_n) \\ &= \big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \ldots, (\mathcal{R}_a(\Phi_n))(x_n)\big) \in \mathbb{R}^{[\sum_{j=1}^{n}\mathcal{O}(\Phi_j)]}\end{split} \tag{3.95}$$

(cf. Definitions 3.6 and 3.19).

Proof of Proposition 3.21. Throughout this proof let $L \in \mathbb{N}$ satisfy that $L = \mathcal{L}(\Phi_1)$, let $l_{j,0}, l_{j,1}, \ldots, l_{j,L} \in \mathbb{N}$, $j \in \{1, 2, \ldots, n\}$, satisfy for all $j \in \{1, 2, \ldots, n\}$ that $\mathcal{D}(\Phi_j) = (l_{j,0}, l_{j,1}, \ldots, l_{j,L})$, let $\big((W_{j,1}, B_{j,1}),$

$(W_{j,2}, B_{j,2}), \ldots, (W_{j,L}, B_{j,L})) \in (\times_{k=1}^{L}(\mathbb{R}^{l_{j,k}\times l_{j,k-1}} \times \mathbb{R}^{l_{j,k}}))$, $j \in \{1, 2, \ldots, n\}$, satisfy for all $j \in \{1, 2, \ldots, n\}$ that

$$\Phi_j = \big((W_{j,1}, B_{j,1}), (W_{j,2}, B_{j,2}), \ldots, (W_{j,L}, B_{j,L})\big), \tag{3.96}$$

let $\alpha_k \in \mathbb{N}$, $k \in \{0, 1, \ldots, L\}$, satisfy for all $k \in \{0, 1, \ldots, L\}$ that $\alpha_k = \sum_{j=1}^{n} l_{j,k}$, let $\big((A_1, b_1), (A_2, b_2), \ldots, (A_L, b_L)\big) \in (\times_{k=1}^{L}(\mathbb{R}^{\alpha_k\times\alpha_{k-1}} \times \mathbb{R}^{\alpha_k}))$ satisfy that

$$\mathbf{P}_n(\Phi) = \big((A_1, b_1), (A_2, b_2), \ldots, (A_L, b_L)\big) \tag{3.97}$$

(cf. Lemma 3.20), let $(x_{j,0}, x_{j,1}, \ldots, x_{j,L-1}) \in (\mathbb{R}^{l_{j,0}} \times \mathbb{R}^{l_{j,1}} \times \ldots \times \mathbb{R}^{l_{j,L-1}})$, $j \in \{1, 2, \ldots, n\}$, satisfy for all $j \in \{1, 2, \ldots, n\}$, $k \in \mathbb{N} \cap (0, L)$ that

$$x_{j,k} = \mathfrak{M}_{a,l_{j,k}}(W_{j,k}x_{j,k-1} + B_{j,k}) \tag{3.98}$$

(cf. Definition 3.5), and let $\mathfrak{x}_0 \in \mathbb{R}^{\alpha_0}, \mathfrak{x}_1 \in \mathbb{R}^{\alpha_1}, \ldots, \mathfrak{x}_{L-1} \in \mathbb{R}^{\alpha_{L-1}}$ satisfy for all $k \in \{0, 1, \ldots, L-1\}$ that $\mathfrak{x}_k = (x_{1,k}, x_{2,k}, \ldots, x_{n,k})$. Observe that (3.97) demonstrates that $\mathcal{I}(\mathbf{P}_n(\Phi)) = \alpha_0$ and $\mathcal{O}(\mathbf{P}_n(\Phi)) = \alpha_L$. Combining this with item (ii) in Lemma 3.7, the fact that for all $k \in \{0, 1, \ldots, L\}$ it holds that $\alpha_k = \sum_{j=1}^{n} l_{j,k}$, the fact that for all $j \in \{1, 2, \ldots, n\}$ it holds that $\mathcal{I}(\Phi_j) = l_{j,0}$, and the fact that for all $j \in \{1, 2, \ldots, n\}$ it holds that $\mathcal{O}(\Phi_j) = l_{j,L}$ ensures that

$$\begin{aligned}\mathcal{R}_a(\mathbf{P}_n(\Phi)) \in C(\mathbb{R}^{\alpha_0}, \mathbb{R}^{\alpha_L}) &= C\big(\mathbb{R}^{[\sum_{j=1}^{n} l_{j,0}]}, \mathbb{R}^{[\sum_{j=1}^{n} l_{j,L}]}\big)\\ &= C\big(\mathbb{R}^{[\sum_{j=1}^{n} \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^{n} \mathcal{O}(\Phi_j)]}\big).\end{aligned} \tag{3.99}$$

This proves item (i). Moreover, observe that (3.92) and (3.97) demonstrate that for all $k \in \{1, 2, \ldots, L\}$ it holds that

$$A_k = \begin{pmatrix} W_{1,k} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,k} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,k} \end{pmatrix} \quad \text{and} \quad b_k = \begin{pmatrix} B_{1,k} \\ B_{2,k} \\ B_{3,k} \\ \vdots \\ B_{n,k} \end{pmatrix}. \tag{3.100}$$

Combining this with (3.5), (3.98), and the fact that for all $k \in \mathbb{N} \cap [0, L)$ it holds that $\mathfrak{x}_k = (x_{1,k}, x_{2,k}, \ldots, x_{n,k})$ implies that for all $k \in \mathbb{N} \cap (0, L)$ it holds that

$$\mathfrak{M}_{a,\alpha_k}(A_k\mathfrak{x}_{k-1} + b_k) = \begin{pmatrix} \mathfrak{M}_{a,l_{1,k}}(W_{1,k}x_{1,k-1} + B_{1,k}) \\ \mathfrak{M}_{a,l_{2,k}}(W_{2,k}x_{2,k-1} + B_{2,k}) \\ \vdots \\ \mathfrak{M}_{a,l_{n,k}}(W_{n,k}x_{n,k-1} + B_{n,k}) \end{pmatrix} = \begin{pmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{pmatrix} = \mathfrak{x}_k. \tag{3.101}$$

This, (3.6), (3.96), (3.97), (3.98), (3.100), the fact that $\mathfrak{x}_0 = (x_{1,0}, x_{2,0}, \ldots, x_{n,0})$, and the fact that $\mathfrak{x}_{L-1} = (x_{1,L-1}, x_{2,L-1}, \ldots, x_{n,L-1})$ ensure that

$$\begin{aligned}&\big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(x_{1,0}, x_{2,0}, \ldots, x_{n,0}) = \big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(\mathfrak{x}_0)\\ &= A_L\mathfrak{x}_{L-1} + b_L = \begin{pmatrix} W_{1,L}x_{1,L-1} + B_{1,L} \\ W_{2,L}x_{2,L-1} + B_{2,L} \\ \vdots \\ W_{n,L}x_{n,L-1} + B_{n,L} \end{pmatrix} = \begin{pmatrix} (\mathcal{R}_a(\Phi_1))(x_{1,0}) \\ (\mathcal{R}_a(\Phi_2))(x_{2,0}) \\ \vdots \\ (\mathcal{R}_a(\Phi_n))(x_{n,0}) \end{pmatrix}.\end{aligned} \tag{3.102}$$

This establishes item (ii). The proof of Proposition 3.21 is thus completed. ☐

A comment from Josh: I think I need this result...

Proposition 3.22. *Let $n, L \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy that $L = \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)$ (cf. Definition 3.1). Then*

(i) it holds that

$$\mathcal{D}\big(\mathbf{P}_n(\Phi)\big) = \big(\textstyle\sum_{j=1}^n \mathbb{D}_0(\Phi_j), \sum_{j=1}^n \mathbb{D}_1(\Phi_j), \dots, \sum_{j=1}^n \mathbb{D}_L(\Phi_j)\big) \tag{3.103}$$

and

(ii) it holds that

$$\mathcal{P}(\mathbf{P}_n(\Phi)) \leq \tfrac{1}{2}\big[\textstyle\sum_{j=1}^n \mathcal{P}(\Phi_j)\big]^2 \tag{3.104}$$

(cf. Definition 3.19).

Proof of Proposition 3.22. Throughout this proof let $l_{j,0}, l_{j,1}, \dots, l_{j,L} \in \mathbb{N}$, $j \in \{1, 2, \dots, n\}$, satisfy for all $j \in \{1, 2, \dots, n\}$, $k \in \{0, 1, \dots, L\}$ that $l_{j,k} = \mathbb{D}_k(\Phi_j)$. Note that Lemma 3.20 assures that

$$\mathcal{D}\big(\mathbf{P}_n(\Phi)\big) = \big(\textstyle\sum_{j=1}^n l_{j,0}, \sum_{j=1}^n l_{j,1}, \dots, \sum_{j=1}^n l_{j,L}\big). \tag{3.105}$$

This establishes item (i). Moreover, observe that (3.105) demonstrates that

$$\begin{aligned}
\mathcal{P}(\mathbf{P}_n(\Phi)) &= \sum_{k=1}^L \Big[\textstyle\sum_{i=1}^n l_{i,k}\Big]\Big[\big(\textstyle\sum_{i=1}^n l_{i,k-1}\big) + 1\Big] \\
&= \sum_{k=1}^L \Big[\textstyle\sum_{i=1}^n l_{i,k}\Big]\Big[\big(\textstyle\sum_{j=1}^n l_{j,k-1}\big) + 1\Big] \\
&\leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^L l_{i,k}(l_{j,k-1} + 1) \leq \sum_{i=1}^n \sum_{j=1}^n \sum_{k,\ell=1}^L l_{i,k}(l_{j,\ell-1} + 1) \\
&= \sum_{i=1}^n \sum_{j=1}^n \Big[\textstyle\sum_{k=1}^L l_{i,k}\Big]\Big[\textstyle\sum_{\ell=1}^L (l_{j,\ell-1} + 1)\Big] \\
&\leq \sum_{i=1}^n \sum_{j=1}^n \Big[\textstyle\sum_{k=1}^L \frac{1}{2} l_{i,k}(l_{i,k-1} + 1)\Big]\Big[\textstyle\sum_{\ell=1}^L l_{j,\ell}(l_{j,\ell-1} + 1)\Big] \\
&= \sum_{i=1}^n \sum_{j=1}^n \tfrac{1}{2}\mathcal{P}(\Phi_i)\mathcal{P}(\Phi_j) = \tfrac{1}{2}\Big[\textstyle\sum_{i=1}^n \mathcal{P}(\Phi_i)\Big]^2.
\end{aligned} \tag{3.106}$$

The proof of Proposition 3.22 is thus completed. □

A comment from Josh: I think I need this result...

Corollary 3.23. *Let $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy that $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_n)$ (cf. Definition 3.1). Then it holds that $\mathcal{P}(\mathbf{P}_n(\Phi)) \leq n^2\mathcal{P}(\Phi_1)$ (cf. Definition 3.19).*

Proof of Corollary 3.23. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$ satisfy that $\mathcal{D}(\Phi_1) = (l_0, l_1, \dots, l_L)$. Note that item (i) in Proposition 3.22 and the fact that $\forall\, j \in \{1, 2, \dots, n\} \colon \mathcal{D}(\Phi_j) = (l_0, l_1, \dots, l_L)$ demonstrate that

$$\begin{aligned}
\mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) &= \sum_{j=1}^L (nl_j)\big((nl_{j-1}) + 1\big) \leq \sum_{j=1}^L (nl_j)\big((nl_{j-1}) + n\big) \\
&= n^2\Bigg[\sum_{j=1}^L l_j(l_{j-1} + 1)\Bigg] = n^2\mathcal{P}(\Phi_1).
\end{aligned} \tag{3.107}$$

The proof of Corollary 3.23 is thus completed. □

3.4.2 Parallelizations of ANNs with different lengths

A comment from Josh: Should we update this definition to match that of the generalized sum?

Definition 3.24 (Parallelization of ANNs with different length)**.** *Let $n \in \mathbb{N}$, $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_n) \in \mathbf{N}^n$ satisfy for all $j \in \{1, 2, \dots, n\}$ that $\mathcal{H}(\Psi_j) = 1$ and $\mathcal{I}(\Psi_j) = \mathcal{O}(\Psi_j)$. Then we denote by*

$$\mathrm{P}_{n,\Psi} \colon \{(\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n \colon (\forall j \in \{1, 2, \dots, n\} \colon \mathcal{O}(\Phi_j) = \mathcal{I}(\Psi_j))\} \to \mathbf{N} \tag{3.108}$$

the function which satisfies for all $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ with $\forall j \in \{1, 2, \dots, n\} \colon \mathcal{O}(\Phi_j) = \mathcal{I}(\Psi_j)$ that

$$\mathrm{P}_{n,\Psi}(\Phi) = \mathbf{P}_n\big(\mathcal{E}_{\max_{k\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_k),\Psi_1}(\Phi_1), \dots, \mathcal{E}_{\max_{k\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_k),\Psi_n}(\Phi_n)\big) \tag{3.109}$$

(cf. Definitions 3.1, 3.15, and 3.19 and Lemma 3.16).

A comment from Josh: I think I need this result...

Corollary 3.25. *Let $a \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $\mathbb{I} = (\mathbb{I}_1, \mathbb{I}_2, \dots, \mathbb{I}_n)$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy for all $j \in \{1, 2, \dots, n\}$, $x \in \mathbb{R}^{\mathcal{O}(\Phi_j)}$ that $\mathcal{H}(\mathbb{I}_j) = 1$, $\mathcal{I}(\mathbb{I}_j) = \mathcal{O}(\mathbb{I}_j) = \mathcal{O}(\Phi_j)$, and $(\mathcal{R}_a(\mathbb{I}_j))(x) = x$ (cf. Definitions 3.1 and 3.6). Then*

(i) it holds that

$$\mathcal{R}_a\big(\mathrm{P}_{n,\mathbb{I}}(\Phi)\big) \in C\big(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\big) \tag{3.110}$$

and

(ii) it holds for all $x_1 \in \mathbb{R}^{\mathcal{I}(\Phi_1)}, x_2 \in \mathbb{R}^{\mathcal{I}(\Phi_2)}, \dots, x_n \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ that

$$\begin{aligned} &\big(\mathcal{R}_a(\mathrm{P}_{n,\mathbb{I}}(\Phi))\big)(x_1, x_2, \dots, x_n) \\ &= \big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \dots, (\mathcal{R}_a(\Phi_n))(x_n)\big) \in \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]} \end{aligned} \tag{3.111}$$

(cf. Definition 3.24).

Proof of Corollary 3.25. Throughout this proof let $L \in \mathbb{N}$ satisfy that $L = \max_{j\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_j)$. Note that item (ii) in Lemma 3.16, the hypothesis that for all $j \in \{1, 2, \dots, n\}$ it holds that $\mathcal{H}(\mathbb{I}_j) = 1$, (3.68), (3.12), and item (ii) in Lemma 3.17 demonstrate

(I) that for all $j \in \{1, 2, \dots, n\}$ it holds that $\mathcal{L}(\mathcal{E}_{L,\mathbb{I}_j}(\Phi_j)) = L$ and $\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}_j}(\Phi_j)) \in C(\mathbb{R}^{\mathcal{I}(\Phi_j)}, \mathbb{R}^{\mathcal{O}(\Phi_j)})$ and

(II) that for all $j \in \{1, 2, \dots, n\}$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_j)}$ it holds that

$$\big(\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}_j}(\Phi_j))\big)(x) = (\mathcal{R}_a(\Phi_j))(x) \tag{3.112}$$

(cf. Definition 3.15). items (i) and (ii) in Proposition 3.21 therefore imply

(A) that

$$\mathcal{R}_a\big(\mathbf{P}_n\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1), \mathcal{E}_{L,\mathbb{I}_2}(\Phi_2), \dots, \mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big)\big) \in C\big(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\big) \tag{3.113}$$

and

(B) that for all $x_1 \in \mathbb{R}^{\mathcal{I}(\Phi_1)}, x_2 \in \mathbb{R}^{\mathcal{I}(\Phi_2)}, \dots, x_n \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ it holds that

$$\begin{aligned} &\big(\mathcal{R}_a\big(\mathbf{P}_n\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1), \mathcal{E}_{L,\mathbb{I}_2}(\Phi_2), \dots, \mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big)\big)\big)(x_1, x_2, \dots, x_n) \\ &= \Big(\big(\mathcal{R}_a\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1)\big)\big)(x_1), \big(\mathcal{R}_a\big(\mathcal{E}_{L,\mathbb{I}_2}(\Phi_2)\big)\big)(x_2), \dots, \big(\mathcal{R}_a\big(\mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big)\big)(x_n)\Big) \\ &= \Big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \dots, (\mathcal{R}_a(\Phi_n))(x_n)\Big) \end{aligned} \tag{3.114}$$

(cf. Definition 3.19). Combining this with (3.109) and the fact that $L = \max_{j\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_j)$ ensures

(C) that
$$\mathcal{R}_a\big(\mathrm{P}_{n,\mathbb{I}}(\Phi)\big) \in C\big(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\big) \tag{3.115}$$
and

(D) that for all $x_1 \in \mathbb{R}^{\mathcal{I}(\Phi_1)}, x_2 \in \mathbb{R}^{\mathcal{I}(\Phi_2)}, \dots, x_n \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ it holds that
$$\begin{aligned}
&\big(\mathcal{R}_a\big(\mathrm{P}_{n,\mathbb{I}}(\Phi)\big)\big)(x_1, x_2, \dots, x_n) \\
&= \big(\mathcal{R}_a\big(\mathbf{P}_n\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1), \mathcal{E}_{L,\mathbb{I}_2}(\Phi_2), \dots, \mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big)\big)\big)(x_1, x_2, \dots, x_n) \\
&= \Big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \dots, (\mathcal{R}_a(\Phi_n))(x_n)\Big).
\end{aligned} \tag{3.116}$$

This establishes items (i) and (ii). The proof of Corollary 3.25 is thus completed. □

A comment from Josh: I think I need this result...

Corollary 3.26. *Let $n, L \in \mathbb{N}$, $\mathfrak{i}_1, \mathfrak{i}_2, \dots, \mathfrak{i}_n \in \mathbb{N}$, $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_n), \Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy for all $j \in \{1, 2, \dots, n\}$ that $\mathcal{D}(\Psi_j) = (\mathcal{O}(\Phi_j), \mathfrak{i}_j, \mathcal{O}(\Phi_j))$ and $L = \max_{k\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_k)$ (cf. Definition 3.1). Then it holds that*
$$\begin{aligned}
&\mathcal{P}\big(\mathrm{P}_{n,\Psi}(\Phi)\big) \\
&\le \tfrac{1}{2}\bigg(\Big[\textstyle\sum_{j=1}^n \big[\max\big\{1, \tfrac{\mathfrak{i}_j}{\mathcal{O}(\Phi_j)}\big\}\big]\, \mathcal{P}(\Phi_j)\, \mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big] \\
&\quad + \Big[\textstyle\sum_{j=1}^n \big((L - \mathcal{L}(\Phi_j) - 1)\, \mathfrak{i}_j\, (\mathfrak{i}_j + 1) + \mathcal{O}(\Phi_j)\, (\mathfrak{i}_j + 1)\big)\, \mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big] \\
&\quad + \Big[\textstyle\sum_{j=1}^n \mathcal{P}(\Phi_j)\, \mathbb{1}_{\{\mathcal{L}(\Phi_j)\}}(L)\Big] \bigg)^2
\end{aligned} \tag{3.117}$$
(cf. Definition 3.24).

Proof of Corollary 3.26. Observe that (3.109), item (ii) in Proposition 3.22, and item (ii) in Lemma 3.16 assure that
$$\begin{aligned}
&\mathcal{P}\big(\mathrm{P}_{n,\Psi}(\Phi)\big) \\
&= \mathcal{P}\big(\mathbf{P}_n\big(\mathcal{E}_{L,\Psi_1}(\Phi_1), \mathcal{E}_{L,\Psi_2}(\Phi_2), \dots, \mathcal{E}_{L,\Psi_n}(\Phi_n)\big)\big) \\
&\le \tfrac{1}{2}\Big[\textstyle\sum_{j=1}^n \mathcal{P}(\mathcal{E}_{L,\Psi_j}(\Phi_j))\Big]^2 \\
&\le \tfrac{1}{2}\bigg(\Big[\textstyle\sum_{j=1}^n \big[\max\big\{1, \tfrac{\mathfrak{i}_j}{\mathcal{O}(\Phi_j)}\big\}\big]\, \mathcal{P}(\Phi_j)\, \mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big] \\
&\quad + \Big[\textstyle\sum_{j=1}^n \big((L - \mathcal{L}(\Phi_j) - 1)\, \mathfrak{i}_j\, (\mathfrak{i}_j + 1) + \mathcal{O}(\Phi_j)\, (\mathfrak{i}_j + 1)\big)\, \mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big] \\
&\quad + \Big[\textstyle\sum_{j=1}^n \mathcal{P}(\Phi_j)\, \mathbb{1}_{\{\mathcal{L}(\Phi_j)\}}(L)\Big] \bigg)^2
\end{aligned} \tag{3.118}$$
(cf. Definitions 3.15 and 3.19). The proof of Corollary 3.26 is thus completed. □

3.5 Linear transformations of ANNs

A comment from Josh: I have restructured this subsection...

3.5.1 Linear transformations as ANNs

A comment from Josh: Is this an appropriate place for these results?

Definition 3.27 (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 3.1 and 3.2).*

Lemma 3.28. *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 3.1, 3.6, and 3.27).

Proof of Lemma 3.28. Note the fact that $\mathbf{A}_{W,B} \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m) \subseteq \mathbf{N}$ ensures that $\mathcal{D}(\mathbf{A}_{W,B}) = (n, m) \in \mathbb{N}^2$. This establishes item (i). Next, observe that the fact that $\mathbf{A}_{W,B} = (W, B) \in (\mathbb{R}^{m\times n} \times \mathbb{R}^m)$ and (3.6) prove 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

$$(\mathcal{R}_a(\mathbf{A}_{W,B}))(x) = Wx + B. \tag{3.119}$$

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

Lemma 3.29. *Let $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

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

$$\mathcal{D}(\mathbf{A}_{W,B} \bullet \Phi) = \big(\mathbb{D}_0(\Phi), \mathbb{D}_1(\Phi), \ldots, \mathbb{D}_{\mathcal{L}(\Phi)-1}(\Phi), m\big) \in \mathbb{N}^{\mathcal{L}(\Phi)+1}, \tag{3.120}$$

(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

$$(\mathcal{R}_a(\mathbf{A}_{W,B} \bullet \Phi))(x) = W((\mathcal{R}_a(\Phi))(x)) + B, \tag{3.121}$$

(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

$$\mathcal{D}(\Phi \bullet \mathbf{A}_{W,B}) = \big(n, \mathbb{D}_1(\Phi), \mathbb{D}_2(\Phi), \ldots, \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)\big) \in \mathbb{N}^{\mathcal{L}(\Phi)+1}, \tag{3.122}$$

(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

$$(\mathcal{R}_a(\Phi \bullet \mathbf{A}_{W,B}))(x) = (\mathcal{R}_a(\Phi))(Wx + B) \tag{3.123}$$

(cf. Definitions 3.6, 3.8, and 3.27).

Proof of Lemma 3.29. Note that Lemma 3.28 demonstrates 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

$$(\mathcal{R}_a(\mathbf{A}_{W,B}))(x) = Wx + B. \tag{3.124}$$

Combining this and Proposition 3.9 establishes items (i), (ii), (iii), (iv), (v), and (vi). The proof of Lemma 3.29 is thus completed. □

Definition 3.30 (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 3.1, 3.2, and 3.13).*

Lemma 3.31. *Let $n \in \mathbb{N}$ (cf. Definition 3.1). 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 3.5, 3.6, and 3.30).

Proof of Lemma 3.31. Note the fact that $\mathfrak{i}_n \in ((\mathbb{R}^{n\times n} \times \mathbb{R}^n) \times (\mathbb{R}^{n\times n} \times \mathbb{R}^n)) \subseteq \mathbf{N}$ ensures that $\mathcal{D}(\mathfrak{i}_n) = (n, n, n) \in \mathbb{N}^3$. This establishes item (i). Next observe 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 (3.6) prove 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

$$(\mathcal{R}_a(\mathfrak{i}_n))(x) = \mathrm{I}_n(\mathfrak{M}_{a,n}(\mathrm{I}_n x + 0)) + 0 = \mathfrak{M}_{a,n}(x). \tag{3.125}$$

This establishes items (ii) and (iii). The proof of Lemma 3.31 is thus completed. □

Lemma 3.32. *Let $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

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

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

(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 3.6 and 3.30).

Proof of Lemma 3.32. Note that Lemma 3.31 demonstrates that for all $n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathcal{R}_a(\mathfrak{i}_n) \in C(\mathbb{R}^n, \mathbb{R}^n)$ and

$$(\mathcal{R}_a(\mathfrak{i}_n))(x) = \mathfrak{M}_{a,n}(x). \tag{3.126}$$

Combining this and Proposition 3.9 establishes items (i), (ii), (iii), (iv), (v), and (vi). The proof of Lemma 3.32 is thus completed. □

3.5.2 Scalar multiplications of ANNs

Definition 3.33 (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*

$$\lambda \circledast \Phi = \mathbf{A}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi \tag{3.127}$$

(cf. Definitions 3.1, 3.8, 3.13, and 3.27).

Lemma 3.34. *Let $\lambda \in \mathbb{R}$, $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

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

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

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

$$(\mathcal{R}_a(\lambda \circledast \Phi))(x) = \lambda\big((\mathcal{R}_a(\Phi))(x)\big) \tag{3.128}$$

(cf. Definitions 3.6 and 3.33).

Proof of Lemma 3.34. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$ satisfy that $L = \mathcal{L}(\Phi)$ and $(l_0, l_1, \ldots, l_L) = \mathcal{D}(\Phi)$. Note that item (i) in Lemma 3.28 proves that

$$\mathcal{D}(\mathbf{A}_{\lambda \operatorname{I}_{\mathcal{O}(\Phi)},0}) = (\mathcal{O}(\Phi), \mathcal{O}(\Phi)) \tag{3.129}$$

(cf. Definitions 3.13 and 3.27). Combining this and item (i) in Proposition 3.9 assures that

$$\mathcal{D}(\lambda \circledast \Phi) = \mathcal{D}(\mathbf{A}_{\lambda \operatorname{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi) = (l_0, l_1, \ldots, l_{L-1}, \mathcal{O}(\Phi)) = \mathcal{D}(\Phi). \tag{3.130}$$

This establishes item (i). Moreover, observe that items (i) and (ii) in Lemma 3.29 demonstrate that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ it holds that $\mathcal{R}_a(\lambda \circledast \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$ and

$$\begin{aligned}(\mathcal{R}_a(\lambda \circledast \Phi))(x) &= (\mathcal{R}_a(\mathbf{A}_{\lambda \operatorname{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi))(x) \\ &= \lambda \operatorname{I}_{\mathcal{O}(\Phi)}\big((\mathcal{R}_a(\Phi))(x)\big) = \lambda\big((\mathcal{R}_a(\Phi))(x)\big).\end{aligned} \tag{3.131}$$

This establishes items (ii) and (iii). The proof of Lemma 3.34 is thus completed. □

3.6 Representations of the identities with rectifier functions

A comment from Josh: This is new...

Definition 3.35 (Identity network)**.** *We denote by $\mathbf{I} \in \mathbf{N}$ the neural network which satisfies that*

$$\mathbf{I} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right), \left(\begin{pmatrix} 1 & -1 \end{pmatrix}, 0\right)\right) \in \big((\mathbb{R}^{2\times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1\times 2} \times \mathbb{R}^1)\big) \tag{3.132}$$

(cf. Definitions 3.1 and 3.2).

A comment from Josh: This is new...

Lemma 3.36. *Let $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$. Then*

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

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

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

(cf. Definitions 3.1, 3.6, and 3.35).

Proof of Lemma 3.36. Throughout this proof let $L = 2$, $l_0 = 1$, $l_1 = 2$, $l_2 = 1$. Note that (3.132) ensures that $\mathcal{D}(\mathbf{I}) = (1, 2, 1) \in \mathbb{N}^3$. This establishes item (i). Next note that (3.132) assures that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_a(\mathbf{I}))(x) = a(x) - a(-x) = \max\{x, 0\} - \max\{-x, 0\} = x. \tag{3.133}$$

This establishes items (ii) and (iii). The proof of Lemma 3.36 is thus completed. □

3.7 Sums of ANNs

3.7.1 Sums of ANNs with the same length

Definition 3.37. *Let $m, n \in \mathbb{N}$. Then we denote by $\mathfrak{S}_{m,n} \in (\mathbb{R}^{m\times(nm)} \times \mathbb{R}^m)$ the pair given by*

$$\mathfrak{S}_{m,n} = \mathbf{A}_{(\mathrm{I}_m \;\; \mathrm{I}_m \;\; \dots \;\; \mathrm{I}_m),0} \tag{3.134}$$

(cf. Definitions 3.13 and 3.27).

A comment from Josh: I think I need this...

Lemma 3.38. *Let $m, n \in \mathbb{N}$. Then*

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

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

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

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

$$(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \dots, x_n) = \textstyle\sum_{k=1}^n x_k \tag{3.135}$$

(cf. Definitions 3.1, 3.6, and 3.37).

Proof of Lemma 3.38. Note that the fact that $\mathfrak{S}_{m,n} \in (\mathbb{R}^{m\times(nm)} \times \mathbb{R}^m)$ ensures that $\mathfrak{S}_{m,n} \in \mathbf{N}$ and $\mathcal{D}(\mathfrak{S}_{m,n}) = (nm, m) \in \mathbb{N}^2$. This establishes items (i) and (ii). Next observe that items (ii) and (iii) in Lemma 3.28 prove that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x_1, x_2, \dots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$ and

$$\begin{aligned}(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \dots, x_n) &= \big(\mathcal{R}_a\big(\mathbf{A}_{(\mathrm{I}_m \;\; \mathrm{I}_m \;\; \dots \;\; \mathrm{I}_m),0}\big)\big)(x_1, x_2, \dots, x_n) \\ &= (\mathrm{I}_m \;\; \mathrm{I}_m \;\; \dots \;\; \mathrm{I}_m)(x_1, x_2, \dots, x_n) = \textstyle\sum_{k=1}^n x_k\end{aligned} \tag{3.136}$$

(cf. Definitions 3.13 and 3.27). This establishes items (iii) and (iv). The proof of Lemma 3.38 is thus completed. □

A comment from Josh: I think I need this...

Lemma 3.39. *Let $m, n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \{\Psi \in \mathbf{N} \colon \mathcal{O}(\Psi) = nm\}$ (cf. Definition 3.1). Then*

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

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

$$\big(\mathcal{R}_a(\mathfrak{S}_{m,n} \bullet \Phi)\big)(x) = \textstyle\sum_{k=1}^n y_k \tag{3.137}$$

(cf. Definitions 3.6, 3.8, and 3.37).

Proof of Lemma 3.39. Note that Lemma 3.38 ensures that for all $x_1, x_2, \dots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \dots, x_n) = \textstyle\sum_{k=1}^n x_k. \tag{3.138}$$

Combining this and item (v) in Proposition 3.9 establishes items (i) and (ii). The proof of Lemma 3.39 is thus completed. □

A comment from Josh: I think I need this...

Lemma 3.40. *Let $n \in \mathbb{N}$, $a \in C(\mathbb{R},\mathbb{R})$, $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

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

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

$$\big(\mathcal{R}_a(\Phi \bullet \mathfrak{S}_{\mathcal{I}(\Phi),n})\big)(x_1, x_2, \ldots, x_n) = (\mathcal{R}_a(\Phi))(\textstyle\sum_{k=1}^n x_k) \tag{3.139}$$

(cf. Definitions 3.6, 3.8, and 3.37).

Proof of Lemma 3.40. Note that Lemma 3.38 demonstrates that for all $m \in \mathbb{N}$, $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathfrak{S}_{m,n}))(x_1, x_2, \ldots, x_n) = \textstyle\sum_{k=1}^n x_k. \tag{3.140}$$

Combining this and item (v) in Proposition 3.9 establishes items (i) and (ii). The proof of Lemma 3.40 is thus completed. □

Definition 3.41 (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 3.42 (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 pair given by*

$$\mathfrak{T}_{m,n} = \mathbf{A}_{(\mathrm{I}_m \;\; \mathrm{I}_m \;\; \ldots \;\; \mathrm{I}_m)^*, 0} \tag{3.141}$$

(cf. Definitions 3.13, 3.27, and 3.41).

A comment from Josh: I think I need this...

Lemma 3.43. *Let $m, n \in \mathbb{N}$. Then*

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

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

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

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

$$(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) = (x, x, \ldots, x) \tag{3.142}$$

(cf. Definitions 3.1, 3.6, and 3.42).

Proof of Lemma 3.43. Note that the fact that $\mathfrak{T}_{m,n} \in (\mathbb{R}^{(nm) \times m} \times \mathbb{R}^{nm})$ ensures that $\mathfrak{T}_{m,n} \in \mathbf{N}$ and $\mathcal{D}(\mathfrak{T}_{m,n}) = (m, nm) \in \mathbb{N}^2$. This establishes items (i) and (ii). Next observe that items (ii) and (iii) in Lemma 3.28 prove that for all $a \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{nm})$ and

$$\begin{aligned}(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) &= \big(\mathcal{R}_a\big(\mathbf{A}_{(\mathrm{I}_m \;\; \mathrm{I}_m \;\; \ldots \;\; \mathrm{I}_m)^*, 0}\big)\big)(x) \\ &= (\mathrm{I}_m \;\; \mathrm{I}_m \;\; \ldots \;\; \mathrm{I}_m)^* x = (x, x, \ldots, x)\end{aligned} \tag{3.143}$$

(cf. Definitions 3.13 and 3.27). This establishes items (iii) and (iv). The proof of Lemma 3.43 is thus completed. □

A comment from Josh: I think I need this...

Lemma 3.44. *Let $n \in \mathbb{N}$, $a \in C(\mathbb{R},\mathbb{R})$, $\Phi \in \mathbf{N}$ (cf. Definition 3.1). Then*

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

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

$$\big(\mathcal{R}_a(\mathfrak{T}_{\mathcal{O}(\Phi),n} \bullet \Phi)\big)(x) = \big((\mathcal{R}_a(\Phi))(x), (\mathcal{R}_a(\Phi))(x), \dots, (\mathcal{R}_a(\Phi))(x)\big) \tag{3.144}$$

(cf. Definitions 3.6, 3.8, and 3.42).

Proof of Lemma 3.44. Note that Lemma 3.43 ensures that for all $m \in \mathbb{N}$, $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{nm})$ and

$$(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) = (x, x, \dots, x). \tag{3.145}$$

Combining this and item (v) in Proposition 3.9 establishes items (i) and (ii). The proof of Lemma 3.44 is thus completed. □

A comment from Josh: I think I need this...

Lemma 3.45. *Let* $m, n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \{\Psi \in \mathbf{N} \colon \mathcal{I}(\Psi) = nm\}$ *(cf. Definition 3.1). Then*

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

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

$$\big(\mathcal{R}_a(\Phi \bullet \mathfrak{T}_{m,n})\big)(x) = (\mathcal{R}_a(\Phi))(x, x, \dots, x) \tag{3.146}$$

(cf. Definitions 3.6, 3.8, and 3.42).

Proof of Lemma 3.45. Observe that Lemma 3.43 demonstrates that for all $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathfrak{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{nm})$ and

$$(\mathcal{R}_a(\mathfrak{T}_{m,n}))(x) = (x, x, \dots, x). \tag{3.147}$$

Combining this and item (v) in Proposition 3.9 establishes items (i) and (ii). The proof of Lemma 3.45 is thus completed. □

Definition 3.46 (Sums of DNNs with the same length)**.** *Let* $n \in \mathbb{N}$, $\Phi_1, \Phi_2, \dots, \Phi_n \in \mathbf{N}$ *satisfy for all* $k \in \{1, 2, \dots, n\}$ *that* $\mathcal{L}(\Phi_k) = \mathcal{L}(\Phi_1)$, $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_1)$, *and* $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_1)$. *Then we denote by* $\oplus_{k\in\{1,2,\dots,n\}}\Phi_k$ *(we denote by* $\Phi_1 \oplus \Phi_2 \oplus \dots \oplus \Phi_n$*) the neural network given by*

$$\oplus_{k\in\{1,2,\dots,n\}} \Phi_k = \big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n} \bullet \big[\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)\big] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big) \in \mathbf{N} \tag{3.148}$$

(cf. Definitions 3.1, 3.2, 3.8, 3.19, 3.37, and 3.42).

A comment from Josh: I think I need this...

Lemma 3.47. *Let* $n \in \mathbb{N}$, $\Phi_1, \Phi_2, \dots, \Phi_n \in \mathbf{N}$ *satisfy for all* $k \in \{1, 2, \dots, n\}$ *that* $\mathcal{L}(\Phi_k) = \mathcal{L}(\Phi_1)$, $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_1)$, *and* $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_1)$ *(cf. Definition 3.1). Then*

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

(ii) *it holds that*

$$\begin{aligned} &\mathcal{D}(\oplus_{k\in\{1,2,\dots,n\}}\Phi_k) \\ &= \big(\mathcal{I}(\Phi_1), \textstyle\sum_{k=1}^n \mathbb{D}_1(\Phi_k), \sum_{k=1}^n \mathbb{D}_2(\Phi_k), \dots, \sum_{k=1}^n \mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k), \mathcal{O}(\Phi_1)\big), \end{aligned} \tag{3.149}$$

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

$$\big(\mathcal{R}_a(\oplus_{k\in\{1,2,\dots,n\}}\Phi_k)\big)(x) = \sum_{k=1}^{n}(\mathcal{R}_a(\Phi_k))(x) \tag{3.150}$$

(cf. Definitions 3.6 and 3.46).

Proof of Lemma 3.47. First, note that Lemma 3.20 proves that

$$\begin{aligned}&\mathcal{D}\big(\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)\big)\\ &= \big(\textstyle\sum_{k=1}^{n}\mathbb{D}_0(\Phi_k), \sum_{k=1}^{n}\mathbb{D}_1(\Phi_k), \dots, \sum_{k=1}^{n}\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k), \sum_{k=1}^{n}\mathbb{D}_{\mathcal{L}(\Phi_1)}(\Phi_k)\big)\\ &= \big(n\mathcal{I}(\Phi_1), \textstyle\sum_{k=1}^{n}\mathbb{D}_1(\Phi_k), \sum_{k=1}^{n}\mathbb{D}_2(\Phi_k), \dots, \sum_{k=1}^{n}\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k), n\mathcal{O}(\Phi_1)\big)\end{aligned} \tag{3.151}$$

(cf. Definition 3.19). Moreover, observe that item (ii) in Lemma 3.38 ensures that

$$\mathcal{D}\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n}\big) = (n\mathcal{O}(\Phi_1), \mathcal{O}(\Phi_1)) \tag{3.152}$$

(cf. Definition 3.37). This, (3.151), and item (i) in Proposition 3.9 demonstrate that

$$\begin{aligned}&\mathcal{D}\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n} \bullet \big[\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)\big]\big)\\ &= \big(n\mathcal{I}(\Phi_1), \textstyle\sum_{k=1}^{n}\mathbb{D}_1(\Phi_k), \sum_{k=1}^{n}\mathbb{D}_2(\Phi_k), \dots, \sum_{k=1}^{n}\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k), \mathcal{O}(\Phi_1)\big).\end{aligned} \tag{3.153}$$

Next note that item (ii) in Lemma 3.43 assures that

$$\mathcal{D}\big(\mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big) = (\mathcal{I}(\Phi_1), n\mathcal{I}(\Phi_1)) \tag{3.154}$$

(cf. Definition 3.42). Combining this, (3.153), and item (i) in Proposition 3.9 proves that

$$\begin{aligned}&\mathcal{D}(\oplus_{k\in\{1,2,\dots,n\}}\Phi_k)\\ &= \mathcal{D}\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n} \bullet \big[\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)\big] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\\ &= \big(\mathcal{I}(\Phi_1), \textstyle\sum_{k=1}^{n}\mathbb{D}_1(\Phi_k), \sum_{k=1}^{n}\mathbb{D}_2(\Phi_k), \dots, \sum_{k=1}^{n}\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_k), \mathcal{O}(\Phi_1)\big).\end{aligned} \tag{3.155}$$

This establishes items (i) and (ii). Next observe that Lemma 3.45 and (3.151) ensure that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_1)}$ it holds that $\mathcal{R}_a([\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_1),n}) \in C(\mathbb{R}^{\mathcal{I}(\Phi_1)}, \mathbb{R}^{n\mathcal{O}(\Phi_1)})$ and

$$\begin{aligned}&\big(\mathcal{R}_a\big([\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\big)(x)\\ &= \big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)\big)\big)(x, x, \dots, x).\end{aligned} \tag{3.156}$$

Combining this with item (ii) in Proposition 3.21 proves that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_1)}$ it holds that

$$\begin{aligned}&\big(\mathcal{R}_a\big([\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\big)(x)\\ &= \big((\mathcal{R}_a(\Phi_1))(x), (\mathcal{R}_a(\Phi_2))(x), \dots, (\mathcal{R}_a(\Phi_n))(x)\big) \in \mathbb{R}^{n\mathcal{O}(\Phi_1)}.\end{aligned} \tag{3.157}$$

Lemma 3.39, (3.152), and Lemma 3.11 therefore demonstrate that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_1)}$ it holds that $\mathcal{R}_a(\oplus_{k\in\{1,2,\dots,n\}}\Phi_k) \in C(\mathbb{R}^{\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$ and

$$\begin{aligned}&\big(\mathcal{R}_a(\oplus_{k\in\{1,2,\dots,n\}}\Phi_k)\big)(x)\\ &= \big(\mathcal{R}_a\big(\mathfrak{S}_{\mathcal{O}(\Phi_1),n} \bullet [\mathbf{P}_n(\Phi_1,\Phi_2,\dots,\Phi_n)] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_1),n}\big)\big)(x) = \sum_{k=1}^{n}(\mathcal{R}_a(\Phi_k))(x).\end{aligned} \tag{3.158}$$

This establishes items (iii) and (iv). The proof of Lemma 3.47 is thus completed. □

3.7.2 Sums of ANNs with different lengths

A comment from Josh: This has been updated...

Definition 3.48 (Sums of ANNs with different lengths)**.** *Let $u \in \mathbb{N}$, $v \in \mathbb{N} \cap (u, \infty)$, $\Phi_u, \Phi_{u+1}, \ldots, \Phi_v, \Psi \in \mathbf{N}$ satisfy that $\mathcal{I}(\Phi_u) = \mathcal{I}(\Phi_{u+1}) = \ldots = \mathcal{I}(\Phi_v)$, $\mathcal{O}(\Phi_u) = \mathcal{O}(\Phi_{u+1}) = \ldots = \mathcal{O}(\Phi_v) = \mathcal{I}(\Psi) = \mathcal{O}(\Psi)$, and $\mathcal{H}(\Psi) = 1$. Then we denote by $\boxplus_{k=u,\Psi}^{v} \Phi_k$ (we denote by $\Phi_u \boxplus_\Psi \Phi_{u+1} \boxplus_\Psi \cdots \boxplus_\Psi \Phi_v$) the tuple given by*

$$\underset{k=u,\Psi}{\overset{v}{\boxplus}} \Phi_k = \overset{v}{\underset{k=u}{\oplus}} \mathcal{E}_{\max_{k \in \{u,u+1,\ldots,v\}} \mathcal{L}(\Phi_k),\Psi}(\Phi_k) \tag{3.159}$$

(cf. Definitions 3.1, 3.42, and 3.46).

A comment from Josh: I think I need this...

Lemma 3.49. *Let $a \in C(\mathbb{R}, \mathbb{R})$, $u \in \mathbb{N}$, $v \in \mathbb{N} \cap (u, \infty)$, $\Phi_u, \Phi_{u+1}, \ldots, \Phi_v \in \mathbf{N}$ $\Psi = (\Psi_u, \Psi_{u+1}, \ldots, \Psi_v) \in \mathbf{N}^{v-u+1}$ satisfy for all $k \in \{u, u+1, \ldots, v\}$ that $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_u)$, $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_u)$, $\mathcal{H}(\Psi_k) = 1$, $\mathcal{I}(\Psi_k) = \mathcal{O}(\Psi_k) = \mathcal{O}(\Phi_u)$, and $(\mathcal{R}_a(\Psi_k))(x) = x$ (cf. Definitions 3.1 and 3.6). Then*

(i) it holds that $\mathcal{L}(\boxplus_{k=u,\Psi}^{v} \Phi_k) = \mathcal{L}(\Phi_u)$,

(ii) it holds that

$$\begin{aligned} &\mathcal{D}\big(\boxplus_{k=u,\Psi}^{v} \Phi_k\big) \\ &= \Big(\mathcal{I}(\Phi_1), \textstyle\sum_{k=1}^{n} \mathbb{D}_1(\mathcal{E}_{\max_{k \in \{u,u+1,\ldots,v\}} \mathcal{L}(\Phi_k),\Psi}(\Phi_k)), \sum_{k=1}^{n} \mathbb{D}_2(\mathcal{E}_{\max_{k \in \{u,u+1,\ldots,v\}} \mathcal{L}(\Phi_k),\Psi}(\Phi_k)) \\ &\quad \ldots, \textstyle\sum_{k=1}^{n} \mathbb{D}_{\max_{k \in \{u,u+1,\ldots,v\}} \mathcal{L}(\Phi_k)-1}(\mathcal{E}_{\max_{k \in \{u,u+1,\ldots,v\}} \mathcal{L}(\Phi_k),\Psi}(\Phi_k)), \mathcal{O}(\Phi_1)\Big), \end{aligned} \tag{3.160}$$

(iii) it holds that $\mathcal{R}_a(\boxplus_{k=u,\Psi}^{v} \Phi_k) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$, and

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

$$\mathcal{R}_a\big(\boxplus_{k=u,\Psi}^{v} \Phi_k\big) = \sum_{k=u}^{v} \mathcal{R}_a(\Phi_k), \tag{3.161}$$

(cf. Definition 3.48).

Proof of Lemma 3.49. **A comment from Josh: Add proof...** □

3.8 Linear combinations of ANNs

A comment from Josh: Should I prove a "simpler" result for linear combinations?

3.8.1 Linear combinations of ANNs with the same length

A comment from Josh: I'm not sure if I need this...

Lemma 3.50. *Let $n \in \mathbb{N}$, $h_1, h_2, \ldots, h_n \in \mathbb{R}$, $\Phi_1, \Phi_2, \ldots, \Phi_n \in \mathbf{N}$ satisfy that $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_n)$, let $A_k \in \mathbb{R}^{\mathcal{I}(\Phi_1) \times (n\mathcal{I}(\Phi_1))}$, $k \in \{1, 2, \ldots, n\}$, satisfy for all $k \in \{1, 2, \ldots, n\}$, $x = (x_i)_{i \in \{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\phi_1)}$ that $A_k x = x_k$, and let $\Psi \in \mathbf{N}$ satisfy that*

$$\Psi = \oplus_{k \in \{1,2,\ldots,n\}}(h_k \circledast (\Phi_k \bullet \mathbf{A}_{A_k,0})) \tag{3.162}$$

(cf. Definitions 3.1, 3.27, 3.33, and 3.46). Then

(i) it holds that

$$\mathcal{D}(\Psi) = (n\mathcal{I}(\Phi_1), n\mathbb{D}_1(\Phi_1), n\mathbb{D}_2(\Phi_1), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_1), \mathcal{O}(\Phi_1)), \tag{3.163}$$

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

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

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

$$(\mathcal{R}_a(\Psi))(x) = \sum_{k=1}^{n} h_k(\mathcal{R}_a(\Phi_k))(x_k) \tag{3.164}$$

(cf. Definition 3.6).

Proof of Lemma 3.50. First, note that item (i) in Lemma 3.28 ensures for all $k \in \{1, 2, \ldots, n\}$ that

$$\mathcal{D}(\mathbf{A}_{A_k,0}) = (n\mathcal{I}(\Phi_1), \mathcal{I}(\Phi_1)) \in \mathbb{N}^2. \tag{3.165}$$

This and item (i) in Proposition 3.9 prove for all $k \in \{1, 2, \ldots, n\}$ that

$$\mathcal{D}(\Phi_k \bullet \mathbf{A}_{A_k,0}) = (n\mathcal{I}(\Phi_1), \mathbb{D}_1(\Phi_k), \mathbb{D}_2(\Phi_k), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_k)}(\Phi_k)). \tag{3.166}$$

Item (i) in Lemma 3.34 therefore demonstrates for all $k \in \{1, 2, \ldots, n\}$ that

$$\begin{aligned} \mathcal{D}(h_k \circledast (\Phi_k \bullet \mathbf{A}_{A_k,0})) &= \mathcal{D}(\Phi_k \bullet \mathbf{A}_{A_k,0}) \\ &= (n\mathcal{I}(\Phi_1), \mathbb{D}_1(\Phi_k), \mathbb{D}_2(\Phi_k), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_k)-1}(\Phi_k), \mathcal{O}(\Phi_k)) \\ &= (n\mathcal{I}(\Phi_1), \mathbb{D}_1(\Phi_1), \mathbb{D}_2(\Phi_1), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_1), \mathcal{O}(\Phi_1)). \end{aligned} \tag{3.167}$$

Combining this with item (ii) in Lemma 3.47 ensures that

$$\begin{aligned} \mathcal{D}(\Psi) &= \mathcal{D}\big(\oplus_{k\in\{1,2,\ldots,n\}} (h_k \circledast (\Phi_k \bullet \mathfrak{W}_{A_k}))\big) \\ &= (n\mathcal{I}(\Phi_1), n\mathbb{D}_1(\Phi_1), n\mathbb{D}_2(\Phi_1), \ldots, n\mathbb{D}_{\mathcal{L}(\Phi_1)-1}(\Phi_1), \mathcal{O}(\Phi_1)). \end{aligned} \tag{3.168}$$

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

$$\mathcal{P}(\Psi) \leq n^2\mathcal{P}(\Phi_1). \tag{3.169}$$

This establishes item (ii). Moreover, observe that items (v) and (vi) in Lemma 3.29 assure for all $k \in \{1, 2, \ldots, n\}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x = (x_i)_{i\in\{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_1)}$ that $\mathcal{R}_a(\Phi_k \bullet \mathbf{A}_{A_k,0}) \in C(\mathbb{R}^{n\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_k)})$ and

$$\big(\mathcal{R}_a(\Phi_k \bullet \mathbf{A}_{A_k,0})\big)(x) = (\mathcal{R}_a(\Phi))(A_k x) = (\mathcal{R}_a(\Phi))(x_k). \tag{3.170}$$

Combining this with items (ii) and (iii) in Lemma 3.34 proves for all $k \in \{1, 2, \ldots, n\}$, $a \in C(\mathbb{R}, \mathbb{R})$, $x = (x_i)_{i\in\{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_1)}$ that $\mathcal{R}_a(h_k \circledast (\Phi_k \bullet \mathbf{A}_{A_k,0})) \in C(\mathbb{R}^{n\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$ and

$$\big(\mathcal{R}_a(h_k \circledast (\Phi_k \bullet \mathbf{A}_{A_k,0}))\big)(x) = h_k(\mathcal{R}_a(\Phi))(x_k). \tag{3.171}$$

Items (iii) and (iv) in Lemma 3.47 and (3.167) hence ensure for all $a \in C(\mathbb{R}, \mathbb{R})$, $x = (x_i)_{i\in\{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_1)}$ that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^{n\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$ and

$$\begin{aligned} (\mathcal{R}_a(\Psi))(x) &= \big(\mathcal{R}_a\big(\oplus_{k\in\{1,2,\ldots,n\}} (h_k \circledast (\Phi_k \bullet \mathbf{A}_{A_k,0}))\big)\big)(x) \\ &= \sum_{k=1}^{n} \big(\mathcal{R}_a(h_k \circledast (\Phi_k \bullet \mathbf{A}_{A_k,0}))\big)(x) = \sum_{k=1}^{n} h_k(\mathcal{R}_a(\Phi_k))(x_k). \end{aligned} \tag{3.172}$$

This establishes items (iii) and (iv). The proof of Lemma 3.50 is thus completed. □

3.8.2 Linear combinations of ANNs with different lengths

A comment from Josh: I think I need this result...

Lemma 3.51. *Let $a \in C(\mathbb{R},\mathbb{R})$, $n \in \mathbb{N}$, $h_1, h_2, \ldots, h_n \in \mathbb{R}$, $\Phi_1, \Phi_2, \ldots, \Phi_n \in \mathbf{N}$ satisfy for all $k \in \{1, 2, \ldots, n\}$ that $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_1)$, let $A_k \in \mathbb{R}^{\mathcal{I}(\Phi_u)\times(c\mathcal{I}(\Phi_u))}$, $k \in \{1, 2, \ldots, n\}$, satisfy for all $k \in \{1, 2, \ldots, n\}$, $x = (x_i)_{i\in\{1,2,\ldots,n\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_u)}$ that $A_k x = x_k$, and let $\Psi \in \mathbf{N}$ satisfy that*

$$\Psi = \underset{k=u,\mathbf{I}}{\overset{v}{\boxplus}} (h_k \circledast (\Phi_k \bullet \mathbf{A}_{A_k,0})) \tag{3.173}$$

(cf. Definitions 3.1, 3.6, 3.27, 3.33, 3.35, and 3.48). Then

(i) it holds that

$$\mathcal{D}(\Psi) = \textbf{A comment from Josh: Add vector...}, \tag{3.174}$$

(ii) it holds that $\mathcal{P}(\Psi) \leq$ **A comment from Josh: Add value...**,

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

(iv) it holds for all $x = (x_k)_{k\in\{u,u+1,\ldots,v\}} \in \mathbb{R}^{n\mathcal{I}(\Phi_u)}$ that

$$(\mathcal{R}_a(\Psi))(x) = \sum_{k=u}^{v} h_k(\mathcal{R}_a(\Phi_k))(x_k). \tag{3.175}$$

Proof of Lemma 3.51. **A comment from Josh: Add proof....** □

4 ANN representations for MLP approximations

4.1 ANN representations for MLP approximations

Lemma 4.1. *Let $\alpha, \beta, M \in [0, \infty)$, $U_n \in [0, \infty)$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}$ that*

$$U_n \leq \alpha M^n + \sum_{k=0}^{n-1} M^{n-k} \left(\max\{\beta, U_k\} + \mathbb{1}_{\mathbb{N}}(k) \max\{\beta, U_{\max\{k-1,0\}}\}\right). \tag{4.1}$$

Then it holds for all $n \in \mathbb{N}$ that $U_n \leq (2M + 1)^n \max\{\alpha, \beta\}$.

Proof of Lemma 4.1. We prove Lemma 4.1 by induction on $n \in \mathbb{N}_0$. Throughout this proof let $S_n \in [0, \infty)$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$ that

$$S_n = M^n + \sum_{k=0}^{n-1} M^{n-k} \left((2M + 1)^k + \mathbb{1}_{\mathbb{N}}(k) \, (2M + 1)^{\max\{k-1,0\}}\right). \tag{4.2}$$

For the base case $n = 0$ note that (4.1) implies that $U_0 \leq \alpha \leq \max\{\alpha, \beta\}$. This proves the base case $n = 0$. For the induction step from $n \in \mathbb{N}_0$ to $n + 1 \in \mathbb{N}$ let $n \in \mathbb{N}_0$ and assume that for all $k \in \{0, 1, \ldots, n\}$ it holds that $U_k \leq (2M + 1)^k \max\{\alpha, \beta\}$. Note that (4.1) yields that

$$\begin{aligned} U_{n+1} &\leq \alpha M^{n+1} + \sum_{k=0}^{n} M^{n+1-k} \left(\max\{\beta, U_k\} + \mathbb{1}_{\mathbb{N}}(k) \max\{\beta, U_{\max\{k-1,0\}}\}\right) \\ &\leq \alpha M^{n+1} + \sum_{k=0}^{n} M^{n+1-k} \Big(\max\{\beta, (2M + 1)^k \max\{\alpha, \beta\}\} \\ &\qquad\qquad + \mathbb{1}_{\mathbb{N}}(k) \max\{\beta, (2M + 1)^{\max\{k-1,0\}} \max\{\alpha, \beta\}\}\Big) \\ &\leq \alpha M^{n+1} + \max\{\alpha, \beta\} \sum_{k=0}^{n} M^{n+1-k} \left((2M + 1)^k + \mathbb{1}_{\mathbb{N}}(k) \, (2M + 1)^{\max\{k-1,0\}}\right) \\ &\leq \max\{\alpha, \beta\} S_{n+1}. \end{aligned} \tag{4.3}$$

Note that by (4.2) and the assumption that $M \in [0,\infty)$ it follows that

$$
\begin{aligned}
S_{n+1} &= M^{n+1} + \sum_{k=0}^{n} M^{n+1-k} \left((2M+1)^k + \mathbb{1}_{\mathbb{N}}(k)\,(2M+1)^{\max\{k-1,0\}}\right) \\
&= M^{n+1} + \sum_{k=0}^{n} M^{n+1-k}(2M+1)^k + \sum_{k=1}^{n} M^{n+1-k}(2M+1)^{k-1} \\
&= M^{n+1} + M\left[\frac{(2M+1)^{n+1} - M^{n+1}}{M+1}\right] + M\left[\frac{(2M+1)^{n} - M^{n}}{M+1}\right] \\
&= M^{n+1} + \frac{M(2M+1)^{n+1}}{M+1} + \frac{M(2M+1)^{n}}{M+1} - M\left[\frac{M^{n+1}+M^{n}}{M+1}\right] \\
&\leq M^{n+1} + \frac{M(2M+1)^{n+1}}{M+1} + \frac{(2M+1)^{n+1}}{M+1} - M^{n+1}\left[\frac{M+1}{M+1}\right] \\
&= (2M+1)^{n+1}.
\end{aligned}
\tag{4.4}
$$

Combining (4.3) and (4.4) completes the induction step, which establishes (4.1). The proof of Lemma 4.1 is thus completed. ☐

Lemma 4.2. *Let $\Theta = (\bigcup_{n\in\mathbb{N}} \mathbb{Z}^n)$, $d, M \in \mathbb{N}$, $T \in (0,\infty)$, $f \in C(\mathbb{R},\mathbb{R})$, $g \in C(\mathbb{R}^d,\mathbb{R})$, $\mathbf{F}, \mathbf{G} \in \mathbf{N}$ satisfy that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) = f$, and $\mathcal{R}_{\mathfrak{r}}(\mathbf{G}) = g$, let $\mathfrak{u}^\theta \in [0,1]$, $\theta \in \Theta$, and $\mathcal{U}^\theta\colon [0,T] \to [0,T]$, $\theta \in \Theta$, satisfy for all $t \in [0,T]$, $\theta \in \Theta$ that $\mathcal{U}^\theta_t = t + (T-t)\mathfrak{u}^\theta$, let $W^\theta\colon [0,T] \to \mathbb{R}^d$, $\theta \in \Theta$, for every $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$ let $Y^\theta_{t,s} \in \mathbb{R}^d$ satisfy $Y^\theta_{t,s} = W^\theta_s - W^\theta_t$, and let $U^\theta_n\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that*

$$
\begin{aligned}
U^\theta_n(t,x) = \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n}\left[\sum_{k=1}^{M^n} g\!\left(x + Y^{(\theta,0,-k)}_{t,T}\right)\right] \\
+ \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}}\left[\sum_{k=1}^{M^{n-i}} \left((f\circ U^{(\theta,i,k)}_i) - \mathbb{1}_{\mathbb{N}}(i)\,(f\circ U^{(\theta,-i,k)}_{\max\{i-1,0\}})\right)\!\left(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}\right)\right]
\end{aligned}
\tag{4.5}
$$

(cf. Definitions 3.1, 3.4, and 3.6). Then

(i) there exist unique $\mathbf{U}^\theta_{n,t} \in \mathbf{N}$, $t \in [0,T]$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, which satisfy 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 $\theta \in \Theta$, $t \in [0,T]$ that $\mathbf{U}^\theta_{0,t} = ((0\ 0\ \dots\ 0), 0) \in \mathbb{R}^{1\times d}\times\mathbb{R}^1$,

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

$$
\begin{aligned}
\mathbf{U}^\theta_{n,t} = &\left[\bigoplus_{k=1}^{M^n}\left(\frac{1}{M^n} \circledast \left(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,0,-k)}_{t,T}}\right)\right)\right] \\
&\boxplus_{\mathbf{I}} \left[\mathop{\boxplus}_{i=0,\mathbf{I}}^{n-1}\left[\left(\frac{(T-t)}{M^{n-i}}\right) \circledast \left(\mathop{\boxplus}_{k=1,\mathbf{I}}^{M^{n-i}}\left(\left(\mathbf{F}\bullet\mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)\right]\right] \\
&\boxplus_{\mathbf{I}} \left[\mathop{\boxplus}_{i=0,\mathbf{I}}^{n-1}\left[\left(\frac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\right) \circledast \left(\mathop{\boxplus}_{k=1,\mathbf{I}}^{M^{n-i}}\left(\left(\mathbf{F}\bullet\mathbf{U}^{(\theta,-i,k)}_{\max\{i-1,0\},\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)\right]\right],
\end{aligned}
\tag{4.6}
$$

A comment from Josh: Would a different construction yield better estimates?

(iv) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$ that $\mathcal{L}(\mathbf{U}^\theta_{n,t}) = n\mathcal{H}(\mathbf{F}) + \max\{1, \mathbb{1}_{\mathbb{N}}(n)\,\mathcal{L}(\mathbf{G})\}$,

(v) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$ that $|||\mathcal{D}(\mathbf{U}^\theta_{n,t})||| \leq (2M+1)^n \max\{2, |||\mathcal{D}(\mathbf{F})|||, |||\mathcal{D}(\mathbf{G})|||\}$,

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

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

$$\mathcal{P}(\mathbf{U}^\theta_{n,t}) \leq 2n\mathcal{H}(\mathbf{F}) + \max\{1, \mathbb{1}_\mathbb{N}(n)\mathcal{L}(\mathbf{G})\} \left[(2M+1)^n \max\{2, \|\!|\mathcal{D}(\mathbf{F})|\!\|, \|\!|\mathcal{D}(\mathbf{G})|\!\|\}\right]^2 \tag{4.7}$$

A comment from Josh: I need to double-check the above... Can it be improved?

(cf. Definitions 3.8, 3.13, 3.27, 3.33, 3.35, 3.37, and 3.48).

A comment from Josh: The proof is still being update/finalized...
A comment from Josh: I cannot do what I did in (4.8)...

Proof of Lemma 4.2. Throughout this proof let $\kappa^n_\ell \in \mathbb{N}$, $\ell \in \{1, 2, \ldots, \mathcal{L}(\mathbf{F} \bullet \mathbf{U}^0_{n,0}) - 1\}$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $\ell \in \{1, 2, \ldots, \mathcal{L}(\mathbf{F} \bullet \mathbf{U}^0_{n,0}) - 1\}$ that

$$\begin{aligned} \kappa^n_\ell &= M^{n+1}\mathbb{D}_\ell(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}}(\mathbf{G})) + \sum_{i=0}^{n} M^{n+1-i}\mathbb{D}_\ell(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}}(\mathbf{F} \bullet \mathbf{U}^0_{i,0})) \\ &\quad + \sum_{i=0}^{n} M^{n+1-i}\mathbb{D}_\ell(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}}(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-1,0\},0})). \end{aligned} \tag{4.8}$$

We prove items (i), (ii), (iii), (iv), (v), and (vi) by induction on $n \in \mathbb{N}_0$. For the base case $n = 0$ note that the fact that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ we have $U^\theta_n(t,x) = 0$ and **A comment from Josh: Add lemma** imply that there exist unique $\mathbf{U}^\theta_{0,t} \in \mathbf{N}$, $t \in [0,T]$, $\theta \in \Theta$, such that it holds for all $\theta \in \Theta$, $t \in [0,T]$ that $\mathbf{U}^\theta_{0,t} = ((0\ 0\ \ldots\ 0), 0) \in \mathbb{R}^{1\times d} \times \mathbb{R}^1$ and for all $\theta_1, \theta_2 \in \Theta$, $t_1, t_2 \in [0,T]$ that $\mathcal{D}(\mathbf{U}^{\theta_1}_{0,t_1}) = \mathcal{D}(\mathbf{U}^{\theta_2}_{0,t_2})$. Moreover, by the assumption that $\mathcal{R}_a(\mathbf{G}) \in C(\mathbb{R}^d, \mathbb{R})$ it follows that for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that $\mathcal{L}(\mathbf{U}^\theta_{0,t}) = 1 = \max\{1, 0\}$, $\|\!|\mathcal{D}(\mathbf{U}^\theta_{0,t})|\!\| = d \leq \|\!|\mathcal{D}(\mathbf{G})|\!\| \leq \max\{\|\!|\mathcal{D}(\mathbf{F})|\!\|, \|\!|\mathcal{D}(\mathbf{G})|\!\|\}$, and $(\mathcal{R}_\mathfrak{r}(\mathbf{U}^\theta_{0,t})(x) = U^\theta_0(t,x)$. This proves the base case $n = 0$. For the induction step from $n \in \mathbb{N}_0$ to $n+1 \in \mathbb{N}_0$ let $n \in \mathbb{N}_0$ and assume that items (i), (ii), (iii), (iv), (v), and (vi) hold true for all $k \in \{0, 1, \ldots, n\}$. **A comment from Josh: Add reference...** shows that for all $\theta \in \Theta$, $t \in [0,T]$ that

$$\mathcal{D}\Big(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, Y^\theta_{t,T}}\Big) = \mathcal{D}(\mathbf{G}). \tag{4.9}$$

(4.9) and **A comment from Josh: Add reference...** demonstrate that for all $\theta \in \Theta$, $t \in [0,T]$ that

$$\begin{aligned} &\mathcal{D}\Bigg(\mathop{\boxplus}_{k=1,\mathbf{I}}^{M^{n+1}} \Bigg(\frac{1}{M^{n+1}} \circledast \Big(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,0,-k)}_{t,T}}\Big)\Bigg)\Bigg) \\ &= \mathcal{D}\big(\mathcal{I}(\mathbf{G}), M^{n+1}\mathbb{D}_1(\mathbf{G}), M^{n+1}\mathbb{D}_2(\mathbf{G}), \ldots, M^{n+1}\mathbb{D}_{\mathcal{L}(\mathbf{G})-1}(\mathbf{G}), \mathcal{O}(\mathbf{G})\big). \end{aligned} \tag{4.10}$$

Next, the induction hypothesis implies for all $\theta \in \Theta$, $t \in [0,T]$, $\ell \in \{0, 1, \ldots, n\}$ that $\mathcal{D}(\mathbf{U}^\theta_{\ell,t}) = \mathcal{D}(\mathbf{U}^0_{\ell,0})$. This and **A comment from Josh: Add reference...** imply that for all $\theta, \eta \in \Theta$, $t \in [0,T]$, $\ell \in \{0, 1, \ldots, n\}$ it holds that

$$\mathcal{D}\Big(\mathbf{U}^\eta_{\ell,\mathcal{U}^\theta_t} \bullet \mathbf{A}_{\mathrm{I}_d, Y^\theta_{t,\mathcal{U}^\theta_t}}\Big) = \mathcal{D}\Big(\mathbf{U}^\eta_{\ell,\mathcal{U}^\theta_t}\Big) = \mathcal{D}(\mathbf{U}^0_{\ell,0}). \tag{4.11}$$

This, Lemma 3.11, **A comment from Josh: Add lemma**, and the induction hypothesis then yield for all $\theta, \eta \in \Theta$, $t \in [0,T]$, $\ell \in \{0, 1, \ldots, n\}$ it holds that

$$\begin{aligned} &\mathcal{D}\Big(\Big(\mathbf{F} \bullet \mathbf{U}^\eta_{\ell,\mathcal{U}^\theta_t}\Big) \bullet \mathbf{A}_{\mathrm{I}_d, Y^\theta_{t,\mathcal{U}^\theta_t}}\Big) = \mathcal{D}\Big(\mathbf{F} \bullet \Big(\mathbf{U}^\eta_{\ell,\mathcal{U}^\theta_t} \bullet \mathbf{A}_{\mathrm{I}_d, Y^\theta_{t,\mathcal{U}^\theta_t}}\Big)\Big) = \mathcal{D}(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}) \\ &= \Big(\mathcal{I}(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}), \mathbb{D}_1(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}), \mathbb{D}_2(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}), \ldots, \mathbb{D}_{\mathcal{L}(\mathbf{U}^0_{\ell,0})+\mathcal{L}(\mathbf{F})-1}(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}), \mathcal{O}(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0})\Big) \\ &= \Big(d, \mathbb{D}_1(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}), \mathbb{D}_2(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}), \ldots, \mathbb{D}_{\mathcal{L}(\mathbf{U}^0_{\ell,0})+\mathcal{L}(\mathbf{F})-1}(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}), 1\Big). \end{aligned} \tag{4.12}$$

This, A comment from Josh: Add reference..., and the induction hypothesis shows that for all $i \in \{0, 1, \ldots, n\}$, $\theta \in \Theta$, $t \in [0, T]$ that Note: The following is incorrect and needs to be updated...

$$\begin{aligned} &\mathcal{D}\Bigg(\boxplus_{k=1,\mathbf{I}}^{M^{n+1-i}} \Bigg(\Big(\mathbf{F} \bullet \mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}_t^{(\theta,i,k)}} \Big) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}_t^{(\theta,i,k)}}} \Bigg) \Bigg) = \mathcal{D}\Bigg(\boxplus_{k=1,\mathbf{I}}^{M^{n+1-i}} \big(\mathbf{F} \bullet \mathbf{U}^0_{i,0}\big) \Bigg) \\ &= \Big(d, M^{n+1-i}\mathbb{D}_1\big(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}\big), M^{n+1-i}\mathbb{D}_2\big(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}\big), \ldots, M^{n+1-i}\mathbb{D}_{\mathcal{L}(\mathbf{U}^0_{\ell,0})+\mathcal{L}(\mathbf{F})-1}\big(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}\big), 1 \Big). \end{aligned} \tag{4.13}$$

Note that by the induction hypothesis we have for all $i \in \{0, 1, \ldots, n\}$, $\theta \in \Theta$, $t \in [0, T]$ that

$$\mathcal{L}(\mathbf{U}^\theta_{i,t}) = \mathcal{L}(\mathbf{U}^0_{i,0}) = i\mathcal{H}(\mathbf{F}) + \max\{1, \mathbb{1}_{\mathbb{N}}(i)\, \mathcal{L}(\mathbf{G})\} \tag{4.14}$$

and, as a consequence of (4.14), for all $i, j \in \{0, 1, \ldots, n\}$, $\theta \in \Theta$, $t \in [0, T]$ with $i \leq j$ it follows that

$$\mathcal{L}(\mathbf{U}^\theta_{i,t}) = \mathcal{L}(\mathbf{U}^0_{i,0}) \leq \mathcal{L}(\mathbf{U}^0_{j,0}) = \mathcal{L}(\mathbf{U}^\theta_{j,t}). \tag{4.15}$$

(4.15) and Definition 3.8 then imply that for all $i, j \in \{0, 1, \ldots, n\}$, $\theta \in \Theta$, $t \in [0, T]$ with $i \leq j$ it follows that

$$\mathcal{L}(\mathbf{F} \bullet \mathbf{U}^\theta_{i,t}) = \mathcal{L}(\mathbf{F} \bullet \mathbf{U}^0_{i,0}) \leq \mathcal{L}(\mathbf{F} \bullet \mathbf{U}^0_{j,0}) = \mathcal{L}(\mathbf{F} \bullet \mathbf{U}^\theta_{j,t}). \tag{4.16}$$

This, (4.13), Definition 3.15, A comment from Josh: Add reference... and the induction hypothesis show that for all $\ell \in \{0, 1, \ldots, n\}$, $\theta \in \Theta$, $t \in [0, T]$ that Note: The following needs to be corrected...

$$\begin{aligned} &\mathcal{D}\left(\boxplus_{i=0,\mathbf{I}}^{n} \left[\left(\frac{(T-t)}{M^{n+1-i}} \right) \circledast \left(\boxplus_{k=1,\mathbf{I}}^{M^{n+1-i}} \left(\Big(\mathbf{F} \bullet \mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}_t^{(\theta,i,k)}} \Big) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}_t^{(\theta,i,k)}}} \right) \right) \right] \right) \\ &= \mathcal{D}\left(\boxplus_{i=0,\mathbf{I}}^{n} \left[\left(\frac{(T-t)}{M^{n+1-i}} \right) \circledast \left(\boxplus_{k=1,\mathbf{I}}^{M^{n+1-i}} \big(\mathbf{F} \bullet \mathbf{U}^0_{i,0}\big) \right) \right] \right) \\ &= \Bigg(d, \sum_{i=0}^{n} M^{n+1-i}\mathbb{D}_1\Big(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}} \big(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}\big) \Big), \sum_{i=0}^{n} M^{n+1-i}\mathbb{D}_2\Big(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}} \big(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}\big) \Big), \\ &\quad \ldots, \sum_{i=0}^{n} M^{n+1-i}\mathbb{D}_{\mathcal{L}(\mathbf{U}^0_{\ell,0})+\mathcal{L}(\mathbf{F})-1}\Big(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}} \big(\mathbf{F} \bullet \mathbf{U}^0_{\ell,0}\big) \Big), 1 \Bigg) \end{aligned} \tag{4.17}$$

This, (4.10), A comment from Josh: Add reference..., and the induction hypothesis ensure that for all $\theta \in \Theta$, $t \in [0, T]$ that Note: This needs to be corrected...

$$\begin{aligned} &\mathcal{D}(\mathbf{U}^\theta_{n+1,t}) \\ &= \mathcal{D}\Bigg(\left[\bigoplus_{k=1}^{M^{n+1}} \left(\frac{1}{M^{n+1}} \circledast \Big(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,0,-k)}_{t,T}} \Big) \right) \right] \\ &\quad \boxplus_{\mathbf{I}} \left[\boxplus_{i=0,\mathbf{I}}^{n} \left[\left(\frac{(T-t)}{M^{n+1-i}} \right) \circledast \left(\boxplus_{k=1,\mathbf{I}}^{M^{n+1-i}} \left(\Big(\mathbf{F} \bullet \mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}_t^{(\theta,i,k)}} \Big) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}_t^{(\theta,i,k)}}} \right) \right) \right] \right] \\ &\quad \boxplus_{\mathbf{I}} \left[\boxplus_{i=0,\mathbf{I}}^{n} \left[\left(\frac{(t-T)\, \mathbb{1}_{\mathbb{N}}(i)}{M^{n+1-i}} \right) \circledast \left(\boxplus_{k=1,\mathbf{I}}^{M^{n+1-i}} \left(\Big(\mathbf{F} \bullet \mathbf{U}^{(\theta,-i,k)}_{\max\{i-1,0\},\mathcal{U}_t^{(\theta,i,k)}} \Big) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}_t^{(\theta,i,k)}}} \right) \right) \right] \right] \Bigg) \\ &= \Big(d, \kappa_1^{n+1}, \kappa_2^{n+1}, \ldots, \kappa^{n+1}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0})-1}, 1 \Big). \end{aligned} \tag{4.18}$$

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

$$\mathcal{D}\big(\mathbf{U}^{\theta_1}_{n+1,t_1}\big) = \mathcal{D}\big(\mathbf{U}^{\theta_2}_{n+1,t_2}\big). \tag{4.19}$$

By **A comment from Josh: Add reference...**, (4.18), and the induction hypothesis it follows that for all $\theta \in \Theta$, $t \in [0,T]$ that

$$\begin{aligned} \mathcal{L}(\mathbf{U}^\theta_{n+1,t}) &= \mathcal{L}(\mathbf{F} \bullet \mathbf{U}^\theta_{n,t}) = \mathcal{L}(\mathbf{U}^\theta_{n,t}) + \mathcal{L}(\mathbf{F}) - 1 \\ &= (n\mathcal{H}(\mathbf{F}) + \max\{1, \mathbb{1}_{\mathbb{N}}(n)\, \mathcal{L}(\mathbf{G})\}) + \mathcal{H}(\mathbf{F}) = (n+1)\mathcal{H}(\mathbf{F}) + \max\{1, \mathbb{1}_{\mathbb{N}}(n)\, \mathcal{L}(\mathbf{G})\}. \end{aligned} \tag{4.20}$$

Furthermore, (4.8) and (4.18), **A comment from Josh: Add reference...**, Lemma 4.1, and the induction hypothesis ensure for all $\theta \in \Theta$, $t \in [0,T]$ that **Note: The following needs to be corrected...**

$$\begin{aligned} \|\!|\mathcal{D}(\mathbf{U}^\theta_{n+1,t})|\!\| &\leq \max\left\{d, \kappa_1, \kappa_2, \ldots, \kappa_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0})-1}, 1\right\} \\ &\leq \max_{\ell \in \{0,1,\ldots,\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0})\}} \Bigg(M^{n+1}\mathbb{D}_\ell(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}}(\mathbf{G})) + \sum_{i=0}^{n} M^{n+1-i}\mathbb{D}_\ell(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}}(\mathbf{F} \bullet \mathbf{U}^0_{i,0})) \\ &\quad + \sum_{i=0}^{n} M^{n+1-i}\mathbb{D}_\ell(\mathcal{E}_{\mathcal{L}(\mathbf{F}\bullet\mathbf{U}^0_{n,0}),\mathfrak{I}}(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-1,0\},0})) \Bigg) \\ &\leq M^{n+1}\max\{2, \|\!|\mathcal{D}(\mathbf{G})|\!\|\} \\ &\quad + \sum_{i=0}^{n} M^{n+1-i}\left(\max\{2, \|\!|\mathcal{D}(\mathbf{F} \bullet \mathbf{U}^0_{i,0})|\!\|\} + \max\{2, \|\!|\mathcal{D}(\mathbf{F} \bullet \mathbf{U}^0_{\max\{i-1,0\},0})|\!\|\}\right) \\ &\leq M^{n+1}\max\{2, \|\!|\mathcal{D}(\mathbf{G})|\!\|\} \\ &\quad + \sum_{i=0}^{n} M^{n+1-i}\left(\max\{2, \|\!|\mathcal{D}(\mathbf{F})|\!\|, \|\!|\mathcal{D}(\mathbf{U}^0_{i,0})|\!\|\} + \max\{2, \|\!|\mathcal{D}(\mathbf{F})|\!\|, \|\!|\mathcal{D}(\mathbf{U}^0_{\max\{i-1,0\},0})|\!\|\}\right) \\ &\leq (2M+1)^{n+1}\max\{2, \|\!|\mathcal{D}(\mathbf{F})|\!\|, \|\!|\mathcal{D}(\mathbf{G})|\!\|\}. \end{aligned} \tag{4.21}$$

Finally, **A comment from Josh: Add references...**, and the induction hypothesis assure for all $\theta \in \Theta$,

$t \in [0,T]$, $x \in \mathbb{R}^d$ that **Note: This should all be correct...**

$$
\begin{aligned}
&(\mathcal{R}_{\mathfrak{r}}(\mathbf{U}^{\theta}_{n+1,t}))(x) \\
&= \left(\mathcal{R}_{\mathfrak{r}}\left(\bigoplus_{k=1}^{M^n}\left(\frac{1}{M^n} \circledast \left(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,0,-k)}_{t,T}}\right)\right)\right)\right)(x) \\
&\quad + \left(\mathcal{R}_{\mathfrak{r}}\left(\boxplus_{i=0,\mathbf{I}}^{n-1}\left[\left(\frac{(T-t)}{M^{n-i}}\right) \circledast \left(\boxplus_{k=1,\mathbf{I}}^{M^{n-i}}\left(\left(\mathbf{F} \bullet \mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)\right]\right)\right)(x) \\
&\quad + \left(\mathcal{R}_{\mathfrak{r}}\left(\boxplus_{i=0,\mathbf{I}}^{n-1}\left[\left(\frac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\right) \circledast \left(\boxplus_{k=1,\mathbf{I}}^{M^{n-i}}\left(\left(\mathbf{F} \bullet \mathbf{U}^{(\theta,-i,k)}_{\max\{i-1,0\},\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)\right]\right)\right)(x) \\
&= \sum_{k=1}^{M^n} \frac{1}{M^n}\left(\mathcal{R}_{\mathfrak{r}}\left(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,0,-k)}_{t,T}}\right)\right)(x) \\
&\quad + \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}}\left(\mathcal{R}_{\mathfrak{r}}\left(\boxplus_{k=1,\mathbf{I}}^{M^{n-i}}\left(\left(\mathbf{F} \bullet \mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)\right)(x) \\
&\quad + \sum_{i=0}^{n-1} \frac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\left(\mathcal{R}_{\mathfrak{r}}\left(\boxplus_{k=1,\mathbf{I}}^{M^{n-i}}\left(\left(\mathbf{F} \bullet \mathbf{U}^{(\theta,-i,k)}_{\max\{i-1,0\},\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)\right)(x) \\
&= \frac{1}{M^n}\left[\sum_{k=1}^{M^n} g\left(x + Y^{(\theta,0,-k)}_{t,T}\right)\right] \\
&\quad + \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}}\left[\sum_{k=1}^{M^{n-i}}\left(\mathcal{R}_{\mathfrak{r}}\left(\left(\mathbf{F} \bullet \mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)(x)\right] \\
&\quad + \sum_{i=0}^{n-1} \frac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\left[\sum_{k=1}^{M^{n-i}}\left(\mathcal{R}_{\mathfrak{r}}\left(\left(\mathbf{F} \bullet \mathbf{U}^{(\theta,-i,k)}_{\max\{i-1,0\},\mathcal{U}^{(\theta,i,k)}_t}\right) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}}\right)\right)\right](x) \\
&= \frac{1}{M^n}\left[\sum_{k=1}^{M^n} g\left(x + Y^{(\theta,0,-k)}_{t,T}\right)\right] + \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}}\left[\sum_{k=1}^{M^{n-i}}\left(f \circ U^{(\theta,i,k)}_i\right)\left(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}\right)\right] \\
&\quad + \sum_{i=0}^{n-1} \frac{(T-t)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\left[\sum_{k=1}^{M^{n-i}}\left((f \circ U^{(\theta,-i,k)}_{\max\{i-1,0\}}\right)\left(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}\right)\right] = U^{\theta}_n(t,x).
\end{aligned}
\tag{4.22}
$$

This completes the induction step. The proof of Lemma 4.2 is thus completed. □

4.2 ANN representations for the PDE nonlinearity

4.2.1 Linear interpolation with ANNs

A comment from Josh: I left this material here... Should I move it to Section 3.5.1?

Lemma 4.3. *Let $\alpha, \beta, h \in \mathbb{R}$. Then*

(i) there exists a unique $\mathbf{H} \in ((\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \times (\mathbb{R}^{1\times 1} \times \mathbb{R}^1)) \subseteq \mathbf{N}$ which satisfies that $\mathbf{H} = h \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) = ((\alpha,\beta),(h,0))$,

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

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

(iv) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbf{H}))(x) = h\max\{\alpha x + \beta, 0\} = h\,[\alpha x + \beta]\,\mathbb{1}_{(-\beta/\alpha,\infty)}(x)$

(cf. Definitions 3.1, 3.8, 3.27, 3.30, and 3.33).

Proof of Lemma 4.3. Note that Definition 3.27 and Lemma 3.28 ensure that for all $x \in \mathbb{R}$ it holds that $\mathbf{A}_{\alpha,\beta} = (\alpha, \beta) \in (\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \subseteq \mathbf{N}$, $\mathcal{D}(\mathbf{A}_{\alpha,\beta}) = (1, 1) \in \mathbb{N}^2$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta}) \in C(\mathbb{R}, \mathbb{R})$, and $(\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta}))(x) = \alpha x + \beta$. By (3.6), Definitions 3.8 and 3.30, and Lemmas 3.31 and 3.32 it follows for all $x \in \mathbb{R}$ that $\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta} = ((\alpha, \beta), (1, 0)) \in ((\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \times (\mathbb{R}^{1\times 1} \times \mathbb{R}^1)) \subseteq \mathbf{N}$, $\mathcal{D}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) = (1, 1, 1) \in \mathbb{N}^3$, $\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) \in C(\mathbb{R}, \mathbb{R})$, and

$$(\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}))(x) = \max\{\alpha x + \beta, 0\} = [\alpha x + \beta]\mathbb{1}_{(-\beta/\alpha,\infty)}(x). \tag{4.23}$$

This, Definition 3.33, and Lemma 3.34 ensure that there exists a unique $\mathbf{H} \in ((\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \times (\mathbb{R}^{1\times 1} \times \mathbb{R}^1)) \subseteq \mathbf{N}$ which satisfies for all $x \in \mathbb{R}$ that $\mathbf{H} = h \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) = ((\alpha, \beta), (h, 0))$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{H}) \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{D}(\mathbf{H}) = (1, 1, 1) \in \mathbb{N}^3$, and

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{H}))(x) = h((\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}))(x)) = h[\alpha x + \beta]\mathbb{1}_{(-\beta/\alpha,\infty)}(x). \tag{4.24}$$

This establishes items (i), (ii), (iii), and (iv). The proof of Lemma 4.3 is thus completed. □

Lemma 4.4. *Let $K \in \mathbb{N}$, $h_0, h_1, \ldots, h_K, \mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \ldots < \mathfrak{r}_K$. Then*

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

$$\Psi = \mathbf{A}_{1,h_0} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{(h_{\min\{k+1,K\}} - h_k)}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \tfrac{(h_k - h_{\max\{k-1,0\}})}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right), \tag{4.25}$$

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

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

(iv) it holds for all $k \in \{0, 1, \ldots, K\}$ that $(\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_k) = h_k$, and

(v) it holds for all $k \in \{1, 2, \ldots, K\}$, $x \in \mathbb{R}$ that

$$(\mathcal{R}_{\mathfrak{r}}(\Psi))(x) = \begin{cases} h_0 & : x \in (-\infty, \mathfrak{r}_0] \\ h_{k-1} + \left(\frac{h_k - h_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(x - \mathfrak{r}_{k-1}) & : x \in (\mathfrak{r}_{k-1}, \mathfrak{r}_k] \\ h_K & : x \in (\mathfrak{r}_K, \infty) \end{cases} \tag{4.26}$$

(cf. Definitions 3.1, 3.27, 3.30, 3.33, and 3.37).

Proof of Lemma 4.4. Throughout this proof let $c_0, c_1, \ldots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that

$$c_k = \frac{(h_{\min\{k+1,K\}} - h_k)}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \frac{(h_k - h_{\max\{k-1,0\}})}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})}. \tag{4.27}$$

Observe that Lemma 4.3 assures for each $k \in \{0, 1, \ldots, K\}$ that

(I) there exists a unique $\Phi_k \in ((\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \times (\mathbb{R}^{1\times 1} \times \mathbb{R}^1)) \subseteq \mathbf{N}$ which satisfies that $\Phi_k = c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) = ((1, -\mathfrak{r}_k), (c_k, 0))$,

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

(III) it holds that $\mathcal{D}(\Phi_k) = (1, 1, 1) \in \mathbb{N}^3$,

(IV) it holds that $(\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = c_k[x - \mathfrak{r}_k]\mathbb{1}_{(\mathfrak{r}_k,\infty)}(x)$.

Item (I) and Lemmas 3.29 and 3.47 guarantee the existence of a unique $\Psi \in ((\mathbb{R}^{1\times 1}\times\mathbb{R}^{1})\times(\mathbb{R}^{1\times 1}\times\mathbb{R}^{1})) \subseteq \mathbf{N}$ such that

$$\Psi = \mathbf{A}_{1,h_0} \bullet \left(\bigoplus_{k=0}^{K} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k})) \right). \tag{4.28}$$

This establishes item (i). Items (II) and (III), Lemmas 3.29 and 3.47, and (4.28) assure that $\mathcal{R}_{\mathfrak{r}}(\Psi) \in C(\mathbb{R},\mathbb{R})$ and $\mathcal{D}(\Psi) = (1, K+1, 1) \in \mathbb{N}^3$. This establishes items (ii) and (iii). Next, observe that item (IV), (4.27), and Lemmas 3.29 and 3.47 assure for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\Psi))(x) = h_0 + \sum_{k=0}^{K} (\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = h_0 + \sum_{k=0}^{K} c_k \max\{x - \mathfrak{r}_k, 0\}. \tag{4.29}$$

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

$$(\mathcal{R}_{\mathfrak{r}}(\Psi))(x) = h_0 + 0 = h_0. \tag{4.30}$$

In addition, observe that the fact that $\forall\, k \in \{1, 2, \ldots, K\} \colon \mathfrak{r}_{k-1} < \mathfrak{r}_k$ and the fact that $\forall\, k \in \{1, 2, \ldots, K\} \colon \sum_{n=0}^{k-1} c_n = \frac{h_k - h_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}$ show that for all $k \in \{1, 2, \ldots, K\}$, $x \in (\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ it holds that

$$\begin{aligned} (\mathcal{R}_{\mathfrak{r}}(\Psi))(x) - (\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_k) &= \sum_{n=0}^{K} c_n \left(\max\{x - \mathfrak{r}_n, 0\} - \max\{\mathfrak{r}_k - \mathfrak{r}_n, 0\} \right) \\ &= \sum_{n=0}^{k} c_n [(x - \mathfrak{r}_n) - (\mathfrak{r}_k - \mathfrak{r}_n)] = \sum_{n=0}^{k} c_n (x - \mathfrak{r}_k) \\ &= \left(\tfrac{h_k - h_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(x - \mathfrak{r}_{k-1}). \end{aligned} \tag{4.31}$$

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

$$(\mathcal{R}_{\mathfrak{r}}(\Psi))(x) = h_{k-1} + \left(\tfrac{h_k - h_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(x - \mathfrak{r}_{k-1}). \tag{4.32}$$

We now prove (4.32) by induction on $k \in \{1, 2, \ldots, K\}$. For the base case $k = 1$ observe that (4.30) and (4.31) demonstrate that for all $x \in (\mathfrak{r}_0, \mathfrak{r}_1]$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\Psi))(x) = (\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_0) + (\mathcal{R}_{\mathfrak{r}}(\Psi))(x) - (\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_0) = h_0 + \left(\tfrac{h_1 - h_0}{\mathfrak{r}_1 - \mathfrak{r}_0}\right)(x - \mathfrak{r}_0). \tag{4.33}$$

This proves (4.32) in the base case $k = 1$. For the induction step note that (4.31) implies for all $k \in \{2, 3, \ldots, K\}$, $x \in (\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ with $\forall\, y \in (\mathfrak{r}_{k-2}, \mathfrak{r}_{k-1}] \colon (\mathcal{R}_{\mathfrak{r}}(\Psi))(y) = h_{k-2} + \left(\tfrac{h_{k-1} - h_{k-2}}{\mathfrak{r}_{k-1} - \mathfrak{r}_{k-2}}\right)(y - \mathfrak{r}_{k-2})$ it holds that

$$\begin{aligned} (\mathcal{R}_{\mathfrak{r}}(\Psi))(x) &= (\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_{k-1}) + (\mathcal{R}_{\mathfrak{r}}(\Psi))(x) - (\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_{k-1}) \\ &= h_{k-2} + \left(\tfrac{h_{k-1} - h_{k-2}}{\mathfrak{r}_{k-1} - \mathfrak{r}_{k-2}}\right)(\mathfrak{r}_{k-1} - \mathfrak{r}_{k-2}) + \left(\tfrac{h_k - h_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(t - \tau_{k-1}) = h_{k-1} + \left(\tfrac{h_k - h_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(x - \mathfrak{r}_{k-1}). \end{aligned} \tag{4.34}$$

Induction thus proves (4.32). Furthermore, observe that (4.32), the fact that $\forall\, k \in \{0, 1, \ldots, K\} \colon \mathfrak{r}_k \le \mathfrak{r}_K$, and the fact that $\sum_{n=0}^{K} c_k = 0$ imply that for all $x \in (\mathfrak{r}_K, \infty)$ it holds that

$$\begin{aligned} (\mathcal{R}_{\mathfrak{r}}(\Psi))(x) - (\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_K) &= \left[\sum_{n=0}^{K} c_k \left(\max\{x - \mathfrak{r}_n, 0\} - \max\{\mathfrak{r}_K - \mathfrak{r}_n, 0\} \right) \right] \\ &= \sum_{n=0}^{K} c_n [(x - \mathfrak{r}_n) - (\mathfrak{r}_K - \mathfrak{r}_n)] = \sum_{n=0}^{K} c_k (x - \mathfrak{r}_K) = 0. \end{aligned} \tag{4.35}$$

This and (4.32) show that for all $x \in (\mathfrak{r}_K, \infty)$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\Psi))(x) = (\mathcal{R}_{\mathfrak{r}}(\Psi))(\mathfrak{r}_K) = h_{K-1} + \left(\tfrac{h_K - h_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}}\right)(\mathfrak{r}_K - \mathfrak{r}_{K-1}) = h_K. \tag{4.36}$$

Combining this, (4.30), and (4.32) establishes item (v). Moreover, note that item (v) implies item (iv). The proof of Lemma 4.4 is thus completed. □

4.2.2 ANN approximations of one-dimensional functions

Definition 4.5 (Modulus of continuity)**.** *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*

$$w_f(h)=\sup\bigl(\bigl\{|f(x)-f(y)|\in[0,\infty)\colon \bigl(x,y\in\mathbb{R} \text{ with } |x-y|\leq h\bigr)\bigr\}\cup\{0\}\bigr) \tag{4.37}$$

and we call w_f the modulus of continuity of f.

Lemma 4.6. *Let $f\colon\mathbb{R}\to\mathbb{R}$ 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 $x,y\in\mathbb{R}$ that $|f(x)-f(y)|\leq w_f(|x-y|)$, and

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

(cf. Definition 4.5).

Proof of Lemma 4.6. First, observe that items (i), (ii), and (iv) are an immediate consequence of the definition of w_f. Next, note that item (iii) follows directly from the definition of a globally bounded set. Finally, by Definition 4.5 and the triangle inequality it holds for all $h,\mathfrak{h}\in[0,\infty]$ that

$$\begin{aligned}
w_f(h+\mathfrak{h}) &= \sup\bigl(\bigl\{|f(x)-f(y)|\in[0,\infty)\colon \bigl(x,y\in\mathbb{R}\text{ with }|x-y|\leq(h+\mathfrak{h})\bigr)\bigr\}\cup\{0\}\bigr)\\
&\leq \sup\Bigl(\Bigl\{\Bigl|f(x)-f(x-h\tfrac{x-y}{|x-y|})\Bigr|\in[0,\infty)\colon \bigl(x,y\in\mathbb{R}\text{ with }|x-y|\leq(h+\mathfrak{h})\bigr)\Bigr\}\cup\{0\}\Bigr)\\
&\quad+\sup\Bigl(\Bigl\{\Bigl|f(x-h\tfrac{x-y}{|x-y|})-f(y)\Bigr|\in[0,\infty)\colon \bigl(x,y\in\mathbb{R}\text{ with }|x-y|\leq(h+\mathfrak{h})\bigr)\Bigr\}\cup\{0\}\Bigr)\\
&\leq \sup\Bigl(\Bigl\{\Bigl|f(x)-f(x-h\tfrac{x-y}{|x-y|})\Bigr|\in[0,\infty)\colon \bigl(x,y\in\mathbb{R}\text{ with }|x-y|\leq h\bigr)\Bigr\}\cup\{0\}\Bigr)\\
&\quad+\sup\Bigl(\Bigl\{\Bigl|f(x-h\tfrac{x-y}{|x-y|})-f(y)\Bigr|\in[0,\infty)\colon \bigl(x,y\in\mathbb{R}\text{ with }|x-y|\leq \mathfrak{h}\bigr)\Bigr\}\cup\{0\}\Bigr)\\
&= w_f(h)+w_f(\mathfrak{h}).
\end{aligned} \tag{4.38}$$

This establishes item (v). The proof of Lemma 4.6 is thus completed. ☐

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

(i) there exists a unique $\mathbf{F}\in\mathbf{N}$ which satisfies

$$\mathbf{F}=\mathbf{A}_{1,f(\mathfrak{r}_0)}\bullet\left(\bigoplus_{k=0}^{K}\left(\left(\tfrac{(f(\mathfrak{r}_{\min\{k+1,K\}})-f(\mathfrak{r}_k))}{(\mathfrak{r}_{\min\{k+1,K\}}-\mathfrak{r}_{\min\{k,K-1\}})}-\tfrac{(f(\mathfrak{r}_k)-f(\mathfrak{r}_{\max\{k-1,0\}}))}{(\mathfrak{r}_{\max\{k,1\}}-\mathfrak{r}_{\max\{k-1,0\}})}\right)\circledast(\mathfrak{i}_1\bullet\mathbf{A}_{1,-\mathfrak{r}_k})\right)\right), \tag{4.39}$$

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

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

(iv) it holds for all $k\in\{0,1,\ldots,K\}$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{r}_k)=f(\mathfrak{r}_k)$,

(v) it holds for all $k\in\{1,2,\ldots,K\}$, $x\in\mathbb{R}$ that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x)=\begin{cases} f(\mathfrak{r}_0) & \colon x\in(-\infty,\mathfrak{r}_0]\\ f(\mathfrak{r}_{k-1})+\left(\frac{f(\mathfrak{r}_k)-f(\mathfrak{r}_{k-1})}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)(x-\mathfrak{r}_{k-1}) & \colon x\in(\mathfrak{r}_{k-1},\mathfrak{r}_k]\\ f(\mathfrak{r}_K) & \colon x\in(\mathfrak{r}_K,\infty)\end{cases}, \tag{4.40}$$

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

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \leq \left(\max_{k \in \{1,2,\ldots,K\}} \left(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \right) \right) |x - y|, \tag{4.41}$$

(vii) *it holds that* $\sup_{x \in [\mathfrak{x}_0, \mathfrak{x}_K]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq w_f(\max_{k \in \{1,2,\ldots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}|)$, *and*

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

(cf. Definitions 3.1, 3.4, 3.6, 3.8, 3.27, 3.30, 3.37, and 4.5).

Proof of Lemma 4.7. Throughout this proof let $c_0, c_1, \ldots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that

$$c_k = \frac{(f(\mathfrak{x}_{\min\{k+1,K\}}) - f(\mathfrak{x}_k))}{(\mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_{\min\{k,K-1\}})} - \frac{(f(\mathfrak{x}_k) - f(\mathfrak{x}_{\max\{k-1,0\}}))}{(\mathfrak{x}_{\max\{k,1\}} - \mathfrak{x}_{\max\{k-1,0\}})} \tag{4.42}$$

and let $L \in [0, \infty]$ satisfy that

$$L = \max_{k \in \{1,2,\ldots,K\}} \left(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \right) \tag{4.43}$$

(cf. Definition 4.5). Then Lemma 4.4 assures that

(I) there exists a unique $\mathbf{F} \in \mathbf{N}$ which satisfies

$$\mathbf{F} = \mathbf{A}_{1, f(\mathfrak{x}_0)} \bullet \left(\bigoplus_{k=0}^{K} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1, -\mathfrak{x}_k})) \right), \tag{4.44}$$

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

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

(IV) it holds for all $k \in \{0, 1, \ldots, K\}$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_k) = f(\mathfrak{x}_k)$, and

(V) it holds for all $x \in \mathbb{R}$, $k \in \{1, 2, \ldots, K\}$ that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = \begin{cases} f(\mathfrak{x}_0) & : x \in (-\infty, \mathfrak{x}_0] \\ f(\mathfrak{x}_{k-1}) + \left(\frac{f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} \right)(x - \mathfrak{x}_{k-1}) & : x \in (\mathfrak{x}_{k-1}, \mathfrak{x}_k] \\ f(\mathfrak{x}_K) & : x \in (\mathfrak{x}_K, \infty) \end{cases} \tag{4.45}$$

(cf. Definitions 3.1, 3.27, 3.30, 3.33, and 3.37). This establishes items (i), (ii), (iii), (iv), and (v). Next, observe that for all $k \in \{1, 2, \ldots, K\}$, $x, y \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| = \left| \left(\frac{f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} \right)(x - y) \right| \leq \left(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \right) |x - y| \tag{4.46}$$

(cf. Definition 4.5 and Lemma 4.6). This, item (iv), Definition 4.5, and Lemma 4.6 assure that for all $k, l \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$, $y \in [\mathfrak{x}_{l-1}, \mathfrak{x}_l]$ with $k < l$ it holds that

$$\begin{aligned}
&|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \\
&\leq |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_k)| + |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_k) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_{l-1})| + |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_{l-1}) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \\
&= |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_k)| + |f(\mathfrak{x}_k) - f(\mathfrak{x}_{l-1})| + |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_{l-1}) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \\
&\leq \left(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \right) |x - \mathfrak{x}_k| + w_f(|\mathfrak{x}_k - \mathfrak{x}_{l-1}|) + \left(\frac{w_f(|\mathfrak{x}_l - \mathfrak{x}_{l-1}|)}{|\mathfrak{x}_l - \mathfrak{x}_{l-1}|} \right) |\mathfrak{x}_{l-1} - y| \\
&\leq \left(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \right) |x - \mathfrak{x}_k| + \left(\sum_{j=k+1}^{l-1} w_f(|\mathfrak{x}_j - \mathfrak{x}_{j-1}|) \right) + \left(\frac{w_f(|\mathfrak{x}_l - \mathfrak{x}_{l-1}|)}{|\mathfrak{x}_l - \mathfrak{x}_{l-1}|} \right) |\mathfrak{x}_{l-1} - y| \\
&\leq L \left((\mathfrak{x}_k - x) + \left(\sum_{j=k+1}^{l-1} (\mathfrak{x}_j - \mathfrak{x}_{j-1}) \right) + (y - \mathfrak{x}_{l-1}) \right) = L|x - y|.
\end{aligned} \tag{4.47}$$

Combining this and (4.46) shows that for all $x, y \in [\mathfrak{r}_0, \mathfrak{r}_K]$ it holds that $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y)| \leq L|x - y|$. This, the fact that for all $x, y \in (-\infty, \mathfrak{r}_0]$ it holds that $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y)| = 0 \leq L|x - y|$, the fact that for all $x, y \in [\mathfrak{r}_K, \infty)$ it holds that $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y)| = 0 \leq L|x - y|$, and the triangle inequality hence demonstrate that for all $x, y \in \mathbb{R}$ it holds that $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y)| \leq L|x - y|$. This establishes item (vi). Moreover, note that Definition 4.5, Lemma 4.6, item (iv), and the fact that for all $k \in \{1, 2, \ldots, K\}$, $x \in \mathbb{R}$ that

$$\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} + \frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} = 1 \tag{4.48}$$

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

$$\begin{aligned}
|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)| &= \left|(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} + \frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) f(x)\right| \\
&\leq |f(\mathfrak{r}_{k-1}) - f(x)| \left|\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right| + |f(\mathfrak{r}_k) - f(x)| \left|\frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right| \\
&\leq w_f(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|) \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} + \frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) \\
&= w_f(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|) \leq w_f(\max_{j \in \{1,2,\ldots,K\}} |\mathfrak{r}_j - \mathfrak{r}_{j-1}|).
\end{aligned} \tag{4.49}$$

This establishes item (vii). By Definition 3.1 and item (ii) it follows that

$$\mathcal{P}(\Psi) = (K + 1)(1 + 1) + 1((K + 1) + 1) = 3K + 4. \tag{4.50}$$

This establishes item (viii). 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 $\mathfrak{r}_0 < \mathfrak{r}_1 < \ldots < \mathfrak{r}_K$ 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 $|f(x) - f(y)| \leq L|x - y|$. Then*

(i) there exists a unique $\mathbf{F} \in \mathbf{N}$ which satisfies

$$\mathbf{F} = \mathbf{A}_{1, f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{(f(\mathfrak{r}_{\min\{k+1,K\}}) - f(\mathfrak{r}_k))}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \tfrac{(f(\mathfrak{r}_k) - f(\mathfrak{r}_{\max\{k-1,0\}}))}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right), \tag{4.51}$$

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

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

(iv) it holds for all $k \in \{0, 1, \ldots, K\}$ that $(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{r}_k) = f(\mathfrak{r}_k)$,

(v) it holds for all $k \in \{1, 2, \ldots, K\}$, $x \in \mathbb{R}$ that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = \begin{cases} f(\mathfrak{r}_0) & : x \in (-\infty, \mathfrak{r}_0] \\ f(\mathfrak{r}_{k-1}) + \left(\frac{f(\mathfrak{r}_k) - f(\mathfrak{r}_{k-1})}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(x - \mathfrak{r}_{k-1}) & : x \in (\mathfrak{r}_{k-1}, \mathfrak{r}_k] \\ f(\mathfrak{r}_K) & : x \in (\mathfrak{r}_K, \infty) \end{cases}, \tag{4.52}$$

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

(vii) it holds that $\sup_{x \in [\mathfrak{r}_0, \mathfrak{r}_K]} |(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - f(x)| \leq L(\max_{k \in \{1,2,\ldots,K\}} |\mathfrak{r}_k - \mathfrak{r}_{k-1}|)$, and

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

(cf. Definitions 3.1, 3.4, 3.6, 3.8, 3.27, 3.30, and 3.37).

Proof of Lemma 4.8. Throughout this proof let $c_0, c_1, \dots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \dots, K\}$ that

$$c_k = \frac{(f(\mathfrak{x}_{\min\{k+1,K\}}) - f(\mathfrak{x}_k))}{(\mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_{\min\{k,K-1\}})} - \frac{(f(\mathfrak{x}_k) - f(\mathfrak{x}_{\max\{k-1,0\}}))}{(\mathfrak{x}_{\max\{k,1\}} - \mathfrak{x}_{\max\{k-1,0\}})}. \tag{4.53}$$

Then Lemma 4.7 assures that

(I) there exists a unique $\mathbf{F} \in \mathbf{N}$ which satisfies

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{x}_0)} \bullet \left(\overset{K}{\underset{k=0}{\oplus}} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{x}_k})) \right), \tag{4.54}$$

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

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

(IV) it holds for all $k \in \{0, 1, \dots, K\}$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{x}_k) = f(\mathfrak{x}_k)$,

(V) it holds for all $x \in \mathbb{R}$, $k \in \{1, 2, \dots, K\}$ that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = \begin{cases} f(\mathfrak{x}_0) & : x \in (-\infty, \mathfrak{x}_0] \\ f(\mathfrak{x}_{k-1}) + \left(\frac{f(\mathfrak{x}_k) - f(\mathfrak{x}_{k-1})}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\right)(x - \mathfrak{x}_{k-1}) & : x \in (\mathfrak{x}_{k-1}, \mathfrak{x}_k] \\ f(\mathfrak{x}_K) & : x \in (\mathfrak{x}_K, \infty) \end{cases}, \tag{4.55}$$

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

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \le \left(\max_{k \in \{1,2,\dots,K\}} \left(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \right) \right) |x - y|, \tag{4.56}$$

(VII) it holds that $\sup_{x \in [\mathfrak{x}_0, \mathfrak{x}_K]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \le w_f(\max_{k \in \{1,2,\dots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}|)$, and

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

(cf. Definitions 3.1, 3.27, 3.30, 3.33, and 3.37). This establishes items (i), (ii), (iii), (iv), (v), and (viii). Observe that the fact that for all $x, y \in [\mathfrak{x}_0, \mathfrak{x}_K]$ it holds that $|f(x) - f(y)| \le L|x - y|$ and Definition 4.5 imply for all $x, y \in [\mathfrak{x}_0, \mathfrak{x}_K]$ it holds that

$$\begin{aligned} w_f(|x - y|) &= \sup\bigl(\bigl\{|f(u) - f(v)| \in [0, \infty) \colon (u, v \in \mathbb{R} \text{ with } |u - v| \le |x - y|)\bigr\} \cup \{0\}\bigr) \\ &\le \sup\bigl(\bigl\{L|u - v| \in [0, \infty) \colon (u, v \in \mathbb{R} \text{ with } |u - v| \le |x - y|)\bigr\} \cup \{0\}\bigr) \le L|x - y|. \end{aligned} \tag{4.57}$$

This and item (IV) imply for all $x, y \in [\mathfrak{x}_0, \mathfrak{x}_K]$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \le \left(\max_{k \in \{1,2,\dots,K\}} \left(\frac{w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)}{|\mathfrak{x}_k - \mathfrak{x}_{k-1}|} \right) \right) |x - y| \le L|x - y|. \tag{4.58}$$

This, the fact that for all $x, y \in (-\infty, \mathfrak{x}_0]$ it holds that $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| = 0 \le L|x - y|$, the fact that for all $x, y \in [\mathfrak{x}_K, \infty)$ it holds that $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| = 0 \le L|x - y|$, and the triangle inequality hence demonstrate for all $x, y \in \mathbb{R}$ it holds that $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(y)| \le L|x - y|$. This establishes item (vi). Note that Lemma 4.6, (4.57), and item (VII) assure for all $k \in \{0, 1, \dots, K\}$, $x \in [\mathfrak{x}_0, \mathfrak{x}_K]$ it holds that

$$\begin{aligned} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| &\le w_f\left(\max_{k \in \{1,2,\dots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}| \right) \le \max_{k \in \{1,2,\dots,K\}} (w_f(|\mathfrak{x}_k - \mathfrak{x}_{k-1}|)) \\ &\le \max_{k \in \{1,2,\dots,K\}} (L|\mathfrak{x}_k - \mathfrak{x}_{k-1}|) = L\left(\max_{k \in \{1,2,\dots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}| \right). \end{aligned} \tag{4.59}$$

This establishes item (vii). The proof of Lemma 4.8 it thus completed. ☐

Corollary 4.9. *Let $K \in \mathbb{N}$, $L, a, \mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K \in \mathbb{R}$, $b \in (a, \infty)$ satisfy for all $k \in \{0, 1, \dots, K\}$ that $\mathfrak{r}_k = a + \frac{k(b-a)}{K}$ and let $f \colon [a, b] \to \mathbb{R}$ satisfy for all $x, y \in [a, b]$ that $|f(x) - f(y)| \le L|x - y|$. Then*

(i) there exists a unique $\mathbf{F} \in \mathbf{N}$ which satisfies

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right), \tag{4.60}$$

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

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

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

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

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

(cf. Definitions 3.1, 3.4, 3.6, 3.8, 3.27, 3.30, and 3.37).

Proof of Corollary 4.9. Note that for all $k \in \{0, 1, \dots, K\}$ it holds that $\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}} = \mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}} = (b - a)K^{-1}$ and

$$\frac{(f(\mathfrak{r}_{\min\{k+1,K\}}) - f(\mathfrak{r}_k))}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \frac{(f(\mathfrak{r}_k) - f(\mathfrak{r}_{\max\{k-1,0\}}))}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} = \frac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)}. \tag{4.61}$$

This and items (i), (ii), (iii), (vi), and (viii) of Lemma 4.8 prove items (i), (ii), (iii), (iv), and (vi). Moreover, note that item (vii) of Lemma 4.8 demonstrates that for all $x \in [a, b]$ it holds that

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

This establishes item (v). The proof of Corollary 4.9 is thus completed. □

Lemma 4.10. *Let $L, a \in \mathbb{R}$, $b \in (a, \infty)$, $\xi \in [a, b]$ and let $f \colon [a, b] \to \mathbb{R}$ satisfy for all $x, y \in [a, b]$ that $|f(x) - f(y)| \le L|x - y|$. Then*

(i) there exists a unique $\mathbf{F} \in \mathbf{N}$ which satisfies $\mathbf{F} = \mathbf{A}_{1,f(\xi)} \bullet (0 \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\xi}))$,

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

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

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

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

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

(cf. Definitions 3.1, 3.4, 3.6, 3.8, 3.27, and 3.30).

Proof of Lemma 4.10. Note that Definitions 3.8 and 3.27, items (i) and (ii) of Lemma 3.28, items (i) and (ii) of Lemma 3.29, and items (i), (ii), and (iii) of Lemma 4.3 establish items (i), (ii), and (iii). By Definitions 3.4, 3.6, 3.8, 3.27, and 3.30 it follows for all $x \in \mathbb{R}$ that

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

This establishes item (iv). 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 $|f(x) - f(y)| \leq L|x-y|$ assure that for all $x \in [a,b]$ it holds that

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

This establishes item (v). Moreover, note that Definition 3.1 and item (iii) assure that

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

This establishes item (vi). The proof of Lemma 4.10 it thus completed. ☐

Corollary 4.11. *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 $\mathfrak{r}_k = a + \frac{k(b-a)}{\max\{K,1\}}$ and let $f\colon [a,b] \to \mathbb{R}$ satisfy for all $x, y \in [a,b]$ that $|f(x) - f(y)| \leq L|x-y|$. Then*

(i) there exists a unique $\mathbf{F} \in \mathbf{N}$ which satisfies

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

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

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

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

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

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

(cf. Definitions 3.1, 3.4, 3.6, 3.8, 3.27, 3.30, and 3.37).

Proof of Corollary 4.11. Note that the fact that $K \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$ implies that $\frac{L(b-a)}{\max\{K,1\}} \leq \varepsilon$. This, items (i), (ii), (iii), (iv), and (v) of Corollary 4.9, and items (i), (ii), (iii), (iv), and (v) of Lemma 4.10 establish items (i), (ii), (iii), (iv), and (v). In addition, note that for all $k \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$ it holds that $k - 1 \leq \frac{L(b-a)}{\varepsilon}$. This and the fact that $K \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$ therefore ensure that $K \leq 1 + \frac{L(b-a)}{\varepsilon}$. Item (vi) of Corollary 4.9 and item (vi) of Lemma 4.10 hence assure that

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

This establishes item (vi). The proof of Corollary 4.11 is thus completed. ☐

Corollary 4.12. *Let $\varepsilon \in (0,\infty)$, $L, a \in \mathbb{R}$, $b \in (a,\infty)$ and let $f\colon [a,b] \to \mathbb{R}$ satisfy for all $x, y \in [a,b]$ that $|f(x) - f(y)| \leq L|x-y|$. 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 $\mathbb{D}_1(\mathbf{F}) \leq L(b-a)\varepsilon^{-1} + 2$,

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

(v) it holds that $\sup_{x\in[a,b]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq \varepsilon$, and

(vi) it holds that $\mathcal{P}(\mathbf{F}) = 3(\mathbb{D}_1(\mathbf{F})) + 1 \leq 3L(b-a)\varepsilon^{-1} + 7$

(cf. Definitions 3.1, 3.4, and 3.6).

Proof of Corollary 4.12. Throughout this proof let $K \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$, $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \dots, K\}$ that $\mathfrak{r}_k = a + \frac{k(b-a)}{\max\{K,1\}}$ and let $\mathbf{F} \in \mathbf{N}$ satisfy that

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}})-2f(\mathfrak{r}_k)+f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right) \tag{4.68}$$

(cf. Definitions 3.8, 3.27, 3.30, and 3.37). By items (i), (ii), (iv), (v), and (vi) of Corollary 4.11 the neural network $\mathbf{F}$ satisfies items (i), (ii), (iv), (v), and (vi). Next, note that for all $k \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$ it holds that $k - 1 \leq \frac{L(b-a)}{\varepsilon}$. This and the fact that $K \in \mathbb{N}_0 \cap [\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1)$ therefore ensures that $K \leq 1 + \frac{L(b-a)}{\varepsilon}$. Combining this with item (iii) of Corollary 4.11 establishes item (iii). The proof of Corollary 4.12 is thus completed. □

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

(i) there exists a unique $\mathbf{F} \in \mathbf{N}$ which satisfies

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}})-2f(\mathfrak{r}_k))+f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right), \tag{4.69}$$

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

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

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

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

(vi) it holds for all $x \in (-\infty, a) \cup (b, \infty)$ that $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq L(\min\{a - x, x - b\})$, and

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

(cf. Definitions 3.1, 3.4, 3.6, 3.8, 3.27, 3.30, 3.37, and 4.5).

Proof of Corollary 4.13. First, observe that items (i), (ii), (iii), (iv), (v), and (vi) of Corollary 4.11 establish items (i), (ii), (iii), (iv), (v), and (vii). Note that the triangle inequality, item (iv) of Lemma 4.8, and the fact that for all $x, y \in \mathbb{R}$ it holds that $|f(x) - f(y)| \leq L|x - y|$ imply that for all $x \in (-\infty, a)$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| = |f(a) - f(x)| \leq L|x - a| = L(a - x). \tag{4.70}$$

In addition, note that triangle inequality, item (iv) of Lemma 4.8, and the fact that for all $x, y \in \mathbb{R}$ it holds that $|f(x) - f(y)| \leq L|x - y|$ imply that for all $x \in (b, \infty)$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| = |f(b) - f(x)| \leq L|x - b| = L(x - b). \tag{4.71}$$

It then follows from (4.70) and (4.71) that for all $x \in (-\infty, a) \cup (b, \infty)$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq L(\min\{a - x, x - b\}). \tag{4.72}$$

This establishes item (vi). The proof of Corollary 4.13 is thus completed. □

Corollary 4.14. *Let $\varepsilon \in (0, \infty)$, $q \in (1, \infty)$, $L, a \in \mathbb{R}$, $b \in (a, \infty)$ and let $f \colon \mathbb{R} \to \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leq L|x - y|$. 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 $\mathbb{D}_1(\mathbf{F}) \leq (\frac{2L}{\varepsilon})^{q/(q-1)} + 2$,

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

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

(vi) it holds that $\mathcal{P}(\mathbf{F}) = 3(\mathbb{D}_1(\mathbf{F})) + 1 \leq \frac{3(2L)^{q/(q-1)}}{\varepsilon^{q/(q-1)}} + 7$

(cf. Definitions 3.1, 3.4, and 3.6).

Proof of Corollary 4.14. Throughout this proof let $K \in \mathbb{N}_0$, $R \in (0, \infty)$ satisfy that $2L = \varepsilon R^{q-1}$ and $K \in [\frac{2LR}{\varepsilon}, \frac{2LR}{\varepsilon} + 1)$, let $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \dots, K\}$ that $\mathfrak{r}_k = -R + \frac{2kR}{\max\{K,1\}}$, and let $\mathbf{F} \in \mathbf{N}$ satisfy that

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{2R} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right) \tag{4.73}$$

(cf. Definitions 3.8, 3.27, 3.30, and 3.37). Items (ii), (iii), and (iv) of Corollary 4.13 then establish items (i), (ii), and (iv). Next, note that for all $k \in \mathbb{N}_0 \cap [\frac{2LR}{\varepsilon}, \frac{2LR}{\varepsilon} + 1)$ it holds that $k - 1 \leq \frac{2LR}{\varepsilon}$. This and the fact that $K \in \mathbb{N}_0 \cap [\frac{2LR}{\varepsilon}, \frac{2LR}{\varepsilon} + 1)$ therefore ensures that $K \leq 1 + \frac{2LR}{\varepsilon}$. This, the fact that $2L = \varepsilon R^{q-1}$, and item (iii) then imply that

$$\mathbb{D}_1(\mathbf{F}) = K + 1 \leq 2 + \frac{2LR}{\varepsilon} = 2 + \frac{2L}{\varepsilon}\left(\frac{2L}{\varepsilon}\right)^{1/(q-1)} = 2 + \left(\frac{2L}{\varepsilon}\right)^{q/(q-1)}. \tag{4.74}$$

This establishes item (iii). By item (v) of Corollary 4.13 we have that it holds for all $x \in [-R, R]$ that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq \frac{L(b-a)}{\max\{K, 1\}} \leq \varepsilon \leq \varepsilon(\max\{1, |x|^q\}). \tag{4.75}$$

By item (vi) of Corollary 4.13 we have that it holds for all $x \in (-\infty, -R) \cup (R, \infty)$ that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq L(\min\{-R - x, x - R\}). \tag{4.76}$$

This and the fact that $2L = \varepsilon R^{q-1}$ imply that for all $x \in (-\infty, -R) \cup (R, \infty)$ it holds that

$$\begin{split} \frac{|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)|}{\max\{1, |x|^q\}} &\leq \frac{L(\min\{-R - x, x - R\})}{\max\{1, |x|^q\}} \leq \frac{L(|x| + R)}{\max\{1, |x|^q\}} \\ &\leq \frac{2L|x|}{\max\{1, |x|^q\}} \leq \frac{2L|x|}{|x|^q} \leq \frac{2L}{|x|^{q-1}} \leq \frac{2L}{R^{q-1}} = \varepsilon. \end{split} \tag{4.77}$$

Combining this with (4.75) then establishes item (v). Item (vi) of Corollary 4.9, the fact that $2L = \varepsilon R^{q-1}$, and the fact that $K \in \mathbb{N}_0 \cap [\frac{2LR}{\varepsilon}, \frac{2LR}{\varepsilon} + 1)$ imply that

$$\mathcal{P}(\mathbf{F}) \leq 3K + 4 \leq \frac{6LR}{\varepsilon} + 7 = \frac{3(2L)^{q/(q-1)}}{\varepsilon^{q/(q-1)}} + 7. \tag{4.78}$$

This establishes item (vi). The proof of Corollary 4.14 is thus completed. □

5 ANN Approximation Results

5.1 ANN approximations for products

A comment from Josh: I need this result...

Lemma 5.1. *Let* $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$*,* $(A_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{4\times 4}$*,* $\mathbb{A}, B \in \mathbb{R}^{4\times 1}$*,* $(C_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{1\times 4}$ *satisfy for all* $k \in \mathbb{N}$ *that*

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \qquad \mathbb{A} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \qquad C_k = \begin{pmatrix} -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \tag{5.1}$$

and $c_k = 2^{1-2k}$*. Then*

(i) there exist unique $\Phi_k \in \mathbf{N}$*,* $k \in \mathbb{N}$*, which satisfy for all* $k \in [2,\infty) \cap \mathbb{N}$ *that* $\Phi_1 = (\mathbf{A}_{C_1,0} \bullet \mathfrak{i}_4) \bullet \mathbf{A}_{\mathbb{A},\mathbb{B}}$ *and that*

$$\Phi_k = (\mathbf{A}_{C_k,0} \bullet \mathfrak{i}_4) \bullet (\mathbf{A}_{A_{k-1},B} \bullet \mathfrak{i}_4) \bullet (\mathbf{A}_{A_{k-2},B} \bullet \mathfrak{i}_4) \bullet \ldots \bullet (\mathbf{A}_{A_1,B} \bullet \mathfrak{i}_4) \bullet \mathbf{A}_{\mathbb{A},\mathbb{B}}, \tag{5.2}$$

(ii) it holds for all $k \in \mathbb{N}$ *that* $\mathcal{R}_{\mathfrak{r}}(\Phi_k) \in C(\mathbb{R},\mathbb{R})$*,*

(iii) it holds for all $k \in \mathbb{N}$ *that* $\mathcal{D}(\Phi_k) = (1,4,4,\ldots,4,1) \in \mathbb{N}^{k+2}$*,*

(iv) it holds for all $k \in \mathbb{N}$*,* $x \in \mathbb{R}\backslash[0,1]$ *that* $(\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = \mathfrak{r}(x)$*,*

(v) it holds for all $k \in \mathbb{N}$*,* $x \in [0,1]$ *that* $|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x)| \leq 2^{-2k-2}$*, and*

(vi) it holds for all $k \in \mathbb{N}$ *that* $\mathcal{P}(\Phi_k) = 20k - 7$

(cf. Definitions 3.1, 3.4, 3.6, 3.27, and 3.30).

A comment from Josh: I'm still updating this proof...

Proof of Lemma 5.1. Throughout this proof let $(\alpha_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$, $(\beta_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ satisfy for all $k \in \mathbb{N}$ that $\alpha_k = -c_k$ and $\beta_k = 2c_k$, let $g_k \colon \mathbb{R} \to [0,1]$, $k \in \mathbb{N}$, be the functions which satisfy for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ that

$$g_1(x) = \begin{cases} 2x & : x \in [0, \frac{1}{2}) \\ 2 - 2x & : x \in [\frac{1}{2}, 1] \\ 0 & : x \in \mathbb{R}\backslash[0,1] \end{cases} \tag{5.3}$$

and $g_{k+1}(x) = g_1(g_k(x))$, let $f_k \colon [0,1] \to [0,1]$, $k \in \mathbb{N}_0$, be the functions which satisfy for all $k \in \mathbb{N}_0$, $n \in \{0,1,\ldots,2^k-1\}$, $x \in \left[\frac{n}{2^k}, \frac{n+1}{2^k}\right)$ that $f_k(1) = 1$ and

$$f_k(x) = \left[\tfrac{2n+1}{2^k}\right] x - \tfrac{(n^2+n)}{2^{2k}}, \tag{5.4}$$

and let $r_k = (r_{k,1}, r_{k,2}, r_{k,3}, r_{k,4}) \colon \mathbb{R} \to \mathbb{R}^4$, $k \in \mathbb{N}$, be the functions which satisfy for all $x \in \mathbb{R}$, $k \in \mathbb{N}$ that

$$r_1(x) = (r_{1,1}(x), r_{1,2}(x), r_{1,3}(x), r_{1,4}(x)) = \mathfrak{M}_{\mathfrak{r},4}\big(x, x - \tfrac{1}{2}, x - 1, x\big) \tag{5.5}$$

and

$$r_{k+1}(x) = (r_{k+1,1}(x), r_{k+1,2}(x), r_{k+1,3}(x), r_{k+1,4}(x)) = \mathfrak{M}_{\mathfrak{r},4}\big(A_{k+1} r_k(x) + b_{k+1}\big). \tag{5.6}$$

Note that **A comment from Josh: Add stuff...**. This establishes items (i), (ii), and (iii). Note that (3.5), (5.3), and (5.5) and the fact that for all $x \in \mathbb{R}$ it holds that $\mathfrak{r}(x) = \max\{x, 0\}$ show that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} 2r_{1,1}(x) - 4r_{1,2}(x) + 2r_{1,3}(x) &= 2\mathfrak{r}(x) - 4\mathfrak{r}(x - \tfrac{1}{2}) + 2\mathfrak{r}(x-1) \\ &= 2\max\{x,0\} - 4\max\{x - \tfrac{1}{2}, 0\} + 2\max\{x-1,0\} = g_1(x). \end{aligned} \tag{5.7}$$

Furthermore, observe that (3.5) and (5.5), the fact that for all $x \in \mathbb{R}$ it holds that $\mathfrak{r}(x) = \max\{x, 0\}$, and the fact that for all $x \in [0,1]$ it holds that $f_0(x) = x = \max\{x, 0\}$ imply that for all $x \in \mathbb{R}$ it holds that

$$r_{1,4}(x) = \max\{x, 0\} = \begin{cases} f_0(x) & : x \in [0,1] \\ \max\{x, 0\} & : x \in \mathbb{R}\backslash[0,1] \end{cases}. \tag{5.8}$$

Next we claim that for all $k \in \mathbb{N}$ it holds that

$$\big(\forall\, x \in \mathbb{R}\colon\; 2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) = g_k(x)\big) \tag{5.9}$$

and

$$\left(\forall\, x \in \mathbb{R}\colon\; r_{k,4}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x, 0\} & : x \in \mathbb{R}\backslash[0,1] \end{cases}\right). \tag{5.10}$$

We now prove (5.9) and (5.10) by induction on $k \in \mathbb{N}$. Note that (5.7) and (5.8) prove (5.9) and (5.10) in the base case $k = 1$. For the induction step $\mathbb{N} \ni k \to k+1 \in \mathbb{N} \cap [2, \infty)$ assume that there exists $k \in \mathbb{N}$ such that for all $x \in \mathbb{R}$ it holds that

$$2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) = g_k(x) \tag{5.11}$$

$$\text{and} \qquad r_{k,4}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x, 0\} & : x \in \mathbb{R}\backslash[0,1] \end{cases}. \tag{5.12}$$

Observe that (3.5), (5.1), (5.6), (5.7), and (5.11) ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} g_{k+1}(x) &= g_1(g_k(x)) = g_1(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)) \\ &= 2\,\mathfrak{r}\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)\big) \\ &\quad - 4\,\mathfrak{r}\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) - \tfrac{1}{2}\big) \\ &\quad + 2\,\mathfrak{r}\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) - 1\big) \\ &= 2r_{k+1,1}(x) - 4r_{k+1,2}(x) + 2r_{k+1,3}(x). \end{aligned} \tag{5.13}$$

In addition, observe that (3.5), (5.1), (5.6), and (5.11) demonstrate that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} r_{k+1,4}(x) &= \mathfrak{r}\big((-2)^{3-2(k+1)} r_{k,1}(x) + 2^{4-2(k+1)} r_{k,2}(x) + (-2)^{3-2(k+1)} r_{k,3}(x) + r_{k,4}(x)\big) \\ &= \mathfrak{r}\big((-2)^{1-2k} r_{k,1}(x) + 2^{2-2k} r_{k,2}(x) + (-2)^{1-2k} r_{k,3}(x) + r_{k,4}(x)\big) \\ &= \mathfrak{r}\big(2^{-2k}\big[-2r_{k,1}(x) + 2^2 r_{k,2}(x) - 2r_{k,3}(x)\big] + r_{k,4}(x)\big) \\ &= \mathfrak{r}\big(-\big[2^{-2k}\big]\big[2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)\big] + r_{k,4}(x)\big) = \mathfrak{r}\big(-\big[2^{-2k}\big] g_k(x) + r_{k,4}(x)\big). \end{aligned} \tag{5.14}$$

Combining this with 5.12, **A comment from Josh: Add reference...**, the fact that for all $x \in \mathbb{R}$ it holds that $\mathfrak{r}(x) = \max\{x, 0\}$, and the fact that for all $x \in [0,1]$ it holds that $f_k(x) \geq 0$ shows that for all $x \in [0,1]$ it holds that

$$\begin{aligned} r_{k+1,4}(x) &= \mathfrak{r}\big(-\big[2^{-2k} g_k(x)\big] + f_{k-1}(x)\big) = \mathfrak{r}\bigg(-\big(2^{-2k} g_k(x)\big) + x - \bigg[\sum_{j=1}^{k-1} \big(2^{-2j} g_j(x)\big)\bigg]\bigg) \\ &= \mathfrak{r}\bigg(x - \bigg[\sum_{j=1}^{k} 2^{-2j} g_j(x)\bigg]\bigg) = \mathfrak{r}(f_k(x)) = f_k(x). \end{aligned} \tag{5.15}$$

Next note that (5.12) and (5.14), **A comment from Josh: Add reference...**, and the hypothesis that for all $x \in \mathbb{R}$ it holds that $a(x) = \max\{x, 0\}$ prove that for all $x \in \mathbb{R}\backslash[0,1]$ it holds that

$$r_{k+1,4}(x) = a\Big(-\big(2^{-2k} g_k(x)\big) + r_{k,4}(x)\Big) = a(\max\{x, 0\}) = \max\{x, 0\}. \tag{5.16}$$

Combining (5.13) and (5.15) hence proves (5.9) and (5.10) in the case $k+1$. Induction thus establishes (5.9) and (5.10). Next note that (3.6), (5.1), (5.5), (5.6), and (5.9) **A comment from Josh: Add definition references for the new network... A comment from Josh: Add lemma references...** assure that for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ it holds that $\mathcal{R}_{\mathfrak{r}}(\Phi_k) \in C(\mathbb{R}, \mathbb{R})$ and

$$\begin{aligned}
&(\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) \\
&= \big(\mathcal{R}_{\mathfrak{r}}\big((\mathbf{A}_{C_k,0} \bullet \mathfrak{i}_4) \bullet \big(\mathbf{A}_{A_{k-1},B} \bullet \mathfrak{i}_4\big) \bullet \big(\mathbf{A}_{A_{k-2},B} \bullet \mathfrak{i}_4\big) \bullet \cdots \bullet (\mathbf{A}_{A_1,B} \bullet \mathfrak{i}_4) \bullet \mathbf{A}_{\mathbb{A},\mathbb{B}}\big)\big)(x) \\
&= (-2)^{1-2k} r_{k,1}(x) + 2^{2-2k} r_{k,2}(x) + (-2)^{1-2k} r_{k,3}(x) + r_{k,4}(x) \\
&= (-2)^{2-2k} \Big(\Big[\tfrac{r_{k,1}(x)+r_{k,3}(x)}{(-2)}\Big] + r_{k,2}(x)\Big) + r_{k,4}(x) \\
&= 2^{2-2k} \Big(\Big[\tfrac{r_{k,1}(x)+r_{k,3}(x)}{(-2)}\Big] + r_{k,2}(x)\Big) + r_{k,4}(x) \\
&= 2^{-2k} \big(4 r_{k,2}(x) - 2 r_{k,1}(x) - 2 r_{k,3}(x)\big) + r_{k,4}(x) \\
&= -\big[2^{-2k}\big]\big[2 r_{k,1}(x) - 4 r_{k,2}(x) + 2 r_{k,3}(x)\big] + r_{k,4}(x) = -\big[2^{-2k}\big] g_k(x) + r_{k,4}(x).
\end{aligned} \tag{5.17}$$

Combining this with 5.10 and **A comment from Josh: Add reference...** shows that for all $k \in \mathbb{N}$, $x \in [0,1]$ it holds that

$$\begin{aligned}
(\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) &= -\big(2^{-2k} g_k(x)\big) + f_{k-1}(x) = -\big(2^{-2k} g_k(x)\big) + x - \Big[\sum_{j=1}^{k-1} 2^{-2j} g_j(x)\Big] \\
&= x - \Big[\sum_{j=1}^{k} 2^{-2j} g_j(x)\Big] = f_k(x).
\end{aligned} \tag{5.18}$$

A comment from Josh: Add reference... therefore implies that for all $k \in \mathbb{N}$, $x \in [0,1]$ it holds that

$$|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x)| \le 2^{-2k-2}. \tag{5.19}$$

This establishes item (v). Moreover, observe that **A comment from Josh: Add reference...**, (5.10) and (5.17) ensure that for all $k \in \mathbb{N}$, $x \in \mathbb{R} \backslash [0,1]$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = -2^{-2k} g_k(x) + r_{k,4}(x) = r_{k,4}(x) = \max\{x, 0\} = \mathfrak{r}(x). \tag{5.20}$$

This establishes item (iv). Note that item (iii) ensures for all $k \in \mathbb{N}$ that $\mathcal{L}(\Phi_k) = k+1$ and

$$\mathcal{P}(\Phi_k) = 4(1+1) + \left[\sum_{j=2}^{k} 4(4+1)\right] + (4+1) = 8 + 20(k-1) + 5 = 20k - 7. \tag{5.21}$$

This establishes item (vi). The proof of Lemma 5.1 is thus completed. □

A comment from Josh: I need this result...

Corollary 5.2. *Let $\varepsilon \in (0,\infty)$, $M = \min\big(\big[\frac{1}{2}\log_2(\varepsilon^{-1}) - 1, \infty\big) \cap \mathbb{N}\big)$, $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$, $(A_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{4\times 4}$, $\mathbb{A}, B \in \mathbb{R}^{4\times 1}$, $(C_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{1\times 4}$ satisfy for all $k \in \mathbb{N}$ that*

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \qquad \mathbb{A} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \qquad C_k = \begin{pmatrix} -c_k & 2c_k & -c_k & 1 \end{pmatrix}, \tag{5.22}$$

and $c_k = 2^{1-2k}$. Then

(i) there exists a unique $\Phi \in \mathbf{N}$ which satisfies that

$$\Phi = \begin{cases} (\mathbf{A}_{C_1,0} \bullet \mathfrak{i}_4) \bullet \mathbf{A}_{\mathbb{A},\mathbb{B}} & : M = 1 \\ (\mathbf{A}_{C_M,0} \bullet \mathfrak{i}_4) \bullet \big(\mathbf{A}_{A_{M-1},B} \bullet \mathfrak{i}_4\big) \bullet \big(\mathbf{A}_{A_{M-2},B} \bullet \mathfrak{i}_4\big) & \\ \qquad \bullet \ldots \bullet (\mathbf{A}_{A_1,B} \bullet \mathfrak{i}_4) \bullet \mathbf{A}_{\mathbb{A},\mathbb{B}} & : M \in [2,\infty) \cap \mathbb{N} \end{cases}, \tag{5.23}$$

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

(iii) *it holds that* $\mathcal{D}(\Phi) = (1, 4, 4, \ldots, 4, 1) \in \mathbb{N}^{M+2}$,

(iv) *it holds for all* $x \in \mathbb{R}\backslash[0,1]$ *that* $(\mathcal{R}_{\mathfrak{r}}(\Phi))(x) = \mathfrak{r}(x)$,

(v) *it holds for all* $x \in [0,1]$ *that* $|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)| \leq 2^{-2M-2} \leq \varepsilon$,

(vi) *it holds that* $\mathcal{L}(\Phi) = M + 1 \leq \max\{\frac{1}{2}\log_2(\varepsilon^{-1}) + 1, 2\}$, *and*

(vii) *it holds that* $\mathcal{P}(\Phi) = 20M - 7 \leq \max\{10\log_2(\varepsilon^{-1}) - 7, 13\}$

(cf. Definitions 3.1, 3.4, 3.6, 3.27, and 3.30).

Note: Double-check proof...

Proof of Corollary 5.2. Note that items (i), (ii), and (iv) of Lemma 5.1 establish items (i), (ii), and (iv). Next note the fact that $M = \min(\mathbb{N} \cap [\frac{1}{2}\log_2(\varepsilon^{-1}) - 1, \infty))$ assures that

$$M = \min\bigl(\mathbb{N} \cap \bigl[\tfrac{1}{2}\log_2(\varepsilon^{-1}) - 1, \infty\bigr)\bigr) \geq \min\bigl(\bigl[\max\bigl\{1, \tfrac{1}{2}\log_2(\varepsilon^{-1}) - 1\bigr\}, \infty\bigr)\bigr) \geq \tfrac{1}{2}\log_2(\varepsilon^{-1}) - 1. \tag{5.24}$$

This and item (v) of Lemma 5.1 demonstrate that for all $x \in [0,1]$ it holds that

$$|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)| \leq 2^{-2M-2} = 2^{-2(M+1)} \leq 2^{-\log_2(\varepsilon^{-1})} = \varepsilon. \tag{5.25}$$

This establishes item (v). Furthermore, the fact that $M = \min(\mathbb{N} \cap [\frac{1}{2}\log_2(\varepsilon^{-1}) - 1, \infty))$ and item (iii) of Lemma 5.1 assure that

$$\mathcal{L}(\Phi) = M + 1 \leq \max\{\tfrac{1}{2}\log_2(\varepsilon^{-1}) + 1, 2\}. \tag{5.26}$$

This establishes item (vi). This and item (vi) of Lemma 5.1 show that

$$\mathcal{P}(\Phi_M) \leq 20M - 7 \leq 20\max\{\tfrac{1}{2}\log_2(\varepsilon^{-1}), 2\} - 7 = \max\{10\log_2(\varepsilon^{-1}) - 7, 13\}. \tag{5.27}$$

This establishes item (vii). The proof of Corollary 5.2 is thus completed. □

A comment from Josh: I need this result...

Lemma 5.3. *Let* $\delta, \varepsilon \in (0, \infty)$, $\alpha \in (0, \infty)$, $q \in (2, \infty)$, $\Phi \in \mathbf{N}$ *satisfy that* $\delta = 2^{-2/(q-2)}\varepsilon^{q/(q-2)}$, $\alpha = (\varepsilon/2)^{1/(q-2)}$, $\Phi \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{L}(\Phi) \leq \max\{\frac{1}{2}\log_2(\delta^{-1}) + 1, 2\}$, $\mathcal{P}(\Phi) \leq \max\{10\log_2(\delta^{-1}) - 7, 13\}$, $\sup_{x \in \mathbb{R}\backslash[0,1]} |(\mathcal{R}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$, *and* $\sup_{x \in [0,1]} |x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)| \leq \delta$ *(cf. Definitions 3.1, 3.4, and 3.6). Then*

(i) *there exists a unique* $\Psi \in \mathbf{N}$ *which satisfies* $\Psi = (\mathbf{A}_{\alpha^{-2},0} \bullet \Phi \bullet \mathbf{A}_{\alpha,0}) \oplus (\mathbf{A}_{\alpha^{-2},0} \bullet \Phi \bullet \mathbf{A}_{-\alpha,0})$,

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

(iii) *it holds that* $(\mathcal{R}_{\mathfrak{r}}(\Psi))(0) = 0$,

(iv) *it holds for all* $x \in \mathbb{R}$ *that* $0 \leq (\mathcal{R}_{\mathfrak{r}}(\Psi))(x) \leq \varepsilon + |x|^2$,

(v) *it holds for all* $x \in \mathbb{R}$ *that* $|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Psi))(x)| \leq \varepsilon\max\{1, |x|^q\}$,

(vi) *it holds that* $\mathcal{L}(\Psi) \leq \max\bigl\{1 + \frac{1}{(q-2)} + \frac{q}{2(q-2)}\log_2(\varepsilon^{-1}), 2\bigr\}$, *and*

(vii) *it holds that* $\mathcal{P}(\Psi) \leq \max\bigl\{\bigl[\frac{40q}{(q-2)}\bigr]\log_2(\varepsilon^{-1}) + \frac{80}{(q-2)} - 28, 52\bigr\}$

(cf. Definitions 3.8, 3.19, 3.27, and 3.37).

Note: Double-check proof...

Proof of Lemma 5.3. **A comment from Josh: Add references...** This establishes items (i) and (ii). Next, note that **A comment from Josh: Add references...** ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_{\mathfrak{r}}(\Psi))(x) &= (\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha^{-2},0} \bullet \Phi \bullet \mathbf{A}_{\alpha,0}) \oplus (\mathbf{A}_{\alpha^{-2},0} \bullet \Phi \bullet \mathbf{A}_{-\alpha,0}))(x) \\ &= (\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha^{-2},0} \bullet \Phi \bullet \mathbf{A}_{\alpha,0}))(x) + (\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha^{-2},0} \bullet \Phi \bullet \mathbf{A}_{-\alpha,0}))(x) \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)} \left[(\mathcal{R}_{\mathfrak{r}}(\Phi))((\tfrac{\varepsilon}{2})^{1/(q-2)}x) + (\mathcal{R}_{\mathfrak{r}}(\Phi))(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x)\right].\end{aligned} \tag{5.28}$$

This, the assumption that $\Phi \in C(\mathbb{R}, \mathbb{R})$, and the assumption that $\sup_{x\in\mathbb{R}\backslash[0,1]} |(\mathcal{R}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ ensure that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\Psi))(0) = (\tfrac{\varepsilon}{2})^{-2/(q-2)} [(\mathcal{R}_{\mathfrak{r}}(\Phi))(0) + (\mathcal{R}_{\mathfrak{r}}(\Phi))(0)] = (\tfrac{\varepsilon}{2})^{-2/(q-2)} [\mathfrak{r}(0) + \mathfrak{r}(0)] = 0. \tag{5.29}$$

This establishes item (iii). Next, observe that the assumption that $\mathcal{R}_{\mathfrak{r}}(\Phi) \in C(\mathbb{R}, \mathbb{R})$ and the assumption that $\sup_{x\in\mathbb{R}\backslash[0,1]} |(\mathcal{R}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ ensure that for all $x \in \mathbb{R}\backslash[-1, 1]$ it holds that

$$\begin{aligned}[\mathcal{R}_{\mathfrak{r}}(\Phi)](x) + [\mathcal{R}_{\mathfrak{r}}(\Phi)](-x) = \mathfrak{r}(x) + \mathfrak{r}(-x) &= \max\{x, 0\} + \max\{-x, 0\} \\ &= \max\{x, 0\} - \min\{x, 0\} = |x|.\end{aligned} \tag{5.30}$$

The assumption that for all $\sup_{x\in\mathbb{R}\backslash[0,1]} |(\mathcal{R}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ and the assumption that $\sup_{x\in[0,1]} |x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)| \le \delta$ show that

$$\begin{aligned}&\sup_{x\in[-1,1]} \left|x^2 - \big([\mathcal{R}_{\mathfrak{r}}(\Phi)](x) + [\mathcal{R}_{\mathfrak{r}}(\Phi)](-x)\big)\right| \\ &= \max\Big\{ \sup_{x\in[-1,0]} \left|x^2 - \big(\mathfrak{r}(x) + [\mathcal{R}_{\mathfrak{r}}(\Phi)](-x)\big)\right|, \sup_{x\in[0,1]} \left|x^2 - \big([\mathcal{R}_{\mathfrak{r}}(\Phi)](x) + \mathfrak{r}(-x)\big)\right| \Big\} \\ &= \max\Big\{ \sup_{x\in[-1,0]} \left|(-x)^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(-x)\right|, \sup_{x\in[0,1]} \left|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)\right| \Big\} \\ &= \sup_{x\in[0,1]} \left|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)\right| \le \delta.\end{aligned} \tag{5.31}$$

Next observe that (5.28) and (5.30) prove that for all $x \in \mathbb{R}\backslash[-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$\begin{aligned}0 \le [\mathcal{R}_{\mathfrak{r}}(\Psi)](x) &= (\tfrac{\varepsilon}{2})^{-2/(q-2)} \Big([\mathcal{R}_{\mathfrak{r}}(\Phi)]\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) + [\mathcal{R}_{\mathfrak{r}}(\Phi)]\big(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x\big)\Big) \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)} \left|(\tfrac{\varepsilon}{2})^{1/(q-2)}x\right| = (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x| \le |x|^2.\end{aligned} \tag{5.32}$$

The triangle inequality therefore ensures that for all $x \in \mathbb{R}\backslash[-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$\begin{aligned}\left|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Psi))(x)\right| &= \left|x^2 - (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|\right| \le \big(|x|^2 + (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|\big) \\ &= \big(|x|^q|x|^{-(q-2)} + (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|^q|x|^{-(q-1)}\big) \\ &\le \big(|x|^q(\tfrac{\varepsilon}{2})^{(q-2)/(q-2)} + (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|^q(\tfrac{\varepsilon}{2})^{(q-1)/(q-2)}\big) \\ &= (\tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2})|x|^q = \varepsilon|x|^q \le \varepsilon \max\{1, |x|^q\}.\end{aligned} \tag{5.33}$$

Next note that (5.28) and (5.31), and the fact that $\delta = 2^{-2/(q-2)}\varepsilon^{q/(q-2)}$ demonstrate that for all $x \in [-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$\begin{aligned}&\left|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Psi))(x)\right| \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)} \left|\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x\big)^2 - \Big([\mathcal{R}_{\mathfrak{r}}(\Phi)]\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) + [\mathcal{R}_{\mathfrak{r}}(\Phi)]\big(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x\big)\Big)\right| \\ &\le (\tfrac{\varepsilon}{2})^{-2/(q-2)} \left[\sup_{y\in[-1,1]} \left|y^2 - \big([\mathcal{R}_{\mathfrak{r}}(\Phi)](y) + [\mathcal{R}_{\mathfrak{r}}(\Phi)](-y)\big)\right|\right] \\ &\le (\tfrac{\varepsilon}{2})^{-2/(q-2)}\delta = (\tfrac{\varepsilon}{2})^{-2/(q-2)}2^{-2/(q-2)}\varepsilon^{q/(q-2)} = \varepsilon \le \varepsilon \max\{1, |x|^q\}.\end{aligned} \tag{5.34}$$

Combining this and (5.33) implies that for all $x \in \mathbb{R}$ it holds that

$$\left|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Psi))(x)\right| \leq \varepsilon \max\left\{1, |x|^q\right\}. \tag{5.35}$$

This establishes item (v). In addition, note that (5.34) ensures that for all $x \in [-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\Psi))(x)| \leq \left|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Psi))(x)\right| + |x|^2 \leq \varepsilon + |x|^2. \tag{5.36}$$

This and (5.32) show for all $x \in \mathbb{R}$ that

$$|(\mathcal{R}_{\mathfrak{r}}(\Psi))(x)| \leq \varepsilon + |x|^2. \tag{5.37}$$

This establishes item (iv). Furthermore, observe that the fact that $\delta = 2^{-2/(q-2)}\varepsilon^{q/(q-2)}$ ensures that

$$\log_2(\delta^{-1}) = \log_2(2^{2/(q-2)}\varepsilon^{-q/(q-2)}) = \tfrac{2}{(q-2)} + \left[\left[\tfrac{q}{(q-2)}\right] \log_2(\varepsilon^{-1})\right]. \tag{5.38}$$

A comment from Josh: Finish proof... Note: I need parameter estimate results for the sum of neural networks... In addition observe that **A comment from Josh: Add references...** demonstrate that

$$\begin{aligned}\mathcal{L}(\Psi) &= \mathcal{L}(\mathbf{A}_{\alpha^{-2},0} \bullet \Phi \bullet \mathbf{A}_{\alpha,0}) = \mathcal{L}(\Phi) \leq \max\{\tfrac{1}{2}\log_2(\delta^{-1}) + 1, 2\} \\ &= \max\left\{1 + \tfrac{2}{(q-2)} + \left[\left[\tfrac{q}{(q-2)}\right] \log_2(\varepsilon^{-1})\right], 2\right\}.\end{aligned} \tag{5.39}$$

This establishes item (vi). The proof of Lemma 5.3 is thus completed. ☐

A comment from Josh: I need this...

Lemma 5.4. *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$, $\mathbb{A}_1, \mathbb{A}_2, \mathbb{A}_3 \in \mathbb{R}^{1\times 2}$, $\Phi \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}$ that $\delta = \varepsilon(2^{q-1} + 1)^{-1}$, $\mathbb{A}_1 = \begin{pmatrix}1 & 1\end{pmatrix}$, $\mathbb{A}_2 = \begin{pmatrix}1 & 0\end{pmatrix}$, $\mathbb{A}_3 = \begin{pmatrix}0 & 1\end{pmatrix}$, $\Phi \in C(\mathbb{R}, \mathbb{R})$, $(\mathcal{R}_{\mathfrak{r}}(\Phi))(0) = 0$, $0 \leq (\mathcal{R}_{\mathfrak{r}}(\Phi))(x) \leq \delta + |x|^2$, $|x^2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)| \leq \delta \max\{1, |x|^q\}$, $\mathcal{L}(\Phi) \leq \max\left\{1 + \frac{1}{(q-2)} + \frac{q}{2(q-2)}\log_2(\delta^{-1}), 2\right\}$, and $\mathcal{P}(\Phi) \leq \max\left\{\left[\frac{40q}{(q-2)}\right]\log_2(\delta^{-1}) + \frac{80}{(q-2)} - 28, 52\right\}$ (cf. Definitions 3.1, 3.4, and 3.6). Then*

(i) there exists a unique $\Gamma \in \mathbf{N}$ which satisfies that

$$\Gamma = \left(\tfrac{1}{2} \circledast (\Phi \bullet \mathbf{A}_{\mathbb{A}_1,0})\right) \oplus \left((-\tfrac{1}{2}) \circledast (\Phi \bullet \mathbf{A}_{\mathbb{A}_2,0})\right) \oplus \left((-\tfrac{1}{2}) \circledast (\Phi \bullet \mathbf{A}_{\mathbb{A}_3,0})\right) \tag{5.40}$$

(ii) it holds that $\mathcal{R}_{\mathfrak{r}}(\Gamma) \in C(\mathbb{R}^2, \mathbb{R})$,

(iii) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_{\mathfrak{r}}(\Gamma))(x, 0) = (\mathcal{R}_a(\Gamma))(0, x) = 0$,

(iv) it holds for all $x, y \in \mathbb{R}$ that $|xy - (\mathcal{R}_{\mathfrak{r}}(\Gamma))(x, y)| \leq \varepsilon \max\{1, |x|^q, |y|^q\}$,

(v) it holds that $\mathcal{P}(\Gamma) \leq \frac{360q}{(q-2)}[\log_2(\varepsilon^{-1}) + q + 1] - 252$, and

(vi) it holds that $\mathcal{L}(\Gamma) \leq \frac{q}{(q-2)}[\log_2(\varepsilon^{-1}) + q]$

(cf. Definitions 3.8, 3.27, 3.33, and 3.37).

Proof of Lemma 5.4. **A comment from Josh: Add proof...** ☐

A comment from Josh: I need this...

Lemma 5.5. *Let $\varepsilon \in (0, \infty)$, $q \in (2, \infty)$. Then there exists $\Gamma \in \mathbf{N}$ such that*

(i) it holds that $\mathcal{R}_{\mathfrak{r}}(\Gamma) \in C(\mathbb{R}^2, \mathbb{R})$,

(ii) it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_{\mathfrak{r}}(\Gamma))(x, 0) = (\mathcal{R}_a(\Gamma))(0, x) = 0$,

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

(iv) *it holds that* $\mathcal{P}(\Gamma) \leq \frac{360q}{(q-2)}[\log_2(\varepsilon^{-1}) + q + 1] - 252$, *and*

(v) *it holds that* $\mathcal{L}(\Gamma) \leq \frac{q}{(q-2)}[\log_2(\varepsilon^{-1}) + q]$

(cf. Definitions 3.1, 3.4, and 3.6).

Proof of Lemma 5.5. **A comment from Josh: Add proof...** □

5.2 Linear interpolation of MLP approximations

5.2.1 Properties of linear interpolations

A comment from Josh: Should I move these results to Section 4.2.2?

Definition 5.6 (Linear interpolation function)**.** *Let* $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K, f_0, f_1, \ldots, f_K \in \mathbb{R}$ *satisfy* $\mathfrak{r}_0 < \mathfrak{r}_1 < \ldots < \mathfrak{r}_K$. *Then we denote by* $\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K} \colon \mathbb{R} \to \mathbb{R}$ *the function which satisfies for all* $k \in \{1, 2, \ldots, K\}$, $r \in (-\infty, \mathfrak{r}_0)$, $s \in [\mathfrak{r}_K, \infty)$, $t \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k)$ *that* $(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(r) = f_0$, $(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(s) = f_K$, *and*

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(t) = f_{k-1} + \left(\tfrac{t - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(f_k - f_{k-1}). \tag{5.41}$$

A comment from Josh: I need this...

Lemma 5.7. *Let* $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K, f_0, f_1, \ldots, f_K \in \mathbb{R}$ *satisfy* $\mathfrak{r}_0 < \mathfrak{r}_1 < \ldots < \mathfrak{r}_K$. *Then*

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

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

$$\left|(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - (\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y)\right| = \left(\tfrac{|f_k - f_{k-1}|}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\right)|x - y|, \tag{5.42}$$

and

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

$$\left|(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - (\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y)\right| \leq \left[\max_{k \in \{1, 2, \ldots, K\}} \left(\tfrac{|f_k - f_{k-1}|}{|\tau_k - \tau_{k-1}|}\right)\right]|x - y| \tag{5.43}$$

(cf. Definition 5.6).

Note: This needs to be double-checked...

Proof of Lemma 5.7. Throughout this proof let $L \in \mathbb{R}$ satisfy that $L = \max_{k \in \{1, 2, \ldots, K\}}(\frac{|f_k - f_{k-1}|}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|})$. Note that Definition 5.6 implies item (i). Next observe that for all $k \in \{1, 2, \ldots, K\}$, $x, y \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ it holds that

$$\left|(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - (\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y)\right| = \left(\tfrac{|f_k - f_{k-1}|}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\right)|x - y|. \tag{5.44}$$

This establishes item (ii). The triangle inequality and item (ii) assure that for all $k, l \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$, $y \in [\mathfrak{r}_{l-1}, \mathfrak{r}_l]$ with $k < l$ it holds that

$$\begin{aligned}
&\left|(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - (\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y)\right| \\
&\leq \left|(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - f_k\right| + |f_k - f_{l-1}| + \left|f_{l-1} - (\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y)\right| \\
&= \left|(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(x) - (\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(\mathfrak{r}_k)\right| + |f_k - f_{l-1}| + \left|(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(\mathfrak{r}_{l-1}) - (\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K})(y)\right| \\
&\leq \left(\tfrac{|f_k - f_{k-1}|}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\right)|x - \mathfrak{r}_k| + \sum_{j=k+1}^{l-1} |f_j - f_{j-1}| + \left(\tfrac{|f_l - f_{l-1}|}{|\mathfrak{r}_l - \mathfrak{r}_{l-1}|}\right)|\mathfrak{r}_{l-1} - y| \\
&\leq L\left[|x - \mathfrak{r}_k| + \sum_{j=k+1}^{l-1} |\mathfrak{r}_j - \mathfrak{r}_{j-1}| + |\mathfrak{r}_{l-1} - y|\right] = L|x - y|.
\end{aligned} \tag{5.45}$$

Combining this and item (ii) shows that for all $x, y \in [\mathfrak{r}_0, \mathfrak{r}_K]$ it holds that

$$\big|(\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x) - (\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(y)\big| \leq L|x-y|. \tag{5.46}$$

This, the fact that for all $x, y \in (-\infty, \mathfrak{r}_0]$ it holds that $|(\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x) - (\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(y)| = 0 \leq L|x-y|$, the fact that for all $x, y \in [\mathfrak{r}_K, \infty)$ it holds that $|(\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x) - (\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(y)| = 0 \leq L|x-y|$, and the triangle inequality hence demonstrate for all $x, y \in \mathbb{R}$ it holds that $|(\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x) - (\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(y)| = 0 \leq L|x-y|$. This establishes item (iii). The proof of Lemma 5.7 is thus completed. □

5.2.2 Properties of linear interpolations employing a perturbed product function

A comment from Josh: Do we want a different name?

Definition 5.8 (Perturbed linear interpolation function)**.** *Let $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K, f_0, f_1, \dots, f_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \dots < \mathfrak{r}_K$ and let $\mathfrak{p}\colon \mathbb{R}^2 \to \mathbb{R}$ be a function. Then we denote by $\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K}\colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $k \in \{1, 2, \dots, K\}$, $r \in (-\infty, \mathfrak{r}_0)$, $s \in [\mathfrak{r}_K, \infty)$, $t \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k)$ that $(\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(r) = f_0$, $(\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(s) = f_K$, and*

$$(\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(t) = f_{k-1} + \mathfrak{p}\big(\tfrac{t-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}, f_k - f_{k-1}\big). \tag{5.47}$$

A comment from Josh: I need this...

Lemma 5.9. *Let $\varepsilon, q \in [0, \infty)$, $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K, f_0, f_1, \dots, f_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \dots < \mathfrak{r}_K$ and let $\mathfrak{p}\colon \mathbb{R}^2 \to \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|xy - \mathfrak{p}(x,y)| \leq \varepsilon \max\{1, |x|^q, |y|^q\}\mathbb{1}_{\mathbb{R}\setminus\{0\}}(xy)$. Then*

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

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

$$|(\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x) - (\mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x)| \leq \varepsilon \max\{1, |f_k - f_{k-1}|^q\}, \tag{5.48}$$

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

$$\big|(\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x) - (\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(y)\big| \leq \big(\tfrac{|f_k - f_{k-1}|}{|\mathfrak{r}_k - \mathfrak{r}_{k-1}|}\big)|x-y| + \varepsilon \max\{1, |f_k - f_{k-1}|^q\}, \tag{5.49}$$

and

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

$$|(\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(x) - f_k| \leq |f_k - f_{k-1}| + \varepsilon \max\{1, |f_k - f_{k-1}|^q\} \tag{5.50}$$

(cf. Definitions 5.6 and 5.8).

Proof of Lemma 5.9. **A comment from Josh: Add proof...** □

A comment from Josh: I need this...
Note: Should this be combined with the previous lemma?

Lemma 5.10. *Let $\varepsilon, q \in [0, \infty)$, $K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K, f_0, f_1, \dots, f_K, g_0, g_1, \dots, g_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \dots < \mathfrak{r}_K$ and let $\mathfrak{p}\colon \mathbb{R}^2 \to \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}$ that $|xy - \mathfrak{p}(x,y)| \leq \varepsilon \max\{1, |x|^q, |y|^q\}\mathbb{1}_{\mathbb{R}\setminus\{0\}}(xy)$. Then it holds for all $k \in \{1, 2, \dots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ that*

$$\big|(\mathscr{P}^{\mathfrak{p},f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(t) - (\mathscr{P}^{\mathfrak{p},g_0,g_1,\dots,g_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K})(t)\big| \leq \max_{j\in\{k-1,k\}} |f_j - g_j| + 2\varepsilon \max\bigg\{1, 2^q\bigg(\max_{j\in\{k-1,k\}} |f_j - g_j|^q\bigg)\bigg\} \tag{5.51}$$

(cf. Definition 5.8).

Proof of Lemma 5.10. **A comment from Josh: Add proof...** □

5.3 Linear interpolation of ANNs

5.3.1 ANN representation of the perturbed linear interpolation function

A comment from Josh: This is new...

Lemma 5.11. *Let $\varepsilon \in (0,\infty)$, $q \in (2,\infty)$, $d, K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K \in \mathbb{R}$, $f_0, f_1, \dots, f_K \in C(\mathbb{R}^d, \mathbb{R})$, $\mathbf{P}, \mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_K \in \mathbf{N}$ satisfy for all $k \in \{0, 1, \dots, K\}$, $x, y \in \mathbb{R}$ that $\mathfrak{r}_0 < \mathfrak{r}_1 < \dots < \mathfrak{r}_K$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{P}) \in C(\mathbb{R}^2, \mathbb{R})$, $|xy - (\mathcal{R}_{\mathfrak{r}}(\mathbf{P}))(x,y)| \leq \varepsilon \max\{1, |x|^q, |y|^q\} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(xy)$, and $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_k) = f_k$ (cf. Definitions 3.1, 3.4, and 3.6). Then*

(i) there exists a unique $\mathbf{G} \in \mathbf{N}$ which satisfies

$$\begin{aligned} \mathbf{G} = \mathbf{F}_0 \boxplus_{\mathbf{I}} \Bigg[\mathop{\boxplus}_{k=0,\mathbf{I}}^{K} \Bigg(\mathbf{P} \bullet \Bigg[\mathrm{P}_{2,(\mathbf{I},\mathbf{I})} \Big((\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}), \Big(\Big(\tfrac{1}{\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}}} \Big) \circledast \Big(\mathbf{F}_{\min\{k+1,K\}} \\ \boxplus_{\mathbf{I}} ((-1) \circledast \mathbf{F}_k) \Big) \boxplus_{\mathbf{I}} \Big(\Big(\tfrac{1}{\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}}} \Big) \circledast (\mathbf{F}_{\max\{k-1,0\}} \boxplus_{\mathbf{I}} ((-1) \circledast \mathbf{F}_k)) \Big) \Big) \Big) \Bigg] \Bigg) \Bigg], \end{aligned} \tag{5.52}$$

(ii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{G}) \in C(\mathbb{R}^{d+1}, \mathbb{R})$, and

(iii) it holds for all $t \in \mathbb{R}$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}))(t,x) = (\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}), f_0(x), f_1(x), \dots, f_K(x)}_{\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K})(t)$

(cf. Definitions 3.8, 3.24, 3.27, 3.33, 3.35, 3.37, and 5.8).

Proof of Lemma 5.11. **A comment from Josh: Add proof...** □

A comment from Josh: This is new...

Lemma 5.12. *Let $\varepsilon \in (0,\infty)$, $q \in (2,\infty)$, $d, K \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K \in \mathbb{R}$, $f_0, f_1, \dots, f_K \in C(\mathbb{R}^d, \mathbb{R})$, $\mathbf{P}, \mathbf{F}_0, \mathbf{F}_1, \dots, \mathbf{F}_K \in \mathbf{N}$ satisfy for all $k \in \{0, 1, \dots, K\}$, $x, y \in \mathbb{R}$ that $\mathfrak{r}_0 < \mathfrak{r}_1 < \dots < \mathfrak{r}_K$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{P}) \in C(\mathbb{R}^2, \mathbb{R})$, $|xy - (\mathcal{R}_{\mathfrak{r}}(\mathbf{P}))(x,y)| \leq \varepsilon \max\{1, |x|^q, |y|^q\} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(xy)$, and $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_k) = f_k$ (cf. Definitions 3.1, 3.4, and 3.6). Then*

(i) there exists a unique $\mathbf{G} \in \mathbf{N}$ which satisfies

$$\begin{aligned} \mathbf{G} = \mathbf{F}_0 \boxplus_{\mathbf{I}} \Bigg[\mathop{\boxplus}_{k=0,\mathbf{I}}^{K} \Bigg(\mathbf{P} \bullet \Bigg[\mathrm{P}_{2,(\mathbf{I},\mathbf{I})} \Big((\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}), \Big(\Big(\tfrac{1}{\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}}} \Big) \circledast \Big(\mathbf{F}_{\min\{k+1,K\}} \\ \boxplus_{\mathbf{I}} ((-1) \circledast \mathbf{F}_k) \Big) \boxplus_{\mathbf{I}} \Big(\Big(\tfrac{1}{\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}}} \Big) \circledast (\mathbf{F}_{\max\{k-1,0\}} \boxplus_{\mathbf{I}} ((-1) \circledast \mathbf{F}_k)) \Big) \Big) \Big) \Bigg] \Bigg) \Bigg], \end{aligned} \tag{5.53}$$

(ii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{G}) \in C(\mathbb{R}^{d+1}, \mathbb{R})$,

(iii) it holds for all $t \in \mathbb{R}$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}))(t,x) = (\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}), f_0(x), f_1(x), \dots, f_K(x)}_{\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K})(t)$,

(iv) it holds that $\mathcal{L}(\mathbf{G}) \leq \mathcal{L}(\mathbf{P}) + \max\{1, \mathcal{H}(\mathbf{F}_0), \mathcal{H}(\mathbf{F}_1), \dots, \mathcal{H}(\mathbf{F}_K)\}$, and

(v) it holds that $\mathcal{P}(\mathbf{G}) \leq$ **A comment from Josh: Add value...**,

(cf. Definitions 3.8, 3.24, 3.27, 3.33, 3.35, 3.37, and 5.8).

Proof of Lemma 5.12. **A comment from Josh: Add proof...** □

5.3.2 ANN representation of the perturbed linear interpolation of MLP approximations

A comment from Josh: This is new...

Lemma 5.13. *Let* $\Theta = \left(\bigcup_{n\in\mathbb{N}} \mathbb{Z}^n\right)$, $\varepsilon \in (0,\infty)$, $q \in (2,\infty)$, $d, K, M \in \mathbb{N}$, $\mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K, T \in \mathbb{R}$, $\mathbf{F}, \mathbf{G}, \mathbf{P} \in \mathbf{N}$ *satisfy for all* $x, y \in \mathbb{R}$ *that* $0 = \mathfrak{r}_0 < \mathfrak{r}_1 < \ldots < \mathfrak{r}_K = T$, $\mathcal{R}_\mathfrak{r}(\mathbf{F}) \in C(\mathbb{R},\mathbb{R})$, $\mathcal{R}_\mathfrak{r}(\mathbf{G}) \in C(\mathbb{R}^d,\mathbb{R})$, $\mathcal{R}_\mathfrak{r}(\mathbf{P}) \in C(\mathbb{R}^2,\mathbb{R})$, *and* $|xy - (\mathcal{R}_\mathfrak{r}(\mathbf{P}))(x,y)| \leq \varepsilon \max\{1, |x|^q, |y|^q\} \mathbb{1}_{\mathbb{R}\setminus\{0\}}(xy)$, *let* $\mathfrak{u}^\theta \in [0,1]$, $\theta \in \Theta$, *let* $\mathcal{U}^\theta \colon [0,T] \to [0,T]$, $\theta \in \Theta$, *satisfy for all* $t \in [0,T]$, $\theta \in \Theta$ *that* $\mathcal{U}^\theta_t = t + (T-t)\mathfrak{u}^\theta$, *let* $W^\theta \colon [0,T] \to \mathbb{R}^d$, $\theta \in \Theta$, *for every* $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$ *let* $Y^\theta_{t,s} \in \mathbb{R}^d$ *satisfy* $Y^\theta_{t,s} = W^\theta_s - W^\theta_t$, *and let* $U^\theta_n \colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, *satisfy for all* $n \in \mathbb{N}_0$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ *that*

$$\begin{aligned} U^\theta_n(t,x) = {} & \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n} \left[\sum_{k=1}^{M^n} (\mathcal{R}_a(\mathbf{G}))\Big(x + Y^{(\theta,0,-k)}_{t,T}\Big) \right] \\ & + \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}} \left[\sum_{k=1}^{M^{n-i}} \Big(\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,i,k)}_i\Big)\Big(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}\Big) \right] \\ & - \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}} \left[\sum_{k=1}^{M^{n-i}} \Big(\mathbb{1}_{\mathbb{N}}(i)\, (\mathcal{R}_a(\mathbf{F}) \circ U^{(\theta,-i,k)}_{\max\{i-1,0\}})\Big)\Big(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}\Big) \right] \end{aligned} \tag{5.54}$$

(cf. Definitions 3.1, 3.4, and 3.6). Then

(i) there exist unique $\mathbf{U}^\theta_{n,t} \in \mathbf{N}$, $t \in [0,T]$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, *which 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*

$$\begin{aligned} \mathbf{U}^\theta_{n,t} = {} & \left[\bigoplus_{k=1}^{M^n} \left(\frac{1}{M^n} \circledast \Big(\mathbf{G} \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,0,-k)}_{t,T}}\Big)\right)\right] \\ & \boxplus_{\mathbf{I}} \left[\mathop{\boxplus}_{i=0,\mathbf{I}}^{n-1} \left[\left(\frac{(T-t)}{M^{n-i}}\right) \circledast \left(\mathop{\boxplus}_{k=1,\mathbf{I}}^{M^{n-i}} \left(\Big(\mathbf{F} \bullet \mathbf{U}^{(\theta,i,k)}_{i,\mathcal{U}^{(\theta,i,k)}_t}\Big) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}} \right)\right)\right]\right] \\ & \boxplus_{\mathbf{I}} \left[\mathop{\boxplus}_{i=0,\mathbf{I}}^{n-1} \left[\left(\frac{(t-T)\,\mathbb{1}_{\mathbb{N}}(i)}{M^{n-i}}\right) \circledast \left(\mathop{\boxplus}_{k=1,\mathbf{I}}^{M^{n-i}} \left(\Big(\mathbf{F} \bullet \mathbf{U}^{(\theta,-i,k)}_{\max\{i-1,0\},\mathcal{U}^{(\theta,i,k)}_t}\Big) \bullet \mathbf{A}_{\mathrm{I}_d, Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t}} \right)\right)\right]\right], \end{aligned} \tag{5.55}$$

(ii) there exist unique $\mathbf{\Phi}^\theta_n \in \mathbf{N}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, *which satisfy for all* $\theta \in \Theta$, $n \in \mathbb{N}_0$ *that*

$$\begin{aligned} \mathbf{\Phi}^\theta_n = \mathbf{U}^\theta_{n,\mathfrak{r}_0} \boxplus_{\mathbf{I}} \Bigg[\mathop{\boxplus}_{k=0,\mathbf{I}}^{K} \Bigg(\mathbf{P} \bullet \Bigg[\mathrm{P}_{2,(\mathbf{I},\mathbf{I})} \Big((\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}), \Big(\Big(\tfrac{1}{\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}}}\Big) \circledast \Big(\mathbf{U}^\theta_{n,\mathfrak{r}_{\min\{k+1,K\}}} \\ \boxplus_{\mathbf{I}} ((-1) \circledast \mathbf{U}^\theta_{n,\mathfrak{r}_k})\Big) \boxplus_{\mathbf{I}} \Big(\Big(\tfrac{1}{\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}}}\Big) \circledast \Big(\mathbf{U}^\theta_{n,\mathfrak{r}_{\max\{k-1,0\}}} \boxplus_{\mathbf{I}} ((-1) \circledast \mathbf{U}^\theta_{n,\mathfrak{r}_k})\Big)\Big)\Big)\Bigg]\Bigg)\Bigg], \end{aligned} \tag{5.56}$$

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

$$(\mathcal{R}_\mathfrak{r}(\mathbf{\Phi}^\theta_n))(t,x) = \big(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}), U^\theta_n(\mathfrak{r}_0,x), U^\theta_n(\mathfrak{r}_1,x), \dots, U^\theta_n(\mathfrak{r}_K,x)}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K}\big)(t), \tag{5.57}$$

(iv) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ *that* $\mathcal{L}(\mathbf{\Phi}^\theta_n) \leq$ **A comment from Josh: Add value...**, *and*

(v) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ *that* $\mathcal{P}(\mathbf{\Phi}^\theta_n) \leq$ **A comment from Josh: Add value...**

(cf. Definitions 3.8, 3.13, 3.24, 3.27, 3.30, 3.33, 3.35, 3.48, and 5.8).

Proof of Lemma 5.13. **A comment from Josh: Add proof...** □

6 ANN approximations for PDEs

6.1 ANN approximations with specific polynomial convergence rates

A comment from Josh: Some of the details regarding the constants will be updated...
Note: I may need different assumptions on α and β...

Theorem 6.1. *Let $r, L, c, C, \alpha, \beta \in [0, \infty)$, $p, \mathfrak{q} \in [1, \infty)$, $q \in (2, \infty)$, $T \in (0, \infty)$, $f \in C(\mathbb{R}, \mathbb{R})$, for every $d \in \mathbb{N}$ let $\nu_d \colon \mathcal{B}(\mathbb{R}^{d+1}) \to [0, 1]$ be a probability measure on $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$, let $g_d \in C^1(\mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, let $\mathbf{G}_{d,\varepsilon} \in \mathbf{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0, \infty)$, and assume for all $d \in \mathbb{N}$, $v, w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$ that $|f(v) - f(w)| \leq L|v - w|$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}) \in C^1(\mathbb{R}^d, \mathbb{R})$, $\varepsilon(\|(\nabla g_d)(x)\| + |(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)|) + |g_d(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| \leq \varepsilon C d^p (1 + \|x\|)^{pq}$, $\mathcal{L}(\mathbf{G}_{d,\varepsilon}) \leq C d^p \varepsilon^{-\beta}$, $|||\mathcal{D}(\mathbf{G}_{d,\varepsilon})||| \leq C d^p \varepsilon^{-\alpha}$, and $(\int_{\mathbb{R}^{d+1}} \|y\|^{pq\mathfrak{q}} \, \nu_d(dy))^{1/(pq\mathfrak{q})} \leq C d^r$ (cf. Definitions 2.1, 3.1, 3.3, 3.4, and 3.6). Then*

(i) there exist unique $u_d \in C([0, T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, which satisfy for every $d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$, every probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and every standard Brownian motion $\mathbf{W} \colon [0, T] \times \Omega \to \mathbb{R}^d$ that $\sup_{s \in [0,T]} \sup_{y \in \mathbb{R}^d} (\frac{|u_d(s,y)|}{1 + \|y\|^{pq}}) < \infty$ and

$$u_d(t, x) = \mathbb{E}[g_d(x + \mathbf{W}_{T-t})] + \int_t^T \mathbb{E}[f(u_d(s, x + \mathbf{W}_{s-t}))] \, ds \tag{6.1}$$

and

(ii) there exist $(\mathbf{U}_{d,\varepsilon})_{(d,\varepsilon) \in (\mathbb{N} \times (0,1])} \in \mathbf{N}$ and $\eta, c \in (0, \infty)$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{U}_{d,\varepsilon}) \in C(\mathbb{R}^{d+1}, \mathbb{R})$, $\mathcal{P}(\mathbf{U}_{d,\varepsilon}) \leq c d^\eta \varepsilon^{-c}$, and

$$\left(\int_{[0,T] \times \mathbb{R}^d} |u_d(t, y) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{U}_{d,\varepsilon}))(t, y)|^{\mathfrak{q}} \, \nu_d(dt, dy) \right)^{1/\mathfrak{q}} \leq \varepsilon. \tag{6.2}$$

A comment from Josh: I will be more explicit on the bounds after the proof is complete...

Proof of Theorem 6.1. Throughout this proof let $M \in \mathbb{N}$ satisfy that $M = \inf\{m \in \mathbb{N} \colon {}^{(1+2LT)}/\sqrt{m} < 1\}$, let $\mathfrak{C}_d \in \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$ that $\mathfrak{C}_d = e^{LT}[C d^p (1 + \sqrt{d + 2}) + L(T + C d^p e^{LT})(T + 1)]$, let $\mathfrak{D} \in \mathbb{R}$ satisfy $\mathfrak{D} = e^{LT}(T + 1) C e^{M/2}$, let $\mathfrak{E}_d \in \mathbb{R}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$ that $\mathfrak{E}_d = C d^p (e^{LT}(T + 1))^{q+1}((C d^p)^q + 1)$ **A comment from Josh: I may need to fix this constant...**, and assume without loss of generality that

$$\max\{|f(0)| + 1, \textbf{A comment from Josh: Add value...}\} \leq C. \tag{6.3}$$

Note that the triangle inequality and the fact that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$ it holds that $\varepsilon |(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| + |g_d(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| \leq \varepsilon C d^p (1 + \|x\|)^{pq}$ imply for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$ that

$$|g_d(x)| \leq |g_d(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| + |(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| \leq \varepsilon C d^p (1 + \|x\|)^{pq} + C d^p (1 + \|x\|)^{pq}. \tag{6.4}$$

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

$$|g_d(x)| \leq (C + 1) d^p (1 + \|x\|)^{pq}. \tag{6.5}$$

A comment from Josh: Add reference..., the fact that for all $v, w \in \mathbb{R}$ it holds that $|f(v) - f(w)| \leq L|v - w|$, and (6.5) hence establish item (i). It thus remains to prove item (ii). To that end, note that Corollary 4.14 ensures that there exists $\mathbf{F}_\varepsilon \in \mathbf{N}$, $\varepsilon \in (0, 1]$, which satisfy for all $v, w \in \mathbb{R}$, $\varepsilon \in (0, 1]$ that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon) \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{H}(\mathbf{F}_\varepsilon) = 1$, $\mathbb{D}_1(\mathbf{F}_\varepsilon) \leq (\frac{2L}{\varepsilon})^{q/(q-1)} + 2$, $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(v) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(w)| \leq L|v - w|$, $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(v) - f(v)| \leq \varepsilon(\max\{1, |v|^q\})$, and

$$\mathcal{P}(\mathbf{F}_\varepsilon) \leq \frac{3(2L)^{q/(q-1)}}{\varepsilon^{q/(q-1)}} + 7. \tag{6.6}$$

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

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(0)| \leq |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(0) - f(0)| + |f(0)| \leq \varepsilon + |f(0)| \leq C. \tag{6.7}$$

A comment from Josh: Add reference..., the fact that for all $v, w \in \mathbb{R}$, $\varepsilon \in (0,1]$ it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon) \in C(\mathbb{R}, \mathbb{R})$, $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(v) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(w)| \leq L|v-w|$, and the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $x \in \mathbb{R}^d$ it holds that $|(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| \leq Cd^p(1+\|x\|)^{pq}$ ensure that there exist unique $\mathfrak{u}_{d,\varepsilon} \in C([0,T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, which satisfy for every $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $t \in [0,T]$, $x \in \mathbb{R}^d$, every probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and every standard Brownian motion $\mathbf{W}^d \colon [0,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, that $\sup_{s \in [0,T]} \sup_{y \in \mathbb{R}^d} \big(\frac{|\mathfrak{u}_{d,\varepsilon}(s,y)|}{1+\|y\|^{pq}}\big) < \infty$ and

$$\mathfrak{u}_{d,\varepsilon}(t,x) = \mathbb{E}[(\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x + \mathbf{W}_{T-t})] + \int_t^T \mathbb{E}[(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(\mathfrak{u}_{d,\varepsilon}(s, x + \mathbf{W}_{s,t}))]\, ds. \tag{6.8}$$

Next, 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 independent uniformly distributed 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 $\mathcal{U}^\theta_t = t + (T-t)\mathfrak{u}^\theta$, for every $d \in \mathbb{N}$ let $W^{\theta,d} \colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, be independent standard Brownian motions, assume for every $d \in \mathbb{N}$, $\theta \in \Theta$ that $\mathcal{U}^\theta$ and $W^{\theta,d}$ are independent, for every $d \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$ let $Y^{\theta,d}_{t,s} \colon \Omega \to \mathbb{R}^d$ satisfy $Y^{\theta,d}_{t,s} = W^{\theta,d}_s - W^{\theta,d}_t$, and let $U^\theta_{n,d,\delta} \colon [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $n \in \mathbb{N}_0$, $d \in \mathbb{N}$, $\delta \in (0,1]$, $\theta \in \Theta$, satisfy for all $n \in \mathbb{N}_0$, $d \in \mathbb{N}$, $\delta \in (0,1]$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} U^\theta_{n,d,\delta}(t,x) &= \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n}\left[\sum_{k=1}^{M^n} (\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\delta}))\Big(x + Y^{(\theta,0,-k),d}_{t,T}\Big)\right] \\ &\quad + \sum_{i=0}^{n-1} \frac{(T-t)}{M^{n-i}} \left[\sum_{k=1}^{M^{n-i}} \Big((\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\delta))\Big(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t,d}, U^{(\theta,i,k)}_{i,d,\delta}\Big(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k),d}_{t,\mathcal{U}^{(\theta,i,k)}_t}\Big)\Big) \right. \\ &\quad \left. - \mathbb{1}_{\mathbb{N}}(i)\, (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\delta))\Big(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k)}_{t,\mathcal{U}^{(\theta,i,k)}_t,d}, U^{(\theta,-i,k)}_{\max\{i-1,0\},d,\delta}\Big(\mathcal{U}^{(\theta,i,k)}_t, x + Y^{(\theta,i,k),d}_{t,\mathcal{U}^{(\theta,i,k)}_t}\Big)\Big)\Big)\right], \end{aligned} \tag{6.9}$$

let $\mathbf{P}_\gamma \in \mathbf{N}$, $\gamma \in (0,1]$, satisfy for all $\gamma \in (0,1]$, $v, w \in \mathbb{R}$ that $\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma) \in C(\mathbb{R}^2, \mathbb{R})$ and $|vw - (\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma))(v,w)| \leq \gamma \max\{1, |v|^q, |w|^q\}$, let $E_{d,\lambda} \colon \mathbb{R}^d \to [1,\infty)$, $d \in \mathbb{N}$, $\lambda \in [1,\infty)$, be the function which satisfies for all $d \in \mathbb{N}$, $\lambda \in [1,\infty)$, $x \in \mathbb{R}^d$ that

$$E_{d,\lambda}(x) = \sup_{s \in [0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}^d_s\|\big)^\lambda\Big], \tag{6.10}$$

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

$$c_d = \left[1 + \left(\int_{\mathbb{R}^{d+1}} \|y\|^{pq\mathfrak{q}} \nu_d(dy)\right)^{1/(pq\mathfrak{q})} + \big(\mathbb{E}\big[\|\mathbf{W}^d_T\|^{pq}\big]\big)^{1/(pq)}\right]^{pq}, \tag{6.11}$$

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

$$K_{d,\varepsilon} = \inf\left\{k \in \mathbb{N} \colon \frac{c_d \mathfrak{C}_d T^{1/2}}{k^{1/2}} \leq \frac{\varepsilon}{8}\right\}, \tag{6.12}$$

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

$$N_{d,\varepsilon} = \inf\left\{n \in \mathbb{N} \colon \left[\mathfrak{D}\left(\frac{1 + 2LT}{M^{1/2}}\right)^n\right] \leq \frac{\varepsilon}{4}\right\}, \tag{6.13}$$

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

$$B_{d,\varepsilon} = \max\{2, \|\!|\mathcal{D}(\mathbf{F}_\varepsilon)\|\!|, \|\!|\mathcal{D}(\mathbf{G}_{d,\varepsilon})\|\!|\}, \tag{6.14}$$

let $\mathfrak{r}_{K,k} \in [0,T]$, $K \in \mathbb{N}$, $k \in \{0,1,\dots,K\}$, satisfy for all $K \in \mathbb{N}$, $k \in \{0,1,\dots,K\}$ that $\mathfrak{r}_{K,k} = \frac{kT}{K}$, let **A comment from Josh: Add constant assumptions...** let $\delta_{d,\varepsilon} \in (0,1]$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that $\delta_{d,\varepsilon} = \frac{\varepsilon}{4c_d\mathfrak{C}_d}$, and let $\gamma_{d,\varepsilon} \in (0,1]$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that $\gamma_{d,\varepsilon} = \frac{\varepsilon}{10c_d}$. Observe that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $1 \leq E_{d,p}(x) \leq E_{d,pq}(x)$, $1 \leq E_{d,q}(x) \leq E_{d,pq}(x)$, and

$$\begin{aligned}
\left(\int_{[0,T]\times\mathbb{R}^d} [E_{d,pq}(x)]^{\mathfrak{q}}\,\nu_d(dt,dx)\right)^{1/\mathfrak{q}} &= \left(\int_{[0,T]\times\mathbb{R}^d} \left[\sup_{s\in[0,T]} \mathbb{E}\Big[\big(1+\|x+\mathbf{W}^d_s\|\big)^{pq}\Big]\right]^{\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\leq \left(\int_{[0,T]\times\mathbb{R}^d} \left[1+\|x\|+\big(\mathbb{E}\big[\|\mathbf{W}^d_T\|^{pq}\big]\big)^{1/(pq)}\right]^{pq\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\leq \left[1+\left(\int_{\mathbb{R}^{d+1}} \|y\|^{pq\mathfrak{q}}\nu_d(dy)\right)^{1/(pq\mathfrak{q})} + \big(\mathbb{E}\big[\|\mathbf{W}^d_T\|^{pq}\big]\big)^{1/(pq)}\right]^{pq} = c_d.
\end{aligned} \tag{6.15}$$

Further note that the fact that for all $d \in \mathbb{N}$ the random variable $\|\mathbf{W}^d_T/\sqrt{T}\|^{\mathfrak{q}}$ is a chi-squared distributed random variable with d degrees of freedom and Jensen's inequality imply that for all $d \in \mathbb{N}$ it holds that

$$\big(\mathbb{E}\big[\|\mathbf{W}^d_T\|^{pq}\big]\big)^2 \leq \mathbb{E}\Big[\|\mathbf{W}^d_T\|^{2pq}\Big] = (2T)^{pq}\left[\frac{\Gamma(\frac{d}{2}+pq)}{\Gamma(\frac{d}{2})}\right] = (2T)^{pq}\left[\prod_{k=0}^{pq-1}\left(\frac{d}{2}+k\right)\right]. \tag{6.16}$$

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

$$\big(\mathbb{E}\big[\|\mathbf{W}^d_T\|^{pq}\big]\big)^{1/(pq)} = \big(\mathbb{E}\big[\|\mathbf{W}^d_T\|^{pq}\big]\big)^{2/(2pq)} \leq \sqrt{2T}\left[\prod_{k=0}^{pq-1}\left(\frac{d}{2}+k\right)\right]^{1/(2pq)} \leq \sqrt{2T\left(\frac{d}{2}+pq-1\right)}. \tag{6.17}$$

This, together with the fact that for all $d \in \mathbb{N}$ it holds that $(\int_{\mathbb{R}^{d+1}} \|y\|^{pq\mathfrak{q}}\nu_d(dy))^{1/(pq\mathfrak{q})} \leq Cd^r$ implies that there exist $\overline{C} \in (0,\infty)$ such that for all $d \in \mathbb{N}$ it holds that

$$c_d \leq \overline{C}\left(\frac{1+d^r+\sqrt{d}}{3}\right)^{pq} \leq \overline{C}d^{\max\{rpq,rp/2\}}. \tag{6.18}$$

Note that the triangle inequality implies that for all $n \in \mathbb{N}_0$, $d, K \in \mathbb{N}$, $\delta, \gamma \in (0,1]$ it holds that

$$\begin{aligned}
&\left(\int_{[0,T]\times\mathbb{R}^d} \mathbb{E}\left[\left|u_d(t,x) - (\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),U^0_{n,d,\delta}(\mathfrak{r}_{K,0},x),U^0_{n,d,\delta}(\mathfrak{r}_{K,1},x),\ldots,U^0_{n,d,\delta}(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right|^{\mathfrak{q}}\right] \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&= \left(\sum_{k=1}^K \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} \mathbb{E}\left[\left|u_d(t,x) - (\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),U^0_{n,d,\delta}(\mathfrak{r}_{K,0},x),U^0_{n,d,\delta}(\mathfrak{r}_{K,1},x),\ldots,U^0_{n,d,\delta}(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right|^{\mathfrak{q}}\right] \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\leq \left(\sum_{k=1}^K \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} |u_d(t,x) - u_d(\mathfrak{r}_{K,k},x)|^{\mathfrak{q}} \, \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\quad + \left(\sum_{k=1}^K \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} \left|u_d(\mathfrak{r}_{K,k},x) - (\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),u_d(\mathfrak{r}_{K,0},x),u_d(\mathfrak{r}_{K,1},x),\ldots,u_d(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right|^{\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\quad + \left(\sum_{k=1}^K \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} \left|(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),u_d(\mathfrak{r}_{K,0},x),u_d(\mathfrak{r}_{K,1},x),\ldots,u_d(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right.\right. \\
&\qquad\qquad \left.\left. -(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right|^{\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\quad + \left(\sum_{k=1}^K \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} \mathbb{E}\left[\left|(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right.\right.\right. \\
&\qquad\qquad \left.\left.\left. -(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),U^0_{n,d,\delta}(\mathfrak{r}_{K,0},x),U^0_{n,d,\delta}(\mathfrak{r}_{K,1},x),\ldots,U^0_{n,d,\delta}(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right|^{\mathfrak{q}}\right] \nu_d(dt,dx)\right)^{1/\mathfrak{q}}.
\end{aligned} \tag{6.19}$$

The fact that for all $v, w \in \mathbb{R}$ it holds that $|f(v) - f(w)| \leq L|v - w|$, the fact that $f(0) \leq C \leq Cd^p(1 + \|x\|)^{pq}$, the fact that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $\|(\nabla g_d)(x)\| \leq Cd^p(1 + \|x\|)^{pq}$, and Corollary 2.7 (with $L = L$, $C = C$, $p = pq$, $f = f$, $g = g_d$, $\mathbf{W} = \mathbf{W}^d$, $\mathfrak{t} = t$, $t = \mathfrak{r}_{K,k-1}$ in the notation of Corollary 2.7) imply for all $d, K \in \mathbb{N}$, $\gamma \in (0,1]$, $k \in \{1, 2, \ldots, K\}$ it holds that

$$\begin{aligned}
&\int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} |u_d(t,x) - u_d(\mathfrak{r}_{K,k},x)|^{\mathfrak{q}} \, \nu_d(dt,dx) \\
&\leq \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} \left[\sqrt{t - \mathfrak{r}_{K,k-1}}\big(\mathfrak{C}_d E_{d,p}(x)\big)\right]^{\mathfrak{q}} \nu_d(dt,dx) \\
&\leq \left(\frac{T}{K}\right)^{\mathfrak{q}/2} (\mathfrak{C}_d)^{\mathfrak{q}} \left[\int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} [E_{d,pq}(x)]^{\mathfrak{q}} \, \nu_d(dt,dx)\right].
\end{aligned} \tag{6.20}$$

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

$$\begin{aligned}
&\left(\sum_{k=1}^K \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} |u_d(t,x) - u_d(\mathfrak{r}_{K,k},x)|^{\mathfrak{q}} \, \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\leq \mathfrak{C}_d \left(\frac{T}{K}\right)^{1/2} \left(\sum_{k=1}^K \left[\int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} [E_{d,pq}(x)]^{\mathfrak{q}} \, \nu_d(dt,dx)\right]\right)^{1/\mathfrak{q}} \\
&= \mathfrak{C}_d \left(\frac{T}{K}\right)^{1/2} \left(\int_{[0,T]\times\mathbb{R}^d} [E_{d,pq}(x)]^{\mathfrak{q}} \, \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \leq (\mathfrak{C}_d c_d) \left(\frac{T}{K}\right)^{1/2}.
\end{aligned} \tag{6.21}$$

(6.20), the triangle inequality, Jensen's inequality, Definition 5.8, Lemma 5.9 (with **A comment from Josh: Add stuff...** in the notation of Lemma 5.9), and Corollary 2.7 (with $L = L$, $C = C$, $p = pq$,

$f = f$, $g = g_d$, $\mathbf{W} = \mathbf{W}^d$, $\mathfrak{t} = \mathfrak{x}_{K,k}$, $t = \mathfrak{x}_{K,k-1}$ in the notation of Corollary 2.7) imply for all $d, K \in \mathbb{N}$, $\gamma \in (0,1]$, $k \in \{1,2,\ldots,K\}$ it holds that

$$\begin{aligned}
&\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} \left| u_d(\mathfrak{x}_{K,k},x) - (\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),u_d(\mathfrak{x}_{K,0},x),u_d(\mathfrak{x}_{K,1},x),\ldots,u_d(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\right|^{\mathfrak{q}} \nu_d(dt,dx) \\
&\leq \int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} \big||u_d(\mathfrak{x}_{K,k},x) - u_d(\mathfrak{x}_{K,k-1},x)| + \gamma\max\{1, |u_d(\mathfrak{x}_{K,k},x) - u_d(\mathfrak{x}_{K,k-1},x)|^q\}\big|^{\mathfrak{q}} \nu_d(dt,dx) \\
&\leq \int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} \left|\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} E_{d,p}(x) + \gamma\max\left\{1, \left(\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} E_{d,p}(x)\right)^q\right\}\right|^{\mathfrak{q}} \nu_d(dt,dx).
\end{aligned} \tag{6.22}$$

This ensures for all $d, K \in \mathbb{N}$, $\gamma \in (0,1]$, $k \in \{1,2,\ldots,K\}$ it holds that

$$\begin{aligned}
&\left(\sum_{k=1}^K \int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} \left| u_d(\mathfrak{x}_{K,k},x) - (\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),u_d(\mathfrak{x}_{K,0},x),u_d(\mathfrak{x}_{K,1},x),\ldots,u_d(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\right|^{\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\leq \left(\sum_{k=1}^K \int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} \left|\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} E_{d,p}(x) + \gamma\max\left\{1, \left(\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} E_{d,p}(x)\right)^q\right\}\right|^{\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&= \left(\int_{[0,T]\times\mathbb{R}^d} \left|\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} E_{d,p}(x) + \gamma\max\left\{1, \left(\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} E_{d,p}(x)\right)^q\right\}\right|^{\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\leq \left[\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} + \gamma\max\left\{1, (\mathfrak{C}_d)^q\left(\frac{T}{K}\right)^{q/2}\right\}\right]\left(\int_{[0,T]\times\mathbb{R}^d} [E_{d,pq}(x)]^{\mathfrak{q}} \nu_d(dt,dx)\right)^{1/\mathfrak{q}} \\
&\leq \left[\mathfrak{C}_d\left(\frac{T}{K}\right)^{1/2} + \gamma\max\left\{1, (\mathfrak{C}_d)^q\left(\frac{T}{K}\right)^{q/2}\right\}\right] c_d.
\end{aligned} \tag{6.23}$$

The fact that for all $d \in \mathbb{N}$, $v \in \mathbb{R}$, $x \in \mathbb{R}^d$ $\varepsilon \in (0,1]$ it holds that $|f(v) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon))(v)| \leq \varepsilon(\max\{1,|v|^q\})$ and $|g_d(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{G}_{d,\varepsilon}))(x)| \leq \varepsilon C d^p(1+\|x\|)^{pq}$ implies that for all $d \in \mathbb{N}$, $v \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
\max\{|f(v) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}_{d,\varepsilon}))(v)|, |g_d(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{G}_{d,\varepsilon}))(x)|\} &\leq \max\{\varepsilon(\max\{1,|v|^q\}, \varepsilon C d^p(1+\|x\|)^{pq}\} \\
&\leq \varepsilon C d^p((1+\|x\|)^{pq} + |v|^q).
\end{aligned} \tag{6.24}$$

Note that Lemma 5.10 (with $\varepsilon = \gamma$, $q = q$, $K = K$, $\mathfrak{p} = \mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma)$, and for all $k \in \{0,1,\ldots,K\}$ that $\mathfrak{x}_k = \mathfrak{x}_{K,k}$, $f_k = u_d(\mathfrak{x}_{K,k},x)$, $g_k = \mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,k},x)$ in the notation of Lemma 5.10), and the triangle inequality imply for all $d, K \in \mathbb{N}$, $\delta \in (0,1]$, $k \in \{1,2,\ldots,K\}$ it holds that

$$\begin{aligned}
&\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} \Big|(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),u_d(\mathfrak{x}_{K,0},x),u_d(\mathfrak{x}_{K,1},x),\ldots,u_d(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t) \\
&\qquad -(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\Big|^{\mathfrak{q}} \nu_d(dt,dx) \\
&\leq \int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} \Big|\max_{j\in\{k-1,k\}} |u_d(\mathfrak{x}_{K,j},x) - \mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,j},x)| \\
&\qquad +2\gamma\max\left\{1, 2^q\left(\max_{j\in\{k-1,k\}} |u_d(\mathfrak{x}_{K,j},x) - \mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,j},x)|\right)\right\}\Big|^{\mathfrak{q}} \nu_d(dt,dx)
\end{aligned} \tag{6.25}$$

This, (6.5), (6.7), the fact that for all $d \in \mathbb{N}$, $v, w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ it holds that $|f(v) - f(w)| \leq L|v-w|$, $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon))(v) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon))(w)| \leq L|v-w|$, $|f(0)| \leq C$, $\varepsilon|(\mathcal{R}_\mathfrak{r}(\mathbf{G}_{d,\varepsilon}))(x)| + |g_d(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{G}_{d,\varepsilon}))(x)| \leq \varepsilon C d^p(1+\|x\|)^{pq}$, and Corollary 2.4 (with $f_1 = f$, $f_2 = \mathcal{R}_\mathfrak{r}(\mathbf{F}_\delta)$, $g_1 = g_d$, $g_2 = \mathcal{R}_\mathfrak{r}(\mathbf{G}_{d,\delta})$, $L = L$, $B = \delta C d^p$,

$C = Cd^p$, $\mathbf{W} = \mathbf{W}^d$ in the notation of Corollary 2.4) imply for all $d, K \in \mathbb{N}$, $\delta \in (0,1]$, $k \in \{1,2,\ldots,K\}$ it holds that

$$\begin{aligned}\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d} &\Big|(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),u_d(\mathfrak{x}_{K,0},x),u_d(\mathfrak{x}_{K,1},x),\ldots,u_d(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\\ &\quad-(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\Big|^{\mathfrak{q}}\,\nu_d(dt,dx)\\ &\leq \int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d}\big|[\delta\mathfrak{E}_d E_{d,p}(x)] + 2\gamma\max\{1,[\delta\mathfrak{E}_d E_{d,p}(x)]^q\}\big|^{\mathfrak{q}}\,\nu_d(dt,dx)\end{aligned} \tag{6.26}$$

This, Jensen's inequality, and the triangle inequality imply for all $d, K \in \mathbb{N}$, $\delta \in (0,1]$ it holds that

$$\begin{aligned}&\left(\sum_{k=1}^K\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d}\Big|(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),u_d(\mathfrak{x}_{K,0},x),u_d(\mathfrak{x}_{K,1},x),\ldots,u_d(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\right.\\ &\qquad\left.-(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\Big|^{\mathfrak{q}}\,\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\ &\leq\left(\sum_{k=1}^K\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d}\big|[\delta\mathfrak{E}_d E_{d,p}(x)] + 2\gamma\max\{1,[\delta\mathfrak{E}_d E_{d,p}(x)]^q\}\big|^{\mathfrak{q}}\,\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\ &=\left(\int_{[0,T]\times\mathbb{R}^d}\big|[\delta\mathfrak{E}_d E_{d,p}(x)] + 2\gamma\max\{1,[\delta\mathfrak{E}_d E_{d,p}(x)]^q\}\big|^{\mathfrak{q}}\,\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\ &\leq[\delta\mathfrak{E}_d + 2\gamma\max\{2,(2\delta\mathfrak{E}_d)^q\}]\left(\int_{[0,T]\times\mathbb{R}^d}[E_{d,pq}(x)]^{\mathfrak{q}}\,\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\ &\leq[\delta\mathfrak{E}_d + 2\gamma\max\{2,(2\delta\mathfrak{E}_d)^q\}]c_d.\end{aligned} \tag{6.27}$$

Note that Lemma 5.10 (with $\varepsilon = \gamma$, $q = q$, $K = K$, $\mathfrak{p} = \mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma)$, and for all $k \in \{0,1,\ldots,K\}$ that $\mathfrak{x}_k = \mathfrak{x}_{K,k}$, $f_k = u_{d,\delta}(\mathfrak{x}_{K,k},x)$, $g_k = U^0_{n,d,\delta}(\mathfrak{x}_{K,k},x)$ in the notation of Lemma 5.10), and the triangle inequality imply for all $d, K \in \mathbb{N}$, $\delta, \gamma \in (0,1]$, $k \in \{1,2,\ldots,K\}$ it holds that

$$\begin{aligned}\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d}&\mathbb{E}\Big[\Big|(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\\ &\quad-(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),U^0_{n,d,\delta}(\mathfrak{x}_{K,0},x),U^0_{n,d,\delta}(\mathfrak{x}_{K,1},x),\ldots,U^0_{n,d,\delta}(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\Big|^{\mathfrak{q}}\Big]\,\nu_d(dt,dx)\\ \leq\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d}&\Big|\max_{j\in\{k-1,k\}}|\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,j},x)-U^0_{n,d,\delta}(\mathfrak{x}_{K,j},x)|\\ &\quad+2\gamma\max\Big\{1,2^q\Big(\max_{j\in\{k-1,k\}}|\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,j},x)-U^0_{n,d,\delta}(\mathfrak{x}_{K,j},x)|^q\Big)\Big\}\Big|^{\mathfrak{q}}\,\nu_d(dt,dx)\end{aligned} \tag{6.28}$$

This, the fact that for all $v, w \in \mathbb{R}$, $\delta \in (0,1]$ it holds that $|(\mathcal{R}_\mathfrak{r}(\mathbf{F}_\delta))(v) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}_\delta))(w)| \leq L|v-w|$, the fact that for all $d \in \mathbb{N}$, $\delta \in (0,1]$ it holds that $|(\mathcal{R}_\mathfrak{r}(\mathbf{G}_{d,\delta}))(x)| \leq Cd^p(1+\|x\|)^{pq}$, (6.8), and Lemma 2.8 (with $M = M$, $L = L$, $C = Cd^p$, $p = p$, $f = \mathcal{R}_\mathfrak{r}(\mathbf{F}_\delta)$, $g = \mathcal{R}_\mathfrak{r}(\mathbf{G}_{d,\delta})$, $u = \mathfrak{u}_{d,\delta}$, $U^\theta_n = U^0_{n,d,\delta}$ in the notation of Lemma 2.8) imply for all $d, K \in \mathbb{N}$, $\delta, \gamma \in (0,1]$, $k \in \{1,2,\ldots,K\}$ it holds that

$$\begin{aligned}\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d}&\mathbb{E}\Big[\Big|(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\\ &\quad-(\mathscr{P}^{\mathcal{R}_\mathfrak{r}(\mathbf{P}_\gamma),U^0_{n,d,\delta}(\mathfrak{x}_{K,0},x),U^0_{n,d,\delta}(\mathfrak{x}_{K,1},x),\ldots,U^0_{n,d,\delta}(\mathfrak{x}_{K,K},x)}_{\mathfrak{x}_{K,0},\mathfrak{x}_{K,1},\ldots,\mathfrak{x}_{K,K}})(t)\Big|^{\mathfrak{q}}\Big]\,\nu_d(dt,dx)\\ \leq\int_{[\mathfrak{x}_{K,k-1},\mathfrak{x}_{K,k}]\times\mathbb{R}^d}&\left|\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)E_{d,p}(x)+2\gamma\max\left\{1,\left[2\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)E_{d,p}(x)\right]^q\right\}\right|^{\mathfrak{q}}\nu_d(dt,dx).\end{aligned} \tag{6.29}$$

This, Jensen's inequality, and the triangle inequality imply for all $d, K \in \mathbb{N}$, $\delta, \gamma \in (0,1]$ it holds that

$$\begin{aligned}
&\left(\sum_{k=1}^{K} \int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d} \mathbb{E}\Big[\Big|(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,0},x),\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,1},x),\ldots,\mathfrak{u}_{d,\delta}(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\right.\\
&\qquad\left.-(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_\gamma),U^0_{n,d,\delta}(\mathfrak{r}_{K,0},x),U^0_{n,d,\delta}(\mathfrak{r}_{K,1},x),\ldots,U^0_{n,d,\delta}(\mathfrak{r}_{K,K},x)}_{\mathfrak{r}_{K,0},\mathfrak{r}_{K,1},\ldots,\mathfrak{r}_{K,K}})(t)\Big|^{\mathfrak{q}}\Big]\,\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\
&\le\left(\sum_{k=1}^{K}\int_{[\mathfrak{r}_{K,k-1},\mathfrak{r}_{K,k}]\times\mathbb{R}^d}\left|\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)E_{d,p}(x)\right.\right.\\
&\qquad\left.\left.+2\gamma\max\left\{1,\left[2\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)E_{d,p}(x)\right]^q\right\}\right|^{\mathfrak{q}}\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\
&=\left(\int_{[0,T]\times\mathbb{R}^d}\left|\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)E_{d,p}(x)+2\gamma\max\left\{1,\left[2\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)E_{d,p}(x)\right]^q\right\}\right|^{\mathfrak{q}}\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\
&\le\left[\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)+2\gamma\max\left\{2,(2\mathfrak{D})^q\left(\frac{(1+2LT)^n}{M^{n/2}}\right)^q\right\}\right]\left(\int_{[0,T]\times\mathbb{R}^d}[E_{d,pq}(x)]^{\mathfrak{q}}\,\nu_d(dt,dx)\right)^{1/\mathfrak{q}}\\
&\le\left[\mathfrak{D}\left(\frac{(1+2LT)^n}{M^{n/2}}\right)+2\gamma\max\left\{2,(2\mathfrak{D})^q\left(\frac{(1+2LT)^n}{M^{n/2}}\right)^q\right\}\right]c_d.
\end{aligned}\tag{6.30}$$

Combining (6.19) with (6.21), (6.23), (6.27), (6.30), and Fubini's theorem then imply that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\int_{[0,T]\times\mathbb{R}^d}\left|u_d(t,x)-(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_{\gamma_{d,\varepsilon}}),U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},0},x),U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},1},x),\ldots,U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},K_{d,\varepsilon}},x)}_{\mathfrak{r}_{K_{d,\varepsilon},0},\mathfrak{r}_{K_{d,\varepsilon},1},\ldots,\mathfrak{r}_{K_{d,\varepsilon},K_{d,\varepsilon}}})(t)\right|^{\mathfrak{q}}\nu_d(dt,dx)\right]\\
&=\int_{[0,T]\times\mathbb{R}^d}\mathbb{E}\Big[\Big|u_d(t,x)\\
&\quad-(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_{\gamma_{d,\varepsilon}}),U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},0},x),U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},1},x),\ldots,U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},K_{d,\varepsilon}},x)}_{\mathfrak{r}_{K_{d,\varepsilon},0},\mathfrak{r}_{K_{d,\varepsilon},1},\ldots,\mathfrak{r}_{K_{d,\varepsilon},K_{d,\varepsilon}}})(t)\Big|^{\mathfrak{q}}\Big]\,\nu_d(dt,dx)\\
&\le\left[2\mathfrak{C}_d\left(\frac{T}{K_{d,\varepsilon}}\right)^{1/2}+\delta\mathfrak{E}_d+2\gamma_{d,\varepsilon}\max\{2,(2\delta\mathfrak{E}_d)^q\}+\gamma_{d,\varepsilon}\max\left\{1,(\mathfrak{C}_d)^q\left(\frac{T}{K_{d,\varepsilon}}\right)^{q/2}\right\}\right.\\
&\quad\left.+\mathfrak{D}\left(\frac{(1+2LT)^{N_{d,\varepsilon}}}{M^{N_{d,\varepsilon}/2}}\right)+2\gamma_{d,\varepsilon}\max\left\{2,(2\mathfrak{D})^q\left(\frac{(1+2LT)^{N_{d,\varepsilon}}}{M^{N_{d,\varepsilon}/2}}\right)^q\right\}\right]^{\mathfrak{q}}(c_d)^{\mathfrak{q}}\\
&\le\left[2\mathfrak{C}_d\left(\frac{T}{K_{d,\varepsilon}}\right)^{1/2}+\delta\mathfrak{E}_d+9\gamma_{d,\varepsilon}+2\gamma_{d,\varepsilon}(2\delta\mathfrak{E}_d)^q+\gamma_{d,\varepsilon}(\mathfrak{C}_d)^q\left(\frac{T}{K_{d,\varepsilon}}\right)^{q/2}\right.\\
&\quad\left.+\mathfrak{D}\left(\frac{(1+2LT)^{N_{d,\varepsilon}}}{M^{N_{d,\varepsilon}/2}}\right)+2\gamma_{d,\varepsilon}(2\mathfrak{D})^q\left(\frac{(1+2LT)^{N_{d,\varepsilon}}}{M^{N_{d,\varepsilon}/2}}\right)^q\right]^{\mathfrak{q}}(c_d)^{\mathfrak{q}}\\
&\le\left[\frac{\varepsilon}{4}+\frac{\varepsilon}{4}+9\gamma_{d,\varepsilon}+2\gamma_{d,\varepsilon}\left(\frac{\varepsilon}{2}\right)^q+\gamma_{d,\varepsilon}\left(\frac{\varepsilon}{2}\right)^q+\frac{\varepsilon}{4}+2\gamma_{d,\varepsilon}\left(\frac{\varepsilon}{2}\right)^q\right]^{\mathfrak{q}}\le\left[\frac{3\varepsilon}{4}+10\gamma_{d,\varepsilon}\right]^{\mathfrak{q}}=\varepsilon^{\mathfrak{q}}.
\end{aligned}\tag{6.31}$$

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

$$\begin{aligned}
&\int_{[0,T]\times\mathbb{R}^d}\left|u_d(t,x)-(\mathscr{P}^{\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_{\gamma_{d,\varepsilon}}),U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},0},x),U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},1},x),\ldots,U^0_{N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(\mathfrak{r}_{K_{d,\varepsilon},K_{d,\varepsilon}},x)}_{\mathfrak{r}_{K_{d,\varepsilon},0},\mathfrak{r}_{K_{d,\varepsilon},1},\ldots,\mathfrak{r}_{K_{d,\varepsilon},K_{d,\varepsilon}}})(t)\right|^{\mathfrak{q}}\nu_d(dt,dx)\\
&\le\varepsilon^{\mathfrak{q}}.
\end{aligned}\tag{6.32}$$

A comment from Josh: The rest of the proof will come after I finish the ANN interpolation details...

The proof of Theorem 6.1 is thus completed. □

6.2 ANN approximations with general polynomial convergence rates

A comment from Josh: This needs to be updated...

Corollary 6.2. *Let $T, \kappa \in (0,\infty)$, $\mathfrak{q} \in [1,\infty)$, $f \in C(\mathbb{R},\mathbb{R})$, 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 $|f(v) - f(w)| \leq \kappa|v - w|$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, $\varepsilon(\|(\nabla_x u_d)(T,x)\| + |u_d(t,x)|) + |u_d(T,x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{G}_{d,\varepsilon}))(x)| \leq \varepsilon\kappa d^\kappa(1 + \|x\|^\kappa)$, $\mathcal{P}(\mathbf{G}_{d,\varepsilon}) \leq \kappa d^\kappa \varepsilon^{-\kappa}$, and*

$$(\tfrac{\partial}{\partial t} u_d)(t,x) + \tfrac{1}{2}(\Delta_x u_d)(t,x) + f_d(u_d(t,x)) = 0 \tag{6.33}$$

(cf. Definitions 2.1, 3.1, 3.3, 3.4, and 3.6). Then there exist $c \in (0,\infty)$, $\mathbf{U}_{d,\varepsilon} \in \mathbf{N}$, $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, which satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that $\mathcal{R}_{\mathfrak{r}}(\mathbf{U}_{d,\varepsilon}) \in C(\mathbb{R}^{d+1}, \mathbb{R})$, $\mathcal{P}(\mathbf{U}_{d,\varepsilon}) \leq c d^c \varepsilon^{-c}$, and

$$\left[\int_{[0,T]\times[0,1]^d} |u_d(y) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{U}_{d,\varepsilon}))(y)|^{\mathfrak{q}}\, dy\right]^{1/\mathfrak{q}} \leq \varepsilon. \tag{6.34}$$

Proof of Corollary 6.2. **A comment from Josh: Add proof....** □

Acknowledgments

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

References

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

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

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

[4] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *J. Mach. Learn. Res. 20*, 74 (2019), 1–25.

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

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

[7] 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. *Minor revision requested from SIAM J. Math. Data Sci., arXiv:1809.03062* (2018), 35 pages.

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

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

[10] 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.

[11] 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.

[12] Elbrächter, D., Grohs, P., Jentzen, A., and Schwab, C. DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing. *arXiv:1809.07669* (2018), 50 pages.

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

[14] 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).

[15] 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.

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

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

[18] 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.

[19] 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.

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

[21] Han, J., Jentzen, A., and E, W. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences 115*, 34 (2018), 8505–8510.

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

[23] 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.

[24] 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.

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

[26] 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).

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

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

[29] 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.

[30] 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.

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

[32] 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.

[33] 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. *arXiv:2005.05409* (2020), 40 pages.

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

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

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

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