Mathematical Introduction to Deep Learning

Arnulf Jentzen, Benno Kuckuck,
and Philippe von Wurstemberger

May 6, 2022

Preface

These lecture notes are far away from being complete and remain under construction. In particular, these lecture notes do not yet contain a suitable comparison of the presented material with existing results, arguments, and notions in the literature. The presented material consists of slightly modified extracts from joint works with Sebastian Becker (Zenai AG), Christian Beck (ETH Zurich), Philipp Grohs (University of Vienna), Fabian Hornung (ETH Zurich and Karlsruhe Institute of Technology), Martin Hutzenthaler (University of Duisburg-Essen), Nor Jaafari (Zenai AG), Benno Kuckuck (University of Münster), Thomas Kruse (University of Giessen), Tuan Nguyen (University of Duisburg-Essen), Philippe von Wurstemberger (ETH Zurich), and Philipp Zimmermann (ETH Zurich).

Shenzhen, January 2022

Arnulf Jentzen

Contents

Chapter 1

Introduction

1.1 Introductory comments on deep supervised learning

Very roughly speaking, the field *deep learning* can be divided into three subfields, deep *supervised learning*, deep *unsupervised learning*, and deep *reinforcement learning*. Algorithms in deep supervised learning seem often to be most accessible for a mathematical analysis. In the following we briefly sketch in a special situation some ideas of deep supervised learning.

Let $d, M \in \mathbb{N} = \{1, 2, 3, \dots\}$, $\mathcal{E} \in C(\mathbb{R}^d, \mathbb{R})$, $x_1, x_2, \dots, x_{M+1} \in \mathbb{R}^d$, $y_1, y_2, \dots, y_M \in \mathbb{R}$ satisfy for all $m \in \{1, 2, \dots, M\}$ that

$$y_m = \mathcal{E}(x_m). \tag{1.1}$$

In the framework described in the previous sentence we think of $M \in \mathbb{N}$ as the number of available input-output data pairs, we think of $d \in \mathbb{N}$ as the dimension of the input data, we think of $\mathcal{E} \colon \mathbb{R}^d \to \mathbb{R}$ as an unknown function which relates input and output data through (1.1), we think of $x_1, x_2, \dots, x_{M+1} \in \mathbb{R}^d$ as the available known input data, and we think of $y_1, y_2, \dots, y_M \in \mathbb{R}$ as the available known output data. The key question in the context of supervised learning is then that one intends to approximately compute the output $\mathcal{E}(x_{M+1})$ of the $(M+1)$-th input data x_{M+1} without using explicit knowledge of the function $\mathcal{E} \colon \mathbb{R}^d \to \mathbb{R}$ but instead by using the knowledge of the M input-output data pairs $(x_1, y_1) = (x_1, \mathcal{E}(x_1)), (x_2, y_2) = (x_2, \mathcal{E}(x_2)), \dots, (x_M, y_M) = (x_M, \mathcal{E}(x_M)) \in \mathbb{R}^d \times \mathbb{R}$. To accomplish this, one considers the optimization problem of approximately computing global minima of the function $\Phi \colon C(\mathbb{R}^d, \mathbb{R}) \to [0, \infty)$ which satisfies for all $\phi \in C(\mathbb{R}^d, \mathbb{R})$ that

$$\Phi(\phi) = \sum_{m=1}^{M} |\phi(x_m) - y_m|^2. \tag{1.2}$$

Observe that (1.1) ensures that $\Phi(\mathcal{E}) = 0$ and, in particular, we have that the unknown function $\mathcal{E} \colon \mathbb{R}^d \to \mathbb{R}$ in (1.1) above is a global minimizer of the function $\Phi \colon C(\mathbb{R}^d, \mathbb{R}) \to [0, \infty)$. The optimization problem of approximately computing minima of the function Φ is not suitable for discrete numerical computations on a computer as the function Φ is defined on the infinite dimensional vector space $C(\mathbb{R}^d, \mathbb{R})$. To overcome this we introduce a spatially discretized version of this optimization problem. More specifically, let $\mathfrak{d} \in \mathbb{N}$, let $\psi = (\psi_\theta)_{\theta \in \mathbb{R}^\mathfrak{d}} \colon \mathbb{R}^\mathfrak{d} \to C(\mathbb{R}^d, \mathbb{R})$ be a function, and let $\Psi \colon \mathbb{R}^\mathfrak{d} \to [0, \infty)$ satisfy $\Psi = \Phi \circ \psi$. We think of the set

$$\big\{\psi_\theta \colon \theta \in \mathbb{R}^\mathfrak{d}\big\} \subseteq C(\mathbb{R}^d, \mathbb{R}) \tag{1.3}$$

as a parametrized set of functions which we employ to approximate the infinite dimensional Banach space $C(\mathbb{R}^d, \mathbb{R})$ and we think of the function $\mathbb{R}^\mathfrak{d} \ni \theta \mapsto \psi_\theta \in C(\mathbb{R}^d, \mathbb{R})$ as the parametrization function corresponding to this set. Taking the set in (1.3) and its parametrization function $\mathbb{R}^\mathfrak{d} \ni \theta \mapsto \psi_\theta \in C(\mathbb{R}^d, \mathbb{R})$ into account, we then intend to approximately compute minima of the function Φ restricted to the set $\{\psi_\theta \colon \theta \in \mathbb{R}^\mathfrak{d}\}$, that is, we consider the optimization problem of approximately computing minima of the function

$$\big\{\psi_\theta \colon \theta \in \mathbb{R}^\mathfrak{d}\big\} \ni \phi \mapsto \Phi(\phi) = \left[\sum_{m=1}^{M} |\phi(x_m) - y_m|^2\right] \in [0, \infty). \tag{1.4}$$

Employing the parametrization function $\mathbb{R}^\mathfrak{d} \ni \theta \mapsto \psi_\theta \in C(\mathbb{R}^d, \mathbb{R})$ one can also reformulate this optimization problem as the optimization problem of approximately computing minima of the function

$$\mathbb{R}^\mathfrak{d} \ni \theta \mapsto \Psi(\theta) = \Phi(\psi_\theta) = \left[\sum_{m=1}^{M} |\psi_\theta(x_m) - y_m|^2\right] \in [0, \infty) \tag{1.5}$$

and this optimization is now accessible for discrete numerical computations. In the context of deep supervised learning algorithms, one would choose the parametrization function $\mathbb{R}^\mathfrak{d} \ni \theta \mapsto \psi_\theta \in C(\mathbb{R}^d, \mathbb{R})$ as deep neural network parametrizations and one would then apply a stochastic gradient descent optimization algorithm to the optimization problem in (1.5) to approximately compute minima of (1.5).

Chapter 2

Basics on artificial neural networks (ANNs)

In this chapter we present two approaches on how artificial neural networks (ANNs) can be described in a rigorous mathematical way.

2.1 Vectorized description of ANNs

2.1.1 Affine functions

Definition 2.1.1 (Affine functions)**.** *Let $\mathfrak{d}, m, n \in \mathbb{N}$, $s \in \mathbb{N}_0$, $\theta = (\theta_1, \theta_2, \ldots, \theta_\mathfrak{d}) \in \mathbb{R}^\mathfrak{d}$ satisfy $\mathfrak{d} \geq s + mn + m$. Then we denote by $\mathcal{A}^{\theta,s}_{m,n} \colon \mathbb{R}^n \to \mathbb{R}^m$ the function which satisfies for all $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ that*

$$\begin{aligned}
\mathcal{A}^{\theta,s}_{m,n}(x) &= \begin{pmatrix} \theta_{s+1} & \theta_{s+2} & \cdots & \theta_{s+n} \\ \theta_{s+n+1} & \theta_{s+n+2} & \cdots & \theta_{s+2n} \\ \theta_{s+2n+1} & \theta_{s+2n+2} & \cdots & \theta_{s+3n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{s+(m-1)n+1} & \theta_{s+(m-1)n+2} & \cdots & \theta_{s+mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \theta_{s+mn+1} \\ \theta_{s+mn+2} \\ \theta_{s+mn+3} \\ \vdots \\ \theta_{s+mn+m} \end{pmatrix} \\
&= \Big(\Big[\textstyle\sum_{k=1}^n x_k \theta_{s+k} \Big] + \theta_{s+mn+1}, \Big[\textstyle\sum_{k=1}^n x_k \theta_{s+n+k} \Big] + \theta_{s+mn+2}, \ldots, \Big[\textstyle\sum_{k=1}^n x_k \theta_{s+(m-1)n+k} \Big] + \theta_{s+mn+m} \Big)
\end{aligned} \tag{2.1}$$

and we call $\mathcal{A}^{\theta,s}_{m,n}$ the affine function from $\mathbb{R}^n$ to $\mathbb{R}^m$ associated to (θ, s).

2.1.2 Vectorized description of ANNs

Definition 2.1.2 (Vectorized description of ANNs)**.** *Let $\mathfrak{d}, L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ satisfy*

$$\mathfrak{d} \geq \sum_{k=1}^{L} l_k(l_{k-1} + 1) \tag{2.2}$$

and let $\Psi_k \colon \mathbb{R}^{l_k} \to \mathbb{R}^{l_k}$, $k \in \{1, 2, \ldots, L\}$, be functions. Then we denote by $\mathcal{N}^{\theta, l_0}_{\Psi_1, \Psi_2, \ldots, \Psi_L} \colon \mathbb{R}^{l_0} \to \mathbb{R}^{l_L}$ the function which satisfies for all $x \in \mathbb{R}^{l_0}$ that

$$\begin{split}\big(\mathcal{N}^{\theta, l_0}_{\Psi_1, \Psi_2, \ldots, \Psi_L}\big)(x) = \big(\Psi_L \circ \mathcal{A}^{\theta, \sum_{k=1}^{L-1} l_k(l_{k-1}+1)}_{l_L, l_{L-1}} \circ \Psi_{L-1} \circ \mathcal{A}^{\theta, \sum_{k=1}^{L-2} l_k(l_{k-1}+1)}_{l_{L-1}, l_{L-2}} \circ \ldots \\ \ldots \circ \Psi_2 \circ \mathcal{A}^{\theta, l_1(l_0+1)}_{l_2, l_1} \circ \Psi_1 \circ \mathcal{A}^{\theta, 0}_{l_1, l_0}\big)(x)\end{split} \tag{2.3}$$

(cf. Definition 2.1.1) and we call $\mathcal{N}^{\theta, l_0}_{\Psi_1, \Psi_2, \ldots, \Psi_L}$ the realization of the fully connected feedforward ANN associated to θ with $L+1$ layers with dimensions $(l_0, l_1, \ldots, l_L)$ and activation functions $(\Psi_1, \Psi_2, \ldots, \Psi_L)$ (we call $\mathcal{N}^{\theta, l_0}_{\Psi_1, \Psi_2, \ldots, \Psi_L}$ the fully connected feedforward ANN associated to θ with $L+1$ layers with dimensions $(l_0, l_1, \ldots, l_L)$ and activation functions $(\Psi_1, \Psi_2, \ldots, \Psi_L)$, we call $\mathcal{N}^{\theta, l_0}_{\Psi_1, \Psi_2, \ldots, \Psi_L}$ the ANN associated to θ with $L+1$ layers with dimensions $(l_0, l_1, \ldots, l_L)$ and activation functions $(\Psi_1, \Psi_2, \ldots, \Psi_L)$, we call $\mathcal{N}^{\theta, l_0}_{\Psi_1, \Psi_2, \ldots, \Psi_L}$ the fully connected feedforward ANN associated to θ with $L-1$ hidden layers with dimensions $(l_1, l_2, \ldots, l_{L-1})$ and activation functions $(\Psi_1, \Psi_2, \ldots, \Psi_{L-1})$, input dimension l_0, output dimension l_L, and output activation function Ψ_L).

2.1.3 Weights and biases of ANNs

Remark 2.1.3. *Let $L \in \{2, 3, 4, \ldots\}$, $v_0, v_1, \ldots, v_{L-1} \in \mathbb{N}_0$, $l_0, l_1, \ldots, l_L$, $\mathfrak{d} \in \mathbb{N}$, $\theta = (\theta_1, \theta_2, \ldots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ satisfy for all $k \in \{0, 1, \ldots, L-1\}$ that*

$$\mathfrak{d} \geq \sum_{i=1}^{L} l_i(l_{i-1} + 1) \qquad \text{and} \qquad v_k = \sum_{i=1}^{k} l_i(l_{i-1} + 1), \tag{2.4}$$

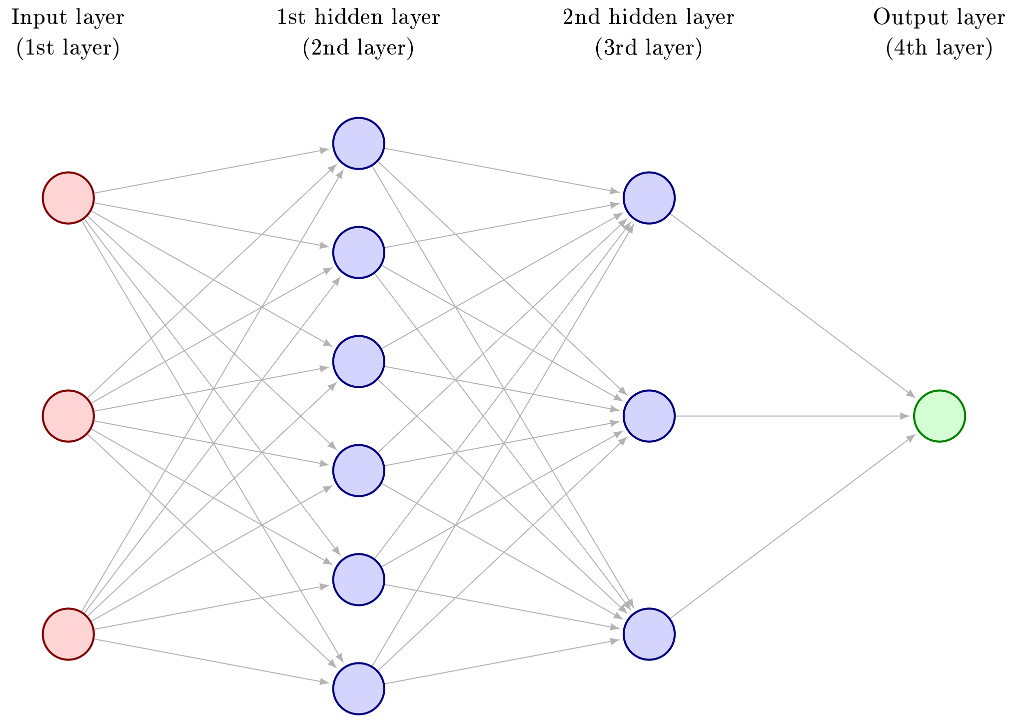

Figure 2.1: Graphical illustration of an artificial neural network. The ANN has 2 hidden layers and length $L = 3$ with 3 neurons in the input layer (corresponding to $l_0 = 3$), 6 neurons in the first hidden layer (corresponding to $l_1 = 6$), 3 neurons in the second hidden layer (corresponding to $l_2 = 3$), and one neuron in the output layer (corresponding to $l_3 = 1$). In this situation we have an ANN with 39 weights and 10 biases adding up to 49 parameters overall. The realization of this ANN is a function from $\mathbb{R}^3$ to $\mathbb{R}$.

let $W_k \in \mathbb{R}^{l_k \times l_{k-1}}$, $k \in \{1, 2, \ldots, L\}$, *and* $b_k \in \mathbb{R}^{l_k}$, $k \in \{1, 2, \ldots, L\}$, *satisfy for all* $k \in \{1, 2, \ldots, L\}$ *that*

$$W_k = \underbrace{\begin{pmatrix} \theta_{v_{k-1}+1} & \theta_{v_{k-1}+2} & \cdots & \theta_{v_{k-1}+l_{k-1}} \\ \theta_{v_{k-1}+l_{k-1}+1} & \theta_{v_{k-1}+l_{k-1}+2} & \cdots & \theta_{v_{k-1}+2l_{k-1}} \\ \theta_{v_{k-1}+2l_{k-1}+1} & \theta_{v_{k-1}+2l_{k-1}+2} & \cdots & \theta_{v_{k-1}+3l_{k-1}} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{v_{k-1}+(l_k-1)l_{k-1}+1} & \theta_{v_{k-1}+(l_k-1)l_{k-1}+2} & \cdots & \theta_{v_{k-1}+l_k l_{k-1}} \end{pmatrix}}_{weights} \tag{2.5}$$

$$\text{and} \qquad b_k = \underbrace{\left(\theta_{v_{k-1}+l_k l_{k-1}+1}, \theta_{v_{k-1}+l_k l_{k-1}+2}, \ldots, \theta_{v_{k-1}+l_k l_{k-1}+l_k}\right)}_{biases}, \tag{2.6}$$

and let $\Psi_k \colon \mathbb{R}^{l_k} \to \mathbb{R}^{l_k}$, $k \in \{1, 2, \ldots, L\}$, *be functions. Then*

(i) it holds that

$$\mathcal{N}^{\theta,l_0}_{\Psi_1,\Psi_2,\ldots,\Psi_L} = \Psi_L \circ \mathcal{A}^{\theta,v_{L-1}}_{l_L,l_{L-1}} \circ \Psi_{L-1} \circ \mathcal{A}^{\theta,v_{L-2}}_{l_{L-1},l_{L-2}} \circ \Psi_{L-2} \circ \ldots \circ \mathcal{A}^{\theta,v_1}_{l_2,l_1} \circ \Psi_1 \circ \mathcal{A}^{\theta,v_0}_{l_1,l_0} \tag{2.7}$$

and

(ii) it holds for all $k \in \{1, 2, \ldots, L\}$, $x \in \mathbb{R}^{l_{k-1}}$ *that* $\mathcal{A}^{\theta,v_{k-1}}_{l_k,l_{k-1}}(x) = W_k x + b_k$

(cf. Definitions 2.1.1 and 2.1.2).

2.1.4 Activation functions

2.1.4.1 Multidimensional versions

To describe multidimensional activation functions, we frequently employ the concept of the multidimensional version of a function. This concept is the subject of the next notion.

Definition 2.1.4 (Multidimensional versions of one dimensional functions)**.** *Let* $T \in \mathbb{N}$, $d_1, d_2, \ldots, d_T \in \mathbb{N}$ *and let* $\psi \colon \mathbb{R} \to \mathbb{R}$ *be a function. Then we denote by* $\mathfrak{M}_{\psi,d_1,d_2,\ldots,d_T} \colon \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_T} \to \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_T}$ *the function which satisfies for all* $x = (x_{k_1,k_2,\ldots,k_T})_{(k_1,k_2,\ldots,k_T) \in (\times_{t=1}^T \{1,2,\ldots,d_t\})} \in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_T}$, $y = (y_{k_1,k_2,\ldots,k_T})_{(k_1,k_2,\ldots,k_T) \in (\times_{t=1}^T \{1,2,\ldots,d_t\})} \in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_T}$ *with* $\forall\, k_1 \in \{1, 2, \ldots, d_1\}$, $k_2 \in \{1, 2, \ldots, d_2\}$, $\ldots$, $k_T \in \{1, 2, \ldots, d_T\} \colon y_{k_1,k_2,\ldots,k_T} = \psi(x_{k_1,k_2,\ldots,k_T})$ *that*

$$\mathfrak{M}_{\psi,d_1,d_2,\ldots,d_T}(x) = y \tag{2.8}$$

and we call $\mathfrak{M}_{\psi,d_1,d_2,\ldots,d_T}$ *the* $d_1 \times d_2 \times \ldots \times d_T$*-dimensional version of* ψ*.*

2.1.4.2 Single hidden layer ANNs

Example 2.1.5. *Let $\mathcal{I}, \mathcal{H} \in \mathbb{N}$, $\theta = (\theta_1, \theta_2, \dots, \theta_{\mathcal{H}\mathcal{I}+2\mathcal{H}+1}) \in \mathbb{R}^{\mathcal{H}\mathcal{I}+2\mathcal{H}+1}$, $x = (x_1, x_2, \dots, x_{\mathcal{I}}) \in \mathbb{R}^{\mathcal{I}}$ and let $\psi \colon \mathbb{R} \to \mathbb{R}$ be a function. Then*

$$\begin{aligned}
&\mathcal{N}^{\theta,\mathcal{I}}_{\mathfrak{M}_{\psi,\mathcal{H}},\mathrm{id}_{\mathbb{R}}}(x) \\
&= \Big((\mathrm{id}_{\mathbb{R}}) \circ \mathcal{A}^{\theta,\mathcal{H}\mathcal{I}+\mathcal{H}}_{1,\mathcal{H}} \circ \mathfrak{M}_{\psi,\mathcal{H}} \circ \mathcal{A}^{\theta,0}_{\mathcal{H},\mathcal{I}}\Big)(x) \\
&= \mathcal{A}^{\theta,\mathcal{H}\mathcal{I}+\mathcal{H}}_{1,\mathcal{H}}\big(\mathfrak{M}_{\psi,\mathcal{H}}\big(\mathcal{A}^{\theta,0}_{\mathcal{H},\mathcal{I}}(x)\big)\big) \\
&= \left[\sum_{k=1}^{\mathcal{H}} \theta_{\mathcal{H}\mathcal{I}+\mathcal{H}+k}\, \psi\!\left(\left[\sum_{i=1}^{\mathcal{I}} x_i \theta_{(k-1)\mathcal{I}+i}\right] + \theta_{\mathcal{H}\mathcal{I}+k}\right)\right] + \theta_{\mathcal{H}\mathcal{I}+2\mathcal{H}+1}.
\end{aligned} \tag{2.9}$$

(cf. Definitions 2.1.1, 2.1.2, and 2.1.4).

2.1.4.3 The rectifier function

In this subsection we formulate the rectifier function which is maybe the most commonly used activation function in deep learning applications (cf., for example, Le Cun, Bengio, & Hinton [7]).

Definition 2.1.6 (Rectifier function)**.** *We denote by $\mathfrak{r} \colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that*

$$\mathfrak{r}(x) = \max\{x, 0\}. \tag{2.10}$$

and we call $\mathfrak{r}$ the rectifier function.

Definition 2.1.7 (Multidimensional rectifier functions)**.** *Let $d \in \mathbb{N}$. Then we denote by $\mathfrak{R}_d \colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by*

$$\mathfrak{R}_d = \mathfrak{M}_{\mathfrak{r},d} \tag{2.11}$$

(cf. Definitions 2.1.4 and 2.1.6) and we call $\mathfrak{R}_d$ the d-dimensional rectifier function.

Proposition 2.1.8 (An ANN with the rectifier function as the activation function)**.** *Let $W_1 = w_1 = 1$, $W_2 = w_2 = -1$, $b_1 = b_2 = B = 0$. Then it holds for all $x \in \mathbb{R}$ that*

$$x = W_1 \max\{w_1 x + b_1, 0\} + W_2 \max\{w_2 x + b_2, 0\} + B. \tag{2.12}$$

Proof of Proposition 2.1.8. Observe that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}
&W_1 \max\{w_1 x + b_1, 0\} + W_2 \max\{w_2 x + b_2, 0\} + B \\
&= \max\{w_1 x + b_1, 0\} - \max\{w_2 x + b_2, 0\} = \max\{x, 0\} - \max\{-x, 0\} \\
&= \max\{x, 0\} + \min\{x, 0\} = x.
\end{aligned} \tag{2.13}$$

The proof of Proposition 2.1.8 is thus complete. □

Exercise 2.1.1 (Real identity). *Prove or disprove the following statement: There exist $\mathfrak{d}, L \in \mathbb{N}$, $l_1, l_2, \ldots, l_L \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ with $\mathfrak{d} \geq 2l_1 + \big[\sum_{k=2}^{L} l_k(l_{k-1}+1)\big] + l_L + 1$ such that for all $x \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\ldots,\mathfrak{R}_{l_L},\mathrm{id}_{\mathbb{R}}}\big)(x) = x \tag{2.14}$$

(cf. Definitions 2.1.2 and 2.1.7).

The statement of the next lemma, Lemma 2.1.9, provides a partial answer to Exercise 2.1.1. Lemma 2.1.9 follows from an application of Proposition 2.1.8 and the detailed proof of Lemma 2.1.9 is left as an exercise.

Lemma 2.1.9 (Real identity). *Let $\theta = (1, -1, 0, 0, 1, -1, 0) \in \mathbb{R}^7$. Then it holds for all $x \in \mathbb{R}$ that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_2,\mathrm{id}_{\mathbb{R}}}\big)(x) = x \tag{2.15}$$

(cf. Definitions 2.1.2 and 2.1.7).

Exercise 2.1.2 (Absolute value). *Prove or disprove the following statement: There exist $\mathfrak{d}, L \in \mathbb{N}$, $l_1, l_2, \ldots, l_L \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ with $\mathfrak{d} \geq 2l_1 + \big[\sum_{k=2}^{L} l_k(l_{k-1}+1)\big] + l_L + 1$ such that for all $x \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\ldots,\mathfrak{R}_{l_L},\mathrm{id}_{\mathbb{R}}}\big)(x) = |x| \tag{2.16}$$

(cf. Definitions 2.1.2 and 2.1.7).

Exercise 2.1.3 (Exponential). *Prove or disprove the following statement: There exist $\mathfrak{d}, L \in \mathbb{N}$, $l_1, l_2, \ldots, l_L \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ with $\mathfrak{d} \geq 2l_1 + \big[\sum_{k=2}^{L} l_k(l_{k-1}+1)\big] + l_L + 1$ such that for all $x \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\ldots,\mathfrak{R}_{l_L},\mathrm{id}_{\mathbb{R}}}\big)(x) = e^x \tag{2.17}$$

(cf. Definitions 2.1.2 and 2.1.7).

Exercise 2.1.4 (Two-dimensional maximum). *Prove or disprove the following statement: There exist $\mathfrak{d}, L \in \mathbb{N}$, $l_1, l_2, \ldots, l_L \in \mathbb{N}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ with $\mathfrak{d} \geq 3l_1 + \big[\sum_{k=2}^{L} l_k(l_{k-1}+1)\big] + l_L + 1$ such that for all $x, y \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,2}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\ldots,\mathfrak{R}_{l_L},\mathrm{id}_{\mathbb{R}}}\big)(x, y) = \max\{x, y\} \tag{2.18}$$

(cf. Definitions 2.1.2 and 2.1.7).

Exercise 2.1.5 (Real identity with two hidden layers). *Prove or disprove the following statement: There exist $\mathfrak{d}, l_1, l_2 \in \mathbb{N}$, $\theta\mathbb{R}^{\mathfrak{d}}$ with $\mathfrak{d} \geq 2l_1 + l_1 l_2 + 2l_2 + 1$ such that for all $x \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\mathrm{id}_{\mathbb{R}}}\big)(x) = x \tag{2.19}$$

(cf. Definitions 2.1.2 and 2.1.7).

The statement of the next lemma, Lemma 2.1.10, provides a partial answer to Exercise 2.1.5. The proof of Lemma 2.1.10 is left as an exercise.

Lemma 2.1.10 (Real identity with two hidden layers)**.** *Let $\theta = (1,\ -1,\ 0,\ 0,\ 1,\ -1,\ -1,\ 1,\ 0,\ 0,\ 1,\ -1,\ 0) \in \mathbb{R}^{13}$. Then it holds for all $x \in \mathbb{R}$ that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_2,\mathfrak{R}_2,\mathrm{id}_\mathbb{R}}\big)(x) = x \tag{2.20}$$

(cf. Definitions 2.1.2 and 2.1.7).

Exercise 2.1.6 (Three-dimensional maximum)**.** *Prove or disprove the following statement: There exist $\mathfrak{d}, L \in \mathbb{N}$, $l_1, l_2, \ldots, l_L \in \mathbb{N}$, $\theta \in \mathbb{R}^\mathfrak{d}$ with $\mathfrak{d} \geq 4l_1 + \big[\sum_{k=2}^L l_k(l_{k-1}+1)\big] + l_L + 1$ such that for all $x, y, z \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,3}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\ldots,\mathfrak{R}_{l_L},\mathrm{id}_\mathbb{R}}\big)(x,y,z) = \max\{x,y,z\} \tag{2.21}$$

(cf. Definition 2.1.2 and Definition 2.1.7).

Exercise 2.1.7 (Multidimensional maxima)**.** *Prove or disprove the following statement: For every $k \in \mathbb{N}$ there exist $\mathfrak{d}, L \in \mathbb{N}$, $l_1, l_2, \ldots, l_L \in \mathbb{N}$, $\theta \in \mathbb{R}^\mathfrak{d}$ with $\mathfrak{d} \geq (k+1)l_1 + \big[\sum_{k=2}^L l_k(l_{k-1}+1)\big] + l_L + 1$ such that for all $x_1, x_2, \ldots, x_k \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,k}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\ldots,\mathfrak{R}_{l_L},\mathrm{id}_\mathbb{R}}\big)(x_1,x_2,\ldots,x_k) = \max\{x_1,x_2,\ldots,x_k\} \tag{2.22}$$

(cf. Definitions 2.1.2 and 2.1.7).

Exercise 2.1.8 (Hat function)**.** *Prove or disprove the following statement: There exist $\mathfrak{d}, l \in \mathbb{N}$, $\theta \in \mathbb{R}^\mathfrak{d}$ with $\mathfrak{d} \geq 3l + 1$ such that for all $x \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_l,\mathrm{id}_\mathbb{R}}\big)(x) = \begin{cases} 1 & : x \leq 2 \\ x-1 & : 2 < x \leq 3 \\ 5-x & : 3 < x \leq 4 \\ 1 & : x > 4 \end{cases} \tag{2.23}$$

(cf. Definition 2.1.2 and Definition 2.1.7).

Exercise 2.1.9. *Prove or disprove the following statement: There exist $\mathfrak{d}, l \in \mathbb{N}$, $\theta \in \mathbb{R}^\mathfrak{d}$ with $\mathfrak{d} \geq 3l + 1$ such that for all $x \in \mathbb{R}$ it holds that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_l,\mathrm{id}_\mathbb{R}}\big)(x) = \begin{cases} -2 & : x \leq 1 \\ 2x-4 & : 1 < x \leq 3 \\ 2 & : x > 3 \end{cases} \tag{2.24}$$

(cf. Definition 2.1.2 and Definition 2.1.7).

Exercise 2.1.10. *Prove or disprove the following statement: There exist $\mathfrak{d}, l \in \mathbb{N}$, $\theta \in \mathbb{R}^\mathfrak{d}$ with $\mathfrak{d} \geq 3l + 1$ such that for all $x \in [0,1]$ it holds that*

$$\big(\mathcal{N}^{\theta,1}_{\mathfrak{R}_l,\mathrm{id}_\mathbb{R}}\big)(x) = x^2 \tag{2.25}$$

(cf. Definition 2.1.2 and Definition 2.1.7).

2.1.4.4 Clipping functions

Definition 2.1.11 (Clipping function)**.** *Let $u \in [-\infty, \infty)$, $v \in (u, \infty]$. Then we denote by $\mathfrak{c}_{u,v} \colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that*

$$\mathfrak{c}_{u,v}(x) = \max\{u, \min\{x, v\}\}. \tag{2.26}$$

and we call $\mathfrak{c}_{u,v}$ the (u, v)-clipping function.

Definition 2.1.12 (Multidimensional clipping functions)**.** *Let $d \in \mathbb{N}$, $u \in [-\infty, \infty)$, $v \in (u, \infty]$. Then we denote by $\mathfrak{C}_{u,v,d} \colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by*

$$\mathfrak{C}_{u,v,d} = \mathfrak{M}_{\mathfrak{c}_{u,v},d} \tag{2.27}$$

(cf. Definitions 2.1.4 and 2.1.11) and we call $\mathfrak{C}_{u,v,d}$ the d-dimensional (u, v)-clipping function.

2.1.4.5 The softplus function

Definition 2.1.13 (Softplus function)**.** *We denote by $\mathfrak{s} \colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that*

$$\mathfrak{s}(x) = \ln(1 + \exp(x)) \tag{2.28}$$

and we call $\mathfrak{s}$ the softplus function.

The next result, Lemma 2.1.14 below, presents a few elementary properties of the softplus function.

Lemma 2.1.14 (Properties of the softplus function)**.** *It holds*

(i) for all $x \in [0, \infty)$ that $x \le \mathfrak{s}(x) \le x + 1$,

(ii) that $\lim_{x\to-\infty} \mathfrak{s}(x) = 0$,

(iii) that $\lim_{x\to\infty} \mathfrak{s}(x) = \infty$, and

(iv) that $\mathfrak{s}(0) = \ln(2)$

(cf. Definition 2.1.13).

Proof of Lemma 2.1.14. Observe that the fact that $2 \le \exp(1)$ ensures that for all $x \in [0, \infty)$ it holds that

$$\begin{aligned} x = \ln(\exp(x)) &\le \ln(1 + \exp(x)) = \ln(\exp(0) + \exp(x)) \\ &\le \ln(\exp(x) + \exp(x)) = \ln(2 \exp(x)) \le \ln(\exp(1) \exp(x)) \\ &= \ln(\exp(x + 1)) = x + 1. \end{aligned} \tag{2.29}$$

The proof of Lemma 2.1.14 is thus complete. □

Note that Lemma 2.1.14 ensures that $\mathfrak{s}(0) = \ln(2) = 0.693\ldots$ (cf. Definition 2.1.13). In the next step we introduce the multidimensional version of the softplus function (cf. Definitions 2.1.4 and 2.1.13 above).

Definition 2.1.15 (Multidimensional softplus functions). *Let $d \in \mathbb{N}$. Then we denote by $\mathfrak{S}_d \colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by*

$$\mathfrak{S}_d = \mathfrak{M}_{\mathfrak{s},d} \tag{2.30}$$

(cf. Definitions 2.1.4 and 2.1.13) and we call $\mathfrak{S}_d$ the d-dimensional softplus function.

2.1.4.6 The standard logistic function

Definition 2.1.16 (Standard logistic function). *We denote by $\mathfrak{l} \colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that*

$$\mathfrak{l}(x) = \frac{1}{1 + \exp(-x)} = \frac{\exp(x)}{\exp(x) + 1} \tag{2.31}$$

and we call $\mathfrak{l}$ the standard logistic function.

Definition 2.1.17 (Multidimensional standard logistic functions). *Let $d \in \mathbb{N}$. Then we denote by $\mathfrak{L}_d \colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by*

$$\mathfrak{L}_d = \mathfrak{M}_{\mathfrak{l},d} \tag{2.32}$$

(cf. Definitions 2.1.4 and 2.1.16) and we call $\mathfrak{L}_d$ the d-dimensional standard logistic function.

2.1.4.7 Derivative of the standard logistic function

Proposition 2.1.18 (Logistic differential equation). *It holds that $\mathfrak{l} \colon \mathbb{R} \to \mathbb{R}$ is infinitely often differentiable and it holds for all $x \in \mathbb{R}$ that*

$$\mathfrak{l}(0) = 1/2, \qquad \mathfrak{l}'(x) = \mathfrak{l}(x)(1 - \mathfrak{l}(x)) = \mathfrak{l}(x) - [\mathfrak{l}(x)]^2, \qquad \textit{and} \tag{2.33}$$

$$\mathfrak{l}''(x) = \mathfrak{l}(x)(1 - \mathfrak{l}(x))(1 - 2\,\mathfrak{l}(x)) = 2[\mathfrak{l}(x)]^3 - 3[\mathfrak{l}(x)]^2 + \mathfrak{l}(x) \tag{2.34}$$

(cf. Definition 2.1.16).

Proof of Proposition 2.1.18. Observe that (2.31) ensures that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} \mathfrak{l}'(x) &= \frac{\exp(-x)}{(1 + \exp(-x))^2} = \mathfrak{l}(x)\left(\frac{\exp(-x)}{1 + \exp(-x)}\right) \\ &= \mathfrak{l}(x)\left(\frac{1 + \exp(-x) - 1}{1 + \exp(-x)}\right) = \mathfrak{l}(x)\left(1 - \frac{1}{1 + \exp(-x)}\right) \\ &= \mathfrak{l}(x)(1 - \mathfrak{l}(x)). \end{aligned} \tag{2.35}$$

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

$$\begin{aligned}
\mathfrak{l}''(x) &= \big[\mathfrak{l}(x)(1-\mathfrak{l}(x))\big]' = \mathfrak{l}'(x)(1-\mathfrak{l}(x)) + \mathfrak{l}(x)(1-\mathfrak{l}(x))' \\
&= \mathfrak{l}'(x)(1-\mathfrak{l}(x)) - \mathfrak{l}(x)\,\mathfrak{l}'(x) = \mathfrak{l}'(x)(1-2\,\mathfrak{l}(x)) \\
&= \mathfrak{l}(x)(1-\mathfrak{l}(x))(1-2\,\mathfrak{l}(x)) \\
&= \big(\mathfrak{l}(x)-[\mathfrak{l}(x)]^2\big)(1-2\,\mathfrak{l}(x)) = \mathfrak{l}(x) - [\mathfrak{l}(x)]^2 - 2[\mathfrak{l}(x)]^2 + 2[\mathfrak{l}(x)]^3 \\
&= 2[\mathfrak{l}(x)]^3 - 3[\mathfrak{l}(x)]^2 + \mathfrak{l}(x).
\end{aligned} \tag{2.36}$$

The proof of Proposition 2.1.18 is thus complete. □

2.1.4.8 Integral of the standard logistic function

Lemma 2.1.19 (Primitive of the standard logistic function)**.** *It holds for all $x \in \mathbb{R}$ that*

$$\int_{-\infty}^{x} \mathfrak{l}(y)\,\mathrm{d}y = \int_{-\infty}^{x} \left(\frac{1}{1+e^{-y}}\right)\mathrm{d}y = \ln(1+\exp(x)) = \mathfrak{s}(x) \tag{2.37}$$

(cf. Definitions 2.1.13 and 2.1.16).

Proof of Lemma 2.1.19. Observe that (2.28) implies that for all $x \in \mathbb{R}$ it holds that

$$\mathfrak{s}'(x) = \left[\frac{1}{1+\exp(x)}\right]\exp(x) = \mathfrak{l}(x). \tag{2.38}$$

The fundamental theorem of calculus hence shows that for all $w, x \in \mathbb{R}$ with $w \leq x$ it holds that

$$\int_{w}^{x} \underbrace{\mathfrak{l}(y)}_{\geq 0}\,\mathrm{d}y = \mathfrak{s}(x) - \mathfrak{s}(w). \tag{2.39}$$

Combining this with the fact that $\lim_{w\to-\infty} \mathfrak{s}(w) = 0$ establishes (2.37). The proof of Lemma 2.1.19 is thus complete. □

2.1.4.9 The hyperbolic tangent function

Definition 2.1.20 (Hyperbolic tangent)**.** *We denote by* $\tanh\colon \mathbb{R} \to \mathbb{R}$ *the function which satisfies for all $x \in \mathbb{R}$ that*

$$\tanh(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)} \tag{2.40}$$

and we call tanh *the hyperbolic tangent.*

Definition 2.1.21 (Multidimensional hyperbolic tangent functions)**.** *Let $d \in \mathbb{N}$. Then we denote by $\mathfrak{T}_d\colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by*

$$\mathfrak{T}_d = \mathfrak{M}_{\tanh,d} \tag{2.41}$$

(cf. Definitions 2.1.4 and 2.1.20) and we call $\mathfrak{T}_d$ the d-dimensional hyperbolic tangent.

Lemma 2.1.22. *It holds for all $x \in \mathbb{R}$ that*

$$\tanh(x) = 2\,\mathfrak{l}(2x) - 1 \tag{2.42}$$

(cf. Definitions 2.1.16 and 2.1.20).

Proof of Lemma 2.1.22. Observe that (2.31) and (2.40) ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} 2\,\mathfrak{l}(2x) - 1 &= 2\left(\frac{\exp(2x)}{\exp(2x)+1}\right) - 1 = \frac{2\exp(2x) - (\exp(2x)+1)}{\exp(2x)+1} \\ &= \frac{\exp(2x)-1}{\exp(2x)+1} = \frac{\exp(x)(\exp(x)-\exp(-x))}{\exp(x)(\exp(x)+\exp(-x))} \\ &= \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)} = \tanh(x). \end{aligned} \tag{2.43}$$

The proof of Lemma 2.1.22 is thus complete. □

2.1.4.10 The Heaviside function

Definition 2.1.23 (Heaviside function)**.** *We denote by $\mathfrak{h}\colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that*

$$\mathfrak{h}(x) = \mathbb{1}_{[0,\infty)}(x) = \begin{cases} 1 & : x \geq 0 \\ 0 & : x < 0 \end{cases} \tag{2.44}$$

and we call $\mathfrak{h}$ the Heaviside function (we call $\mathfrak{h}$ the Heaviside step function, we call $\mathfrak{h}$ the unit step function).

Definition 2.1.24 (Multidimensional Heaviside functions)**.** *Let $d \in \mathbb{N}$. Then we denote by $\mathfrak{H}_d\colon \mathbb{R}^d \to \mathbb{R}^d$ the function given by*

$$\mathfrak{H}_d = \mathfrak{M}_{\mathfrak{h},d} \tag{2.45}$$

(cf. Definitions 2.1.4 and 2.1.23) and we call $\mathfrak{H}_d$ the d-dimensional Heaviside function (we call $\mathfrak{H}_d$ the d-dimensional Heaviside step function, we call $\mathfrak{H}_d$ the d-dimensional unit step function).

2.1.4.11 The softmax function

Definition 2.1.25 (The softmax function)**.** *Let $d \in \mathbb{N}$. Then we denote by $\mathcal{S}_d = (\mathcal{S}_d^{(1)}, \mathcal{S}_d^{(2)}, \ldots, \mathcal{S}_d^{(d)})\colon \mathbb{R}^d \to \mathbb{R}^d$ the function which satisfies for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that*

$$\begin{aligned} \mathcal{S}_d(x) &= \big(\mathcal{S}_d^{(1)}(x), \mathcal{S}_d^{(2)}(x), \ldots, \mathcal{S}_d^{(d)}(x)\big) \\ &= \left(\frac{\exp(x_1)}{\left(\sum_{i=1}^d \exp(x_i)\right)}, \frac{\exp(x_2)}{\left(\sum_{i=1}^d \exp(x_i)\right)}, \ldots, \frac{\exp(x_d)}{\left(\sum_{i=1}^d \exp(x_i)\right)}\right) \end{aligned} \tag{2.46}$$

and we call $\mathcal{S}_d$ the d-dimensional softmax function.

Lemma 2.1.26. *Let $d \in \mathbb{N}$. Then*

(i) it holds for all $x \in \mathbb{R}^d$, $k \in \{1, 2, \dots, d\}$ that $\mathscr{S}_d^{(k)}(x) \in (0, 1]$ and

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

$$\sum_{k=1}^{d} \mathscr{S}_d^{(k)}(x) = 1. \tag{2.47}$$

(cf. Definition 2.1.25).

Proof of Lemma 2.1.26. Observe that (2.46) demonstrates that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\sum_{k=1}^{d} \mathscr{S}_d^{(k)}(x) = \sum_{k=1}^{d} \frac{\exp(x_k)}{\left(\sum_{i=1}^{d} \exp(x_i)\right)} = \frac{\sum_{k=1}^{d} \exp(x_k)}{\sum_{i=1}^{d} \exp(x_i)} = 1. \tag{2.48}$$

The proof of Lemma 2.1.26 is thus complete. □

2.1.5 Rectified clipped ANNs

Definition 2.1.27 (Rectified clipped ANNs)**.** *Let $L, \mathfrak{d} \in \mathbb{N}$, $u \in [-\infty, \infty)$, $v \in (u, \infty]$, $\mathbf{l} = (l_0, l_1, \dots, l_L) \in \mathbb{N}^{L+1}$, $\theta \in \mathbb{R}^{\mathfrak{d}}$ satisfy*

$$\mathfrak{d} \geq \sum_{k=1}^{L} l_k(l_{k-1} + 1). \tag{2.49}$$

Then we denote by $\mathscr{N}_{u,v}^{\theta,\mathbf{l}} \colon \mathbb{R}^{l_0} \to \mathbb{R}^{l_L}$ the function which satisfies for all $x \in \mathbb{R}^{l_0}$ that

$$\mathscr{N}_{u,v}^{\theta,\mathbf{l}}(x) = \begin{cases} \big(\mathcal{N}^{\theta,l_0}_{\mathfrak{C}_{u,v,l_L}}\big)(x) & : L = 1 \\ \big(\mathcal{N}^{\theta,l_0}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\dots,\mathfrak{R}_{l_{L-1}},\mathfrak{C}_{u,v,l_L}}\big)(x) & : L > 1 \end{cases} \tag{2.50}$$

(cf. Definitions 2.1.2, 2.1.7, and 2.1.12).

2.2 Structured description of ANNs

2.2.1 Structured description of ANNs

Definition 2.2.1 (Structured description of ANNs)**.** *We denote by $\mathbf{N}$ the set given by*

$$\mathbf{N} = \textstyle\bigcup_{L\in\mathbb{N}} \bigcup_{l_0,l_1,\dots,l_L\in\mathbb{N}} \left(\bigtimes_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \right), \tag{2.51}$$

we denote by $\mathcal{P} \colon \mathbf{N} \to \mathbb{N}$, $\mathcal{L} \colon \mathbf{N} \to \mathbb{N}$, $\mathcal{I} \colon \mathbf{N} \to \mathbb{N}$, $\mathcal{O} \colon \mathbf{N} \to \mathbb{N}$, $\mathcal{H} \colon \mathbf{N} \to \mathbb{N}_0$, $\mathcal{D} \colon \mathbf{N} \to \left(\bigcup_{L=2}^{\infty} \mathbb{N}^L\right)$, and $\mathbb{D}_n \colon \mathbf{N} \to \mathbb{N}_0$, $n \in \mathbb{N}_0$, the functions which satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in$

$\mathbb{N}$, $\Phi \in \left(\bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\right)$, $n \in \mathbb{N}_0$ that $\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1}+1)$, $\mathcal{L}(\Phi) = L$, $\mathcal{I}(\Phi) = l_0$, $\mathcal{O}(\Phi) = l_L$, $\mathcal{H}(\Phi) = L-1$, $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$, and

$$\mathbb{D}_n(\Phi) = \begin{cases} l_n & : n \leq L \\ 0 & : n > L, \end{cases} \tag{2.52}$$

and for every $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in \left(\bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\right)$ *we denote by* $\mathcal{W}_{(\cdot),\Phi} = (\mathcal{W}_{n,\Phi})_{n \in \{1,2,\ldots,L\}} \colon \{1,2,\ldots,L\} \to (\bigcup_{m,k \in \mathbb{N}} \mathbb{R}^{m \times k})$ *and* $\mathcal{B}_{(\cdot),\Phi} = (\mathcal{B}_{n,\Phi})_{n \in \{1,2,\ldots,L\}} \colon \{1,2,\ldots,L\} \to (\bigcup_{m \in \mathbb{N}} \mathbb{R}^{m})$ *the functions which satisfy for all* $n \in \{1,2,\ldots,L\}$ *that* $\mathcal{W}_{n,\Phi} = W_n$ *and* $\mathcal{B}_{n,\Phi} = B_n$.

Definition 2.2.2. *We say that* Φ *is an ANN if and only if it holds that* $\Phi \in \mathbf{N}$.

2.2.2 Realizations of ANNs

Definition 2.2.3 (Realization associated to an ANN)**.** *Let* $a \in C(\mathbb{R}, \mathbb{R})$. *Then we denote by* $\mathcal{R}_a \colon \mathbf{N} \to \left(\bigcup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)\right)$ *the function which satisfies for all* $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in \left(\bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\right)$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_L \in \mathbb{R}^{l_L}$ *with* $\forall\, k \in \{1,2,\ldots,L\} \colon x_k = \mathfrak{M}_{a \mathbb{1}_{(0,L)}(k) + \mathrm{id}_{\mathbb{R}} \mathbb{1}_{\{L\}}(k), l_k}(W_k x_{k-1} + B_k)$ *that*

$$\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}) \qquad \text{and} \qquad (\mathcal{R}_a(\Phi))(x_0) = x_L \tag{2.53}$$

(cf. Definitions 2.1.4 and 2.2.1).

Lemma 2.2.4. *Let* $\Phi \in \mathbf{N}$ *(cf. Definition 2.2.1). 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 2.2.3).

Proof of Lemma 2.2.4. Note that the assumption that $\Phi \in \mathbf{N} = \bigcup_{L \in \mathbb{N}} \bigcup_{(l_0, l_1, \ldots, l_L) \in \mathbb{N}^{L+1}} (\bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$ ensures that there exist $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$ such that

$$\Phi \in \left(\bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\right). \tag{2.54}$$

Observe that (2.54) assures that

$$\mathcal{L}(\Phi) = L, \qquad \mathcal{I}(\Phi) = l_0, \qquad \mathcal{O}(\Phi) = l_L, \tag{2.55}$$

$$\text{and} \qquad \mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L) \in \mathbb{N}^{L+1} = \mathbb{N}^{\mathcal{L}(\Phi)+1}. \tag{2.56}$$

This establishes item (i). Moreover, note that (2.55) and (2.53) 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 2.2.4 is thus complete. $\square$

Exercise 2.2.1. *Prove or disprove the following statement: There exists* $\Phi \in \mathbf{N}$ *such that*

$$\mathcal{R}_{\tanh}(\Phi) = \mathfrak{l} \tag{2.57}$$

(cf. Definitions 2.1.16, 2.1.20, 2.2.1, and 2.2.3).

2.2.3 Compositions of ANNs

2.2.3.1 Standard compositions of ANNs

Definition 2.2.5 (Composition of ANNs)**.** *We denote by* $(\cdot) \bullet (\cdot) \colon \{(\Phi,\Psi) \in \mathbf{N} \times \mathbf{N} \colon \mathcal{I}(\Phi) = \mathcal{O}(\Psi)\} \to \mathbf{N}$ *the function which satisfies for all* $L, \mathfrak{L} \in \mathbb{N}$, $l_0, l_1, \ldots, l_L, \mathfrak{l}_0, \mathfrak{l}_1, \ldots, \mathfrak{l}_{\mathfrak{L}} \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in \big(\bigtimes_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\big)$, $\Psi = ((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \ldots, (\mathscr{W}_{\mathfrak{L}}, \mathscr{B}_{\mathfrak{L}})) \in \big(\bigtimes_{k=1}^{\mathfrak{L}} (\mathbb{R}^{\mathfrak{l}_k \times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k})\big)$ *with* $l_0 = \mathcal{I}(\Phi) = \mathcal{O}(\Psi) = \mathfrak{l}_{\mathfrak{L}}$ *that*

$$\Phi \bullet \Psi = \begin{cases} \begin{array}{r} \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \ldots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1 \mathscr{W}_{\mathfrak{L}}, W_1 \mathscr{B}_{\mathfrak{L}} + B_1), \\ (W_2, B_2), (W_3, B_3), \ldots, (W_L, B_L)\big) \end{array} & : (L > 1) \wedge (\mathfrak{L} > 1) \\ \big((W_1 \mathscr{W}_1, W_1 \mathscr{B}_1 + B_1), (W_2, B_2), (W_3, B_3), \ldots, (W_L, B_L)\big) & : (L > 1) \wedge (\mathfrak{L} = 1) \\ \big((\mathscr{W}_1, \mathscr{B}_1), (\mathscr{W}_2, \mathscr{B}_2), \ldots, (\mathscr{W}_{\mathfrak{L}-1}, \mathscr{B}_{\mathfrak{L}-1}), (W_1 \mathscr{W}_{\mathfrak{L}}, W_1 \mathscr{B}_{\mathfrak{L}} + B_1)\big) & : (L = 1) \wedge (\mathfrak{L} > 1) \\ \big((W_1 \mathscr{W}_1, W_1 \mathscr{B}_1 + B_1)\big) & : (L = 1) \wedge (\mathfrak{L} = 1) \end{cases} \tag{2.58}$$

(cf. Definition 2.2.1).

2.2.3.2 Elementary properties of standard compositions of ANNs

Lemma 2.2.6. *Let* $\Phi, \Psi \in \mathbf{N}$ *satisfy* $\mathcal{I}(\Phi) = \mathcal{O}(\Psi)$ *(cf. Definition 2.2.1). Then*

(i) it holds that $\mathcal{L}(\Phi \bullet \Psi) = \mathcal{L}(\Phi) + \mathcal{L}(\Psi) - 1$ *and*

(ii) it holds for all $i \in \{1, 2, \ldots, \mathcal{L}(\Phi \bullet \Psi)\}$ *that*

$$(\mathcal{W}_{i,(\Phi \bullet \Psi)}, \mathcal{B}_{i,(\Phi \bullet \Psi)}) = \begin{cases} (\mathcal{W}_{i,\Psi}, \mathcal{B}_{i,\Psi}) & : i < \mathcal{L}(\Psi) \\ (\mathcal{W}_{1,\Phi} \mathcal{W}_{\mathcal{L}(\Psi),\Psi}, \mathcal{W}_{1,\Phi} \mathcal{B}_{\mathcal{L}(\Psi),\Psi} + \mathcal{B}_{1,\Phi}) & : i = \mathcal{L}(\Psi) \\ (\mathcal{W}_{i-\mathcal{L}(\Psi)+1,\Phi}, \mathcal{B}_{i-\mathcal{L}(\Psi)+1,\Phi}) & : i > \mathcal{L}(\Psi). \end{cases} \tag{2.59}$$

Proof of Lemma 2.2.6. Note that (2.58) implies items (i) and (ii). The proof of Lemma 2.2.6 is thus complete. □

Proposition 2.2.7. *Let* $\Phi_1, \Phi_2 \in \mathbf{N}$ *satisfy* $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ *(cf. Definition 2.2.1). Then*

(i) it holds that

$$\mathcal{D}(\Phi_1 \bullet \Phi_2) = (\mathbb{D}_0(\Phi_2), \mathbb{D}_1(\Phi_2), \ldots, \mathbb{D}_{\mathcal{H}(\Phi_2)}(\Phi_2), \mathbb{D}_1(\Phi_1), \mathbb{D}_2(\Phi_1), \ldots, \mathbb{D}_{\mathcal{L}(\Phi_1)}(\Phi_1)), \tag{2.60}$$

(ii) it holds that

$$[\mathcal{L}(\Phi_1 \bullet \Phi_2) - 1] = [\mathcal{L}(\Phi_1) - 1] + [\mathcal{L}(\Phi_2) - 1], \tag{2.61}$$

(iii) *it holds that*

$$\mathcal{H}(\Phi_1 \bullet \Phi_2) = \mathcal{H}(\Phi_1) + \mathcal{H}(\Phi_2), \tag{2.62}$$

(iv) *it holds that*

$$\begin{aligned}\mathcal{P}(\Phi_1 \bullet \Phi_2) &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + \mathbb{D}_1(\Phi_1)(\mathbb{D}_{\mathcal{L}(\Phi_2)-1}(\Phi_2) + 1) \\ &\quad - \mathbb{D}_1(\Phi_1)(\mathbb{D}_0(\Phi_1) + 1) - \mathbb{D}_{\mathcal{L}(\Phi_2)}(\Phi_2)(\mathbb{D}_{\mathcal{L}(\Phi_2)-1}(\Phi_2) + 1) \\ &\leq \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + \mathbb{D}_1(\Phi_1)\mathbb{D}_{\mathcal{H}(\Phi_2)}(\Phi_2),\end{aligned} \tag{2.63}$$

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*

$$\mathcal{R}_a(\Phi_1 \bullet \Phi_2) = [\mathcal{R}_a(\Phi_1)] \circ [\mathcal{R}_a(\Phi_2)] \tag{2.64}$$

(cf. Definitions 2.2.3 and 2.2.5).

Proof of Proposition 2.2.7. 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 $l_{1,0}, l_{1,1}, \ldots, l_{1,\mathcal{L}(\Phi_1)}, l_{2,0}, l_{2,1}, \ldots, l_{2,\mathcal{L}(\Phi_2)} \in \mathbb{N}$, $\big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big) \in (\times_{j=1}^{L_k}(\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}}))$, $k \in \{1, 2\}$, satisfy for all $k \in \{1, 2\}$ that

$$\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big), \tag{2.65}$$

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

$$\forall\, j \in \mathbb{N} \cap (0, L_2) \colon\; x_j = \mathfrak{M}_{a,l_{2,j}}(W_{2,j}x_{j-1} + B_{2,j}) \tag{2.66}$$

(cf. Definition 2.1.4), 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 $y_0 = W_{2,L_2}x_{L_2-1} + B_{2,L_2}$ and

$$\forall\, j \in \mathbb{N} \cap (0, L_1) \colon\; y_j = \mathfrak{M}_{a,l_{1,j}}(W_{1,j}y_{j-1} + B_{1,j}), \tag{2.67}$$

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 $z_0 = x_0$ and

$$\forall\, j \in \mathbb{N} \cap (0, L_3) \colon\; z_j = \mathfrak{M}_{a,l_{3,j}}(W_{3,j}z_{j-1} + B_{3,j}). \tag{2.68}$$

Note that (2.58) ensures that

$$\Phi_3 = \Phi_1 \bullet \Phi_2 = \begin{cases} \big((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \dots, (W_{2,L_2-1}, B_{2,L_2-1}), \\ \quad (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), (W_{1,2}, B_{1,2}), \\ \quad (W_{1,3}, B_{1,3}), \dots, (W_{1,L_1}, B_{1,L_1})\big) & : (L_1 > 1) \wedge (L_2 > 1) \\ \big((W_{1,1}W_{2,1}, W_{1,1}B_{2,1} + B_{1,1}), (W_{1,2}, B_{1,2}), \\ \quad (W_{1,3}, B_{1,3}), \dots, (W_{1,L_1}, B_{1,L_1})\big) & : (L_1 > 1) \wedge (L_2 = 1) \\ \big((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \dots, (W_{2,L_2-1}, B_{2,L_2-1}), \\ \quad (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1})\big) & : (L_1 = 1) \wedge (L_2 > 1) \\ (W_{1,1}W_{2,1}, W_{1,1}B_{2,1} + B_{1,1}) & : (L_1 = 1) \wedge (L_2 = 1). \end{cases} \tag{2.69}$$

Hence, we obtain that

$$\begin{split} [\mathcal{L}(\Phi_1 \bullet \Phi_2) - 1] &= [(L_2 - 1) + 1 + (L_1 - 1)] - 1 \\ &= [L_1 - 1] + [L_2 - 1] = [\mathcal{L}(\Phi_1) - 1] + [\mathcal{L}(\Phi_2) - 1] \end{split} \tag{2.70}$$

$$\text{and} \qquad \mathcal{D}(\Phi_1 \bullet \Phi_2) = (l_{2,0}, l_{2,1}, \dots, l_{2,L_2-1}, l_{1,1}, l_{1,2}, \dots, l_{1,L_1}). \tag{2.71}$$

This establishes items (i), (ii), and (iii). In addition, observe that (2.71) demonstrates that

$$\begin{split} &\mathcal{P}(\Phi_1 \bullet \Phi_2) = \sum_{j=1}^{L_3} l_{3,j}(l_{3,j-1} + 1) \\ &= \left[\sum_{j=1}^{L_2-1} l_{3,j}(l_{3,j-1} + 1)\right] + l_{3,L_2}(l_{3,L_2-1} + 1) + \left[\sum_{j=L_2+1}^{L_3} l_{3,j}(l_{3,j-1} + 1)\right] \\ &= \left[\sum_{j=1}^{L_2-1} l_{2,j}(l_{2,j-1} + 1)\right] + l_{1,1}(l_{2,L_2-1} + 1) + \left[\sum_{j=L_2+1}^{L_3} l_{1,j-L_2+1}(l_{1,j-L_2} + 1)\right] \\ &= \left[\sum_{j=1}^{L_2-1} l_{2,j}(l_{2,j-1} + 1)\right] + \left[\sum_{j=2}^{L_1} l_{1,j}(l_{1,j-1} + 1)\right] + l_{1,1}\big(l_{2,L_2-1} + 1\big) \\ &= \left[\sum_{j=1}^{L_2} l_{2,j}(l_{2,j-1} + 1)\right] + \left[\sum_{j=1}^{L_1} l_{1,j}(l_{1,j-1} + 1)\right] + l_{1,1}(l_{2,L_2-1} + 1) \\ &\quad - l_{2,L_2}(l_{2,L_2-1} + 1) - l_{1,1}(l_{1,0} + 1) \\ &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1}(l_{2,L_2-1} + 1) - l_{2,L_2}(l_{2,L_2-1} + 1) \\ &\quad - l_{1,1}(l_{1,0} + 1) \\ &\leq \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1}l_{2,L_2-1}. \end{split} \tag{2.72}$$

This establishes item (iv). Moreover, observe that (2.69) and the fact that $a \in C(\mathbb{R}, \mathbb{R})$ ensure that

$$\mathcal{R}_a(\Phi_1 \bullet \Phi_2) \in C(\mathbb{R}^{l_{2,0}}, \mathbb{R}^{l_{1,L_1}}) = C(\mathbb{R}^{\mathcal{I}(\Phi_2)}, \mathbb{R}^{\mathcal{O}(\Phi_1)}). \tag{2.73}$$

Next note that (2.70) implies that $L_3 = L_1 + L_2 - 1$. This, (2.69), and (2.71) ensure that

$$(l_{3,0}, l_{3,1}, \dots, l_{3,L_1+L_2-1}) = (l_{2,0}, l_{2,1}, \dots, l_{2,L_2-1}, l_{1,1}, l_{1,2}, \dots, l_{1,L_1}), \tag{2.74}$$

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_2) \colon\ (W_{3,j}, B_{3,j}) = (W_{2,j}, B_{2,j})\big], \tag{2.75}$$

$$(W_{3,L_2}, B_{3,L_2}) = (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), \tag{2.76}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_2, L_1 + L_2) \colon\ (W_{3,j}, B_{3,j}) = (W_{1,j+1-L_2}, B_{1,j+1-L_2})\big]. \tag{2.77}$$

This, (2.66), (2.68), 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 (2.76) and the fact that $y_0 = W_{2,L_2} x_{L_2-1} + B_{2,L_2}$ ensures that

$$\begin{aligned} W_{3,L_2} z_{L_2-1} + B_{3,L_2} &= W_{3,L_2} x_{L_2-1} + B_{3,L_2} \\ &= W_{1,1} W_{2,L_2} x_{L_2-1} + W_{1,1} B_{2,L_2} + B_{1,1} \\ &= W_{1,1}(W_{2,L_2} x_{L_2-1} + B_{2,L_2}) + B_{1,1} = W_{1,1} y_0 + B_{1,1}. \end{aligned} \tag{2.78}$$

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

$$W_{3,j} z_{j-1} + B_{3,j} = W_{1,j+1-L_2} y_{j-L_2} + B_{1,j+1-L_2}. \tag{2.79}$$

We prove (2.79) by induction on $j \in \mathbb{N} \cap [L_2, L_1 + L_2)$. Note that (2.78) establishes (2.79) in the base case $j = L_2$. For the induction step note that the fact that $L_3 = L_1 + L_2 - 1$, (2.67), (2.68), (2.74), and (2.77) imply that for all $j \in \mathbb{N} \cap [L_2, \infty) \cap (0, L_1 + L_2 - 1)$ with

$$W_{3,j} z_{j-1} + B_{3,j} = W_{1,j+1-L_2} y_{j-L_2} + B_{1,j+1-L_2} \tag{2.80}$$

it holds that

$$\begin{aligned} &W_{3,j+1} z_j + B_{3,j+1} = W_{3,j+1} \mathfrak{M}_{a,l_{3,j}}(W_{3,j} z_{j-1} + B_{3,j}) + B_{3,j+1} \\ &= W_{1,j+2-L_2} \mathfrak{M}_{a,l_{1,j+1-L_2}}(W_{1,j+1-L_2} y_{j-L_2} + B_{1,j+1-L_2}) + B_{1,j+2-L_2} \\ &= W_{1,j+2-L_2} y_{j+1-L_2} + B_{1,j+2-L_2}. \end{aligned} \tag{2.81}$$

Induction hence proves (2.79). Next observe that (2.79) and the fact that $L_3 = L_1 + L_2 - 1$ assure that

$$W_{3,L_3} z_{L_3-1} + B_{3,L_3} = W_{3,L_1+L_2-1} z_{L_1+L_2-2} + B_{3,L_1+L_2-1} = W_{1,L_1} y_{L_1-1} + B_{1,L_1}. \tag{2.82}$$

The fact that $\Phi_3 = \Phi_1 \bullet \Phi_2$, (2.66), (2.67), and (2.68) therefore prove that

$$\begin{aligned} [\mathcal{R}_a(\Phi_1 \bullet \Phi_2)](x_0) &= [\mathcal{R}_a(\Phi_3)](x_0) = [\mathcal{R}_a(\Phi_3)](z_0) = W_{3,L_3} z_{L_3-1} + B_{3,L_3} \\ &= W_{1,L_1} y_{L_1-1} + B_{1,L_1} = [\mathcal{R}_a(\Phi_1)](y_0) \\ &= [\mathcal{R}_a(\Phi_1)]\big(W_{2,L_2} x_{L_2-1} + B_{2,L_2}\big) \\ &= [\mathcal{R}_a(\Phi_1)]\big([\mathcal{R}_a(\Phi_2)](x_0)\big) = [(\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))](x_0). \end{aligned} \tag{2.83}$$

Combining this with (2.73) establishes item (v). The proof of Proposition 2.2.7 is thus complete. □

2.2.3.3 Associativity of standard compositions of ANNs

Lemma 2.2.8. *Let $\Phi_1, \Phi_2, \Phi_3 \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ and $\mathcal{I}(\Phi_2) = \mathcal{O}(\Phi_3)$ (cf. Definition 2.2.1). Then it holds that*

$$(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3) \tag{2.84}$$

(cf. Definition 2.2.5).

Proof of Lemma 2.2.8. Throughout this proof let $\Phi_4, \Phi_5, \Phi_6, \Phi_7 \in \mathbf{N}$ satisfy $\Phi_4 = \Phi_1 \bullet \Phi_2$, $\Phi_5 = \Phi_2 \bullet \Phi_3$, $\Phi_6 = \Phi_4 \bullet \Phi_3$, and $\Phi_7 = \Phi_1 \bullet \Phi_5$, let $L_k \in \mathbb{N}$, $k \in \{1, 2, \dots, 7\}$, satisfy for all $k \in \{1, 2, \dots, 7\}$ that $L_k = \mathcal{L}(\Phi_k)$, let $l_{k,0}, l_{k,1}, \dots, l_{k,L_k} \in \mathbb{N}$, $k \in \{1, 2, \dots, 7\}$, and let $\big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \dots, (W_{k,L_k}, B_{k,L_k})\big) \in (\bigtimes_{j=1}^{L_k} (\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}}))$, $k \in \{1, 2, \dots, 7\}$, satisfy for all $k \in \{1, 2, \dots, 7\}$ that

$$\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \dots, (W_{k,L_k}, B_{k,L_k})\big). \tag{2.85}$$

Observe that item (ii) in Proposition 2.2.7 and the fact that for all $k \in \{1, 2, 3\}$ it holds that $\mathcal{L}(\Phi_k) = L_k$ proves that

$$\begin{aligned}
\mathcal{L}(\Phi_6) &= \mathcal{L}((\Phi_1 \bullet \Phi_2) \bullet \Phi_3) = \mathcal{L}(\Phi_1 \bullet \Phi_2) + \mathcal{L}(\Phi_3) - 1 \\
&= \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2) + \mathcal{L}(\Phi_3) - 2 = L_1 + L_2 + L_3 - 2 \\
&= \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2 \bullet \Phi_3) - 1 = \mathcal{L}(\Phi_1 \bullet (\Phi_2 \bullet \Phi_3)) = \mathcal{L}(\Phi_7).
\end{aligned} \tag{2.86}$$

Next note that Lemma 2.2.6, (2.85), and the fact that $\Phi_4 = \Phi_1 \bullet \Phi_2$ imply that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_2) \colon\ (W_{4,j}, B_{4,j}) = (W_{2,j}, B_{2,j})\big], \tag{2.87}$$

$$(W_{4,L_2}, B_{4,L_2}) = (W_{1,1} W_{2,L_2}, W_{1,1} B_{2,L_2} + B_{1,1}), \tag{2.88}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_2, L_1 + L_2) \colon\ (W_{4,j}, B_{4,j}) = (W_{1,j+1-L_2}, B_{1,j+1-L_2})\big]. \tag{2.89}$$

Hence, we obtain that

$$\big[\forall\, j \in \mathbb{N} \cap (L_3 - 1, L_2 + L_3 - 1) \colon (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3})\big], \tag{2.90}$$

$$(W_{4,L_2}, B_{4,L_2}) = (W_{1,1} W_{2,L_2}, W_{1,1} B_{2,L_2} + B_{1,1}), \tag{2.91}$$

and

$$\begin{aligned}
&\big[\forall\, j \in \mathbb{N} \cap (L_2 + L_3 - 1, L_1 + L_2 + L_3 - 1) \colon \\
&\qquad\qquad (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{1,j+2-L_2-L_3}, B_{1,j+2-L_2-L_3})\big].
\end{aligned} \tag{2.92}$$

In addition, observe that Lemma 2.2.6, (2.85), and the fact that $\Phi_5 = \Phi_2 \bullet \Phi_3$ demonstrate that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_3) \colon\ (W_{5,j}, B_{5,j}) = (W_{3,j}, B_{3,j})\big], \tag{2.93}$$

$$(W_{5,L_3}, B_{5,L_3}) = (W_{2,1}W_{3,L_3}, W_{2,1}B_{3,L_3} + B_{2,1}), \tag{2.94}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_3, L_2 + L_3)\colon\ (W_{5,j}, B_{5,j}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3})\big]. \tag{2.95}$$

Moreover, note that Lemma 2.2.6, (2.85), and the fact that $\Phi_6 = \Phi_4 \bullet \Phi_3$ ensure that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_3)\colon\ (W_{6,j}, B_{6,j}) = (W_{3,j}, B_{3,j})\big], \tag{2.96}$$

$$(W_{6,L_3}, B_{6,L_3}) = (W_{4,1}W_{3,L_3}, W_{4,1}B_{3,L_3} + B_{4,1}), \tag{2.97}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_3, L_4 + L_3)\colon\ (W_{6,j}, B_{6,j}) = (W_{4,j+1-L_3}, B_{4,j+1-L_3})\big]. \tag{2.98}$$

Furthermore, observe that Lemma 2.2.6, (2.85), and the fact that $\Phi_7 = \Phi_1 \bullet \Phi_5$ show that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_5)\colon\ (W_{7,j}, B_{7,j}) = (W_{5,j}, B_{5,j})\big], \tag{2.99}$$

$$(W_{7,L_5}, B_{7,L_5}) = (W_{1,1}W_{5,L_5}, W_{1,1}B_{5,L_5} + B_{1,1}), \tag{2.100}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_5, L_1 + L_5)\colon\ (W_{7,j}, B_{7,j}) = (W_{1,j+1-L_5}, B_{1,j+1-L_5})\big]. \tag{2.101}$$

This, the fact that $L_3 \leq L_2 + L_3 - 1 = L_5$, (2.93), and (2.96) imply that for all $j \in \mathbb{N} \cap (0, L_3)$ it holds that

$$(W_{6,j}, B_{6,j}) = (W_{3,j}, B_{3,j}) = (W_{5,j}, B_{5,j}) = (W_{7,j}, B_{7,j}). \tag{2.102}$$

In addition, observe that (2.87), (2.88), (2.93), (2.94), (2.97), (2.99), (2.100), and the fact that $L_5 = L_2 + L_3 - 1$ demonstrate that

$$\begin{aligned}
(W_{6,L_3}, B_{6,L_3}) &= (W_{4,1}W_{3,L_3}, W_{4,1}B_{3,L_3} + B_{4,1}) \\
&= \begin{cases} (W_{2,1}W_{3,L_3}, W_{2,1}B_{3,L_3} + B_{2,1}) & : L_2 > 1 \\ (W_{1,1}W_{2,1}W_{3,L_3}, W_{1,1}W_{2,1}B_{3,L_3} + W_{1,1}B_{2,1} + B_{1,1}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{2,1}W_{3,L_3}, W_{2,1}B_{3,L_3} + B_{2,1}) & : L_2 > 1 \\ (W_{1,1}(W_{2,1}W_{3,L_3}), W_{1,1}(W_{2,1}B_{3,L_3} + B_{2,1}) + B_{1,1}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{5,L_3}, B_{5,L_3}) & : L_2 > 1 \\ (W_{1,1}W_{5,L_3}, W_{1,1}B_{5,L_3} + B_{1,1}) & : L_2 = 1 \end{cases} \\
&= (W_{7,L_3}, B_{7,L_3}).
\end{aligned} \tag{2.103}$$

Next note that the fact that $L_5 = L_2 + L_3 - 1 < L_1 + L_2 + L_3 - 1 = L_3 + L_4$, (2.98), (2.90), (2.95), and (2.99) ensure that for all $j \in \mathbb{N}$ with $L_3 < j < L_5$ it holds that

$$\begin{aligned}
(W_{6,j}, B_{6,j}) &= (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3}) \\
&= (W_{5,j}, B_{5,j}) = (W_{7,j}, B_{7,j}).
\end{aligned} \tag{2.104}$$

Moreover, observe that the fact that $L_5 = L_2 + L_3 - 1 < L_1 + L_2 + L_3 - 1 = L_3 + L_4$, (2.98), (2.103), (2.88), (2.95), and (2.100) prove that

$$\begin{aligned}
(W_{6,L_5}, B_{6,L_5}) &= \begin{cases} (W_{4,L_5+1-L_3}, B_{4,L_5+1-L_3}) & : L_2 > 1 \\ (W_{6,L_3}, B_{6,L_3}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{4,L_2}, B_{4,L_2}) & : L_2 > 1 \\ (W_{7,L_3}, B_{7,L_3}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}) & : L_2 > 1 \\ (W_{7,L_5}, B_{7,L_5}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{1,1}W_{5,L_5}, W_{1,1}B_{5,L_5} + B_{1,1}) & : L_2 > 1 \\ (W_{7,L_5}, B_{7,L_5}) & : L_2 = 1 \end{cases} \\
&= (W_{7,L_5}, B_{7,L_5}).
\end{aligned} \tag{2.105}$$

Furthermore, note that (2.98), (2.92), (2.101), 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

$$\begin{aligned}
(W_{6,j}, B_{6,j}) &= (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{1,j+2-L_2-L_3}, B_{1,j+2-L_2-L_3}) \\
&= (W_{1,j+1-L_5}, B_{1,j+1-L_5}) = (W_{7,j}, B_{7,j}).
\end{aligned} \tag{2.106}$$

Combining this with (2.86), (2.102), (2.103), (2.104), and (2.105) establishes that

$$(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_4 \bullet \Phi_3 = \Phi_6 = \Phi_7 = \Phi_1 \bullet \Phi_5 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3). \tag{2.107}$$

The proof of Lemma 2.2.8 is thus complete. □

2.2.3.4 Powers of ANNs

Definition 2.2.9. *Let $d \in \mathbb{N}$. Then we denote by $\mathrm{I}_d \in \mathbb{R}^{d \times d}$ the identity matrix in $\mathbb{R}^{d \times d}$.*

Definition 2.2.10. *We denote by $(\cdot)^{\bullet n} \colon \{\Phi \in \mathbf{N} \colon \mathcal{I}(\Phi) = \mathcal{O}(\Phi)\} \to \mathbf{N}$, $n \in \mathbb{N}_0$, the functions which satisfy for all $n \in \mathbb{N}_0$, $\Phi \in \mathbf{N}$ with $\mathcal{I}(\Phi) = \mathcal{O}(\Phi)$ that*

$$\Phi^{\bullet n} = \begin{cases} \big(\mathrm{I}_{\mathcal{O}(\Phi)}, (0, 0, \dots, 0)\big) \in \mathbb{R}^{\mathcal{O}(\Phi) \times \mathcal{O}(\Phi)} \times \mathbb{R}^{\mathcal{O}(\Phi)} & : n = 0 \\ \Phi \bullet (\Phi^{\bullet (n-1)}) & : n \in \mathbb{N} \end{cases} \tag{2.108}$$

(cf. Definitions 2.2.1, 2.2.5, and 2.2.9).

2.2.4 Parallelizations of ANNs

2.2.4.1 Parallelizations of ANNs with the same length

Definition 2.2.11 (Parallelization of ANNs)**.** *Let $n \in \mathbb{N}$. Then we denote by*

$$\mathbf{P}_n \colon \big\{(\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n \colon \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)\big\} \to \mathbf{N} \tag{2.109}$$

the function which satisfies for all $L \in \mathbb{N}$, $\Phi_1, \Phi_2, \dots, \Phi_n \in \mathbf{N}$ with $L = \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \cdots = \mathcal{L}(\Phi_n)$ that

$$\begin{split}\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n) = \Bigg(&\left(\begin{pmatrix} \mathcal{W}_{1,\Phi_1} & 0 & 0 & \cdots & 0 \\ 0 & \mathcal{W}_{1,\Phi_2} & 0 & \cdots & 0 \\ 0 & 0 & \mathcal{W}_{1,\Phi_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathcal{W}_{1,\Phi_n} \end{pmatrix}, \begin{pmatrix} \mathcal{B}_{1,\Phi_1} \\ \mathcal{B}_{1,\Phi_2} \\ \mathcal{B}_{1,\Phi_3} \\ \vdots \\ \mathcal{B}_{1,\Phi_n} \end{pmatrix} \right), \\ &\left(\begin{pmatrix} \mathcal{W}_{2,\Phi_1} & 0 & 0 & \cdots & 0 \\ 0 & \mathcal{W}_{2,\Phi_2} & 0 & \cdots & 0 \\ 0 & 0 & \mathcal{W}_{2,\Phi_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathcal{W}_{2,\Phi_n} \end{pmatrix}, \begin{pmatrix} \mathcal{B}_{2,\Phi_1} \\ \mathcal{B}_{2,\Phi_2} \\ \mathcal{B}_{2,\Phi_3} \\ \vdots \\ \mathcal{B}_{2,\Phi_n} \end{pmatrix} \right), \dots, \\ &\left(\begin{pmatrix} \mathcal{W}_{L,\Phi_1} & 0 & 0 & \cdots & 0 \\ 0 & \mathcal{W}_{L,\Phi_2} & 0 & \cdots & 0 \\ 0 & 0 & \mathcal{W}_{L,\Phi_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathcal{W}_{L,\Phi_n} \end{pmatrix}, \begin{pmatrix} \mathcal{B}_{L,\Phi_1} \\ \mathcal{B}_{L,\Phi_2} \\ \mathcal{B}_{L,\Phi_3} \\ \vdots \\ \mathcal{B}_{L,\Phi_n} \end{pmatrix} \right) \Bigg)\end{split} \tag{2.110}$$

(cf. Definition 2.2.1).

Lemma 2.2.12. *Let $n, L \in \mathbb{N}$, $\Phi_1, \Phi_2, \dots, \Phi_n \in \mathbf{N}$ satisfy $L = \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)$ (cf. Definition 2.2.1). Then it holds that*

$$\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n) \in \left(\bigtimes_{k=1}^{L} \left(\mathbb{R}^{(\sum_{j=1}^n \mathbb{D}_k(\Phi_j)) \times (\sum_{j=1}^n \mathbb{D}_{k-1}(\Phi_j))} \times \mathbb{R}^{(\sum_{j=1}^n \mathbb{D}_k(\Phi_j))} \right) \right) \tag{2.111}$$

(cf. Definition 2.2.11).

Proof of Lemma 2.2.12. Note that (2.110) establishes (2.111). The proof of Lemma 2.2.12 is thus complete. □

Proposition 2.2.13. *Let $a \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy $\mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)$ (cf. Definition 2.2.1). Then*

(i) *it holds that*

$$\mathcal{R}_a(\mathbf{P}_n(\Phi)) \in C\big(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\big) \tag{2.112}$$

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*

$$\begin{split}&\big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(x_1, x_2, \ldots, x_n)\\ &= \big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \ldots, (\mathcal{R}_a(\Phi_n))(x_n)\big) \in \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\end{split} \tag{2.113}$$

(cf. Definitions 2.2.3 and 2.2.11).

Proof of Proposition 2.2.13. Throughout this proof let $L \in \mathbb{N}$ satisfy $L = \mathcal{L}(\Phi_1)$, 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 $\big((W_{j,1}, B_{j,1}), (W_{j,2}, B_{j,2}), \ldots, (W_{j,L}, B_{j,L})\big) \in (\times_{k=1}^L (\mathbb{R}^{l_{j,k} \times l_{j,k-1}} \times \mathbb{R}^{l_{j,k}}))$, $j \in \{1, 2, \ldots, n\}$, satisfy for all $j \in \{1, 2, \ldots, n\}$ that

$$\Phi_j = \big((W_{j,1}, B_{j,1}), (W_{j,2}, B_{j,2}), \ldots, (W_{j,L}, B_{j,L})\big), \tag{2.114}$$

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

$$\mathbf{P}_n(\Phi) = \big((A_1, b_1), (A_2, b_2), \ldots, (A_L, b_L)\big) \tag{2.115}$$

(cf. Lemma 2.2.12), 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

$$x_{j,k} = \mathfrak{M}_{a, l_{j,k}}(W_{j,k} x_{j,k-1} + B_{j,k}) \tag{2.116}$$

(cf. Definition 2.1.4), 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 (2.115) 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 2.2.4, 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

$$\begin{split}\mathcal{R}_a(\mathbf{P}_n(\Phi)) \in C(\mathbb{R}^{\alpha_0}, \mathbb{R}^{\alpha_L}) &= C\big(\mathbb{R}^{[\sum_{j=1}^n l_{j,0}]}, \mathbb{R}^{[\sum_{j=1}^n l_{j,L}]}\big)\\ &= C\big(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\big).\end{split} \tag{2.117}$$

This proves item (i). Moreover, observe that (2.110) and (2.115) demonstrate that for all $k \in \{1, 2, \ldots, L\}$ it holds that

$$A_k = \begin{pmatrix} W_{1,k} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,k} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,k} \end{pmatrix} \quad \text{and} \quad b_k = \begin{pmatrix} B_{1,k} \\ B_{2,k} \\ B_{3,k} \\ \vdots \\ B_{n,k} \end{pmatrix}. \tag{2.118}$$

Combining this with (2.8), (2.116), and the fact that for all $k \in \mathbb{N} \cap [0, L)$ it holds that $\mathfrak{x}_k = (x_{1,k}, x_{2,k}, \dots, x_{n,k})$ implies that for all $k \in \mathbb{N} \cap (0, L)$ it holds that

$$\mathfrak{M}_{a,\alpha_k}(A_k \mathfrak{x}_{k-1} + b_k) = \begin{pmatrix} \mathfrak{M}_{a,l_{1,k}}(W_{1,k}x_{1,k-1} + B_{1,k}) \\ \mathfrak{M}_{a,l_{2,k}}(W_{2,k}x_{2,k-1} + B_{2,k}) \\ \vdots \\ \mathfrak{M}_{a,l_{n,k}}(W_{n,k}x_{n,k-1} + B_{n,k}) \end{pmatrix} = \begin{pmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{pmatrix} = \mathfrak{x}_k. \tag{2.119}$$

This, (2.53), (2.114), (2.115), (2.116), (2.118), the fact that $\mathfrak{x}_0 = (x_{1,0}, x_{2,0}, \dots, x_{n,0})$, and the fact that $\mathfrak{x}_{L-1} = (x_{1,L-1}, x_{2,L-1}, \dots, x_{n,L-1})$ ensure that

$$\begin{split} &\big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(x_{1,0}, x_{2,0}, \dots, x_{n,0}) = \big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(\mathfrak{x}_0) \\ &= A_L \mathfrak{x}_{L-1} + b_L = \begin{pmatrix} W_{1,L}x_{1,L-1} + B_{1,L} \\ W_{2,L}x_{2,L-1} + B_{2,L} \\ \vdots \\ W_{n,L}x_{n,L-1} + B_{n,L} \end{pmatrix} = \begin{pmatrix} (\mathcal{R}_a(\Phi_1))(x_{1,0}) \\ (\mathcal{R}_a(\Phi_2))(x_{2,0}) \\ \vdots \\ (\mathcal{R}_a(\Phi_n))(x_{n,0}) \end{pmatrix}. \end{split} \tag{2.120}$$

This establishes item (ii). The proof of Proposition 2.2.13 is thus complete. □

Proposition 2.2.14. *Let $n, L \in \mathbb{N}$, $\Phi_1, \Phi_2, \dots, \Phi_n \in \mathbf{N}$ satisfy $L = \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)$ (cf. Definition 2.2.1). Then*

(i) it holds for all $k \in \mathbb{N}_0$ that

$$\mathbb{D}_k(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) = \mathbb{D}_k(\Phi_1) + \mathbb{D}_k(\Phi_2) + \ldots + \mathbb{D}_k(\Phi_n), \tag{2.121}$$

(ii) it holds that

$$\mathcal{D}\big(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)\big) = \mathcal{D}(\Phi_1) + \mathcal{D}(\Phi_2) + \cdots + \mathcal{D}(\Phi_n), \tag{2.122}$$

and

(iii) it holds that

$$\mathcal{P}\big(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)\big) \le \tfrac{1}{2}\big[\textstyle\sum_{j=1}^n \mathcal{P}(\Phi_j)\big]^2 \tag{2.123}$$

(cf. Definition 2.2.11).

Proof of Proposition 2.2.14. 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 2.2.12

establishes item (i). In addition, observe that item (i) implies item (ii). Moreover, note that item (i) demonstrates that

$$
\begin{aligned}
\mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) &= \sum_{k=1}^{L} \Big[\textstyle\sum_{i=1}^{n} l_{i,k}\Big]\Big[\big(\textstyle\sum_{i=1}^{n} l_{i,k-1}\big) + 1\Big] \\
&= \sum_{k=1}^{L} \Big[\textstyle\sum_{i=1}^{n} l_{i,k}\Big]\Big[\big(\textstyle\sum_{j=1}^{n} l_{j,k-1}\big) + 1\Big] \\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{L} l_{i,k}(l_{j,k-1} + 1) \leq \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k,\ell=1}^{L} l_{i,k}(l_{j,\ell-1} + 1) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \Big[\textstyle\sum_{k=1}^{L} l_{i,k}\Big]\Big[\textstyle\sum_{\ell=1}^{L}(l_{j,\ell-1} + 1)\Big] \\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n} \Big[\textstyle\sum_{k=1}^{L} \tfrac{1}{2} l_{i,k}(l_{i,k-1} + 1)\Big]\Big[\textstyle\sum_{\ell=1}^{L} l_{j,\ell}(l_{j,\ell-1} + 1)\Big] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \tfrac{1}{2}\mathcal{P}(\Phi_i)\mathcal{P}(\Phi_j) = \tfrac{1}{2}\Big[\textstyle\sum_{i=1}^{n} \mathcal{P}(\Phi_i)\Big]^2.
\end{aligned}
\tag{2.124}
$$

The proof of Proposition 2.2.14 is thus complete. □

Corollary 2.2.15. *Let $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_n)$ (cf. Definition 2.2.1). Then*

$$
\big[\tfrac{n^2}{2}\big]\mathcal{P}(\Phi_1) \leq \big[\tfrac{n^2+n}{2}\big]\mathcal{P}(\Phi_1) \leq \mathcal{P}(\mathbf{P}_n(\Phi)) \leq n^2\mathcal{P}(\Phi_1) \leq \tfrac{1}{2}\big[\textstyle\sum_{i=1}^{n} \mathcal{P}(\Phi_i)\big]^2 \tag{2.125}
$$

(cf. Definition 2.2.11).

Proof of Corollary 2.2.15. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$ satisfy $\mathcal{D}(\Phi_1) = (l_0, l_1, \dots, l_L)$. Note that item (ii) in Proposition 2.2.14 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

$$
\mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) = \sum_{j=1}^{L} (n l_j)\big((n l_{j-1}) + 1\big). \tag{2.126}
$$

Hence, we obtain that

$$
\mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) \leq \sum_{j=1}^{L} (n l_j)\big((n l_{j-1}) + n\big) = n^2\left[\sum_{j=1}^{L} l_j(l_{j-1} + 1)\right] = n^2\mathcal{P}(\Phi_1). \tag{2.127}
$$

Furthermore, observe that the assumption that $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_n)$ and the fact that $\mathcal{P}(\Phi_1) \geq l_1(l_0 + 1) \geq 2$ ensure that

$$
n^2\mathcal{P}(\Phi_1) \leq \tfrac{n^2}{2}[\mathcal{P}(\Phi_1)]^2 = \tfrac{1}{2}[n\mathcal{P}(\Phi_1)]^2 = \tfrac{1}{2}\left[\sum_{i=1}^{n} \mathcal{P}(\Phi_1)\right]^2 = \tfrac{1}{2}\left[\sum_{i=1}^{n} \mathcal{P}(\Phi_i)\right]^2. \tag{2.128}
$$

Next note that (2.126) and the fact that for all $a, b \in \mathbb{N}$ it holds that $2(ab+1) = ab+1+(a-1)(b-1)+a+b \geq ab+a+b+1 = (a+1)(b+1)$ show that

$$\begin{aligned} \mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) &\geq \tfrac{1}{2}\left[\sum_{j=1}^{L}(nl_j)(n+1)(l_{j-1}+1)\right] \\ &= \tfrac{n(n+1)}{2}\left[\sum_{j=1}^{L} l_j(l_{j-1}+1)\right] = \left[\tfrac{n^2+n}{2}\right]\mathcal{P}(\Phi_1). \end{aligned} \tag{2.129}$$

This, (2.127), and (2.128) establish (2.125). The proof of Corollary 2.2.15 is thus complete. □

Exercise 2.2.2. *Prove or disprove the following statement: For every $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ with $\mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \dots = \mathcal{L}(\Phi_n)$ it holds that $\mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) \leq n\big[\sum_{i=1}^{n} \mathcal{P}(\Phi_i)\big]$.*

Exercise 2.2.3. *Prove or disprove the following statement: For every $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ with $\mathcal{P}(\Phi_1) = \mathcal{P}(\Phi_2) = \dots = \mathcal{P}(\Phi_n)$ it holds that $\mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n)) \leq n^2\mathcal{P}(\Phi_1)$.*

2.2.5 Representations of the identities with rectifier functions

Definition 2.2.16 (Identity ANNs)**.** *We denote by $\mathfrak{I}_d \in \mathbf{N}$, $d \in \mathbb{N}$, the ANNs which satisfy for all $d \in \mathbb{N}$ that*

$$\mathfrak{I}_1 = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right), \left(\begin{pmatrix} 1 & -1 \end{pmatrix}, 0\right)\right) \in \left((\mathbb{R}^{2\times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1\times 2} \times \mathbb{R}^1)\right) \tag{2.130}$$

and

$$\mathfrak{I}_d = \mathbf{P}_d(\mathfrak{I}_1, \mathfrak{I}_1, \dots, \mathfrak{I}_1) \tag{2.131}$$

(cf. Definitions 2.2.1 and 2.2.11).

Lemma 2.2.17. *Let $d \in \mathbb{N}$. Then*

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

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

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

$$(\mathcal{R}_{\mathfrak{r}}(\mathfrak{I}_d))(x) = x \tag{2.132}$$

(cf. Definitions 2.2.1, 2.2.3, and 2.2.16).

Proof of Lemma 2.2.17. Throughout this proof let $L = 2$, $l_0 = 1$, $l_1 = 2$, $l_2 = 1$. Note that (2.130) ensures that

$$\mathcal{D}(\mathfrak{I}_1) = (1, 2, 1) = (l_0, l_1, l_2). \tag{2.133}$$

This, (2.131), and Proposition 2.2.14 prove that $\mathcal{D}(\mathfrak{I}_d) = (d, 2d, d) \in \mathbb{N}^3$. This establishes item (i). Next note that (2.130) assures that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathfrak{I}_1))(x) = \mathfrak{r}(x) - \mathfrak{r}(-x) = \max\{x, 0\} - \max\{-x, 0\} = x. \tag{2.134}$$

Combining this and Proposition 2.2.13 demonstrates that for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{R}_\mathfrak{r}(\mathfrak{I}_d) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and

$$\begin{aligned}(\mathcal{R}_\mathfrak{r}(\mathfrak{I}_d))(x) &= \big(\mathcal{R}_\mathfrak{r}\big(\mathbf{P}_d(\mathfrak{I}_1, \mathfrak{I}_1, \ldots, \mathfrak{I}_1)\big)\big)(x_1, x_2, \ldots, x_d) \\ &= \big((\mathcal{R}_\mathfrak{r}(\mathfrak{I}_1))(x_1), (\mathcal{R}_\mathfrak{r}(\mathfrak{I}_1))(x_2), \ldots, (\mathcal{R}_\mathfrak{r}(\mathfrak{I}_1))(x_d)\big) \\ &= (x_1, x_2, \ldots, x_d) = x\end{aligned} \tag{2.135}$$

(cf. Definition 2.2.11). This establishes items (ii) and (iii). The proof of Lemma 2.2.17 is thus complete. □

2.2.6 Scalar multiplications of ANNs

2.2.6.1 Affine transformations as ANNs

Definition 2.2.18 (Affine linear transformation ANN)**.** *Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^m$. Then we denote by $\mathbf{A}_{W,B} \in (\mathbb{R}^{m \times n} \times \mathbb{R}^m) \subseteq \mathbf{N}$ the ANN given by $\mathbf{A}_{W,B} = (W, B)$ (cf. Definitions 2.2.1 and 2.2.2).*

Lemma 2.2.19. *Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^m$. Then*

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

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

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

(cf. Definitions 2.2.1, 2.2.3, and 2.2.18).

Proof of Lemma 2.2.19. 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 (2.53) 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

$$(\mathcal{R}_a(\mathbf{A}_{W,B}))(x) = Wx + B. \tag{2.136}$$

This establishes items (ii) and (iii). The proof of Lemma 2.2.19 is thus complete. □

Lemma 2.2.20. *Let $\Phi \in \mathbf{N}$ (cf. Definition 2.2.1). Then*

(i) it holds for all $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathcal{O}(\Phi)}$, $B \in \mathbb{R}^m$ that $\mathcal{D}(\mathbf{A}_{W,B} \bullet \Phi) = (\mathbb{D}_0(\Phi), \mathbb{D}_1(\Phi), \ldots, \mathbb{D}_{\mathcal{H}(\Phi)}(\Phi), m)$

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

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

$$(\mathcal{R}_a(\mathbf{A}_{W,B} \bullet \Phi))(x) = W((\mathcal{R}_a(\Phi))(x)) + B, \tag{2.137}$$

(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* $\mathcal{D}(\Phi \bullet \mathbf{A}_{W,B}) = (n, \mathbb{D}_1(\Phi), \mathbb{D}_2(\Phi), \dots, \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi))$,

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

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

$$(\mathcal{R}_a(\Phi \bullet \mathbf{A}_{W,B}))(x) = (\mathcal{R}_a(\Phi))(Wx + B), \tag{2.138}$$

(cf. Definitions 2.2.3, 2.2.5, and 2.2.18).

Proof of Lemma 2.2.20. Note that Lemma 2.2.19 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

$$(\mathcal{R}_a(\mathbf{A}_{W,B}))(x) = Wx + B. \tag{2.139}$$

Combining this and Proposition 2.2.7 establishes items (i), (ii), (iii), (iv), (v), and (vi). The proof of Lemma 2.2.20 is thus complete. □

2.2.6.2 Scalar multiplications of ANNs

Definition 2.2.21 (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*

$$\lambda \circledast \Phi = \mathbf{A}_{\lambda \mathrm{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi \tag{2.140}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.9, and 2.2.18).

Lemma 2.2.22. *Let* $\lambda \in \mathbb{R}$, $\Phi \in \mathbf{N}$ *(cf. Definition 2.2.1). 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*

$$(\mathcal{R}_a(\lambda \circledast \Phi))(x) = \lambda\big((\mathcal{R}_a(\Phi))(x)\big) \tag{2.141}$$

(cf. Definitions 2.2.3 and 2.2.21).

Proof of Lemma 2.2.22. Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$ satisfy $L = \mathcal{L}(\Phi)$ and $(l_0, l_1, \dots, l_L) = \mathcal{D}(\Phi)$. Note that item (i) in Lemma 2.2.19 proves that

$$\mathcal{D}(\mathbf{A}_{\lambda \operatorname{I}_{\mathcal{O}(\Phi)},0}) = (\mathcal{O}(\Phi), \mathcal{O}(\Phi)) \tag{2.142}$$

(cf. Definitions 2.2.9 and 2.2.18). Combining this and item (i) in Lemma 2.2.20 assures that

$$\mathcal{D}(\lambda \circledast \Phi) = \mathcal{D}(\mathbf{A}_{\lambda \operatorname{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi) = (l_0, l_1, \dots, l_{L-1}, \mathcal{O}(\Phi)) = \mathcal{D}(\Phi). \tag{2.143}$$

This establishes item (i). Moreover, observe that items (ii) and (iii) in Lemma 2.2.20 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

$$\begin{aligned}(\mathcal{R}_a(\lambda \circledast \Phi))(x) &= (\mathcal{R}_a(\mathbf{A}_{\lambda \operatorname{I}_{\mathcal{O}(\Phi)},0} \bullet \Phi))(x) \\ &= \lambda \operatorname{I}_{\mathcal{O}(\Phi)}\big((\mathcal{R}_a(\Phi))(x)\big) = \lambda\big((\mathcal{R}_a(\Phi))(x)\big).\end{aligned} \tag{2.144}$$

This establishes items (ii) and (iii). The proof of Lemma 2.2.22 is thus complete. □

2.2.7 Sums of ANNs with the same length

2.2.7.1 Sums of vectors as ANNs

Definition 2.2.23. *Let $m, n \in \mathbb{N}$. Then we denote by $\mathbb{S}_{m,n} \in (\mathbb{R}^{m \times (mn)} \times \mathbb{R}^m)$ the ANN given by*

$$\mathbb{S}_{m,n} = \mathbf{A}_{(\operatorname{I}_m \ \operatorname{I}_m \ \dots \ \operatorname{I}_m),0} \tag{2.145}$$

(cf. Definitions 2.2.9 and 2.2.18).

Lemma 2.2.24. *Let $m, n \in \mathbb{N}$. Then*

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

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

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

$$(\mathcal{R}_a(\mathbb{S}_{m,n}))(x_1, x_2, \dots, x_n) = \sum_{k=1}^{n} x_k \tag{2.146}$$

(cf. Definitions 2.2.1, 2.2.3, and 2.2.23).

Proof of Lemma 2.2.24. Note that the fact that $\mathbb{S}_{m,n} \in (\mathbb{R}^{m\times(mn)} \times \mathbb{R}^m)$ ensures that $\mathbb{S}_{m,n} \in \mathbf{N}$ and $\mathcal{D}(\mathbb{S}_{m,n}) = (mn, m) \in \mathbb{N}^2$. This establishes items (i) and (ii). Next observe that items (ii) and (iii) in Lemma 2.2.19 prove that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathbb{S}_{m,n}) \in C(\mathbb{R}^{mn}, \mathbb{R}^m)$ and

$$\begin{split}(\mathcal{R}_a(\mathbb{S}_{m,n}))(x_1, x_2, \ldots, x_n) &= \big(\mathcal{R}_a\big(\mathbf{A}_{(\mathrm{I}_m\ \mathrm{I}_m\ \ldots\ \mathrm{I}_m),0}\big)\big)(x_1, x_2, \ldots, x_n) \\ &= (\mathrm{I}_m\ \ \mathrm{I}_m\ \ \ldots\ \ \mathrm{I}_m)(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} x_k\end{split} \tag{2.147}$$

(cf. Definition 2.2.9 and Definition 2.2.18). This establishes items (iii) and (iv). The proof of Lemma 2.2.24 is thus complete. □

Lemma 2.2.25. *Let $m, n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \mathbf{N}$ satisfy $\mathcal{O}(\Phi) = nm$ (cf. Definition 2.2.1). Then*

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

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

$$\big(\mathcal{R}_a(\mathbb{S}_{m,n} \bullet \Phi)\big)(x) = \sum_{k=1}^{n} y_k \tag{2.148}$$

(cf. Definitions 2.2.3, 2.2.5, and 2.2.23).

Proof of Lemma 2.2.25. Note that Lemma 2.2.24 ensures that for all $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathbb{S}_{m,n}) \in C(\mathbb{R}^{nm}, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathbb{S}_{m,n}))(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} x_k. \tag{2.149}$$

Combining this and item (v) in Proposition 2.2.7 establishes items (i) and (ii). The proof of Lemma 2.2.25 is thus complete. □

Lemma 2.2.26. *Let $n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \mathbf{N}$ (cf. Definition 2.2.1). Then*

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

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

$$\big(\mathcal{R}_a(\Phi \bullet \mathbb{S}_{\mathcal{I}(\Phi),n})\big)(x_1, x_2, \ldots, x_n) = (\mathcal{R}_a(\Phi))\left(\sum_{k=1}^{n} x_k\right) \tag{2.150}$$

(cf. Definitions 2.2.3, 2.2.5, and 2.2.23).

Proof of Lemma 2.2.26. Note that Lemma 2.2.24 demonstrates that for all $m \in \mathbb{N}$, $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathbb{S}_{m,n}) \in C(\mathbb{R}^{mn}, \mathbb{R}^m)$ and

$$(\mathcal{R}_a(\mathbb{S}_{m,n}))(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} x_k. \tag{2.151}$$

Combining this and item (v) in Proposition 2.2.7 establishes items (i) and (ii). The proof of Lemma 2.2.26 is thus complete. □

2.2.7.2 Concatenation of vectors as ANNs

Definition 2.2.27. *Let $m, n \in \mathbb{N}$, $A \in \mathbb{R}^{m\times n}$. Then we denote by $A^* \in \mathbb{R}^{n\times m}$ the transpose of A.*

Definition 2.2.28. *Let $m, n \in \mathbb{N}$. Then we denote by $\mathbb{T}_{m,n} \in (\mathbb{R}^{(mn)\times m} \times \mathbb{R}^{mn})$ the ANN given by*

$$\mathbb{T}_{m,n} = \mathbf{A}_{(\mathrm{I}_m \ \ \mathrm{I}_m \ \ \dots \ \ \mathrm{I}_m)^*,0} \tag{2.152}$$

(cf. Definitions 2.2.9, 2.2.18, and 2.2.27).

Lemma 2.2.29. *Let $m, n \in \mathbb{N}$. Then*

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

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

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

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

$$(\mathcal{R}_a(\mathbb{T}_{m,n}))(x) = (x, x, \dots, x) \tag{2.153}$$

(cf. Definitions 2.2.1, 2.2.3, and 2.2.28).

Proof of Lemma 2.2.29. Note that the fact that $\mathbb{T}_{m,n} \in (\mathbb{R}^{(mn)\times m} \times \mathbb{R}^{mn})$ ensures that $\mathbb{T}_{m,n} \in \mathbf{N}$ and $\mathcal{D}(\mathbb{T}_{m,n}) = (m, mn) \in \mathbb{N}^2$. This establishes items (i) and (ii). Next observe that item (iii) in Lemma 2.2.19 proves that for all $a \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathbb{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{mn})$ and

$$\begin{aligned}(\mathcal{R}_a(\mathbb{T}_{m,n}))(x) &= \big(\mathcal{R}_a\big(\mathbf{A}_{(\mathrm{I}_m \ \ \mathrm{I}_m \ \ \dots \ \ \mathrm{I}_m)^*,0}\big)\big)(x) \\ &= (\mathrm{I}_m \ \ \mathrm{I}_m \ \ \dots \ \ \mathrm{I}_m)^* x = (x, x, \dots, x)\end{aligned} \tag{2.154}$$

(cf. Definitions 2.2.9 and 2.2.18). This establishes items (iii) and (iv). The proof of Lemma 2.2.29 is thus complete. ☐

Lemma 2.2.30. *Let $n \in \mathbb{N}$, $a \in C(\mathbb{R},\mathbb{R})$, $\Phi \in \mathbf{N}$ (cf. Definition 2.2.1). Then*

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

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

$$\big(\mathcal{R}_a(\mathbb{T}_{\mathcal{O}(\Phi),n} \bullet \Phi)\big)(x) = \big((\mathcal{R}_a(\Phi))(x), (\mathcal{R}_a(\Phi))(x), \dots, (\mathcal{R}_a(\Phi))(x)\big) \tag{2.155}$$

(cf. Definitions 2.2.3, 2.2.5, and 2.2.28).

Proof of Lemma 2.2.30. Note that Lemma 2.2.29 ensures that for all $m \in \mathbb{N}$, $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathbb{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{mn})$ and

$$(\mathcal{R}_a(\mathbb{T}_{m,n}))(x) = (x, x, \dots, x). \tag{2.156}$$

Combining this and item (v) in Proposition 2.2.7 establishes items (i) and (ii). The proof of Lemma 2.2.30 is thus complete. □

Lemma 2.2.31. *Let $m, n \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \mathbf{N}$ satisfy $\mathcal{I}(\Phi) = mn$ (cf. Definition 2.2.1). Then*

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

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

$$\big(\mathcal{R}_a(\Phi \bullet \mathbb{T}_{m,n})\big)(x) = (\mathcal{R}_a(\Phi))(x, x, \dots, x) \tag{2.157}$$

(cf. Definitions 2.2.3, 2.2.5, and 2.2.28).

Proof of Lemma 2.2.31. Observe that Lemma 2.2.29 demonstrates that for all $x \in \mathbb{R}^m$ it holds that $\mathcal{R}_a(\mathbb{T}_{m,n}) \in C(\mathbb{R}^m, \mathbb{R}^{mn})$ and

$$(\mathcal{R}_a(\mathbb{T}_{m,n}))(x) = (x, x, \dots, x). \tag{2.158}$$

Combining this and item (v) in Proposition 2.2.7 establishes items (i) and (ii). The proof of Lemma 2.2.31 is thus complete. □

2.2.7.3 Sums of ANNs

Definition 2.2.32 (Sums of ANNs with the same length)**.** *Let $n \in \mathbb{Z}$, $m \in \{n, n+1, \dots\}$, $\Phi_n, \Phi_{n+1}, \dots, \Phi_m \in \mathbf{N}$ satisfy for all $k \in \{n, n+1, \dots, m\}$ that $\mathcal{L}(\Phi_k) = \mathcal{L}(\Phi_n)$, $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_n)$, and $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_n)$. Then we denote by $\bigoplus_{k=n}^m \Phi_k$ (we denote by $\Phi_n \oplus \Phi_{n+1} \oplus \dots \oplus \Phi_m$) the ANN given by*

$$\bigoplus_{k=n}^{m} \Phi_k = \big(\mathbb{S}_{\mathcal{O}(\Phi_n), m-n+1} \bullet \big[\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \dots, \Phi_m)\big] \bullet \mathbb{T}_{\mathcal{I}(\Phi_n), m-n+1}\big) \in \mathbf{N} \tag{2.159}$$

(cf. Definitions 2.2.1, 2.2.2, 2.2.5, 2.2.11, 2.2.23, and 2.2.28).

Lemma 2.2.33. *Let $n \in \mathbb{Z}$, $m \in \{n, n+1, \dots\}$, $\Phi_n, \Phi_{n+1}, \dots, \Phi_m \in \mathbf{N}$ satisfy for all $k \in \{n, n+1, \dots, m\}$ that $\mathcal{L}(\Phi_k) = \mathcal{L}(\Phi_n)$, $\mathcal{I}(\Phi_k) = \mathcal{I}(\Phi_n)$, and $\mathcal{O}(\Phi_k) = \mathcal{O}(\Phi_n)$ (cf. Definition 2.2.1). Then*

(i) it holds that $\mathcal{L}(\bigoplus_{k=n}^m \Phi_k) = \mathcal{L}(\Phi_n)$,

(ii) *it holds that*

$$\mathcal{D}\bigg(\bigoplus_{k=n}^{m} \Phi_k\bigg) = \bigg(\mathcal{I}(\Phi_n), \sum_{k=n}^{m} \mathbb{D}_1(\Phi_k), \sum_{k=n}^{m} \mathbb{D}_2(\Phi_k), \ldots, \sum_{k=n}^{m} \mathbb{D}_{\mathcal{H}(\Phi_n)}(\Phi_k), \mathcal{O}(\Phi_n)\bigg), \tag{2.160}$$

(iii) *it holds for all $a \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_a(\bigoplus_{k=n}^{m} \Phi_k) \in C(\mathbb{R}^{\mathcal{I}(\Phi_n)}, \mathbb{R}^{\mathcal{O}(\Phi_n)})$, and*

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

$$\bigg(\mathcal{R}_a\bigg(\bigoplus_{k=n}^{m} \Phi_k\bigg)\bigg)(x) = \sum_{k=n}^{m} (\mathcal{R}_a(\Phi_k))(x) \tag{2.161}$$

(cf. Definitions 2.2.3 and 2.2.32).

Proof of Lemma 2.2.33. First, note that Lemma 2.2.12 proves that

$$\begin{aligned}
&\mathcal{D}\big(\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \ldots, \Phi_m)\big) \\
&= \bigg(\sum_{k=n}^{m} \mathbb{D}_0(\Phi_k), \sum_{k=n}^{m} \mathbb{D}_1(\Phi_k), \ldots, \sum_{k=n}^{m} \mathbb{D}_{\mathcal{L}(\Phi_n)-1}(\Phi_k), \sum_{k=n}^{m} \mathbb{D}_{\mathcal{L}(\Phi_n)}(\Phi_k)\bigg) \\
&= \bigg((m-n+1)\mathcal{I}(\Phi_n), \sum_{k=n}^{m} \mathbb{D}_1(\Phi_k), \sum_{k=n}^{m} \mathbb{D}_2(\Phi_k), \ldots, \sum_{k=n}^{m} \mathbb{D}_{\mathcal{L}(\Phi_n)-1}(\Phi_k), (m-n+1)\mathcal{O}(\Phi_n)\bigg)
\end{aligned} \tag{2.162}$$

(cf. Definition 2.2.11). Moreover, observe that item (ii) in Lemma 2.2.24 ensures that

$$\mathcal{D}\big(\mathbb{S}_{\mathcal{O}(\Phi_n), m-n+1}\big) = ((m-n+1)\mathcal{O}(\Phi_n), \mathcal{O}(\Phi_n)) \tag{2.163}$$

(cf. Definition 2.2.23). This, (2.162), and item (i) in Proposition 2.2.7 demonstrate that

$$\begin{aligned}
&\mathcal{D}\big(\mathbb{S}_{\mathcal{O}(\Phi_n), m-n+1} \bullet \big[\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \ldots, \Phi_m)\big]\big) \\
&= \bigg((m-n+1)\mathcal{I}(\Phi_n), \sum_{k=n}^{m} \mathbb{D}_1(\Phi_k), \sum_{k=n}^{m} \mathbb{D}_2(\Phi_k), \ldots, \sum_{k=n}^{m} \mathbb{D}_{\mathcal{L}(\Phi_n)-1}(\Phi_k), \mathcal{O}(\Phi_n)\bigg).
\end{aligned} \tag{2.164}$$

Next note that item (ii) in Lemma 2.2.29 assures that

$$\mathcal{D}\big(\mathbb{T}_{\mathcal{I}(\Phi_n), m-n+1}\big) = (\mathcal{I}(\Phi_n), (m-n+1)\mathcal{I}(\Phi_n)) \tag{2.165}$$

(cf. Definition 2.2.28). Combining this, (2.164), and, item (i) in Proposition 2.2.7 proves that

$$\begin{aligned}
&\mathcal{D}\bigg(\bigoplus_{k=n}^{m} \Phi_k\bigg) \\
&= \mathcal{D}\big(\mathbb{S}_{\mathcal{O}(\Phi_n),(m-n+1)} \bullet \big[\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \ldots, \Phi_m)\big] \bullet \mathbb{T}_{\mathcal{I}(\Phi_n),(m-n+1)}\big) \\
&= \bigg(\mathcal{I}(\Phi_n), \sum_{k=n}^{m} \mathbb{D}_1(\Phi_k), \sum_{k=n}^{m} \mathbb{D}_2(\Phi_k), \ldots, \sum_{k=n}^{m} \mathbb{D}_{\mathcal{L}(\Phi_n)-1}(\Phi_k), \mathcal{O}(\Phi_n)\bigg).
\end{aligned} \tag{2.166}$$

This establishes items (i) and (ii). Next observe that Lemma 2.2.31 and (2.162) ensure that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ it holds that $\mathcal{R}_a([\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \dots, \Phi_m)] \bullet \mathbb{T}_{\mathcal{I}(\Phi_n), m-n+1}) \in C(\mathbb{R}^{\mathcal{I}(\Phi_n)}, \mathbb{R}^{(m-n+1)\mathcal{O}(\Phi_n)})$ and

$$\begin{aligned}&\big(\mathcal{R}_a\big([\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \dots, \Phi_m)] \bullet \mathbb{T}_{\mathcal{I}(\Phi_n), m-n+1}\big)\big)(x) \\ &= \big(\mathcal{R}_a\big(\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \dots, \Phi_m)\big)\big)(x, x, \dots, x).\end{aligned} \tag{2.167}$$

Combining this with item (ii) in Proposition 2.2.13 proves that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ it holds that

$$\begin{aligned}&\big(\mathcal{R}_a\big([\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \dots, \Phi_m)] \bullet \mathbb{T}_{\mathcal{I}(\Phi_n), m-n+1}\big)\big)(x) \\ &= \big((\mathcal{R}_a(\Phi_n))(x), (\mathcal{R}_a(\Phi_{n+1}))(x), \dots, (\mathcal{R}_a(\Phi_m))(x)\big) \in \mathbb{R}^{(m-n+1)\mathcal{O}(\Phi_n)}.\end{aligned} \tag{2.168}$$

Lemma 2.2.25, (2.163), and Lemma 2.2.8 therefore demonstrate that for all $a \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ it holds that $\mathcal{R}_a(\bigoplus_{k=n}^m \Phi_k) \in C(\mathbb{R}^{\mathcal{I}(\Phi_n)}, \mathbb{R}^{\mathcal{O}(\Phi_n)})$ and

$$\begin{aligned}&\left(\mathcal{R}_a\left(\bigoplus_{k=n}^m \Phi_k\right)\right)(x) \\ &= \big(\mathcal{R}_a\big(\mathbb{S}_{\mathcal{O}(\Phi_n), m-n+1} \bullet [\mathbf{P}_{m-n+1}(\Phi_n, \Phi_{n+1}, \dots, \Phi_m)] \bullet \mathbb{T}_{\mathcal{I}(\Phi_n), m-n+1}\big)\big)(x) = \sum_{k=n}^m (\mathcal{R}_a(\Phi_k))(x).\end{aligned} \tag{2.169}$$

This establishes items (iii) and (iv). The proof of Lemma 2.2.33 is thus complete. □

2.2.8 On the connection to the vectorized description of ANNs

Definition 2.2.34. *We denote by* $\mathcal{T} \colon \mathbf{N} \to \big(\bigcup_{d \in \mathbb{N}} \mathbb{R}^d\big)$ *the function which satisfies for all* $L, d \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L)) \in \big(\times_{m=1}^L (\mathbb{R}^{l_m \times l_{m-1}} \times \mathbb{R}^{l_m})\big)$, $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \mathbb{R}^d$, $k \in \{1, 2, \dots, L\}$ *with* $\mathcal{T}(\Phi) = \theta$ *that*

$$d = \mathcal{P}(\Phi), \qquad B_k = \begin{pmatrix} \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_k l_{k-1} + 1} \\ \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_k l_{k-1} + 2} \\ \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_k l_{k-1} + 3} \\ \vdots \\ \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_k l_{k-1} + l_k} \end{pmatrix}, \qquad \textit{and}$$

$$W_k = \begin{pmatrix} \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + 1} & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + 2} & \cdots & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_{k-1}} \\ \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_{k-1} + 1} & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_{k-1} + 2} & \cdots & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + 2l_{k-1}} \\ \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + 2l_{k-1} + 1} & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + 2l_{k-1} + 2} & \cdots & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + 3l_{k-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + (l_k - 1)l_{k-1} + 1} & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + (l_k - 1)l_{k-1} + 2} & \cdots & \theta_{(\sum_{i=1}^{k-1} l_i(l_{i-1}+1)) + l_k l_{k-1}} \end{pmatrix}, \tag{2.170}$$

(cf. Definition 2.2.1).

Lemma 2.2.35. *Let $a, b \in \mathbb{N}$, $W = (W_{i,j})_{(i,j)\in\{1,2,\dots,a\}\times\{1,2,\dots,b\}} \in \mathbb{R}^{a\times b}$, $B = (B_1, B_2, \dots, B_a) \in \mathbb{R}^a$. Then*

$$\mathcal{T}\big(\mathbf{A}_{W,B}\big) = \big(W_{1,1}, W_{1,2}, \dots, W_{1,b}, W_{2,1}, W_{2,2}, \dots, W_{2,b}, \dots, W_{a,1}, W_{a,2}, \dots, W_{a,b}, B_1, B_2, \dots, B_a\big) \tag{2.171}$$

(cf. Definitions 2.2.18 and 2.2.34).

Proof of Lemma 2.2.35. Observe that (2.170) establishes (2.171). The proof of Lemma 2.2.35 is thus complete. □

Lemma 2.2.36. *Let $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, let $W_k = (W_{k,i,j})_{(i,j)\in\{1,2,\dots,l_k\}\times\{1,2,\dots,l_{k-1}\}} \in \mathbb{R}^{l_k\times l_{k-1}}$, $k \in \{1, 2, \dots, L\}$, and let $B_k = (B_{k,1}, B_{k,2}, \dots, B_{k,l_k}) \in \mathbb{R}^{l_k}$, $k \in \{1, 2, \dots, L\}$. Then*

(i) it holds for all $k \in \{1, 2, \dots, L\}$ that

$$\begin{split}\mathcal{T}\big(((W_k, B_k))\big) = \big(&W_{k,1,1}, W_{k,1,2}, \dots, W_{k,1,l_{k-1}}, W_{k,2,1}, W_{k,2,2}, \dots, W_{k,2,l_{k-1}}, \dots, \\ &W_{k,l_k,1}, W_{k,l_k,2}, \dots, W_{k,l_k,l_{k-1}}, B_{k,1}, B_{k,2}, \dots, B_{k,l_k}\big)\end{split} \tag{2.172}$$

and

(ii) it holds that

$$\begin{split}&\mathcal{T}\Big(\big((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L)\big)\Big) \\ &= \Big(W_{1,1,1}, W_{1,1,2}, \dots, W_{1,1,l_0}, \dots, W_{1,l_1,1}, W_{1,l_1,2}, \dots, W_{1,l_1,l_0}, B_{1,1}, B_{1,2}, \dots, B_{1,l_1}, \\ &\quad W_{2,1,1}, W_{2,1,2}, \dots, W_{2,1,l_1}, \dots, W_{2,l_2,1}, W_{2,l_2,2}, \dots, W_{2,l_2,l_1}, B_{2,1}, B_{2,2}, \dots, B_{2,l_2}, \\ &\quad \dots, \\ &\quad W_{L,1,1}, W_{L,1,2}, \dots, W_{L,1,l_{L-1}}, \dots W_{L,l_L,1}, W_{L,l_L,2}, \dots, W_{L,l_L,l_{L-1}}, B_{L,1}, B_{L,2}, \dots, B_{L,l_L}\Big)\end{split} \tag{2.173}$$

(cf. Definition 2.2.34).

Proof of Lemma 2.2.36. Note that Lemma 2.2.35 proves item (i). Moreover, observe that (2.170) establishes item (ii). The proof of Lemma 2.2.36 is thus complete. □

Exercise 2.2.4. *Prove or disprove the following statement: The function $\mathcal{T}$ is injective (cf. Definition 2.2.34).*

Exercise 2.2.5. *Prove or disprove the following statement: The function $\mathcal{T}$ is surjective (cf. Definition 2.2.34).*

Exercise 2.2.6. *Prove or disprove the following statement: The function $\mathcal{T}$ is bijective (cf. Definition 2.2.34).*

Lemma 2.2.37. *Let $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \mathbf{N}$, $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$ satisfy $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$ (cf. Definition 2.2.1). Then it holds for all $x \in \mathbb{R}^{l_0}$ that*

$$(\mathcal{R}_a(\Phi))(x) = \begin{cases} \big(\mathcal{N}^{\mathcal{T}(\Phi), l_0}_{\mathrm{id}_{\mathbb{R}^{l_L}}}\big)(x) & : L = 1 \\ \big(\mathcal{N}^{\mathcal{T}(\Phi), l_0}_{\mathfrak{M}_{a,l_1}, \mathfrak{M}_{a,l_2}, \ldots, \mathfrak{M}_{a,l_{L-1}}, \mathrm{id}_{\mathbb{R}^{l_L}}}\big)(x) & : L > 1 \end{cases} \tag{2.174}$$

(cf. Definitions 2.1.2, 2.1.4, 2.2.3, and 2.2.34).

Proof of Lemma 2.2.37. Throughout this proof let $((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in \big(\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\big)$ satisfy $\Phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L))$. Note that (2.170) shows that for all $k \in \{1, 2, \ldots, L\}$, $x \in \mathbb{R}^{l_{k-1}}$ it holds that

$$W_k x + B_k = \big(\mathcal{A}^{\mathcal{T}(\Phi), \sum_{i=1}^{k-1} l_i(l_{i-1}+1)}_{l_k, l_{k-1}}\big)(x) \tag{2.175}$$

(cf. Definitions 2.1.1 and 2.2.34). This demonstrates that for all $x_0 \in \mathbb{R}^{l_0}$, $x_1 \in \mathbb{R}^{l_1}, \ldots, x_L \in \mathbb{R}^{l_L}$ with $\forall\, k \in \{1, 2, \ldots, L\}\colon x_k = \mathfrak{M}_{a,l_k}(W_k x_{k-1} + B_k)$ it holds that

$$\begin{split} &x_{L-1} = \\ &\begin{cases} x_0 & : L = 1 \\ \big(\mathfrak{M}_{a,l_{L-1}} \circ \mathcal{A}^{\mathcal{T}(\Phi), \sum_{i=1}^{L-2} l_i(l_{i-1}+1)}_{l_{L-1}, l_{L-2}} \circ \mathfrak{M}_{a,l_{L-2}} \circ \mathcal{A}^{\mathcal{T}(\Phi), \sum_{i=1}^{L-3} l_i(l_{i-1}+1)}_{l_{L-2}, l_{L-3}} \circ \ldots \circ \mathfrak{M}_{a,l_1} \circ \mathcal{A}^{\mathcal{T}(\Phi), 0}_{l_1, l_0}\big)(x_0) & : L > 1 \end{cases} \end{split} \tag{2.176}$$

(cf. Definition 2.1.4). Combining this and (2.175) with (2.3) and (2.53) proves that for all $x_0 \in \mathbb{R}^{l_0}$, $x_1 \in \mathbb{R}^{l_1}, \ldots, x_L \in \mathbb{R}^{l_L}$ with $\forall\, k \in \{1, 2, \ldots, L\}\colon x_k = \mathfrak{M}_{a,l_k}(W_k x_{k-1} + B_k)$ it holds that

$$\begin{split} \big(\mathcal{R}_a(\Phi)\big)(x_0) &= W_L x_{L-1} + B_L = \big(\mathcal{A}^{\mathcal{T}(\Phi), \sum_{i=1}^{L-1} l_i(l_{i-1}+1)}_{l_L, l_{L-1}}\big)(x_{L-1}) \\ &= \begin{cases} \big(\mathcal{N}^{\mathcal{T}(\Phi), l_0}_{\mathrm{id}_{\mathbb{R}^{l_L}}}\big)(x_0) & : L = 1 \\ \big(\mathcal{N}^{\mathcal{T}(\Phi), l_0}_{\mathfrak{M}_{a,l_1}, \mathfrak{M}_{a,l_2}, \ldots, \mathfrak{M}_{a,l_{L-1}}, \mathrm{id}_{\mathbb{R}^{l_L}}}\big)(x_0) & : L > 1 \end{cases} \end{split} \tag{2.177}$$

(cf. Definitions 2.1.2 and 2.2.3). The proof of Lemma 2.2.37 is thus complete. □

Corollary 2.2.38. *Let $\Phi \in \mathbf{N}$ (cf. Definition 2.2.1). Then it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that*

$$\big(\mathcal{N}^{\mathcal{T}(\Phi), \mathcal{D}(\Phi)}_{-\infty, \infty}\big)(x) = (\mathcal{R}_{\mathfrak{r}}(\Phi))(x) \tag{2.178}$$

(cf. Definitions 2.1.6, 2.1.27, 2.2.3, and 2.2.34).

Proof of Corollary 2.2.38. Note that Lemma 2.2.37, (2.50), (2.11), and the fact that for all $d \in \mathbb{N}$ it holds that $\mathfrak{C}_{-\infty,\infty,d} = \mathrm{id}_{\mathbb{R}^d}$ establish (2.178) (cf. Definition 2.1.12). The proof of Corollary 2.2.38 is thus complete. □

Chapter 3

Low-dimensional ANN approximation results

3.1 One-dimensional ANN approximation results

3.1.1 Linear interpolation of one-dimensional functions

3.1.1.1 On the modulus of continuity

Definition 3.1.1 (Modulus of continuity)**.** *Let $A \subseteq \mathbb{R}$ be a set and let $f\colon A \to \mathbb{R}$ be a function. Then we denote by $w_f\colon [0,\infty] \to [0,\infty]$ the function which satisfies for all $h \in [0,\infty]$ that*

$$\begin{aligned} w_f(h) &= \sup\bigl(\bigl\{|f(x)-f(y)| \in [0,\infty)\colon (x,y \in A \text{ with } |x-y| \le h)\bigr\} \cup \{0\}\bigr) \\ &= \sup\bigl(\bigl\{r \in \mathbb{R}\colon (\exists x \in A, y \in [x-h,x+h]\colon r = |f(x)-f(y)|)\bigr\} \cup \{0\}\bigr) \end{aligned} \tag{3.1}$$

and we call w_f the modulus of continuity of f.

Lemma 3.1.2. *Let $a \in [-\infty,\infty]$, $b \in [a,\infty]$ and let $f\colon ([a,b]\cap\mathbb{R}) \to \mathbb{R}$ be a function. Then*

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

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

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

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

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

(cf. Definition 3.1.1).

Proof of Lemma 3.1.2. First, observe that (3.1) implies items (i), (ii), (iii), and (iv). Note the fact that for all $h \in [0, \infty)$, $\mathfrak{h} \in [0, h]$, $x \in \mathbb{R}$, $y_1 \in [x - \mathfrak{h}, x + \mathfrak{h}]$, $y_2 \in [x + \mathfrak{h}, x + h + \mathfrak{h}]$, $y_3 \in [x - h - \mathfrak{h}, x - \mathfrak{h}]$ it holds that $|x - y_1| \le h$, $|x - (y_2 - \mathfrak{h})| \le h$, and $|x - (y_3 + \mathfrak{h})| \le h$ shows that for all $h \in [0, \infty)$, $\mathfrak{h} \in [0, h]$, $x, y \in [a, b] \cap \mathbb{R}$ with $|x - y| \le h + \mathfrak{h}$ there exists $z \in [a, b] \cap \mathbb{R}$ such that

$$|x - z| \le h \qquad \text{and} \qquad |y - z| \le \mathfrak{h}. \tag{3.2}$$

Item (iv) therefore implies that for all $h \in [0, \infty)$, $\mathfrak{h} \in [0, h]$, $x, y \in [a, b] \cap \mathbb{R}$ with $|x - y| \le h + \mathfrak{h}$ there exists $z \in [a, b] \cap \mathbb{R}$ such that

$$|f(x) - f(y)| \le |f(x) - f(z)| + |f(y) - f(z)| \le w_f(h) + w_f(\mathfrak{h}) \tag{3.3}$$

(cf. Definition 3.1.1). Combining this with (3.1) proves that for all $h \in [0, \infty)$, $\mathfrak{h} \in [0, h]$ it holds that

$$w_f(h + \mathfrak{h}) \le w_f(h) + w_f(\mathfrak{h}). \tag{3.4}$$

This establishes item (v). The proof of Lemma 3.1.2 is thus complete. □

Lemma 3.1.3. *Let $A \subseteq \mathbb{R}$, $L \in [0, \infty)$ and let $f \colon A \to \mathbb{R}$ satisfy for all $x, y \in A$ that $|f(x) - f(y)| \le L|x - y|$. Then it holds for all $h \in [0, \infty)$ that $w_f(h) \le Lh$.*

Proof of Lemma 3.1.3. Observe that the assumption that for all $x, y \in A$ it holds that $|f(x) - f(y)| \le L|x - y|$ and (3.1) imply that for all $h \in [0, \infty)$ it holds that

$$\begin{aligned} w_f(h) &= \sup\bigl(\bigl\{|f(x) - f(y)| \in [0, \infty) \colon (x, y \in A \text{ with } |x - y| \le h)\bigr\} \cup \{0\}\bigr) \\ &\le \sup\bigl(\bigl\{L|x - y| \in [0, \infty) \colon (x, y \in A \text{ with } |x - y| \le h)\bigr\} \cup \{0\}\bigr) \\ &\le \sup(\{Lh, 0\}) = Lh. \end{aligned} \tag{3.5}$$

The proof of Lemma 3.1.3 is thus complete. □

3.1.1.2 Linear interpolation of one-dimensional functions

Definition 3.1.4 (Linear interpolation operator)**.** *Let $K \in \mathbb{N}$, $\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K, f_0, f_1, \ldots, f_K \in \mathbb{R}$ satisfy $\mathfrak{x}_0 < \mathfrak{x}_1 < \ldots < \mathfrak{x}_K$. Then we denote by $\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K} \colon \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $k \in \{1, 2, \ldots, K\}$, $x \in (-\infty, \mathfrak{x}_0)$, $y \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k)$, $z \in [\mathfrak{x}_K, \infty)$ that $(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(x) = f_0$, $(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(z) = f_K$, and*

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(y) = f_{k-1} + \bigl(\tfrac{y - \mathfrak{x}_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\bigr)(f_k - f_{k-1}). \tag{3.6}$$

Lemma 3.1.5. *Let $K \in \mathbb{N}$, $\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K, f_0, f_1, \ldots, f_K \in \mathbb{R}$ satisfy $\mathfrak{x}_0 < \mathfrak{x}_1 < \ldots < \mathfrak{x}_K$. Then*

(i) it holds for all $k \in \{0, 1, \ldots, K\}$ that

$$(\mathscr{L}^{f_0, f_1, \ldots, f_K}_{\mathfrak{x}_0, \mathfrak{x}_1, \ldots, \mathfrak{x}_K})(\mathfrak{x}_k) = f_k, \tag{3.7}$$

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

$$(\mathscr{L}^{f_0,f_1,\ldots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K})(x) = f_{k-1} + \left(\tfrac{x-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)(f_k - f_{k-1}), \tag{3.8}$$

and

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

$$(\mathscr{L}^{f_0,f_1,\ldots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K})(x) = \left(\tfrac{\mathfrak{r}_k-x}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)f_{k-1} + \left(\tfrac{x-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)f_k. \tag{3.9}$$

(cf. Definition 3.1.4).

Proof of Lemma 3.1.5. Observe that (3.6) implies items (i) and (ii). Moreover, note that item (ii) implies that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ it holds that

$$(\mathscr{L}^{f_0,f_1,\ldots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K})(x) = \left[\left(\tfrac{\mathfrak{r}_k-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right) - \left(\tfrac{x-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)\right]f_{k-1} + \left(\tfrac{x-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)f_k = \left(\tfrac{\mathfrak{r}_k-x}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)f_{k-1} + \left(\tfrac{x-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)f_k. \tag{3.10}$$

This proves item (iii). The proof of Lemma 3.1.5 is thus complete. □

Lemma 3.1.6. *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) it holds for all $x, y \in \mathbb{R}$ *with* $x \neq y$ *that*

$$\left|(\mathscr{L}^{f(\mathfrak{r}_0),f(\mathfrak{r}_1),\ldots,f(\mathfrak{r}_K)}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K})(x) - (\mathscr{L}^{f(\mathfrak{r}_0),f(\mathfrak{r}_1),\ldots,f(\mathfrak{r}_K)}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K})(y)\right| \le \left(\max_{k\in\{1,2,\ldots,K\}}\left(\frac{w_f(\mathfrak{r}_k - \mathfrak{r}_{k-1})}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)\right)|x-y| \tag{3.11}$$

and

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

(cf. Definitions 3.1.1 and 3.1.4).

Proof of Lemma 3.1.6. Throughout this proof let $L \in [0, \infty]$ satisfy

$$L = \max_{k\in\{1,2,\ldots,K\}}\left(\frac{w_f(\mathfrak{r}_k - \mathfrak{r}_{k-1})}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right) \tag{3.12}$$

and let $\mathfrak{l}\colon \mathbb{R} \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}$ that $\mathfrak{l}(x) = (\mathscr{L}^{f(\mathfrak{r}_0),f(\mathfrak{r}_1),\ldots,f(\mathfrak{r}_K)}_{\mathfrak{r}_0,\mathfrak{r}_1,\ldots,\mathfrak{r}_K})(x)$ (cf. Definitions 3.1.1 and 3.1.4). Observe that item (ii) in Lemma 3.1.5, item (iv) in Lemma 3.1.2, and (3.12) assure that for all $k \in \{1, 2, \ldots, K\}$, $x, y \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ with $x \neq y$ it holds that

$$\begin{aligned} |\mathfrak{l}(x) - \mathfrak{l}(y)| &= \left|\left(\tfrac{x-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)(f(\mathfrak{r}_k) - f(\mathfrak{r}_{k-1})) - \left(\tfrac{y-\mathfrak{r}_{k-1}}{\mathfrak{r}_k-\mathfrak{r}_{k-1}}\right)(f(\mathfrak{r}_k) - f(\mathfrak{r}_{k-1}))\right| \\ &= \left|\left(\frac{f(\mathfrak{r}_k) - f(\mathfrak{r}_{k-1})}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)(x-y)\right| \le \left(\frac{w_f(\mathfrak{r}_k - \mathfrak{r}_{k-1})}{\mathfrak{r}_k - \mathfrak{r}_{k-1}}\right)|x-y| \le L|x-y|. \end{aligned} \tag{3.13}$$

This, item (iv) in Lemma 3.1.2, item (i) in Lemma 3.1.5, and (3.12) ensure 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 $k < l$ and $x \neq y$ it holds that

$$\begin{split} &|\mathfrak{l}(x) - \mathfrak{l}(y)| \\ &\leq |\mathfrak{l}(x) - \mathfrak{l}(\mathfrak{r}_k)| + |\mathfrak{l}(\mathfrak{r}_k) - \mathfrak{l}(\mathfrak{r}_{l-1})| + |\mathfrak{l}(\mathfrak{r}_{l-1}) - \mathfrak{l}(y)| \\ &= |\mathfrak{l}(x) - \mathfrak{l}(\mathfrak{r}_k)| + |f(\mathfrak{r}_k) - f(\mathfrak{r}_{l-1})| + |\mathfrak{l}(\mathfrak{r}_{l-1}) - \mathfrak{l}(y)| \\ &\leq |\mathfrak{l}(x) - \mathfrak{l}(\mathfrak{r}_k)| + \left(\sum_{j=k+1}^{l-1} |f(\mathfrak{r}_j) - f(\mathfrak{r}_{j-1})| \right) + |\mathfrak{l}(\mathfrak{r}_{l-1}) - \mathfrak{l}(y)| \\ &\leq |\mathfrak{l}(x) - \mathfrak{l}(\mathfrak{r}_k)| + \left(\sum_{j=k+1}^{l-1} w_f(|\mathfrak{r}_j - \mathfrak{r}_{j-1}|) \right) + |\mathfrak{l}(\mathfrak{r}_{l-1}) - \mathfrak{l}(y)| \\ &\leq L \left((\mathfrak{r}_k - x) + \left(\sum_{j=k+1}^{l-1} (\mathfrak{r}_j - \mathfrak{r}_{j-1}) \right) + (y - \mathfrak{r}_{l-1}) \right) = L|x - y|. \end{split} \tag{3.14}$$

Combining this and (3.13) shows that for all $x, y \in [\mathfrak{r}_0, \mathfrak{r}_K]$ with $x \neq y$ it holds that $|\mathfrak{l}(x) - \mathfrak{l}(y)| \leq L|x - y|$. This, the fact that for all $x, y \in (-\infty, \mathfrak{r}_0]$ with $x \neq y$ it holds that $|\mathfrak{l}(x) - \mathfrak{l}(y)| = 0 \leq L|x - y|$, the fact that for all $x, y \in [\mathfrak{r}_K, \infty)$ with $x \neq y$ it holds that $|\mathfrak{l}(x) - \mathfrak{l}(y)| = 0 \leq L|x - y|$, and the triangle inequality hence demonstrate that for all $x, y \in \mathbb{R}$ with $x \neq y$ it holds that $|\mathfrak{l}(x) - \mathfrak{l}(y)| \leq L|x - y|$. This proves item (i). Moreover, note that (3.1), Lemma 3.1.2, and item (iii) in Lemma 3.1.5 assure that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{r}_{k-1}, \mathfrak{r}_k]$ it holds that

$$\begin{split} |\mathfrak{l}(x) - f(x)| &= \left| \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} \right) f(\mathfrak{r}_k) + \left(\frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} \right) f(\mathfrak{r}_{k-1}) - f(x) \right| \\ &= \left| \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} \right) (f(\mathfrak{r}_k) - f(x)) + \left(\frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} \right) (f(\mathfrak{r}_{k-1}) - f(x)) \right| \\ &\leq \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} \right) |f(\mathfrak{r}_k) - f(x)| + \left(\frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} \right) |f(\mathfrak{r}_{k-1}) - f(x)| \\ &\leq w_f(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|) \left(\frac{\mathfrak{r}_k - x}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} + \frac{x - \mathfrak{r}_{k-1}}{\mathfrak{r}_k - \mathfrak{r}_{k-1}} \right) \\ &= w_f(|\mathfrak{r}_k - \mathfrak{r}_{k-1}|) \leq w_f(\max\nolimits_{j \in \{1,2,\ldots,K\}} |\mathfrak{r}_j - \mathfrak{r}_{j-1}|). \end{split} \tag{3.15}$$

This establishes item (ii). The proof of Lemma 3.1.6 is thus complete. □

Lemma 3.1.7. *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) it holds for all $x, y \in \mathbb{R}$ that

$$\left| (\mathscr{L}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)})(x) - (\mathscr{L}_{\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K}^{f(\mathfrak{r}_0), f(\mathfrak{r}_1), \ldots, f(\mathfrak{r}_K)})(y) \right| \leq L|x - y| \tag{3.16}$$

and

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

(cf. Definition 3.1.4).

Proof of Lemma 3.1.7. Note that the assumption that for all $x,y\in[\mathfrak{x}_0,\mathfrak{x}_K]$ it holds that $|f(x)-f(y)|\leq L|x-y|$, Lemma 3.1.3, and item (i) in Lemma 3.1.6 demonstrate that for all $x,y\in\mathbb{R}$ it holds that

$$\begin{aligned}&\big|(\mathscr{L}^{f(\mathfrak{x}_0),f(\mathfrak{x}_1),\ldots,f(\mathfrak{x}_K)}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K})(x)-(\mathscr{L}^{f(\mathfrak{x}_0),f(\mathfrak{x}_1),\ldots,f(\mathfrak{x}_K)}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K})(y)\big|\\&\leq\left(\max_{k\in\{1,2,\ldots,K\}}\left(\frac{L|\mathfrak{x}_k-\mathfrak{x}_{k-1}|}{|\mathfrak{x}_k-\mathfrak{x}_{k-1}|}\right)\right)|x-y|=L|x-y|.\end{aligned}\tag{3.17}$$

This proves item (i). Moreover, observe that the assumption that for all $x,y\in[\mathfrak{x}_0,\mathfrak{x}_K]$ it holds that $|f(x)-f(y)|\leq L|x-y|$, Lemma 3.1.3, and item (ii) in Lemma 3.1.6 assure that

$$\sup_{x\in[\mathfrak{x}_0,\mathfrak{x}_K]}\big|(\mathscr{L}^{f(\mathfrak{x}_0),f(\mathfrak{x}_1),\ldots,f(\mathfrak{x}_K)}_{\mathfrak{x}_0,\mathfrak{x}_1,\ldots,\mathfrak{x}_K})(x)-f(x)\big|\leq L\left(\max_{k\in\{1,2,\ldots,K\}}|\mathfrak{x}_k-\mathfrak{x}_{k-1}|\right).\tag{3.18}$$

This establishes item (ii). The proof of Lemma 3.1.7 is thus complete. □

3.1.2 Activation functions as ANNs

Definition 3.1.8 (Activation functions as ANNs)**.** *Let* $n\in\mathbb{N}$*. Then we denote by* $\mathfrak{i}_n\in((\mathbb{R}^{n\times n}\times\mathbb{R}^n)\times(\mathbb{R}^{n\times n}\times\mathbb{R}^n))\subseteq\mathbf{N}$ *the ANN given by* $\mathfrak{i}_n=((\mathrm{I}_n,0),(\mathrm{I}_n,0))$ *(cf. Definitions 2.2.1 and 2.2.9).*

Lemma 3.1.9. *Let* $n\in\mathbb{N}$*. Then*

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

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

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

(cf. Definitions 2.1.4, 2.2.1, 2.2.3, and 3.1.8).

Proof of Lemma 3.1.9. 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 (2.53) 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

$$(\mathcal{R}_a(\mathfrak{i}_n))(x)=\mathrm{I}_n(\mathfrak{M}_{a,n}(\mathrm{I}_nx+0))+0=\mathfrak{M}_{a,n}(x).\tag{3.19}$$

This establishes items (ii) and (iii). The proof of Lemma 3.1.9 is thus complete. □

Lemma 3.1.10. *Let* $\Phi\in\mathbf{N}$ *(cf. Definition 2.2.1). Then*

(i) *it holds that* $\mathcal{D}(\mathfrak{i}_{\mathcal{O}(\Phi)} \bullet \Phi) = (\mathbb{D}_0(\Phi), \mathbb{D}_1(\Phi), \mathbb{D}_2(\Phi), \dots, \mathbb{D}_{\mathcal{L}(\Phi)-1}(\Phi), \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi), \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)) \in \mathbb{N}^{\mathcal{L}(\Phi)+2}$,

(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* $a \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathcal{R}_a(\mathfrak{i}_{\mathcal{O}(\Phi)} \bullet \Phi) = \mathfrak{M}_{a,\mathcal{O}(\Phi)} \circ (\mathcal{R}_a(\Phi))$,

(iv) *it holds that* $\mathcal{D}(\Phi \bullet \mathfrak{i}_{\mathcal{I}(\Phi)}) = (\mathbb{D}_0(\Phi), \mathbb{D}_0(\Phi), \mathbb{D}_1(\Phi), \mathbb{D}_2(\Phi), \dots, \mathbb{D}_{\mathcal{L}(\Phi)-1}(\Phi), \mathbb{D}_{\mathcal{L}(\Phi)}(\Phi)) \in \mathbb{N}^{\mathcal{L}(\Phi)+2}$,

(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* $a \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathcal{R}_a(\Phi \bullet \mathfrak{i}_{\mathcal{I}(\Phi)}) = (\mathcal{R}_a(\Phi)) \circ \mathfrak{M}_{a,\mathcal{I}(\Phi)}$

(cf. Definitions 2.1.4, 2.2.3, 2.2.5, and 3.1.8).

Proof of Lemma 3.1.10. Note that Lemma 3.1.9 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

$$(\mathcal{R}_a(\mathfrak{i}_n))(x) = \mathfrak{M}_{a,n}(x) \tag{3.20}$$

(cf. Definitions 2.1.4, 2.2.3, and 3.1.8). Combining this and Proposition 2.2.7 establishes items (i), (ii), (iii), (iv), (v), and (vi). The proof of Lemma 3.1.10 is thus complete. □

3.1.3 Linear interpolation with ANNs

Lemma 3.1.11. *Let* $\alpha, \beta, h \in \mathbb{R}$, $\mathbf{H} \in \mathbf{N}$ *satisfy* $\mathbf{H} = h \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta})$ *(cf. Definitions 2.2.1, 2.2.5, 2.2.18, 2.2.21, and 3.1.8). Then*

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

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

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

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

(cf. Definitions 2.1.6 and 2.2.3).

Proof of Lemma 3.1.11. Note that Lemma 2.2.19 ensures that $\mathbf{A}_{\alpha,\beta} = (\alpha, \beta)$, $\mathcal{D}(\mathbf{A}_{\alpha,\beta}) = (1, 1) \in \mathbb{N}^2$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta}) \in C(\mathbb{R}, \mathbb{R})$, and $\forall\, x \in \mathbb{R}\colon (\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta}))(x) = \alpha x + \beta$ (cf. Definitions 2.1.6 and 2.2.3). Proposition 2.2.7, Lemma 3.1.9, Lemma 3.1.10, (2.10), (2.53), and (2.58) therefore imply that $\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta} = ((\alpha, \beta), (1, 0))$, $\mathcal{D}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) = (1, 1, 1) \in \mathbb{N}^3$, $\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) \in C(\mathbb{R}, \mathbb{R})$, and

$$\forall\, x \in \mathbb{R}\colon (\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}))(x) = \mathfrak{r}(\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\alpha,\beta})(x)) = \max\{\alpha x + \beta, 0\}. \tag{3.21}$$

This, Lemma 2.2.22, and (2.140) ensure that $\mathbf{H} = h \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}) = ((\alpha,\beta),(h,0))$, $\mathcal{D}(\mathbf{H}) = (1,1,1)$, $\mathcal{R}_{\mathfrak{r}}(\mathbf{H}) \in C(\mathbb{R},\mathbb{R})$, and

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{H}))(x) = h((\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{\alpha,\beta}))(x)) = h\max\{\alpha x + \beta, 0\}. \tag{3.22}$$

This establishes items (i), (ii), (iii), and (iv). The proof of Lemma 3.1.11 is thus complete. □

Lemma 3.1.12. *Let $K \in \mathbb{N}$, $f_0, f_1, \dots, f_K, \mathfrak{r}_0, \mathfrak{r}_1, \dots, \mathfrak{r}_K \in \mathbb{R}$ satisfy $\mathfrak{r}_0 < \mathfrak{r}_1 < \ldots < \mathfrak{r}_K$ and let $\mathbf{F} \in \mathbf{N}$ satisfy*

$$\mathbf{F} = \mathbf{A}_{1,f_0} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{(f_{\min\{k+1,K\}} - f_k)}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \tfrac{(f_k - f_{\max\{k-1,0\}})}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right) \tag{3.23}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.18, 2.2.21, 2.2.32, and 3.1.8). Then

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

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

(iii) it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) = \mathscr{L}^{f_0,f_1,\dots,f_K}_{\mathfrak{r}_0,\mathfrak{r}_1,\dots,\mathfrak{r}_K}$, and

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

(cf. Definitions 2.1.6, 2.2.3, and 3.1.4).

Proof of Lemma 3.1.12. Throughout this proof let $c_0, c_1, \dots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0,1,\dots,K\}$ that

$$c_k = \frac{(f_{\min\{k+1,K\}} - f_k)}{(\mathfrak{r}_{\min\{k+1,K\}} - \mathfrak{r}_{\min\{k,K-1\}})} - \frac{(f_k - f_{\max\{k-1,0\}})}{(\mathfrak{r}_{\max\{k,1\}} - \mathfrak{r}_{\max\{k-1,0\}})} \tag{3.24}$$

and let $\Phi_0, \Phi_1, \dots, \Phi_K \in ((\mathbb{R}^{1\times 1} \times \mathbb{R}^1) \times (\mathbb{R}^{1\times 1} \times \mathbb{R}^1)) \subseteq \mathbf{N}$ satisfy for all $k \in \{0,1,\dots,K\}$ that $\Phi_k = c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k})$. Observe that Lemma 3.1.11 assures that for all $k \in \{0,1,\dots,K\}$ it holds that $\mathcal{R}_{\mathfrak{r}}(\Phi_k) \in C(\mathbb{R},\mathbb{R})$, $\mathcal{D}(\Phi_k) = (1,1,1) \in \mathbb{N}^3$, and $\forall\, x \in \mathbb{R}\colon (\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = c_k \max\{x - \mathfrak{r}_k, 0\}$ (cf. Definitions 2.1.6 and 2.2.3). This, Lemma 2.2.20, Lemma 2.2.33, and (3.23) assure that $\mathcal{D}(\mathbf{F}) = (1, K+1, 1) \in \mathbb{N}^3$ and $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}) \in C(\mathbb{R},\mathbb{R})$. This establishes items (i) and (ii). Moreover, note that item (i) and (2.52) imply that

$$\mathcal{P}(\mathbf{F}) = 2(K+1) + (K+2) = 3K + 4. \tag{3.25}$$

This proves item (iv). Next observe that (3.24), Lemma 2.2.20, and Lemma 2.2.33 ensure that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = f_0 + \sum_{k=0}^{K} (\mathcal{R}_{\mathfrak{r}}(\Phi_k))(x) = f_0 + \sum_{k=0}^{K} c_k \max\{x - \mathfrak{r}_k, 0\}. \tag{3.26}$$

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

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = f_0 + 0 = f_0. \tag{3.27}$$

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

$$\sum_{n=0}^{k-1} c_n = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}. \tag{3.28}$$

We now prove (3.28) by induction on $k \in \{1, 2, \ldots, K\}$. For the base case $k = 1$ observe that (3.24) assures that $\sum_{n=0}^{0} c_n = c_0 = \frac{f_1 - f_0}{\mathfrak{x}_1 - \mathfrak{x}_0}$. This proves (3.28) in the base case $k = 1$. For the induction step note that (3.24) ensures that for all $k \in \mathbb{N} \cap (1, \infty) \cap (0, K]$ with $\sum_{n=0}^{k-2} c_n = \frac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}}$ it holds that

$$\sum_{n=0}^{k-1} c_n = c_{k-1} + \sum_{n=0}^{k-2} c_n = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}} - \tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}} + \tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}} = \tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}. \tag{3.29}$$

Induction thus proves (3.28). In addition, observe that (3.26), (3.28), and the fact that $\forall\, k \in \{1, 2, \ldots, K\} \colon \mathfrak{x}_{k-1} < \mathfrak{x}_k$ show that for all $k \in \{1, 2, \ldots, K\}$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ it holds that

$$\begin{split}
(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_{k-1}) &= \sum_{n=0}^{K} c_n (\max\{x - \mathfrak{x}_n, 0\} - \max\{\mathfrak{x}_{k-1} - \mathfrak{x}_n, 0\}) \\
&= \sum_{n=0}^{k-1} c_n [(x - \mathfrak{x}_n) - (\mathfrak{x}_{k-1} - \mathfrak{x}_n)] = \sum_{n=0}^{k-1} c_n (x - \mathfrak{x}_{k-1}) \\
&= \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}).
\end{split} \tag{3.30}$$

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

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = f_{k-1} + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}). \tag{3.31}$$

We now prove (3.31) by induction on $k \in \{1, 2, \ldots, K\}$. For the base case $k = 1$ observe that (3.27) and (3.30) demonstrate that for all $x \in [\mathfrak{x}_0, \mathfrak{x}_1]$ it holds that

$$(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_0) + (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_0) = f_0 + \big(\tfrac{f_1 - f_0}{\mathfrak{x}_1 - \mathfrak{x}_0}\big)(x - \mathfrak{x}_0). \tag{3.32}$$

This proves (3.31) in the base case $k = 1$. For the induction step note that (3.30) implies that for all $k \in \mathbb{N} \cap (1, \infty) \cap [1, K]$, $x \in [\mathfrak{x}_{k-1}, \mathfrak{x}_k]$ with $\forall\, y \in [\mathfrak{x}_{k-2}, \mathfrak{x}_{k-1}] \colon (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(y) = f_{k-2} + \big(\frac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}}\big)(y - \mathfrak{x}_{k-2})$ it holds that

$$\begin{split}
&(\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) = (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_{k-1}) + (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(x) - (\mathcal{R}_\mathfrak{r}(\mathbf{F}))(\mathfrak{x}_{k-1}) \\
&\quad = f_{k-2} + \big(\tfrac{f_{k-1} - f_{k-2}}{\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}}\big)(\mathfrak{x}_{k-1} - \mathfrak{x}_{k-2}) + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}) = f_{k-1} + \big(\tfrac{f_k - f_{k-1}}{\mathfrak{x}_k - \mathfrak{x}_{k-1}}\big)(x - \mathfrak{x}_{k-1}).
\end{split} \tag{3.33}$$

Induction thus proves (3.31). Furthermore, observe that (3.24) and (3.28) ensure that

$$\sum_{n=0}^{K} c_n = c_K + \sum_{n=0}^{K-1} c_n = -\tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}} + \tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}} = 0. \tag{3.34}$$

The fact that $\forall\, k \in \{0, 1, \dots, K\} \colon \mathfrak{r}_k \leq \mathfrak{r}_K$ and (3.26) hence imply that for all $x \in [\mathfrak{r}_K, \infty)$ it holds that

$$\begin{split} (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{r}_K) &= \left[\sum_{n=0}^{K} c_n (\max\{x - \mathfrak{r}_n, 0\} - \max\{\mathfrak{r}_K - \mathfrak{r}_n, 0\})\right] \\ &= \sum_{n=0}^{K} c_n [(x - \mathfrak{r}_n) - (\mathfrak{r}_K - \mathfrak{r}_n)] = \sum_{n=0}^{K} c_n (x - \mathfrak{r}_K) = 0. \end{split} \tag{3.35}$$

This and (3.31) show that for all $x \in [\mathfrak{r}_K, \infty)$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = (\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(\mathfrak{r}_K) = f_{K-1} + (\tfrac{f_K - f_{K-1}}{\mathfrak{r}_K - \mathfrak{r}_{K-1}})(\mathfrak{r}_K - \mathfrak{r}_{K-1}) = f_K. \tag{3.36}$$

Combining this, (3.27), (3.31), and (3.6) establishes item (iii). The proof of Lemma 3.1.12 is thus complete. □

Exercise 3.1.1. *Prove or disprove the following statement: There exists $\Phi \in \mathbf{N}$ such that $\mathcal{P}(\Phi) \leq 16$ and*

$$\sup_{x \in [-2\pi, 2\pi]} \bigl|\cos(x) - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)\bigr| \leq \tfrac{1}{2} \tag{3.37}$$

(cf. Definitions 2.1.6, 2.2.1, and 2.2.3).

Exercise 3.1.2. *Prove or disprove the following statement: There exists $\Phi \in \mathbf{N}$ such that $\mathcal{I}(\Phi) = 4$, $\mathcal{O}(\Phi) = 1$, $\mathcal{P}(\Phi) \leq 60$, and $\forall\, x, y, u, v \in \mathbb{R} \colon (\mathcal{R}_{\mathfrak{r}}(\Phi))(x, y, u, v) = \max\{x, y, u, v\}$ (cf. Definitions 2.1.6, 2.2.1, and 2.2.3).*

Exercise 3.1.3. *Prove or disprove the following statement: For every $m \in \mathbb{N}$ there exists $\Phi \in \mathbf{N}$ such that $\mathcal{I}(\Phi) = 2^m$, $\mathcal{O}(\Phi) = 1$, $\mathcal{P}(\Phi) \leq 3(2^m(2^m + 1))$, and $\forall\, x = (x_1, x_2, \dots, x_{2^m}) \in \mathbb{R} \colon (\mathcal{R}_{\mathfrak{r}}(\Phi))(x) = \max\{x_1, x_2, \dots, x_{2^m}\}$ (cf. Definitions 2.1.6, 2.2.1, and 2.2.3).*

3.1.4 ANN approximations results for one-dimensional functions

Lemma 3.1.13. *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 $\mathfrak{r}_k = a + \frac{k(b-a)}{K}$, let $f \colon [a, b] \to \mathbb{R}$ satisfy for all $x, y \in [a, b]$ that $|f(x) - f(y)| \leq L|x - y|$, and let $\mathbf{F} \in \mathbf{N}$ satisfy*

$$\mathbf{F} = \mathbf{A}_{1, f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1, K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1, 0\}}))}{(b-a)}\right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1, -\mathfrak{r}_k})\right)\right) \tag{3.38}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.18, 2.2.21, 2.2.32, and 3.1.8). Then

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

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

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

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

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

(cf. Definitions 2.1.6, 2.2.3, and 3.1.4).

Proof of Lemma 3.1.13. Note that the fact that $\forall\, k \in \{0, 1, \ldots, K\} \colon \mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_{\min\{k,K-1\}} = \mathfrak{x}_{\max\{k,1\}} - \mathfrak{x}_{\max\{k-1,0\}} = (b-a)K^{-1}$ assures that for all $k \in \{0, 1, \ldots, K\}$ it holds that

$$\tfrac{(f(\mathfrak{x}_{\min\{k+1,K\}}) - f(\mathfrak{x}_k))}{(\mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_{\min\{k,K-1\}})} - \tfrac{(f(\mathfrak{x}_k) - f(\mathfrak{x}_{\max\{k-1,0\}}))}{(\mathfrak{x}_{\max\{k,1\}} - \mathfrak{x}_{\max\{k-1,0\}})} = \tfrac{K(f(\mathfrak{x}_{\min\{k+1,K\}}) - 2f(\mathfrak{x}_k) + f(\mathfrak{x}_{\max\{k-1,0\}}))}{(b-a)}. \tag{3.39}$$

This and items (i), (ii), (iii), and (iv) in Lemma 3.1.12 prove items (i), (ii), (iii), and (vi). Combining item (iii) with the assumption that $\forall\, x, y \in [a,b] \colon |f(x) - f(y)| \leq L|x-y|$ and item (i) in Lemma 3.1.7 establishes item (iv). Moreover, note that item (iii), the assumption that $\forall\, x, y \in [a,b] \colon |f(x) - f(y)| \leq L|x-y|$, item (ii) in Lemma 3.1.7, and the fact that $\forall\, k \in \{1, 2, \ldots, K\} \colon \mathfrak{x}_k - \mathfrak{x}_{k-1} = (b-a)K^{-1}$ demonstrate that for all $x \in [a,b]$ it holds that

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq L\bigg(\max_{k \in \{1,2,\ldots,K\}} |\mathfrak{x}_k - \mathfrak{x}_{k-1}|\bigg) = L(b-a)K^{-1}. \tag{3.40}$$

This establishes item (v). The proof of Lemma 3.1.13 is thus complete. □

Lemma 3.1.14. *Let* $L, a \in \mathbb{R}$, $b \in [a, \infty)$, $\xi \in [a,b]$, *let* $f \colon [a,b] \to \mathbb{R}$ *satisfy for all* $x, y \in [a,b]$ *that* $|f(x) - f(y)| \leq L|x-y|$, *and let* $\mathbf{F} \in \mathbf{N}$ *satisfy* $\mathbf{F} = \mathbf{A}_{1, f(\xi)} \bullet (0 \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\xi}))$ *(cf. Definitions 2.2.1, 2.2.5, 2.2.18, 2.2.21, and 3.1.8). Then*

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

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

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

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

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

(cf. Definitions 2.1.6 and 2.2.3).

Proof of Lemma 3.1.14. Note that items (i) and (ii) in Lemma 2.2.20, and items (ii) and (iii) in Lemma 3.1.11 establish items (i) and (ii). In addition, observe that item (iii) in Lemma 2.2.20 and item (iii) in Lemma 2.2.22 assure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) &= (\mathcal{R}_{\mathfrak{r}}(0 \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\xi})))(x) + f(\xi) \\ &= 0\big((\mathcal{R}_{\mathfrak{r}}(\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\xi}))(x)\big) + f(\xi) = f(\xi)\end{aligned} \tag{3.41}$$

(cf. Definitions 2.1.6 and 2.2.3). This establishes item (iii). Next note that (3.41), 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

$$|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| = |f(\xi) - f(x)| \leq L|x-\xi| \leq L\max\{\xi - a, b - \xi\}. \tag{3.42}$$

This establishes item (iv). Moreover, note that (2.52) and item (i) assure that

$$\mathcal{P}(\mathbf{F}) = 1(1+1) + 1(1+1) = 4. \tag{3.43}$$

This establishes item (v). The proof of Lemma 3.1.14 it thus completed. □

Corollary 3.1.15. *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 $\mathfrak{r}_k = a + \frac{k(b-a)}{\max\{K,1\}}$, let $f\colon [a,b] \to \mathbb{R}$ satisfy for all $x,y \in [a,b]$ that $|f(x)-f(y)| \leq L|x-y|$, and let $\mathbf{F} \in \mathbf{N}$ satisfy*

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} \left(\left(\tfrac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)} \right) \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k}) \right) \right) \tag{3.44}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.18, 2.2.21, 2.2.32, and 3.1.8). Then

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

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

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

(iv) 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

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

(cf. Definitions 2.1.6, 2.2.1, and 2.2.3).

Proof of Corollary 3.1.15. Note that the fact that $K \in \mathbb{N}_0 \cap \big[\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon}+1\big)$ implies that $\frac{L(b-a)}{\max\{K,1\}} \leq \varepsilon$. This, items (i), (ii), (iv), and (v) in Lemma 3.1.13, and items (i), (ii), (iii), and (iv) in Lemma 3.1.14 establish items (i), (ii), (iii), and (iv). Moreover, note that the fact that $K \leq 1 + \frac{L(b-a)}{\varepsilon}$, item (vi) in Lemma 3.1.13, and item (v) in Lemma 3.1.14 assure that

$$\mathcal{P}(\mathbf{F}) = 3K + 4 \leq \frac{3L(b-a)}{\varepsilon} + 7. \tag{3.45}$$

This establishes item (v). The proof of Corollary 3.1.15 is thus complete. □

Definition 3.1.16 (p-norm). *We denote by* $\|\cdot\|_p \colon \left(\bigcup_{d=1}^{\infty} \mathbb{R}^d\right) \to \mathbb{R}$, $p \in [1,\infty]$, *the functions which satisfy for all* $p \in [1,\infty)$, $d \in \mathbb{N}$, $\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \mathbb{R}^d$ *that* $\|\theta\|_p = \left[\sum_{i=1}^{d} |\theta_i|^p\right]^{1/p}$ *and* $\|\theta\|_\infty = \max_{i \in \{1,2,\ldots,d\}} |\theta_i|$.

Corollary 3.1.17. *Let* $\varepsilon \in (0,\infty)$, $L \in [0,\infty)$, $a \in \mathbb{R}$, $b \in [a,\infty)$ *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* $\sup_{x \in [a,b]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq \varepsilon$,

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

(vii) *it holds that* $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\{1, |a|, |b|, 2L, |f(a)|\}$

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, 2.2.34, and 3.1.16).

Proof of Corollary 3.1.17. Throughout this proof assume without loss of generality that $a < b$, let $K \in \mathbb{N}_0 \cap \left[\frac{L(b-a)}{\varepsilon}, \frac{L(b-a)}{\varepsilon} + 1\right)$, $\mathfrak{r}_0, \mathfrak{r}_1, \ldots, \mathfrak{r}_K \in \mathbb{R}$, $c_0, c_1, \ldots, c_K \in \mathbb{R}$ satisfy for all $k \in \{0, 1, \ldots, K\}$ that $\mathfrak{r}_k = a + \frac{k(b-a)}{\max\{K,1\}}$ and

$$c_k = \frac{K(f(\mathfrak{r}_{\min\{k+1,K\}}) - 2f(\mathfrak{r}_k) + f(\mathfrak{r}_{\max\{k-1,0\}}))}{(b-a)}, \tag{3.46}$$

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

$$\mathbf{F} = \mathbf{A}_{1,f(\mathfrak{r}_0)} \bullet \left(\bigoplus_{k=0}^{K} (c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{r}_k})) \right) \tag{3.47}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.18, 2.2.21, 2.2.32, and 3.1.8). Note that Corollary 3.1.15 implies that

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

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

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

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

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

(cf. Definitions 2.1.6 and 2.2.3). This establishes items (i), (iv), and (v). Next note that item (I) and the fact that $K \leq 1 + \frac{L(b-a)}{\varepsilon}$ prove items (ii) and (iii). Next observe that items (I) and (V) imply that

$$\mathcal{P}(\mathbf{F}) = 3K + 4 = 3(K+1) + 1 = 3(\mathbb{D}_1(\mathbf{F})) + 1 \leq \tfrac{3L(b-a)}{\varepsilon} + 7. \tag{3.48}$$

This establishes item (vi). In the next step we observe that Lemma 3.1.11 shows that for all $k \in \{0, 1, \ldots, K\}$ it holds that

$$c_k \circledast (\mathfrak{i}_1 \bullet \mathbf{A}_{1,-\mathfrak{x}_k}) = ((1, -\mathfrak{x}_k), (c_k, 0)). \tag{3.49}$$

Combining this with (2.159), (2.152), (2.145), and Lemma 2.2.6 demonstrates that

$$\mathbf{F} = \left(\left(\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} -\mathfrak{x}_0 \\ -\mathfrak{x}_1 \\ \vdots \\ -\mathfrak{x}_K \end{pmatrix} \right), \left(\begin{pmatrix} c_0 & c_1 & \cdots & c_K \end{pmatrix}, f(\mathfrak{x}_0) \right) \right) \in (\mathbb{R}^{(K+1)\times 1} \times \mathbb{R}^{K+1}) \times (\mathbb{R}^{1\times(K+1)} \times \mathbb{R}). \tag{3.50}$$

Lemma 2.2.36 therefore ensures that

$$\|\mathcal{T}(\mathbf{F})\|_\infty = \max\{|\mathfrak{x}_0|, |\mathfrak{x}_1|, \ldots, |\mathfrak{x}_K|, |c_0|, |c_1|, \ldots, |c_K|, |f(\mathfrak{x}_0)|, 1\} \tag{3.51}$$

(cf. Definitions 2.2.34 and 3.1.16). In addition, note that the assumption that for all $x, y \in [a, b]$ it holds that $|f(x) - f(y)| \leq L|x - y|$ and the fact that $\forall\, k \in \mathbb{N} \cap (0, K+1)\colon \mathfrak{x}_k - \mathfrak{x}_{k-1} = (b-a)[\max\{K, 1\}]^{-1}$ imply that for all $k \in \{0, 1, \ldots, K\}$ it holds that

$$\begin{aligned} |c_k| &\leq \frac{K(|f(\mathfrak{x}_{\min\{k+1,K\}}) - f(\mathfrak{x}_k)| + |f(\mathfrak{x}_{\max\{k-1,0\}})) - f(\mathfrak{x}_k)|}{(b-a)} \\ &\leq \frac{KL(|\mathfrak{x}_{\min\{k+1,K\}} - \mathfrak{x}_k| + |\mathfrak{x}_{\max\{k-1,0\}} - \mathfrak{x}_k|)}{(b-a)} \\ &\leq \frac{2KL(b-a)[\max\{K,1\}]^{-1}}{(b-a)} \leq 2L. \end{aligned} \tag{3.52}$$

This and (3.51) establish item (vii). The proof of Corollary 3.1.17 is thus complete. □

Corollary 3.1.18. *Let $L, a \in \mathbb{R}$, $b \in [a, \infty)$ 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 exist $c \in \mathbb{R}$ and $\mathbf{F} = (\mathbf{F}_\varepsilon)_{\varepsilon \in (0,1]}\colon (0, 1] \to \mathbf{N}$ such that for all $\varepsilon \in (0, 1]$ it holds that*

$$\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon) \in C(\mathbb{R}, \mathbb{R}), \qquad \sup\nolimits_{x\in[a,b]} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(x) - f(x)| \leq \varepsilon, \qquad \mathcal{H}(\mathbf{F}_\varepsilon) = 1, \tag{3.53}$$

$$\|\mathcal{T}(\mathbf{F}_\varepsilon)\|_\infty \leq \max\{1, |a|, |b|, 2L, |f(a)|\}, \qquad \textit{and} \qquad \mathcal{P}(\mathbf{F}_\varepsilon) \leq c\varepsilon^{-1} \tag{3.54}$$

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, 2.2.34, and 3.1.16).

Proof of Corollary 3.1.18. Throughout this proof assume without loss of generality that $L \geq 0$ and let $c = 3L(b-a) + 7$. Observe that for all $\varepsilon \in (0,1]$ it holds that

$$3L(b-a)\varepsilon^{-1} + 7 \leq 3L(b-a)\varepsilon^{-1} + 7\varepsilon^{-1} = c\varepsilon^{-1}. \tag{3.55}$$

This and Corollary 3.1.17 establish that there exists $\mathbf{F} = (\mathbf{F}_\varepsilon)_{\varepsilon\in(0,1]} \colon (0,1] \to \mathbf{N}$ such that for all $\varepsilon \in (0,1]$ it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon) \in C(\mathbb{R},\mathbb{R})$, $\sup_{x\in[a,b]}|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(x) - f(x)| \leq \varepsilon$, $\mathcal{H}(\mathbf{F}_\varepsilon) = 1$, $\|\mathcal{T}(\mathbf{F}_\varepsilon)\|_\infty \leq \max\{1, |a|, |b|, 2L, |f(a)|\}$, and

$$\mathcal{P}(\mathbf{F}_\varepsilon) \leq 3L(b-a)\varepsilon^{-1} + 7 \leq c\varepsilon^{-1} \tag{3.56}$$

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, 2.2.34, and 3.1.16). The proof of Corollary 3.1.18 is thus complete. □

Corollary 3.1.19. *Let $L, a \in \mathbb{R}$, $b \in [a,\infty)$ 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 exist $c \in \mathbb{R}$ and $\mathbf{F} = (\mathbf{F}_\varepsilon)_{\varepsilon\in(0,1]} \colon (0,1] \to \mathbf{N}$ such that for all $\varepsilon \in (0,1]$ it holds that*

$$\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon) \in C(\mathbb{R},\mathbb{R}), \qquad \sup\nolimits_{x\in[a,b]}|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}_\varepsilon))(x) - f(x)| \leq \varepsilon, \qquad \textit{and} \qquad \mathcal{P}(\mathbf{F}_\varepsilon) \leq c\varepsilon^{-1} \tag{3.57}$$

(cf. Definitions 2.1.6, 2.2.1, and 2.2.3).

Proof of Corollary 3.1.19. Observe that Corollary 3.1.18 establishes (3.57). The proof of Corollary 3.1.19 is thus complete. □

Exercise 3.1.4. *Prove or disprove the following statement: There exists $\Phi \in \mathbf{N}$ such that $\mathcal{P}(\Phi) \leq 10$ and*

$$\sup_{x\in[0,10]} \left|\sqrt{x} - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)\right| \leq \tfrac{1}{4} \tag{3.58}$$

(cf. Definitions 2.1.6, 2.2.1, and 2.2.3).

3.2 Multi-dimensional ANN approximation results

3.2.1 Approximations through supremal convolutions

Lemma 3.2.1. *Let (E,δ) be a metric space, let $L \in [0,\infty)$, $D \subseteq E$, $\mathcal{M} \subseteq E$ satisfy $\emptyset \neq \mathcal{M} \subseteq D$, let $f \colon D \to \mathbb{R}$ satisfy for all $x \in D$, $y \in \mathcal{M}$ that $|f(x) - f(y)| \leq L\delta(x,y)$, and let $F \colon E \to \mathbb{R} \cup \{\infty\}$ satisfy for all $x \in E$ that*

$$F(x) = \sup_{y\in\mathcal{M}} [f(y) - L\delta(x,y)]. \tag{3.59}$$

Then

(i) *it holds for all $x \in \mathcal{M}$ that $F(x) = f(x)$,*

(ii) *it holds for all $x \in D$ that $F(x) \leq f(x)$,*

(iii) *it holds for all $x \in E$ that $F(x) < \infty$,*

(iv) *it holds for all $x, y \in E$ that $|F(x) - F(y)| \leq L\delta(x, y)$, and*

(v) *it holds for all $x \in D$ that*

$$|F(x) - f(x)| \leq 2L\left[\inf_{y \in \mathcal{M}} \delta(x, y)\right]. \tag{3.60}$$

Proof of Lemma 3.2.1. First, observe that the assumption that $\forall\, x \in D, y \in \mathcal{M} : |f(x) - f(y)| \leq L\delta(x, y)$ ensures that for all $x \in D$, $y \in \mathcal{M}$ it holds that

$$f(y) + L\delta(x, y) \geq f(x) \geq f(y) - L\delta(x, y). \tag{3.61}$$

Hence, we obtain that for all $x \in D$ it holds that

$$f(x) \geq \sup_{y \in \mathcal{M}} [f(y) - L\delta(x, y)] = F(x). \tag{3.62}$$

This establishes item (ii). Moreover, note that (3.59) implies that for all $x \in \mathcal{M}$ it holds that

$$F(x) \geq f(x) - L\delta(x, x) = f(x). \tag{3.63}$$

This and (3.62) establish item (i). Next observe that (3.61) (applied for all $y, z \in \mathcal{M}$ with $x \curvearrowleft y$, $y \curvearrowleft z$) and the triangle inequality ensure that for all $x \in E$, $y, z \in \mathcal{M}$ it holds that

$$f(y) - L\delta(x, y) \leq f(z) + L\delta(y, z) - L\delta(x, y) \leq f(z) + L\delta(x, z). \tag{3.64}$$

Hence, we obtain that for all $x \in E$, $z \in \mathcal{M}$ it holds that

$$F(x) = \sup_{y \in \mathcal{M}} [f(y) - L\delta(x, y)] \leq f(z) + L\delta(x, z) < \infty. \tag{3.65}$$

This and the assumption that $\mathcal{M} \neq \emptyset$ prove item (iii). Note that item (iii), (3.59), and the triangle inequality show that for all $x, y \in E$ it holds that

$$\begin{aligned}
F(x) - F(y) &= \left[\sup_{v \in \mathcal{M}} (f(v) - L\delta(x, v))\right] - \left[\sup_{w \in \mathcal{M}} (f(w) - L\delta(y, w))\right] \\
&= \sup_{v \in \mathcal{M}} \left[f(v) - L\delta(x, v) - \sup_{w \in \mathcal{M}} (f(w) - L\delta(y, w))\right] \\
&\leq \sup_{v \in \mathcal{M}} [f(v) - L\delta(x, v) - (f(v) - L\delta(y, v))] \\
&= \sup_{v \in \mathcal{M}} (L\delta(y, v) - L\delta(x, v)) \\
&\leq \sup_{v \in \mathcal{M}} (L\delta(y, x) + L\delta(x, v) - L\delta(x, v)) = L\delta(x, y).
\end{aligned} \tag{3.66}$$

This and the fact that for all $x, y \in E$ it holds that $\delta(x,y) = \delta(y,x)$ establish item (iv). Observe that items (i) and (iv), the triangle inequality, and the assumption that $\forall\, x \in D, y \in \mathcal{M}\colon |f(x) - f(y)| \leq L\delta(x,y)$ ensure that for all $x \in D$ it holds that

$$\begin{aligned} |F(x) - f(x)| &= \inf_{y \in \mathcal{M}} |F(x) - F(y) + f(y) - f(x)| \\ &\leq \inf_{y \in \mathcal{M}} (|F(x) - F(y)| + |f(y) - f(x)|) \\ &\leq \inf_{y \in \mathcal{M}} (2L\delta(x,y)) = 2L\left[\inf_{y \in \mathcal{M}} \delta(x,y)\right]. \end{aligned} \tag{3.67}$$

This establishes item (v). The proof of Lemma 3.2.1 is thus complete. □

Corollary 3.2.2. *Let (E, δ) be a metric space, let $L \in [0, \infty)$, $\mathcal{M} \subseteq E$ satisfy $\mathcal{M} \neq \emptyset$, let $f\colon E \to \mathbb{R}$ satisfy for all $x \in E$, $y \in \mathcal{M}$ that $|f(x) - f(y)| \leq L\delta(x,y)$, and let $F\colon E \to \mathbb{R} \cup \{\infty\}$ satisfy for all $x \in E$ that*

$$F(x) = \sup_{y \in \mathcal{M}} [f(y) - L\delta(x,y)]. \tag{3.68}$$

Then

(i) it holds for all $x \in \mathcal{M}$ that $F(x) = f(x)$,

(ii) it holds for all $x \in E$ that $F(x) \leq f(x)$,

(iii) it holds for all $x, y \in E$ that $|F(x) - F(y)| \leq L\delta(x,y)$, and

(iv) it holds for all $x \in E$ that

$$|F(x) - f(x)| \leq 2L\left[\inf_{y \in \mathcal{M}} \delta(x,y)\right]. \tag{3.69}$$

Proof of Corollary 3.2.2. Note that Lemma 3.2.1 establishes items (i), (ii), (iii), and (iv). The proof of Corollary 3.2.2 is thus complete. □

Exercise 3.2.1. *Prove or disprove the following statement: There exists $\Phi \in \mathbf{N}$ such that $\mathcal{I}(\Phi) = 2$, $\mathcal{O}(\Phi) = 1$, $\mathcal{P}(\Phi) \leq 3\,000\,000\,000$, and*

$$\sup_{x,y \in [0,2\pi]} |\sin(x)\sin(y) - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x,y)| \leq \tfrac{1}{5}. \tag{3.70}$$

3.2.2 ANN representations

3.2.2.1 ANN representations for the 1-norm

Definition 3.2.3 (1-norm ANN representations)**.** *We denote by $(\mathbb{L}_d)_{d \in \mathbb{N}} \subseteq \mathbf{N}$ the ANNs which satisfy that*

(i) *it holds that*

$$\mathbb{L}_1 = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 \end{pmatrix} \right) \right) \in (\mathbb{R}^{2\times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1\times 2} \times \mathbb{R}^1) \tag{3.71}$$

and

(ii) *it holds for all $d \in \{2, 3, 4, \ldots\}$ that $\mathbb{L}_d = \mathbb{S}_{1,d} \bullet \mathbf{P}_d(\mathbb{L}_1, \mathbb{L}_1, \ldots, \mathbb{L}_1)$*

(cf. Definitions 2.2.1, 2.2.5, 2.2.11, and 2.2.23).

Proposition 3.2.4. *Let $d \in \mathbb{N}$. Then*

(i) *it holds that $\mathcal{D}(\mathbb{L}_d) = (d, 2d, 1)$,*

(ii) *it holds that $\mathcal{R}_{\mathfrak{r}}(\mathbb{L}_d) \in C(\mathbb{R}^d, \mathbb{R})$, and*

(iii) *it holds for all $x \in \mathbb{R}^d$ that $(\mathcal{R}_{\mathfrak{r}}(\mathbb{L}_d))(x) = \|x\|_1$*

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, 3.1.16, and 3.2.3).

Proof of Proposition 3.2.4. Note that the fact that $\mathcal{D}(\mathbb{L}_1) = (1, 2, 1)$ and Lemma 2.2.12 show that for all $\mathfrak{d} \in \{2, 3, 4, \ldots\}$ it holds that $\mathcal{D}(\mathbf{P}_{\mathfrak{d}}(\mathbb{L}_1, \mathbb{L}_1, \ldots, \mathbb{L}_1)) = (\mathfrak{d}, 2\mathfrak{d}, \mathfrak{d})$ (cf. Definitions 2.2.1, 2.2.11, and 3.2.3). Combining this, Proposition 2.2.7, and Lemma 2.2.19 ensures that for all $\mathfrak{d} \in \{2, 3, 4, \ldots\}$ it holds that $\mathcal{D}(\mathbb{S}_{1,\mathfrak{d}} \bullet \mathbf{P}_{\mathfrak{d}}(\mathbb{L}_1, \mathbb{L}_1, \ldots, \mathbb{L}_1)) = (\mathfrak{d}, 2\mathfrak{d}, 1)$ (cf. Definitions 2.2.5 and 2.2.23). This establishes item (i). Furthermore, observe that (3.71) assures that for all $x \in \mathbb{R}$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbb{L}_1))(x) = \mathfrak{r}(x) + \mathfrak{r}(-x) = \max\{x, 0\} + \max\{-x, 0\} = |x| = \|x\|_1 \tag{3.72}$$

(cf. Definitions 2.1.6, 2.2.3, and 3.1.16). Combining this and Proposition 2.2.13 shows that for all $\mathfrak{d} \in \{2, 3, 4, \ldots\}$, $x = (x_1, x_2, \ldots, x_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\big(\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_{\mathfrak{d}}(\mathbb{L}_1, \mathbb{L}_1, \ldots, \mathbb{L}_1))\big)(x) = (|x_1|, |x_2|, \ldots, |x_{\mathfrak{d}}|). \tag{3.73}$$

This and Lemma 2.2.24 demonstrate that for all $\mathfrak{d} \in \{2, 3, 4, \ldots\}$, $x = (x_1, x_2, \ldots, x_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ it holds that

$$\begin{aligned} (\mathcal{R}_{\mathfrak{r}}(\mathbb{L}_{\mathfrak{d}}))(x) &= \big(\mathcal{R}_{\mathfrak{r}}(\mathbb{S}_{1,\mathfrak{d}} \bullet \mathbf{P}_{\mathfrak{d}}(\mathbb{L}_1, \mathbb{L}_1, \ldots, \mathbb{L}_1))\big)(x) \\ &= \big(\mathcal{R}_{\mathfrak{r}}(\mathbb{S}_{1,\mathfrak{d}})\big)(|x_1|, |x_2|, \ldots, |x_{\mathfrak{d}}|) = \sum_{n=1}^{d} |x_n| = \|x\|_1. \end{aligned} \tag{3.74}$$

This establishes items (ii) and (iii). The proof of Proposition 3.2.4 is thus complete. ☐

Lemma 3.2.5. *Let $d \in \mathbb{N}$. Then*

(i) *it holds that* $\mathcal{B}_{1,\mathbb{L}_d} = 0 \in \mathbb{R}^{2d}$,

(ii) *it holds that* $\mathcal{B}_{2,\mathbb{L}_d} = 0 \in \mathbb{R}$,

(iii) *it holds that* $\mathcal{W}_{1,\mathbb{L}_d} \in \{-1,0,1\}^{(2d)\times d}$,

(iv) *it holds for all* $x \in \mathbb{R}^d$ *that* $\|\mathcal{W}_{1,\mathbb{L}_d} x\|_\infty = \|x\|_\infty$, *and*

(v) *it holds that* $\mathcal{W}_{2,\mathbb{L}_d} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{1\times(2d)}$

(cf. Definitions 2.2.1, 3.1.16, and 3.2.3).

Proof of Lemma 3.2.5. Throughout this proof assume without loss of generality that $d > 1$. Note that the fact that $\mathcal{B}_{1,\mathbb{L}_1} = 0 \in \mathbb{R}^2$, the fact that $\mathcal{B}_{2,\mathbb{L}_1} = 0 \in \mathbb{R}$, the fact that $\mathcal{B}_{1,\mathbb{S}_{1,d}} = 0 \in \mathbb{R}$, and the fact that $\mathbb{L}_d = \mathbb{S}_{1,d} \bullet \mathbf{P}_d(\mathbb{L}_1, \mathbb{L}_1, \ldots, \mathbb{L}_1)$ establish items (i) and (ii) (cf. Definitions 2.2.1, 2.2.5, 2.2.11, 2.2.23, and 3.2.3). In addition, observe that the fact that

$$\mathcal{W}_{1,\mathbb{L}_1} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad \text{and} \qquad \mathcal{W}_{1,\mathbb{L}_d} = \begin{pmatrix} \mathcal{W}_{1,\mathbb{L}_1} & 0 & \cdots & 0 \\ 0 & \mathcal{W}_{1,\mathbb{L}_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{W}_{1,\mathbb{L}_1} \end{pmatrix} \in \mathbb{R}^{(2d)\times d} \tag{3.75}$$

proves item (iii). Next note that (3.75) implies item (iv). Moreover, note that the fact that $\mathcal{W}_{2,\mathbb{L}_1} = (1\ 1)$ and the fact that $\mathbb{L}_d = \mathbb{S}_{1,d} \bullet \mathbf{P}_d(\mathbb{L}_1, \mathbb{L}_1, \ldots, \mathbb{L}_1)$ show that

$$\begin{aligned} \mathcal{W}_{2,\mathbb{L}_d} &= \mathcal{W}_{1,\mathbb{S}_{1,d}} \mathcal{W}_{2,\mathbf{P}_d(\mathbb{L}_1,\mathbb{L}_1,\ldots,\mathbb{L}_1)} \\ &= \underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}}_{\in \mathbb{R}^{1\times d}} \underbrace{\begin{pmatrix} \mathcal{W}_{2,\mathbb{L}_1} & 0 & \cdots & 0 \\ 0 & \mathcal{W}_{2,\mathbb{L}_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{W}_{2,\mathbb{L}_1} \end{pmatrix}}_{\in \mathbb{R}^{d\times(2d)}} \\ &= \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{1\times(2d)}. \end{aligned} \tag{3.76}$$

This establishes item (v). The proof of Lemma 3.2.5 is thus complete. □

3.2.2.2 ANN representations for maxima

Lemma 3.2.6. *There exist unique* $(\phi_d)_{d\in\mathbb{N}} \subseteq \mathbf{N}$ *which satisfy that*

(i) *it holds for all* $d \in \mathbb{N}$ *that* $\mathcal{I}(\phi_d) = d$,

(ii) *it holds for all* $d \in \mathbb{N}$ *that* $\mathcal{O}(\phi_d) = 1$,

(iii) *it holds that* $\phi_1 = \mathbf{A}_{1,0} \in \mathbb{R}^{1\times 1} \times \mathbb{R}^1$,

(iv) *it holds that*

$$\phi_2 = \left(\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 0 \end{pmatrix} \right) \right) \in (\mathbb{R}^{3\times 2} \times \mathbb{R}^3) \times (\mathbb{R}^{1\times 3} \times \mathbb{R}^1), \tag{3.77}$$

(v) *it holds for all* $d \in \{2, 3, 4, \ldots\}$ *that* $\phi_{2d} = \phi_d \bullet \big(\mathbf{P}_d(\phi_2, \phi_2, \ldots, \phi_2)\big)$, *and*

(vi) *it holds for all* $d \in \{2, 3, 4, \ldots\}$ *that* $\phi_{2d-1} = \phi_d \bullet \big(\mathbf{P}_d(\phi_2, \phi_2, \ldots, \phi_2, \mathfrak{I}_1)\big)$

(cf. Definitions 2.2.1, 2.2.5, 2.2.11, 2.2.16, and 2.2.18).

Proof of Lemma 3.2.6. Throughout this proof let $\psi \in \mathbf{N}$ satisfy

$$\psi = \left(\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 0 \end{pmatrix} \right) \right) \in (\mathbb{R}^{3\times 2} \times \mathbb{R}^3) \times (\mathbb{R}^{1\times 3} \times \mathbb{R}^1) \tag{3.78}$$

(cf. Definition 2.2.1). Note that the fact that $\mathcal{I}(\psi) = 2$, the fact that $\mathcal{O}(\psi) = 1$, the fact that $\mathcal{L}(\psi) = \mathcal{L}(\mathfrak{I}_1) = 2$, Lemma 2.2.12, and Lemma 2.2.17 assure that for all $d \in \mathbb{N}$ it holds that $\mathcal{I}(\mathbf{P}_d(\psi, \psi, \ldots, \psi)) = 2d$, $\mathcal{O}(\mathbf{P}_d(\psi, \psi, \ldots, \psi)) = d$, $\mathcal{I}(\mathbf{P}_d(\psi, \psi, \ldots, \psi, \mathfrak{I}_1)) = 2d - 1$, and $\mathcal{O}(\mathbf{P}_d(\psi, \psi, \ldots, \psi, \mathfrak{I}_1)) = d$ (cf. Definitions 2.2.11 and 2.2.16). This, Proposition 2.2.7, and induction establish that there exists unique $\phi_d \in \mathbf{N}$, $d \in \mathbb{N}$, which satisfy that for all $d \in \mathbb{N}$ it holds that $\mathcal{I}(\phi_d) = d$, $\mathcal{O}(\phi_d) = 1$, and

$$\phi_d = \begin{cases} \mathbf{A}_{1,0} & : d = 1 \\ \psi & : d = 2 \\ \phi_{d/2} \bullet \big(\mathbf{P}_{d/2}(\psi, \psi, \ldots, \psi)\big) & : d \in \{4, 6, 8, \ldots\} \\ \phi_{(d+1)/2} \bullet \big(\mathbf{P}_{(d+1)/2}(\psi, \psi, \ldots, \psi, \mathfrak{I}_1)\big) & : d \in \{3, 5, 7, \ldots\}. \end{cases} \tag{3.79}$$

The proof of Lemma 3.2.6 is thus complete. □

Definition 3.2.7 (Maxima ANN representations)**.** *We denote by* $(\mathbb{M}_d)_{d\in\mathbb{N}} \subseteq \mathbf{N}$ *the ANNs which satisfy that*

(i) *it holds for all* $d \in \mathbb{N}$ *that* $\mathcal{I}(\mathbb{M}_d) = d$,

(ii) *it holds for all* $d \in \mathbb{N}$ *that* $\mathcal{O}(\mathbb{M}_d) = 1$,

(iii) *it holds that* $\mathbb{M}_1 = \mathbf{A}_{1,0} \in \mathbb{R}^{1\times 1} \times \mathbb{R}^1$,

(iv) *it holds that*

$$\mathbb{M}_2 = \left(\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 0 \end{pmatrix} \right) \right) \in (\mathbb{R}^{3\times 2} \times \mathbb{R}^3) \times (\mathbb{R}^{1\times 3} \times \mathbb{R}^1), \quad (3.80)$$

(v) *it holds for all* $d \in \{2, 3, 4, \dots\}$ *that* $\mathbb{M}_{2d} = \mathbb{M}_d \bullet \big(\mathbf{P}_d(\mathbb{M}_2, \mathbb{M}_2, \dots, \mathbb{M}_2)\big)$, *and*

(vi) *it holds for all* $d \in \{2, 3, 4, \dots\}$ *that* $\mathbb{M}_{2d-1} = \mathbb{M}_d \bullet \big(\mathbf{P}_d(\mathbb{M}_2, \mathbb{M}_2, \dots, \mathbb{M}_2, \mathfrak{I}_1)\big)$

(cf. Definitions 2.2.1, 2.2.5, 2.2.11, 2.2.16, and 2.2.18 and Lemma 3.2.6).

Definition 3.2.8 (Floor and ceiling of real numbers)**.** *We denote by* $\lceil\cdot\rceil \colon \mathbb{R} \to \mathbb{Z}$ *and* $\lfloor\cdot\rfloor \colon \mathbb{R} \to \mathbb{Z}$ *the functions which satisfy for all* $x \in \mathbb{R}$ *that* $\lceil x \rceil = \min(\mathbb{Z} \cap [x, \infty))$ *and* $\lfloor x \rfloor = \max(\mathbb{Z} \cap (-\infty, x])$.

Exercise 3.2.2. *Prove or disprove the following statement: For all* $n \in \{3, 5, 7, \dots\}$ *it holds that* $\lceil \log_2(n + 1) \rceil = \lceil \log_2(n) \rceil$.

Proposition 3.2.9. *Let* $d \in \mathbb{N}$. *Then*

(i) *it holds that* $\mathcal{H}(\mathbb{M}_d) = \lceil \log_2(d) \rceil$,

(ii) *it holds for all* $i \in \mathbb{N}$ *that* $\mathbb{D}_i(\mathbb{M}_d) \leq 3\lceil \frac{d}{2^i} \rceil$,

(iii) *it holds that* $\mathcal{R}_\mathfrak{r}(\mathbb{M}_d) \in C(\mathbb{R}^d, \mathbb{R})$, *and*

(iv) *it holds for all* $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ *that* $(\mathcal{R}_\mathfrak{r}(\mathbb{M}_d))(x) = \max\{x_1, x_2, \dots, x_d\}$

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, 3.2.7, and 3.2.8).

Proof of Proposition 3.2.9. Throughout this proof assume without loss of generality that $d > 1$. Note that (3.80) ensures that $\mathcal{H}(\mathbb{M}_2) = 1$ (cf. Definitions 2.2.1 and 3.2.7). This and (2.110) demonstrate that for all $\mathfrak{d} \in \{2, 3, 4, \dots\}$ it holds that

$$\mathcal{H}(\mathbf{P}_\mathfrak{d}(\mathbb{M}_2, \mathbb{M}_2, \dots, \mathbb{M}_2)) = \mathcal{H}(\mathbf{P}_\mathfrak{d}(\mathbb{M}_2, \mathbb{M}_2, \dots, \mathbb{M}_2, \mathfrak{I}_1)) = \mathcal{H}(\mathbb{M}_2) = 1 \qquad (3.81)$$

(cf. Definitions 2.2.11 and 2.2.16). Combining this with Proposition 2.2.7 establishes that for all $\mathfrak{d} \in \{3, 4, 5, \dots\}$ it holds that

$$\mathcal{H}(\mathbb{M}_\mathfrak{d}) = \mathcal{H}(\mathbb{M}_{\lceil \mathfrak{d}/2 \rceil}) + 1 \qquad (3.82)$$

(cf. Definition 3.2.8). This assures that for all $\mathfrak{d} \in \{4, 6, 8, \dots\}$ with $\mathcal{H}(\mathbb{M}_{\mathfrak{d}/2}) = \lceil \log_2(\mathfrak{d}/2) \rceil$ it holds that

$$\mathcal{H}(\mathbb{M}_\mathfrak{d}) = \lceil \log_2(\mathfrak{d}/2) \rceil + 1 = \lceil \log_2(\mathfrak{d}) - 1 \rceil + 1 = \lceil \log_2(\mathfrak{d}) \rceil. \qquad (3.83)$$

Moreover, note that (3.82) and the fact that for all $\mathfrak{d} \in \{3,5,7,\ldots\}$ it holds that $\lceil \log_2(\mathfrak{d}+1)\rceil = \lceil \log_2(\mathfrak{d})\rceil$ ensure that for all $\mathfrak{d} \in \{3,5,7,\ldots\}$ with $\mathcal{H}(\mathbb{M}_{\lceil \mathfrak{d}/2\rceil}) = \lceil \log_2(\lceil \mathfrak{d}/2\rceil)\rceil$ it holds that

$$\begin{aligned}\mathcal{H}(\mathbb{M}_{\mathfrak{d}}) &= \big\lceil \log_2(\lceil \mathfrak{d}/2\rceil)\big\rceil + 1 = \big\lceil \log_2(({\mathfrak{d}+1})/2)\big\rceil + 1\\ &= \lceil \log_2(\mathfrak{d}+1) - 1\rceil + 1 = \lceil \log_2(\mathfrak{d}+1)\rceil = \lceil \log_2(\mathfrak{d})\rceil.\end{aligned} \tag{3.84}$$

Combining this and (3.83) demonstrates that for all $\mathfrak{d} \in \{3,4,5,\ldots\}$ with $\forall\, k \in \{2,3,\ldots,\mathfrak{d}-1\}\colon \mathcal{H}(\mathbb{M}_k) = \lceil \log_2(k)\rceil$ it holds that $\mathcal{H}(\mathbb{M}_{\mathfrak{d}}) = \lceil \log_2(\mathfrak{d})\rceil$. The fact that $\mathcal{H}(\mathbb{M}_2) = 1$ and induction hence establish item (i). Next note that the fact that $\mathcal{D}(\mathbb{M}_2) = (2,3,1)$ assure that for all $i \in \mathbb{N}$ it holds that

$$\mathbb{D}_i(\mathbb{M}_2) \leq 3 = 3\big\lceil \tfrac{2}{2^i}\big\rceil. \tag{3.85}$$

Moreover, observe that Proposition 2.2.7 and Lemma 2.2.12 imply that for all $\mathfrak{d} \in \{2,3,4,\ldots\}$, $i \in \mathbb{N}$ it holds that

$$\mathbb{D}_i(\mathbb{M}_{2\mathfrak{d}}) = \begin{cases} 3\mathfrak{d} & : i = 1\\ \mathbb{D}_{i-1}(\mathbb{M}_{\mathfrak{d}}) & : i \geq 2\end{cases} \tag{3.86}$$

and

$$\mathbb{D}_i(\mathbb{M}_{2\mathfrak{d}-1}) = \begin{cases} 3\mathfrak{d} - 1 & : i = 1\\ \mathbb{D}_{i-1}(\mathbb{M}_{\mathfrak{d}}) & : i \geq 2.\end{cases} \tag{3.87}$$

This assures that for all $\mathfrak{d} \in \{2,4,6,\ldots\}$ it holds that

$$\mathbb{D}_1(\mathbb{M}_{\mathfrak{d}}) = 3(\tfrac{\mathfrak{d}}{2}) = 3\big\lceil \tfrac{\mathfrak{d}}{2}\big\rceil. \tag{3.88}$$

Moreover, note that (3.87) ensures that for all $\mathfrak{d} \in \{3,5,7,\ldots\}$ it holds that

$$\mathbb{D}_1(\mathbb{M}_{\mathfrak{d}}) = 3\big\lceil \tfrac{\mathfrak{d}}{2}\big\rceil - 1 \leq 3\big\lceil \tfrac{\mathfrak{d}}{2}\big\rceil. \tag{3.89}$$

This and (3.88) show that for all $\mathfrak{d} \in \{2,3,4,\ldots\}$ it holds that

$$\mathbb{D}_1(\mathbb{M}_{\mathfrak{d}}) \leq 3\big\lceil \tfrac{\mathfrak{d}}{2}\big\rceil. \tag{3.90}$$

In addition, observe that (3.86) demonstrates that for all $\mathfrak{d} \in \{4,6,8,\ldots\}$, $i \in \{2,3,4,\ldots\}$ with $\mathbb{D}_{i-1}(\mathbb{M}_{\mathfrak{d}/2}) \leq 3\big\lceil (\mathfrak{d}/2)\tfrac{1}{2^{i-1}}\big\rceil$ it holds that

$$\mathbb{D}_i(\mathbb{M}_{\mathfrak{d}}) = \mathbb{D}_{i-1}(\mathbb{M}_{\mathfrak{d}/2}) \leq 3\big\lceil (\mathfrak{d}/2)\tfrac{1}{2^{i-1}}\big\rceil = 3\big\lceil \tfrac{\mathfrak{d}}{2^i}\big\rceil. \tag{3.91}$$

Furthermore, note that the fact that for all $\mathfrak{d} \in \{3,5,7,\ldots\}$, $i \in \mathbb{N}$ it holds that $\big\lceil \tfrac{\mathfrak{d}+1}{2^i}\big\rceil = \big\lceil \tfrac{\mathfrak{d}}{2^i}\big\rceil$ and (3.87) assure that for all $\mathfrak{d} \in \{3,5,7,\ldots\}$, $i \in \{2,3,4,\ldots\}$ with $\mathbb{D}_{i-1}(\mathbb{M}_{\lceil \mathfrak{d}/2\rceil}) \leq 3\big\lceil \lceil \mathfrak{d}/2\rceil \tfrac{1}{2^{i-1}}\big\rceil$ it holds that

$$\mathbb{D}_i(\mathbb{M}_{\mathfrak{d}}) = \mathbb{D}_{i-1}(\mathbb{M}_{\lceil \mathfrak{d}/2\rceil}) \leq 3\big\lceil \lceil \mathfrak{d}/2\rceil \tfrac{1}{2^{i-1}}\big\rceil = 3\big\lceil \tfrac{\mathfrak{d}+1}{2^i}\big\rceil = 3\big\lceil \tfrac{\mathfrak{d}}{2^i}\big\rceil. \tag{3.92}$$

This and (3.91) ensure that for all $\mathfrak{d} \in \{3, 4, 5, \ldots\}$, $i \in \{2, 3, 4, \ldots\}$ with $\forall\, k \in \{2, 3, \ldots, \mathfrak{d} - 1\}, j \in \{1, 2, \ldots, i-1\} \colon \mathbb{D}_j(\mathbb{M}_k) \leq 3\lceil \frac{k}{2^j} \rceil$ it holds that

$$\mathbb{D}_i(\mathbb{M}_\mathfrak{d}) \leq 3\lceil \tfrac{\mathfrak{d}}{2^i} \rceil. \tag{3.93}$$

Combining this, (3.85), and (3.90) with induction establishes item (ii). Next observe that (3.80) ensures that for all $x = (x_1, x_2) \in \mathbb{R}^2$ it holds that

$$\begin{aligned} (\mathcal{R}_\mathfrak{r}(\mathbb{M}_2))(x) &= \max\{x_1 - x_2, 0\} + \max\{x_2, 0\} - \max\{-x_2, 0\} \\ &= \max\{x_1 - x_2, 0\} + x_2 = \max\{x_1, x_2\} \end{aligned} \tag{3.94}$$

(cf. Definitions 2.1.6 and 2.2.3). Proposition 2.2.13, Proposition 2.2.7, Lemma 2.2.17, and induction hence imply that for all $\mathfrak{d} \in \{2, 3, 4, \ldots\}$, $x = (x_1, x_2, \ldots, x_\mathfrak{d}) \in \mathbb{R}^\mathfrak{d}$ it holds that $\mathcal{R}_\mathfrak{r}(\mathbb{M}_\mathfrak{d}) \in C(\mathbb{R}^\mathfrak{d}, \mathbb{R})$ and $(\mathcal{R}_\mathfrak{r}(\mathbb{M}_\mathfrak{d}))(x) = \max\{x_1, x_2, \ldots, x_\mathfrak{d}\}$. This establishes items (iii) and (iv). The proof of Proposition 3.2.9 is thus complete. □

Lemma 3.2.10. *Let $d \in \mathbb{N}$, $i \in \{1, 2, \ldots, \mathcal{L}(\mathbb{M}_d)\}$ (cf. Definitions 2.2.1 and 3.2.7). Then*

(i) it holds that $\mathcal{B}_{i,\mathbb{M}_d} = 0 \in \mathbb{R}^{\mathbb{D}_i(\mathbb{M}_d)}$,

(ii) it holds that $\mathcal{W}_{i,\mathbb{M}_d} \in \{-1, 0, 1\}^{\mathbb{D}_i(\mathbb{M}_d) \times \mathbb{D}_{i-1}(\mathbb{M}_d)}$, and

(iii) it holds for all $x \in \mathbb{R}^d$ that $\|\mathcal{W}_{1,\mathbb{M}_d} x\|_\infty \leq 2\|x\|_\infty$

(cf. Definition 3.1.16).

Proof of Lemma 3.2.10. Throughout this proof assume without loss of generality that $d > 2$ (cf. items (iii) and (iv) in Definition 3.2.7) and let $A_1 \in \mathbb{R}^{3\times 2}$, $A_2 \in \mathbb{R}^{1\times 3}$, $C_1 \in \mathbb{R}^{2\times 1}$, $C_2 \in \mathbb{R}^{1\times 2}$ satisfy

$$A_1 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, \qquad C_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \text{and} \qquad C_2 = \begin{pmatrix} 1 & -1 \end{pmatrix}. \tag{3.95}$$

Note that items (iv), (v), and (vi) in Definition 3.2.7 assure that for all $\mathfrak{d} \in \{2, 3, 4, \ldots\}$ it holds that

$$\begin{gathered} \mathcal{W}_{1,\mathbb{M}_{2\mathfrak{d}-1}} = \underbrace{\begin{pmatrix} A_1 & 0 & \cdots & 0 & 0 \\ 0 & A_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_1 & 0 \\ 0 & 0 & \cdots & 0 & C_1 \end{pmatrix}}_{\in \mathbb{R}^{(3\mathfrak{d}-1)\times(2\mathfrak{d}-1)}}, \qquad \mathcal{W}_{1,\mathbb{M}_{2\mathfrak{d}}} = \underbrace{\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_1 \end{pmatrix}}_{\in \mathbb{R}^{(3\mathfrak{d})\times(2\mathfrak{d})}}, \\ \mathcal{B}_{1,\mathbb{M}_{2\mathfrak{d}-1}} = 0 \in \mathbb{R}^{3\mathfrak{d}-1}, \qquad \text{and} \qquad \mathcal{B}_{1,\mathbb{M}_{2\mathfrak{d}}} = 0 \in \mathbb{R}^{3\mathfrak{d}}. \end{gathered} \tag{3.96}$$

This and (3.95) proves item (iii). Furthermore, note that (3.96) and item (iv) in Definition 3.2.7 imply that for all $\mathfrak{d} \in \{2,3,4,\ldots\}$ it holds that $\mathcal{B}_{1,\mathbb{M}_\mathfrak{d}} = 0$. Items (iv), (v), and (vi) in Definition 3.2.7 hence ensure that for all $\mathfrak{d} \in \{2,3,4,\ldots\}$ it holds that

$$\mathcal{W}_{2,\mathbb{M}_{2\mathfrak{d}-1}} = \mathcal{W}_{1,\mathbb{M}_\mathfrak{d}} \underbrace{\begin{pmatrix} A_2 & 0 & \cdots & 0 & 0 \\ 0 & A_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_2 & 0 \\ 0 & 0 & \cdots & 0 & C_2 \end{pmatrix}}_{\in \mathbb{R}^{\mathfrak{d} \times (3\mathfrak{d}-1)}}, \qquad \mathcal{W}_{2,\mathbb{M}_{2\mathfrak{d}}} = \mathcal{W}_{1,\mathbb{M}_\mathfrak{d}} \underbrace{\begin{pmatrix} A_2 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_2 \end{pmatrix}}_{\in \mathbb{R}^{\mathfrak{d} \times (3\mathfrak{d})}},$$
$$\mathcal{B}_{2,\mathbb{M}_{2\mathfrak{d}-1}} = \mathcal{B}_{1,\mathbb{M}_\mathfrak{d}} = 0, \qquad \text{and} \qquad \mathcal{B}_{2,\mathbb{M}_{2\mathfrak{d}}} = \mathcal{B}_{1,\mathbb{M}_\mathfrak{d}} = 0. \tag{3.97}$$

Combining this and item (iv) in Definition 3.2.7 shows that for all $\mathfrak{d} \in \{2,3,4,\ldots\}$ it holds that $\mathcal{B}_{2,\mathbb{M}_\mathfrak{d}} = 0$. Moreover, note that (2.58) demonstrates that for all $\mathfrak{d} \in \{2,3,4,\ldots,\}$, $i \in \{3,4,\ldots,\mathcal{L}(\mathbb{M}_\mathfrak{d})+1\}$ it holds that

$$\mathcal{W}_{i,\mathbb{M}_{2\mathfrak{d}-1}} = \mathcal{W}_{i,\mathbb{M}_{2\mathfrak{d}}} = \mathcal{W}_{i-1,\mathbb{M}_\mathfrak{d}} \qquad \text{and} \qquad \mathcal{B}_{i,\mathbb{M}_{2\mathfrak{d}-1}} = \mathcal{B}_{i,\mathbb{M}_{2\mathfrak{d}}} = \mathcal{B}_{i-1,\mathbb{M}_\mathfrak{d}}. \tag{3.98}$$

This, (3.95), (3.96), (3.97), the fact that for all $\mathfrak{d} \in \{2,3,4,\ldots\}$ it holds that $\mathcal{B}_{2,\mathbb{M}_\mathfrak{d}} = 0$, and induction establish items (i) and (ii). The proof of Lemma 3.2.10 is thus complete. □

3.2.2.3 ANN representations for maximum convolutions

Lemma 3.2.11. *Let $d, K \in \mathbb{N}$, $L \in [0,\infty)$, $\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_K \in \mathbb{R}^d$, $\mathfrak{y} = (\mathfrak{y}_1, \mathfrak{y}_2, \ldots, \mathfrak{y}_K) \in \mathbb{R}^K$, $\Phi \in \mathbf{N}$ satisfy*

$$\Phi = \mathbb{M}_K \bullet \mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}} \bullet \mathbf{P}_K\big(\mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_1}, \mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_2}, \ldots, \mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_K}\big) \bullet \mathbb{T}_{d,K} \tag{3.99}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.9, 2.2.11, 2.2.18, 2.2.28, 3.2.3, and 3.2.7). Then

(i) it holds that $\mathcal{I}(\Phi) = d$,

(ii) it holds that $\mathcal{O}(\Phi) = 1$,

(iii) it holds that $\mathcal{H}(\Phi) = \lceil \log_2(K) \rceil + 1$,

(iv) it holds that $\mathbb{D}_1(\Phi) = 2dK$,

(v) it holds for all $i \in \{2,3,4,\ldots\}$ that $\mathbb{D}_i(\Phi) \le 3\lceil \frac{K}{2^{i-1}} \rceil$,

(vi) it holds that $\|\mathcal{T}(\Phi)\|_\infty \le \max\{1, L, \max_{k\in\{1,2,\ldots,K\}} \|\mathfrak{x}_k\|_\infty, 2\|\mathfrak{y}\|_\infty\}$, and

(vii) it holds for all $x \in \mathbb{R}^d$ that $(\mathcal{R}_\mathfrak{r}(\Phi))(x) = \max_{k\in\{1,2,\ldots,K\}}(\mathfrak{y}_k - L\|x - \mathfrak{x}_k\|_1)$

(cf. Definitions 2.1.6, 2.2.3, 2.2.34, 3.1.16, and 3.2.8).

Proof of Lemma 3.2.11. Throughout this proof let $\Psi_k \in \mathbf{N}$, $k \in \{1, 2, \ldots, K\}$, satisfy for all $k \in \{1, 2, \ldots, K\}$ that $\Psi_k = \mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d, -\mathfrak{x}_k}$, let $\Xi \in \mathbf{N}$ satisfy

$$\Xi = \mathbf{A}_{-L\,\mathrm{I}_K, \mathfrak{y}} \bullet \mathbf{P}_K\big(\Psi_1, \Psi_2, \ldots, \Psi_K\big) \bullet \mathbb{T}_{d,K}, \tag{3.100}$$

and let $|||\cdot|||\colon \bigcup_{m,n\in\mathbb{N}} \mathbb{R}^{m\times n} \to [0, \infty)$ satisfy for all $m, n \in \mathbb{N}$, $M = (M_{i,j})_{i\in\{1,\ldots,m\}, j\in\{1,\ldots,n\}} \in \mathbb{R}^{m\times n}$ that $|||M||| = \max_{i\in\{1,\ldots,m\}, j\in\{1,\ldots,n\}}|M_{i,j}|$. Observe that (3.99) and Proposition 2.2.7 ensure that $\mathcal{O}(\Phi) = \mathcal{O}(\mathbb{M}_K) = 1$ and $\mathcal{I}(\Phi) = \mathcal{I}(\mathbb{T}_{d,K}) = d$. This proves items (i) and (ii). Moreover, observe that the fact that for all $m, n \in \mathbb{N}$, $\mathfrak{W} \in \mathbb{R}^{m\times n}$, $\mathfrak{B} \in \mathbb{R}^m$ it holds that $\mathcal{H}(\mathbf{A}_{\mathfrak{W},\mathfrak{B}}) = 0 = \mathcal{H}(\mathbb{T}_{d,K})$, the fact that $\mathcal{H}(\mathbb{L}_d) = 1$, and Proposition 2.2.7 assure that

$$\mathcal{H}(\Xi) = \mathcal{H}(\mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}}) + \mathcal{H}(\mathbf{P}_K(\Psi_1, \Psi_2, \ldots, \Psi_K)) + \mathcal{H}(\mathbb{T}_{d,K}) = \mathcal{H}(\Psi_1) = \mathcal{H}(\mathbb{L}_d) = 1. \tag{3.101}$$

Proposition 2.2.7 and Proposition 3.2.9 hence ensure that

$$\mathcal{H}(\Phi) = \mathcal{H}(\mathbb{M}_K \bullet \Xi) = \mathcal{H}(\mathbb{M}_K) + \mathcal{H}(\Xi) = \lceil \log_2(K)\rceil + 1 \tag{3.102}$$

(cf. Definition 3.2.8). This establishes item (iii). Next observe that the fact that $\mathcal{H}(\Xi) = 1$, Proposition 2.2.7, and Proposition 3.2.9 assure that for all $i \in \{2, 3, 4, \ldots\}$ it holds that

$$\mathbb{D}_i(\Phi) = \mathbb{D}_{i-1}(\mathbb{M}_K) \leq 3\big\lceil \tfrac{K}{2^{i-1}}\big\rceil. \tag{3.103}$$

This proves item (v). Furthermore, note that Proposition 2.2.7, Proposition 2.2.14, and Proposition 3.2.4 assure that

$$\mathbb{D}_1(\Phi) = \mathbb{D}_1(\Xi) = \mathbb{D}_1(\mathbf{P}_K(\Psi_1, \Psi_2, \ldots, \Psi_K)) = \sum_{i=1}^{K} \mathbb{D}_1(\Psi_i) = \sum_{i=1}^{K} \mathbb{D}_1(\mathbb{L}_d) = 2dK. \tag{3.104}$$

This establishes item (iv). Next observe that (2.58) and Lemma 3.2.10 imply that

$$\begin{split}\Phi = \big(&(\mathcal{W}_{1,\Xi}, \mathcal{B}_{1,\Xi}), (\mathcal{W}_{1,\mathbb{M}_K}\mathcal{W}_{2,\Xi}, \mathcal{W}_{1,\mathbb{M}_K}\mathcal{B}_{2,\Xi}),\\ &(\mathcal{W}_{2,\mathbb{M}_K}, 0), \ldots, (\mathcal{W}_{\mathcal{L}(\mathbb{M}_K),\mathbb{M}_K}, 0)\big).\end{split} \tag{3.105}$$

Moreover, note that the fact that for all $k \in \{1, 2, \ldots, K\}$ it holds that $\mathcal{W}_{1,\Psi_k} = \mathcal{W}_{1,\mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_k}}\mathcal{W}_{1,\mathbb{L}_d} = \mathcal{W}_{1,\mathbb{L}_d}$ assures that

$$\begin{split}\mathcal{W}_{1,\Xi} = \mathcal{W}_{1,\mathbf{P}_K(\Psi_1,\Psi_2,\ldots,\Psi_K)}\mathcal{W}_{1,\mathbb{T}_{d,K}} &= \begin{pmatrix} \mathcal{W}_{1,\Psi_1} & 0 & \cdots & 0 \\ 0 & \mathcal{W}_{1,\Psi_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{W}_{1,\Psi_K} \end{pmatrix} \begin{pmatrix} \mathrm{I}_d \\ \mathrm{I}_d \\ \vdots \\ \mathrm{I}_d \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{W}_{1,\Psi_1} \\ \mathcal{W}_{1,\Psi_2} \\ \vdots \\ \mathcal{W}_{1,\Psi_K} \end{pmatrix} = \begin{pmatrix} \mathcal{W}_{1,\mathbb{L}_d} \\ \mathcal{W}_{1,\mathbb{L}_d} \\ \vdots \\ \mathcal{W}_{1,\mathbb{L}_d} \end{pmatrix}.\end{split} \tag{3.106}$$

Lemma 3.2.5 hence demonstrates that $|||\mathcal{W}_{1,\Xi}||| = 1$. In addition, note that (2.58) implies that

$$\mathcal{B}_{1,\Xi} = \mathcal{W}_{1,\mathbf{P}_K(\Psi_1,\Psi_2,\dots,\Psi_K)}\mathcal{B}_{1,\mathbb{T}_{d,K}} + \mathcal{B}_{1,\mathbf{P}_K(\Psi_1,\Psi_2,\dots,\Psi_K)} = \mathcal{B}_{1,\mathbf{P}_K(\Psi_1,\Psi_2,\dots,\Psi_K)} = \begin{pmatrix} \mathcal{B}_{1,\Psi_1} \\ \mathcal{B}_{1,\Psi_2} \\ \vdots \\ \mathcal{B}_{1,\Psi_K} \end{pmatrix}. \tag{3.107}$$

Furthermore, observe that Lemma 3.2.5 implies that for all $k \in \{1, 2, \dots, K\}$ it holds that

$$\mathcal{B}_{1,\Psi_k} = \mathcal{W}_{1,\mathbb{L}_d}\mathcal{B}_{1,\mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_k}} + \mathcal{B}_{1,\mathbb{L}_d} = -\mathcal{W}_{1,\mathbb{L}_d}\mathfrak{x}_k. \tag{3.108}$$

This, (3.107), and Lemma 3.2.5 show that

$$\|\mathcal{B}_{1,\Xi}\|_\infty = \max_{k\in\{1,2,\dots,K\}}\|\mathcal{B}_{1,\Psi_k}\|_\infty = \max_{k\in\{1,2,\dots,K\}}\|\mathcal{W}_{1,\mathbb{L}_d}\mathfrak{x}_k\|_\infty = \max_{k\in\{1,2,\dots,K\}}\|\mathfrak{x}_k\|_\infty \tag{3.109}$$

(cf. Definition 3.1.16). Combining this, (3.105), Lemma 3.2.10, and the fact that $|||\mathcal{W}_{1,\Xi}||| = 1$ shows that

$$\begin{aligned} \|\mathcal{T}(\Phi)\|_\infty &= \max\{|||\mathcal{W}_{1,\Xi}|||, \|\mathcal{B}_{1,\Xi}\|_\infty, |||\mathcal{W}_{1,\mathbb{M}_K}\mathcal{W}_{2,\Xi}|||, \|\mathcal{W}_{1,\mathbb{M}_K}\mathcal{B}_{2,\Xi}\|_\infty, 1\} \\ &= \max\bigl\{1, \max\nolimits_{k\in\{1,2,\dots,K\}}\|\mathfrak{x}_k\|_\infty, |||\mathcal{W}_{1,\mathbb{M}_K}\mathcal{W}_{2,\Xi}|||, \|\mathcal{W}_{1,\mathbb{M}_K}\mathcal{B}_{2,\Xi}\|_\infty\bigr\} \end{aligned} \tag{3.110}$$

(cf. Definition 2.2.34). Next note that Lemma 3.2.5 ensures that for all $k \in \{1, 2, \dots, K\}$ it holds that $\mathcal{B}_{2,\Psi_k} = \mathcal{B}_{2,\mathbb{L}_d} = 0$. Hence, we obtain that $\mathcal{B}_{2,\mathbf{P}_K(\Psi_1,\Psi_2,\dots,\Psi_K)} = 0$. This implies that

$$\mathcal{B}_{2,\Xi} = \mathcal{W}_{1,\mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}}}\mathcal{B}_{2,\mathbf{P}_K(\Psi_1,\Psi_2,\dots,\Psi_K)} + \mathcal{B}_{1,\mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}}} = \mathcal{B}_{1,\mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}}} = \mathfrak{y}. \tag{3.111}$$

In addition, observe that the fact that for all $k \in \{1, 2, \dots, K\}$ it holds that $\mathcal{W}_{2,\Psi_k} = \mathcal{W}_{2,\mathbb{L}_d}$ assures that

$$\begin{aligned} \mathcal{W}_{2,\Xi} &= \mathcal{W}_{1,\mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}}}\mathcal{W}_{2,\mathbf{P}_K(\Psi_1,\Psi_2,\dots,\Psi_K)} = -L\mathcal{W}_{2,\mathbf{P}_K(\Psi_1,\Psi_2,\dots,\Psi_K)} \\ &= -L\begin{pmatrix} \mathcal{W}_{2,\Psi_1} & 0 & \cdots & 0 \\ 0 & \mathcal{W}_{2,\Psi_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{W}_{2,\Psi_K} \end{pmatrix} = \begin{pmatrix} -L\mathcal{W}_{2,\mathbb{L}_d} & 0 & \cdots & 0 \\ 0 & -L\mathcal{W}_{2,\mathbb{L}_d} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -L\mathcal{W}_{2,\mathbb{L}_d} \end{pmatrix}. \end{aligned} \tag{3.112}$$

Item (v) in Lemma 3.2.5 and Lemma 3.2.10 hence imply that

$$|||\mathcal{W}_{1,\mathbb{M}_K}\mathcal{W}_{2,\Xi}||| = L|||\mathcal{W}_{1,\mathbb{M}_K}||| \le L. \tag{3.113}$$

Moreover, observe that (3.111) and Lemma 3.2.10 assure that

$$\|\mathcal{W}_{1,\mathbb{M}_K}\mathcal{B}_{2,\Xi}\|_\infty \le 2\|\mathcal{B}_{2,\Xi}\|_\infty = 2\|\mathfrak{y}\|_\infty. \tag{3.114}$$

Combining this with (3.110) and (3.113) establishes item (vi). Next observe that Proposition 3.2.4 and Lemma 2.2.20 show that for all $x \in \mathbb{R}^d$, $k \in \{1, 2, \ldots, K\}$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\Psi_k))(x) = \big(\mathcal{R}_{\mathfrak{r}}(\mathbb{L}_d) \circ \mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_k})\big)(x) = \|x - \mathfrak{x}_k\|_1. \tag{3.115}$$

This, Proposition 2.2.13, and Proposition 2.2.7 imply that for all $x \in \mathbb{R}^d$ it holds that

$$\big(\mathcal{R}_{\mathfrak{r}}(\mathbf{P}_K(\Psi_1, \Psi_2, \ldots, \Psi_K) \bullet \mathbb{T}_{d,K})\big)(x) = \big(\|x - \mathfrak{x}_1\|_1, \|x - \mathfrak{x}_2\|_1, \ldots, \|x - \mathfrak{x}_K\|_1\big). \tag{3.116}$$

(cf. Definitions 2.1.6 and 2.2.3). Combining this and Lemma 2.2.20 establishes that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}(\mathcal{R}_{\mathfrak{r}}(\Xi))(x) &= \big(\mathcal{R}_{\mathfrak{r}}(\mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}}) \circ \mathcal{R}_{\mathfrak{r}}(\mathbf{P}_K(\Psi_1, \Psi_2, \ldots, \Psi_K) \bullet \mathbb{T}_{d,K})\big)(x) \\ &= \big(\mathfrak{y}_1 - L\|x - \mathfrak{x}_1\|_1, \mathfrak{y}_2 - L\|x - \mathfrak{x}_2\|_1, \ldots, \mathfrak{y}_K - L\|x - \mathfrak{x}_K\|_1\big).\end{aligned} \tag{3.117}$$

Proposition 2.2.7 and Proposition 3.2.9 hence demonstrate that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}(\mathcal{R}_{\mathfrak{r}}(\Phi))(x) &= \big(\mathcal{R}_{\mathfrak{r}}(\mathbb{M}_K) \circ \mathcal{R}_{\mathfrak{r}}(\Xi)\big)(x) \\ &= (\mathcal{R}_{\mathfrak{r}}(\mathbb{M}_K))\big(\mathfrak{y}_1 - L\|x - \mathfrak{x}_1\|_1, \mathfrak{y}_2 - L\|x - \mathfrak{x}_2\|_1, \ldots, \mathfrak{y}_K - L\|x - \mathfrak{x}_K\|_1\big) \\ &= \max\nolimits_{k \in \{1,2,\ldots,K\}}(\mathfrak{y}_k - L\|x - \mathfrak{x}_k\|_1).\end{aligned} \tag{3.118}$$

This establishes item (vii). The proof of Lemma 3.2.11 is thus complete. □

3.2.3 Explicit approximations through ANNs

Proposition 3.2.12. *Let $d, K \in \mathbb{N}$, $L \in [0, \infty)$, let $E \subseteq \mathbb{R}^d$ be a set, let $\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_K \in E$, let $f\colon E \to \mathbb{R}$ satisfy for all $x, y \in E$ that $|f(x) - f(y)| \leq L\|x - y\|_1$, and let $\mathfrak{y} \in \mathbb{R}^K$, $\Phi \in \mathbf{N}$ satisfy $\mathfrak{y} = (f(\mathfrak{x}_1), f(\mathfrak{x}_2), \ldots, f(\mathfrak{x}_K))$ and*

$$\Phi = \mathbb{M}_K \bullet \mathbf{A}_{-L\,\mathrm{I}_K,\mathfrak{y}} \bullet \mathbf{P}_K\big(\mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_1}, \mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_2}, \ldots, \mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d,-\mathfrak{x}_K}\big) \bullet \mathbb{T}_{d,K} \tag{3.119}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.9, 2.2.11, 2.2.18, 2.2.28, 3.1.16, 3.2.3, and 3.2.7). Then

$$\sup\nolimits_{x \in E}|(\mathcal{R}_{\mathfrak{r}}(\Phi))(x) - f(x)| \leq 2L\big[\sup\nolimits_{x \in E}\big(\min\nolimits_{k \in \{1,2,\ldots,K\}}\|x - \mathfrak{x}_k\|_1\big)\big] \tag{3.120}$$

(cf. Definitions 2.1.6 and 2.2.3).

Proof of Proposition 3.2.12. Throughout this proof let $F\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that

$$F(x) = \max\nolimits_{k \in \{1,2,\ldots,K\}}(f(\mathfrak{x}_k) - L\|x - \mathfrak{x}_k\|_1). \tag{3.121}$$

Observe that Corollary 3.2.2, (3.121), and the assumption that for all $x, y \in E$ it holds that $|f(x) - f(y)| \leq L\|x - y\|_1$ assure that

$$\sup\nolimits_{x \in E}|F(x) - f(x)| \leq 2L\big[\sup\nolimits_{x \in E}\big(\min\nolimits_{k \in \{1,2,\ldots,K\}}\|x - \mathfrak{x}_k\|_1\big)\big]. \tag{3.122}$$

Moreover, note that Lemma 3.2.11 ensures that for all $x \in E$ it holds that $F(x) = (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)$. Combining this and (3.122) establishes (3.120). The proof of Proposition 3.2.12 is thus complete. □

Exercise 3.2.3. *Prove or disprove the following statement: There exists $\Phi \in \mathbf{N}$ such that $\mathcal{I}(\Phi) = 2$, $\mathcal{O}(\Phi) = 1$, $\mathcal{P}(\Phi) < 20$, and*

$$\sup_{v=(x,y)\in[0,2]^2} \left| x^2 + y^2 - 2x - 2y + 2 - (\mathcal{R}_{\mathfrak{r}}(\Phi))(v) \right| \leq \tfrac{3}{8}. \tag{3.123}$$

3.2.4 Analysis of the approximation error

3.2.4.1 Covering number estimates

Definition 3.2.13 (Covering number)**.** *Let (E, δ) be a metric space and let $r \in [0, \infty]$. Then we denote by $\mathcal{C}^{(E,\delta),r} \in \mathbb{N}_0 \cup \{\infty\}$ (we denote by $\mathcal{C}^{E,r} \in \mathbb{N}_0 \cup \{\infty\}$) the extended real number given by*

$$\mathcal{C}^{(E,\delta),r} = \min\!\left(\left\{ n \in \mathbb{N}_0 \colon \left[\exists\, A \subseteq E \colon \begin{pmatrix} (|A| \leq n) \wedge (\forall\, x \in E \colon \\ \exists\, a \in A \colon \delta(a, x) \leq r) \end{pmatrix} \right] \right\} \cup \{\infty\} \right). \tag{3.124}$$

Exercise 3.2.4. *Prove or disprove the following statement: For every metric space (X, d), every $Y \subseteq X$, and every $r \in [0, \infty]$ it holds that $\mathcal{C}^{(Y, d|_{Y\times Y}),r} \leq \mathcal{C}^{(X,d),r}$.*

Exercise 3.2.5. *Prove or disprove the following statement: For every metric space (E, δ) it holds that $\mathcal{C}^{(E,\delta),\infty} = 1$.*

Exercise 3.2.6. *Prove or disprove the following statement: For every metric space (E, δ) and every $r \in [0, \infty)$ with $\mathcal{C}^{(E,\delta),r} < \infty$ it holds that E is bounded. (*Note: *A metric space (E, δ) is bounded if and only if there exists $r \in [0, \infty)$ such that it holds for all $x, y \in E$ that $\delta(x, y) \leq r$.)*

Exercise 3.2.7. *Prove or disprove the following statement: For every bounded metric space (E, δ) and every $r \in [0, \infty]$ it holds that $\mathcal{C}^{(E,\delta),r} < \infty$.*

Lemma 3.2.14. *Let $d \in \mathbb{N}$, $a \in \mathbb{R}$, $b \in (a, \infty)$, $r \in (0, \infty)$ and for every $p \in [1, \infty]$ let $\delta_p \colon ([a, b]^d) \times ([a, b]^d) \to [0, \infty)$ satisfy for all $x, y \in [a, b]^d$ that $\delta_p(x, y) = \|x - y\|_p$ (cf. Definition 3.1.16). Then*

(i) it holds for all $p \in [1, \infty)$ that

$$\mathcal{C}^{([a,b]^d, \delta_p),r} \leq \left(\left\lceil \tfrac{d^{1/p}(b-a)}{2r} \right\rceil \right)^{d} \leq \begin{cases} 1 & \colon r \geq d(b-a)/2 \\ \left(\frac{d(b-a)}{r} \right)^{d} & \colon r < d(b-a)/2 \end{cases} \tag{3.125}$$

and

(ii) it holds that

$$\mathcal{C}^{([a,b]^d, \delta_\infty),r} \leq \left(\left\lceil \tfrac{b-a}{2r} \right\rceil \right)^{d} \leq \begin{cases} 1 & \colon r \geq (b-a)/2 \\ \left(\frac{b-a}{r} \right)^{d} & \colon r < (b-a)/2 \end{cases} \tag{3.126}$$

(cf. Definitions 3.2.8 and 3.2.13).

Proof of Lemma 3.2.14. Throughout this proof let $(\mathfrak{N}_p)_{p\in[1,\infty]} \subseteq \mathbb{N}$ satisfy for all $p \in [1,\infty)$ that

$$\mathfrak{N}_p = \left\lceil \tfrac{d^{1/p}(b-a)}{2r} \right\rceil \qquad \text{and} \qquad \mathfrak{N}_\infty = \left\lceil \tfrac{b-a}{2r} \right\rceil, \tag{3.127}$$

for every $N \in \mathbb{N}$, $i \in \{1,2,\dots,N\}$ let $g_{N,i} \in [a,b]$ be given by

$$g_{N,i} = a + {}^{(i-1/2)(b-a)}\!/_{N} \tag{3.128}$$

and for every $p \in [1,\infty]$ let $A_p \subseteq [a,b]^d$ be given by

$$A_p = \{g_{\mathfrak{N}_p,1}, g_{\mathfrak{N}_p,2}, \dots, g_{\mathfrak{N}_p,\mathfrak{N}_p}\}^d \tag{3.129}$$

(cf. Definition 3.2.8). Observe that it holds for all $N \in \mathbb{N}$, $i \in \{1,2,\dots,N\}$, $x \in [a + {}^{(i-1)(b-a)}\!/_{N}, g_{N,i}]$ that

$$|x - g_{N,i}| = a + \tfrac{(i-1/2)(b-a)}{N} - x \le a + \tfrac{(i-1/2)(b-a)}{N} - \left(a + \tfrac{(i-1)(b-a)}{N}\right) = \tfrac{b-a}{2N}. \tag{3.130}$$

In addition, note that it holds for all $N \in \mathbb{N}$, $i \in \{1,2,\dots,N\}$, $x \in [g_{N,i}, a + {}^{i(b-a)}\!/_{N}]$ that

$$|x - g_{N,i}| = x - \left(a + \tfrac{(i-1/2)(b-a)}{N}\right) \le a + \tfrac{i(b-a)}{N} - \left(a + \tfrac{(i-1/2)(b-a)}{N}\right) = \tfrac{b-a}{2N}. \tag{3.131}$$

Combining (3.130) and (3.131) implies for all $N \in \mathbb{N}$, $i \in \{1,2,\dots,N\}$, $x \in [a + {}^{(i-1)(b-a)}\!/_{N}, a + {}^{i(b-a)}\!/_{N}]$ that $|x - g_{N,i}| \le {}^{(b-a)}\!/_{(2N)}$. This proves that for every $N \in \mathbb{N}$, $x \in [a,b]$ there exists $y \in \{g_{N,1}, g_{N,2}, \dots, g_{N,N}\}$ such that

$$|x - y| \le \tfrac{b-a}{2N}. \tag{3.132}$$

This establishes that for every $p \in [1,\infty)$, $x = (x_1, x_2, \dots, x_d) \in [a,b]^d$ there exists $y = (y_1, y_2, \dots, y_d) \in A_p$ such that

$$\delta_p(x,y) = \|x-y\|_p = \left(\sum_{i=1}^d |x_i - y_i|^p\right)^{1/p} \le \left(\sum_{i=1}^d \tfrac{(b-a)^p}{(2\mathfrak{N}_p)^p}\right)^{1/p} = \tfrac{d^{1/p}(b-a)}{2\mathfrak{N}_p} \le \tfrac{d^{1/p}(b-a)2r}{2d^{1/p}(b-a)} = r. \tag{3.133}$$

Furthermore, (3.132) shows that for every $x = (x_1, x_2, \dots, x_d) \in [a,b]^d$ there exists $y = (y_1, y_2, \dots, y_d) \in A_\infty$ such that

$$\delta_\infty(x,y) = \|x-y\|_\infty = \max_{i\in\{1,2,\dots,d\}} |x_i - y_i| \le \tfrac{b-a}{2\mathfrak{N}_\infty} \le \tfrac{(b-a)2r}{2(b-a)} = r. \tag{3.134}$$

Note that (3.124), (3.129), (3.133), (3.127), and the fact that $\forall\, x \in [0,\infty)\colon \lceil x \rceil \le \mathbb{1}_{(0,1]}(x) + 2x\mathbb{1}_{(1,\infty)}(x) = \mathbb{1}_{(0,r]}(rx) + 2x\mathbb{1}_{(r,\infty)}(rx)$ yield that for all $p \in [1,\infty)$ it holds that

$$\begin{aligned} \mathcal{C}^{([a,b]^d,\delta_p),r} &\le |A_p| = (\mathfrak{N}_p)^d = \left(\left\lceil \tfrac{d^{1/p}(b-a)}{2r} \right\rceil\right)^d \le \left(\left\lceil \tfrac{d(b-a)}{2r} \right\rceil\right)^d \\ &\le \left(\mathbb{1}_{(0,r]}\left(\tfrac{d(b-a)}{2}\right) + \tfrac{2d(b-a)}{2r}\mathbb{1}_{(r,\infty)}\left(\tfrac{d(b-a)}{2}\right)\right)^d \\ &= \mathbb{1}_{(0,r]}\left(\tfrac{d(b-a)}{2}\right) + \left(\tfrac{d(b-a)}{r}\right)^d \mathbb{1}_{(r,\infty)}\left(\tfrac{d(b-a)}{2}\right) \end{aligned} \tag{3.135}$$

(cf. Definition 3.2.13). This proves item (i). In addition, note that (3.124), (3.129), (3.134), (3.127), and the fact that $\forall\, x \in [0,\infty)\colon \lceil x \rceil \leq \mathbb{1}_{(0,r]}(rx) + 2x\mathbb{1}_{(r,\infty)}(rx)$ demonstrate that

$$\mathcal{C}^{([a,b]^d,\delta_\infty),r} \leq |A_\infty| = (\mathfrak{N}_\infty)^d = \left(\left\lceil \tfrac{b-a}{2r} \right\rceil\right)^d \leq \mathbb{1}_{(0,r]}\left(\tfrac{b-a}{2}\right) + \left(\tfrac{b-a}{r}\right)^d \mathbb{1}_{(r,\infty)}\left(\tfrac{b-a}{2}\right). \tag{3.136}$$

This implies item (ii). and thus completes the proof of Lemma 3.2.14. □

3.2.4.2 Convergence rates for the approximation error

Lemma 3.2.15. *Let $d \in \mathbb{N}$, $L, a \in \mathbb{R}$, $b \in (a,\infty)$, let $f\colon [a,b]^d \to \mathbb{R}$ satisfy for all $x, y \in [a,b]^d$ that $|f(x) - f(y)| \leq L\|x-y\|_1$, and let $\mathbf{F} = \mathbf{A}_{0,f((a+b)/2,(a+b)/2,\dots,(a+b)/2)} \in \mathbb{R}^{1\times d} \times \mathbb{R}^1$ (cf. Definitions 2.2.18 and 3.1.16). Then*

(i) it holds that $\mathcal{I}(\mathbf{F}) = d$,

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

(iii) it holds that $\mathcal{H}(\mathbf{F}) = 0$,

(iv) it holds that $\mathcal{P}(\mathbf{F}) = d + 1$,

(v) it holds that $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \sup_{x\in[a,b]^d}|f(x)|$, and

(vi) it holds that $\sup_{x\in[a,b]^d}|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq \frac{dL(b-a)}{2}$

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, and 2.2.34).

Proof of Lemma 3.2.15. Note that the assumption that for all $x, y \in [a,b]^d$ it holds that $|f(x) - f(y)| \leq L\|x-y\|_1$ assures that $L \geq 0$. Next observe that Lemma 2.2.19 assures that for all $x \in \mathbb{R}^d$ it holds that

$$(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) = f\big({(a+b)}/{2}, {(a+b)}/{2}, \dots, {(a+b)}/{2}\big). \tag{3.137}$$

The fact that for all $x \in [a,b]$ it holds that $|x - {(a+b)}/{2}| \leq {(b-a)}/{2}$ and the assumption that for all $x, y \in [a,b]^d$ it holds that $|f(x) - f(y)| \leq L\|x-y\|_1$ hence ensure that for all $x = (x_1, x_2, \dots, x_d) \in [a,b]^d$ it holds that

$$\begin{aligned}
|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| &= \left|f\big({(a+b)}/{2}, {(a+b)}/{2}, \dots, {(a+b)}/{2}\big) - f(x)\right| \\
&\leq L\left\|\big({(a+b)}/{2}, {(a+b)}/{2}, \dots, {(a+b)}/{2}\big) - x\right\|_1 \\
&= L\sum_{i=1}^d \left|{(a+b)}/{2} - x_i\right| \leq \sum_{i=1}^d \frac{L(b-a)}{2} = \frac{dL(b-a)}{2}.
\end{aligned} \tag{3.138}$$

This and the fact that $\|\mathcal{T}(\mathbf{F})\|_\infty = |f({(a+b)}/{2}, {(a+b)}/{2}, \dots, {(a+b)}/{2})| \leq \sup_{x\in[a,b]^d}|f(x)|$ complete the proof of Lemma 3.2.15. □

Proposition 3.2.16. *Let $d \in \mathbb{N}$, $L, a \in \mathbb{R}$, $b \in (a, \infty)$, $r \in (0, d/4)$, let $f \colon [a,b]^d \to \mathbb{R}$ and $\delta \colon [a,b]^d \times [a,b]^d \to \mathbb{R}$ satisfy for all $x, y \in [a,b]^d$ that $|f(x) - f(y)| \leq L\|x - y\|_1$ and $\delta(x,y) = \|x - y\|_1$, and let $K \in \mathbb{N}$, $\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_K \in [a,b]^d$, $\mathfrak{y} \in \mathbb{R}^K$, $\mathbf{F} \in \mathbf{N}$ satisfy $K = \mathcal{C}^{([a,b]^d, \delta), (b-a)r}$, $\sup_{x \in [a,b]^d} \bigl[\min_{k \in \{1,2,\ldots,K\}} \delta(x, \mathfrak{x}_k)\bigr] \leq (b-a)r$, $\mathfrak{y} = (f(\mathfrak{x}_1), f(\mathfrak{x}_2), \ldots, f(\mathfrak{x}_K))$, and*

$$\mathbf{F} = \mathbb{M}_K \bullet \mathbf{A}_{-L\,\mathrm{I}_K, \mathfrak{y}} \bullet \mathbf{P}_K\bigl(\mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d, -\mathfrak{x}_1}, \mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d, -\mathfrak{x}_2}, \ldots, \mathbb{L}_d \bullet \mathbf{A}_{\mathrm{I}_d, -\mathfrak{x}_K}\bigr) \bullet \mathbb{T}_{d,K} \tag{3.139}$$

(cf. Definitions 2.2.1, 2.2.5, 2.2.9, 2.2.11, 2.2.18, 2.2.28, 3.1.16, 3.2.3, 3.2.7, and 3.2.13). Then

(i) it holds that $\mathcal{I}(\mathbf{F}) = d$,

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

(iii) it holds that $\mathcal{H}(\mathbf{F}) \leq \bigl\lceil d \log_2\bigl(\frac{3d}{4r}\bigr)\bigr\rceil + 1$,

(iv) it holds that $\mathbb{D}_1(\mathbf{F}) \leq 2d\bigl(\frac{3d}{4r}\bigr)^d$,

(v) it holds for all $i \in \{2, 3, 4, \ldots\}$ that $\mathbb{D}_i(\mathbf{F}) \leq 3\bigl\lceil\bigl(\frac{3d}{4r}\bigr)^d \frac{1}{2^{i-1}}\bigr\rceil$,

(vi) it holds that $\mathcal{P}(\mathbf{F}) \leq 35\bigl(\frac{3d}{4r}\bigr)^{2d} d^2$,

(vii) it holds that $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\{1, L, |a|, |b|, 2[\sup_{x \in [a,b]^d} |f(x)|]\}$, and

(viii) it holds that $\sup_{x \in [a,b]^d} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq 2L(b-a)r$

(cf. Definitions 2.1.6, 2.2.3, 2.2.34, and 3.2.8).

Proof of Proposition 3.2.16. Note that the assumption that for all $x, y \in [a,b]^d$ it holds that $|f(x) - f(y)| \leq L\|x - y\|_1$ assures that $L \geq 0$. Next observe that (3.139), Lemma 3.2.11, and Proposition 3.2.12 demonstrate that

(I) it holds that $\mathcal{I}(\mathbf{F}) = d$,

(II) it holds that $\mathcal{O}(\mathbf{F}) = 1$,

(III) it holds that $\mathcal{H}(\mathbf{F}) = \lceil \log_2(K) \rceil + 1$,

(IV) it holds that $\mathbb{D}_1(\mathbf{F}) = 2dK$,

(V) it holds for all $i \in \{2, 3, 4, \ldots\}$ that $\mathbb{D}_i(\mathbf{F}) \leq 3\bigl\lceil \frac{K}{2^{i-1}} \bigr\rceil$,

(VI) it holds that $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\{1, L, \max_{k \in \{1,2,\ldots,K\}} \|\mathfrak{x}_k\|_\infty, 2[\max_{k \in \{1,2,\ldots,K\}} |f(\mathfrak{x}_k)|]\}$, and

(VII) it holds that $\sup_{x \in [a,b]^d} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq 2L\bigl[\sup_{x \in [a,b]^d}\bigl(\min_{k \in \{1,2,\ldots,K\}} \delta(x, \mathfrak{x}_k)\bigr)\bigr]$

(cf. Definitions 2.1.6, 2.2.3, 2.2.34, and 3.2.8). Note that items (I) and (II) establish items (i) and (ii). Next observe that item (i) in Lemma 3.2.14 and the fact that $\frac{d}{2r} \geq 2$ imply that

$$K = \mathcal{C}^{([a,b]^d,\delta),(b-a)r} \leq \left(\left\lceil \tfrac{d(b-a)}{2(b-a)r} \right\rceil\right)^d = \left(\left\lceil \tfrac{d}{2r} \right\rceil\right)^d \leq \left(\tfrac{3}{2}\left(\tfrac{d}{2r}\right)\right)^d = \left(\tfrac{3d}{4r}\right)^d. \tag{3.140}$$

Combining this with item (III) assures that

$$\mathcal{H}(\mathbf{F}) = \lceil \log_2(K) \rceil + 1 \leq \left\lceil \log_2\!\left(\left(\tfrac{3d}{4r}\right)^d\right) \right\rceil + 1 = \lceil d \log_2\left(\tfrac{3d}{4r}\right) \rceil + 1. \tag{3.141}$$

This establishes item (iii). Moreover, note that (3.140) and item (IV) imply that

$$\mathbb{D}_1(\mathbf{F}) = 2dK \leq 2d\left(\tfrac{3d}{4r}\right)^d. \tag{3.142}$$

This establishes item (iv). In addition, observe that item (V) and (3.140) establish item (v). Next note that item (III) ensures that for all $i \in \mathbb{N} \cap (1, \mathcal{H}(\mathbf{F})]$ it holds that

$$\tfrac{K}{2^{i-1}} \geq \tfrac{K}{2^{\mathcal{H}(\mathbf{F})-1}} = \tfrac{K}{2^{\lceil \log_2(K) \rceil}} \geq \tfrac{K}{2^{\log_2(K)+1}} = \tfrac{K}{2K} = \tfrac{1}{2}. \tag{3.143}$$

Item (V) and (3.140) hence show that for all $i \in \mathbb{N} \cap (1, \mathcal{H}(\mathbf{F})]$ it holds that

$$\mathbb{D}_i(\mathbf{F}) \leq 3\left\lceil \tfrac{K}{2^{i-1}} \right\rceil \leq \tfrac{3K}{2^{i-2}} \leq \left(\tfrac{3d}{4r}\right)^d \tfrac{3}{2^{i-2}}. \tag{3.144}$$

Furthermore, note that the fact that for all $x \in [a,b]^d$ it holds that $\|x\|_\infty \leq \max\{|a|,|b|\}$ and item (VI) imply that

$$\begin{aligned} \|\mathcal{T}(\mathbf{F})\|_\infty &\leq \max\{1, L, \max\nolimits_{k\in\{1,2,\dots,K\}} \|\mathfrak{x}_k\|_\infty, 2[\max\nolimits_{k\in\{1,2,\dots,K\}} |f(\mathfrak{x}_k)|]\} \\ &\leq \max\{1, L, |a|, |b|, 2[\sup\nolimits_{x\in[a,b]^d} |f(x)|]\}. \end{aligned} \tag{3.145}$$

This establishes item (vii). Moreover, observe that the assumption that

$$\sup\nolimits_{x\in[a,b]^d} \left[\min\nolimits_{k\in\{1,2,\dots,K\}} \delta(x,\mathfrak{x}_k)\right] \leq (b-a)r \tag{3.146}$$

and item (VII) demonstrate that

$$\sup\nolimits_{x\in[a,b]^d} |(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq 2L\left[\sup\nolimits_{x\in[a,b]^d}\left(\min\nolimits_{k\in\{1,2,\dots,K\}} \delta(x,\mathfrak{x}_k)\right)\right] \leq 2L(b-a)r. \tag{3.147}$$

This establishes item (viii). It thus remains to prove item (vi). For this note that items (I) and (II), (3.142), and (3.144) assure that

$$\begin{aligned} \mathcal{P}(\mathbf{F}) &= \sum_{i=1}^{\mathcal{L}(\mathbf{F})} \mathbb{D}_i(\mathbf{F})(\mathbb{D}_{i-1}(\mathbf{F}) + 1) \\ &\leq 2d\left(\tfrac{3d}{4r}\right)^d (d+1) + \left(\tfrac{3d}{4r}\right)^d 3\left(2d\left(\tfrac{3d}{4r}\right)^d + 1\right) \\ &\quad + \left[\sum_{i=3}^{\mathcal{L}(\mathbf{F})-1} \left(\tfrac{3d}{4r}\right)^d \tfrac{3}{2^{i-2}}\left(\left(\tfrac{3d}{4r}\right)^d \tfrac{3}{2^{i-3}} + 1\right)\right] + \left(\tfrac{3d}{4r}\right)^d \tfrac{3}{2^{\mathcal{L}(\mathbf{F})-3}} + 1. \end{aligned} \tag{3.148}$$

Next note that the fact that $\frac{3d}{4r} \geq 3$ ensures that

$$\begin{aligned} &2d\big(\tfrac{3d}{4r}\big)^d(d+1) + \big(\tfrac{3d}{4r}\big)^d 3\big(2d\big(\tfrac{3d}{4r}\big)^d + 1\big) + \big(\tfrac{3d}{4r}\big)^d \tfrac{3}{2^{\mathcal{L}(\mathbf{F})-3}} + 1 \\ &\leq \big(\tfrac{3d}{4r}\big)^{2d}\big(2d(d+1) + 3(2d+1) + \tfrac{3}{2^{1-3}} + 1\big) \\ &\leq \big(\tfrac{3d}{4r}\big)^{2d} d^2(4+9+12+1) = 26\big(\tfrac{3d}{4r}\big)^{2d} d^2. \end{aligned} \tag{3.149}$$

Moreover, observe that the fact that $\frac{3d}{4r} \geq 3$ implies that

$$\begin{aligned} \sum_{i=3}^{\mathcal{L}(\mathbf{F})-1} \big(\tfrac{3d}{4r}\big)^d \tfrac{3}{2^{i-2}}\big(\big(\tfrac{3d}{4r}\big)^d \tfrac{3}{2^{i-3}} + 1\big) &\leq \big(\tfrac{3d}{4r}\big)^{2d} \sum_{i=3}^{\mathcal{L}(\mathbf{F})-1} \tfrac{3}{2^{i-2}}\big(\tfrac{3}{2^{i-3}} + 1\big) \\ &= \big(\tfrac{3d}{4r}\big)^{2d} \sum_{i=3}^{\mathcal{L}(\mathbf{F})-1} \Big[\tfrac{9}{2^{2i-5}} + \tfrac{3}{2^{i-2}}\Big] \\ &= \big(\tfrac{3d}{4r}\big)^{2d} \sum_{i=0}^{\mathcal{L}(\mathbf{F})-4} \Big[\tfrac{9}{2}(4^{-i}) + \tfrac{3}{2}(2^{-i})\Big] \\ &\leq \big(\tfrac{3d}{4r}\big)^{2d}\big(\tfrac{9}{2}\big(\tfrac{1}{1-4^{-1}}\big) + \tfrac{3}{2}\big(\tfrac{1}{1-2^{-1}}\big)\big) = 9\big(\tfrac{3d}{4r}\big)^{2d}. \end{aligned} \tag{3.150}$$

Combining this, (3.148), and (3.149) demonstrates that

$$\mathcal{P}(\mathbf{F}) \leq 26\big(\tfrac{3d}{4r}\big)^{2d} d^2 + 9\big(\tfrac{3d}{4r}\big)^{2d} \leq 35\big(\tfrac{3d}{4r}\big)^{2d} d^2. \tag{3.151}$$

This establishes item (vi). The proof of Proposition 3.2.16 is thus complete. □

Proposition 3.2.17. *Let $d \in \mathbb{N}$, $L, a \in \mathbb{R}$, $b \in (a, \infty)$, $r \in (0, \infty)$ and let $f\colon [a,b]^d \to \mathbb{R}$ satisfy for all $x, y \in [a,b]^d$ that $|f(x) - f(y)| \leq L\|x - y\|_1$ (cf. Definition 3.1.16). Then there exists $\mathbf{F} \in \mathbf{N}$ such that*

(i) it holds that $\mathcal{I}(\mathbf{F}) = d$,

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

(iii) it holds that $\mathcal{H}(\mathbf{F}) \leq \big(\big\lceil d\log_2\big(\tfrac{3d}{4r}\big)\big\rceil + 1\big)\mathbb{1}_{(0,d/4)}(r)$,

(iv) it holds that $\mathbb{D}_1(\mathbf{F}) \leq 2d\big(\tfrac{3d}{4r}\big)^d \mathbb{1}_{(0,d/4)}(r) + \mathbb{1}_{[d/4,\infty)}(r)$,

(v) it holds for all $i \in \{2, 3, 4, \ldots\}$ that $\mathbb{D}_i(\mathbf{F}) \leq 3\big\lceil\big(\tfrac{3d}{4r}\big)^d \tfrac{1}{2^{i-1}}\big\rceil$,

(vi) it holds that $\mathcal{P}(\mathbf{F}) \leq 35\big(\tfrac{3d}{4r}\big)^{2d} d^2 \mathbb{1}_{(0,d/4)}(r) + (d+1)\mathbb{1}_{[d/4,\infty)}(r)$,

(vii) it holds that $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\{1, L, |a|, |b|, 2[\sup_{x\in[a,b]^d}|f(x)|]\}$, and

(viii) *it holds that* $\sup_{x\in[a,b]^d}|(\mathcal{R}_{\mathfrak{r}}(\mathbf{F}))(x) - f(x)| \leq 2L(b-a)r$

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, 2.2.34, and 3.2.8).

Proof of Proposition 3.2.17. Throughout this proof assume without loss of generality that $r < d/4$ (cf. Lemma 3.2.15), let $\delta\colon [a,b]^d \times [a,b]^d \to \mathbb{R}$ satisfy for all $x,y \in [a,b]^d$ that $\delta(x,y) = \|x-y\|_1$, and let $K \in \mathbb{N}\cup\{\infty\}$ satisfy $K = \mathcal{C}^{([a,b]^d,\delta),(b-a)r}$. Note that item (i) in Lemma 3.2.14 assures that $K < \infty$. This and (3.124) ensure that there exist $\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_K \in [a,b]^d$ such that $\sup_{x\in[a,b]^d}\big[\min_{k\in\{1,2,\ldots,K\}} \delta(x,\mathfrak{x}_k)\big] \leq (b-a)r$. Combining this with Proposition 3.2.16 establishes items (i), (ii), (iii), (iv), (v), (vi), (vii), and (viii). The proof of Proposition 3.2.17 is thus complete. □

Proposition 3.2.18. *Let $d \in \mathbb{N}$, $L, a \in \mathbb{R}$, $b \in (a,\infty)$, $\varepsilon \in (0,1]$ and let $f\colon [a,b]^d \to \mathbb{R}$ satisfy for all $x,y \in [a,b]^d$ that $|f(x) - f(y)| \leq L\|x-y\|_1$ (cf. Definition 3.1.16). Then there exists $\mathbf{F} \in \mathbf{N}$ such that*

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

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

(iii) *it holds that* $\mathcal{H}(\mathbf{F}) \leq d\big(\max\big\{\log_2\big(\tfrac{3dL(b-a)}{2}\big), 0\big\} + \log_2(\varepsilon^{-1})\big) + 2$,

(iv) *it holds that* $\mathbb{D}_1(\mathbf{F}) \leq \varepsilon^{-d} d(3d\max\{L(b-a),1\})^d$,

(v) *it holds for all* $i \in \{2,3,4,\ldots\}$ *that* $\mathbb{D}_i(\mathbf{F}) \leq \varepsilon^{-d} 3\big(\tfrac{(3dL(b-a))^d}{2^i} + 1\big)$,

(vi) *it holds that* $\mathcal{P}(\mathbf{F}) \leq \varepsilon^{-2d} 9\big(3d\max\{L(b-a),1\}\big)^{2d} d^2$,

(vii) *it holds that* $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\{1, L, |a|, |b|, 2[\sup_{x\in[a,b]^d}|f(x)|]\}$, *and*

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

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, and 2.2.34).

Proof of Proposition 3.2.18. Throughout this proof assume without loss of generality that $L \neq 0$. Note that the assumption that for all $x,y \in [a,b]^d$ it holds that $|f(x) - f(y)| \leq L\|x-y\|_1$ and the assumption that $L \neq 0$ ensure that $L > 0$. Note that Proposition 3.2.17 shows that there exists $\mathbf{F} \in \mathbf{N}$ which satisfies that

(I) it holds that $\mathcal{I}(\mathbf{F}) = d$,

(II) it holds that $\mathcal{O}(\mathbf{F}) = 1$,

(III) it holds that $\mathcal{H}(\mathbf{F}) \leq \big(\big\lceil d\log_2\big(\tfrac{3dL(b-a)}{2\varepsilon}\big)\big\rceil + 1\big)\mathbb{1}_{(0,d/4)}\big(\tfrac{\varepsilon}{2L(b-a)}\big)$,

(IV) it holds that $\mathbb{D}_1(\mathbf{F}) \leq 2d\big(\tfrac{3dL(b-a)}{2\varepsilon}\big)^d \mathbb{1}_{(0,d/4)}\big(\tfrac{\varepsilon}{2L(b-a)}\big) + \mathbb{1}_{[d/4,\infty)}\big(\tfrac{\varepsilon}{2L(b-a)}\big)$,

(V) it holds for all $i \in \{2, 3, 4, \ldots\}$ that $\mathbb{D}_i(\mathbf{F}) \leq 3\lceil (\frac{3dL(b-a)}{2\varepsilon})^d \frac{1}{2^{i-1}} \rceil$,

(VI) it holds that $\mathcal{P}(\mathbf{F}) \leq 35(\frac{3dL(b-a)}{2\varepsilon})^{2d} d^2 \mathbb{1}_{(0,d/4)}(\frac{\varepsilon}{2L(b-a)}) + (d+1)\mathbb{1}_{[d/4,\infty)}(\frac{\varepsilon}{2L(b-a)})$,

(VII) it holds that $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\{1, L, |a|, |b|, 2[\sup_{x\in[a,b]^d}|f(x)|]\}$, and

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

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, 2.2.34, and 3.2.8). Moreover, observe that item (III) assures that

$$\begin{aligned} \mathcal{H}(\mathbf{F}) &\leq \big(d\big(\log_2\big(\tfrac{3dL(b-a)}{2}\big) + \log_2(\varepsilon^{-1})\big) + 2\big)\mathbb{1}_{(0,d/4)}\big(\tfrac{\varepsilon}{2L(b-a)}\big) \\ &\leq d\big(\max\big\{\log_2\big(\tfrac{3dL(b-a)}{2}\big), 0\big\} + \log_2(\varepsilon^{-1})\big) + 2. \end{aligned} \tag{3.152}$$

In addition, note that item (IV) implies that

$$\mathbb{D}_1(\mathbf{F}) \leq d\big(\tfrac{3d\max\{L(b-a),1\}}{\varepsilon}\big)^d \mathbb{1}_{(0,d/4)}\big(\tfrac{\varepsilon}{2L(b-a)}\big) + \mathbb{1}_{[d/4,\infty)}\big(\tfrac{\varepsilon}{2L(b-a)}\big) \leq \varepsilon^{-d} d(3d\max\{L(b-a),1\})^d. \tag{3.153}$$

Furthermore, observe that item (V) ensures that for all $i \in \{2, 3, 4, \ldots\}$ it holds that

$$\mathbb{D}_i(\mathbf{F}) \leq 3\big(\big(\tfrac{3dL(b-a)}{2\varepsilon}\big)^d \tfrac{1}{2^{i-1}} + 1\big) \leq \varepsilon^{-d} 3\big(\tfrac{(3dL(b-a))^d}{2^i} + 1\big). \tag{3.154}$$

Moreover, note that item (VI) ensures that

$$\begin{aligned} \mathcal{P}(\mathbf{F}) &\leq 9\big(\tfrac{3d\max\{L(b-a),1\}}{\varepsilon}\big)^{2d} d^2 \mathbb{1}_{(0,d/4)}\big(\tfrac{\varepsilon}{2L(b-a)}\big) + (d+1)\mathbb{1}_{[d/4,\infty)}\big(\tfrac{\varepsilon}{2L(b-a)}\big) \\ &\leq \varepsilon^{-2d} 9\big(3d\max\{L(b-a),1\}\big)^{2d} d^2. \end{aligned} \tag{3.155}$$

Combining this, (3.152), (3.153), (3.154), and items (I), (II), (VII), and (VIII) establishes items (i), (ii), (iii), (iv), (v), (vi), (vii), and (viii). The proof of Proposition 3.2.18 is thus complete. □

Corollary 3.2.19. *Let $d \in \mathbb{N}$, $L, a \in \mathbb{R}$, $b \in (a,\infty)$ and let $f\colon [a,b]^d \to \mathbb{R}$ satisfy for all $x, y \in [a,b]^d$ that $|f(x) - f(y)| \leq L\|x - y\|_1$ (cf. Definition 3.1.16). Then there exist $C \in \mathbb{R}$ and $\mathbf{F} = (\mathbf{F}_\varepsilon)_{\varepsilon\in(0,1]}\colon (0,1] \to \mathbf{N}$ such that for all $\varepsilon \in (0,1]$ it holds that*

$$\begin{gathered} \mathcal{H}(\mathbf{F}) \leq C(\log_2(\varepsilon^{-1}) + 1), \qquad \|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\big\{1, L, |a|, |b|, 2\big[\sup_{x\in[a,b]^d}|f(x)|\big]\big\}, \\ \mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R}), \quad \sup_{x\in[a,b]^d}|(\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon))(x) - f(x)| \leq \varepsilon, \quad \text{and} \quad \mathcal{P}(\mathbf{F}_\varepsilon) \leq C\varepsilon^{-2d} \end{gathered} \tag{3.156}$$

(cf. Definitions 2.1.6, 2.2.1, 2.2.3, and 2.2.34).

Proof of Corollary 3.2.19. Throughout this proof let $C = 9\big(3d\max\{L(b-a),1\}\big)^{2d}d^2$. Observe that items (i), (ii), (iii), (vi), (vii), and (viii) in Proposition 3.2.18 and the fact that for all $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned} d\big(\max\big\{\log_2\big(\tfrac{3dL(b-a)}{2}\big),0\big\} + \log_2(\varepsilon^{-1})\big) + 2 &\leq d\big(\big(\tfrac{3dL(b-a)}{2}\big) + \log_2(\varepsilon^{-1})\big) + 2 \\ &\leq d(3dL(b-a)) + 2 + d\log_2(\varepsilon^{-1}) \\ &\leq C(\log_2(\varepsilon^{-1}) + 1) \end{aligned} \tag{3.157}$$

imply that for every $\varepsilon \in (0,1]$ there exists $\mathbf{F}_\varepsilon \in \mathbf{N}$ such that $\mathcal{H}(\mathbf{F}) \leq C(\log_2(\varepsilon^{-1}) + 1)$, $\|\mathcal{T}(\mathbf{F})\|_\infty \leq \max\big\{1, L, |a|, |b|, 2\big[\sup_{x\in[a,b]^d}|f(x)|\big]\big\}$, $\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon) \in C(\mathbb{R}^d,\mathbb{R})$, $\sup_{x\in[a,b]^d}|(\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon))(x) - f(x)| \leq \varepsilon$, and $\mathcal{P}(\mathbf{F}_\varepsilon) \leq C\varepsilon^{-2d}$. The proof of Corollary 3.2.19 is thus complete. □

Corollary 3.2.20. *Let $d \in \mathbb{N}$, $L, a \in \mathbb{R}$, $b \in (a,\infty)$ and let $f\colon [a,b]^d \to \mathbb{R}$ satisfy for all $x, y \in [a,b]^d$ that $|f(x) - f(y)| \leq L\|x-y\|_1$ (cf. Definition 3.1.16). Then there exist $C \in \mathbb{R}$ and $\mathbf{F} = (\mathbf{F}_\varepsilon)_{\varepsilon\in(0,1]}\colon (0,1] \to \mathbf{N}$ such that for all $\varepsilon \in (0,1]$ it holds that*

$$\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon) \in C(\mathbb{R}^d,\mathbb{R}), \quad \sup_{x\in[a,b]^d}|(\mathcal{R}_\mathfrak{r}(\mathbf{F}_\varepsilon))(x) - f(x)| \leq \varepsilon, \quad \text{and} \quad \mathcal{P}(\mathbf{F}_\varepsilon) \leq C\varepsilon^{-2d} \tag{3.158}$$

(cf. Definitions 2.1.6, 2.2.1, and 2.2.3).

Proof of Corollary 3.2.20. Note that Corollary 3.2.19 establishes (3.158). The proof of Corollary 3.2.20 is thus complete. □

Chapter 4

Additional types of ANNs

4.1 Convolutional ANNs

Definition 4.1.1 (Discrete convolutions). *Let $T \in \mathbb{N}$, $a_1, a_2, \ldots, a_T, w_1, w_2, \ldots, w_T, \mathfrak{d}_1, \mathfrak{d}_2, \ldots, \mathfrak{d}_T \in \mathbb{N}$, $A = (A_{i_1,i_2,\ldots,i_T})_{(i_1,i_2,\ldots,i_T) \in (\times_{t=1}^T \{1,2,\ldots,a_t\})} \in \mathbb{R}^{a_1 \times a_2 \times \ldots \times a_T}$, $W = (W_{i_1,i_2,\ldots,i_T})_{(i_1,i_2,\ldots,i_T) \in (\times_{t=1}^T \{1,2,\ldots,w_t\})} \in \mathbb{R}^{w_1 \times w_2 \times \ldots \times w_T}$ satisfy for all $t \in \{1,2,\ldots,T\}$ that $\mathfrak{d}_t = a_t - w_t + 1$. Then we denote by $A * W = ((A * W)_{i_1,i_2,\ldots,i_T})_{(i_1,i_2,\ldots,i_T) \in (\times_{t=1}^T \{1,2,\ldots,\mathfrak{d}_t\})} \in \mathbb{R}^{\mathfrak{d}_1 \times \mathfrak{d}_2 \times \ldots \times \mathfrak{d}_T}$ the tensor which satisfies for all $i_1 \in \{1,2,\ldots,\mathfrak{d}_1\}$, $i_2 \in \{1,2,\ldots,\mathfrak{d}_2\}$, $\ldots$, $i_T \in \{1,2,\ldots,\mathfrak{d}_T\}$ that*

$$(A * W)_{i_1,i_2,\ldots,i_T} = \sum_{r_1=1}^{w_1} \sum_{r_2=1}^{w_2} \cdots \sum_{r_T=1}^{w_T} A_{i_1-1+r_1,i_2-1+r_2,\ldots,i_T-1+r_T} W_{r_1,r_2,\ldots,r_T}. \tag{4.1}$$

Definition 4.1.2 (One tensor). *Let $T \in \mathbb{N}$, $d_1, d_2, \ldots, d_T \in \mathbb{N}$. Then we denote by $\mathbf{I}^{d_1,d_2,\ldots,d_T} = (\mathbf{I}^{d_1,d_2,\ldots,d_T}_{i_1,i_2,\ldots,i_T})_{(i_1,i_2,\ldots,i_T) \in (\times_{t=1}^T \{1,2,\ldots,d_t\})} \in \mathbb{R}^{d_1 \times d_2 \times \ldots \times d_T}$ the tensor which satisfies for all $i_1 \in \{1,2,\ldots,d_1\}$, $i_2 \in \{1,2,\ldots,d_2\}$, $\ldots$, $i_T \in \{1,2,\ldots,d_T\}$ that*

$$\mathbf{I}^{d_1,d_2,\ldots,d_T}_{i_1,i_2,\ldots,i_T} = 1. \tag{4.2}$$

Definition 4.1.3 (Structured description of CNNs). *We denote by $\mathbf{C}$ the set given by*

$$\mathbf{C} = \textstyle\bigcup_{T,L \in \mathbb{N}} \bigcup_{l_0,l_1,\ldots,l_L \in \mathbb{N}} \bigcup_{(c_{k,t})_{(k,t) \in \{1,2,\ldots,L\} \times \{1,2,\ldots,T\}} \subseteq \mathbb{N}} \Big(\times_{k=1}^{L} \big((\mathbb{R}^{c_{k,1} \times c_{k,2} \times \ldots \times c_{k,T}})^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \big) \Big) \tag{4.3}$$

and we call $\mathbf{C}$ the set of convolutional ANNs.

Definition 4.1.4 (Realization associated to a convolutional ANN). *Let $a \in C(\mathbb{R}, \mathbb{R})$. Then we denote by*

$$\mathcal{R}^{\mathbf{C}}_a \colon \mathbf{C} \to \left(\bigcup_{\substack{T,i,o \in \mathbb{N}, \\ c_1,c_2,\ldots,c_T \in \mathbb{N}}} C\left(\bigcup_{\substack{d_1,d_2,\ldots,d_T \in \mathbb{N}, \\ \forall t \in \{1,2,\ldots,T\}\colon d_t \geq c_t}} (\mathbb{R}^{d_1 \times d_2 \times \ldots \times d_T})^i, \bigcup_{d_1,d_2,\ldots,d_T \in \mathbb{N}} (\mathbb{R}^{d_1 \times d_2 \times \ldots \times d_T})^o \right) \right) \tag{4.4}$$

the function which satisfies for all $T, L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $(c_{k,t})_{(k,t)\in\{1,2,\dots,L\}\times\{1,2,\dots,T\}} \subseteq \mathbb{N}$, $\Phi = (((W_{k,n,m})_{(n,m)\in\{1,2,\dots,l_k\}\times\{1,2,\dots,l_{k-1}\}}, (B_{k,n})_{n\in\{1,2,\dots,l_k\}}))_{k\in\{1,2,\dots,L\}} \in \bigtimes_{k=1}^{L}((\mathbb{R}^{c_{k,1}\times c_{k,2}\times\dots\times c_{k,T}})^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})$, $(\mathfrak{d}_{k,t})_{(k,t)\in\{0,1,\dots,L\}\times\{1,2,\dots,T\}} \subseteq \mathbb{N}$, $x_0 = (x_{0,1}, \dots, x_{0,l_0}) \in (\mathbb{R}^{\mathfrak{d}_{0,1}\times\mathfrak{d}_{0,2}\times\dots\times\mathfrak{d}_{0,T}})^{l_0}$, $x_1 = (x_{1,1}, \dots, x_{1,l_1}) \in (\mathbb{R}^{\mathfrak{d}_{1,1}\times\mathfrak{d}_{1,2}\times\dots\times\mathfrak{d}_{1,T}})^{l_1}$, $\dots$, $x_L = (x_{L,1}, \dots, x_{L,l_L}) \in (\mathbb{R}^{\mathfrak{d}_{L,1}\times\mathfrak{d}_{L,2}\times\dots\times\mathfrak{d}_{L,T}})^{l_L}$ *with* $\forall\, k \in \{1,2,\dots,L\}$, $t \in \{1,2,\dots,T\}\colon \mathfrak{d}_{k,t} = \mathfrak{d}_{k-1,t} - c_{k,t} + 1$ *and* $\forall\, k \in \{1,2,\dots,L\}$, $n \in \{1,2,\dots,l_k\}\colon x_{k,n} = \mathfrak{M}_{a\mathbb{1}_{(0,L)}(k)+\mathrm{id}_{\mathbb{R}}\mathbb{1}_{\{L\}}(k),\mathfrak{d}_{k,1},\mathfrak{d}_{k,2},\dots,\mathfrak{d}_{k,T}}\big(B_{k,n}\mathbf{I}^{\mathfrak{d}_{k,1},\mathfrak{d}_{k,2},\dots,\mathfrak{d}_{k,T}} + \sum_{m=1}^{l_{k-1}} x_{k-1,m} * W_{k,n,m}\big)$ *that*

$$\mathcal{R}^{\mathbf{C}}_a(\Phi) \in C\left(\bigcup_{\substack{d_1,d_2,\dots,d_T\in\mathbb{N}\\ \forall\, t\in\{1,2,\dots,T\}\colon d_t-\sum_{k=1}^{L}(c_{k,t}-1)\geq 1}} (\mathbb{R}^{d_1\times d_2\times\dots\times d_T})^{l_0}, \bigcup_{d_1,d_2,\dots,d_T\in\mathbb{N}} (\mathbb{R}^{d_1\times d_2\times\dots\times d_T})^{l_L}\right) \tag{4.5}$$

and $(\mathcal{R}^{\mathbf{C}}_a(\Phi))(x_0) = x_L$ *and for every* $\Phi \in \mathbf{C}$ *we call* $\mathcal{R}^{\mathbf{C}}_a(\Phi)$ *the realization function of the convolutional ANN* Φ *with activation function* a *(for every* $\Phi \in \mathbf{C}$ *we call* $\mathcal{R}^{\mathbf{C}}_a(\Phi)$ *the realization of the convolutional ANN* Φ *with activation* a*) (cf. Definitions 2.1.4, 4.1.1, 4.1.2, and 4.1.3).*

Exercise 4.1.1. *Let*

$$\begin{aligned}\Phi &= (((W_{1,n,m})_{(n,m)\in\{1,2,3\}\times\{1\}}, (B_{1,n})_{n\in\{1,2,3\}}), ((W_{2,n,m})_{(n,m)\in\{1\}\times\{1,2,3\}}, (B_{2,n})_{n\in\{1\}}))\\ &\in ((\mathbb{R}^2)^{3\times 1} \times \mathbb{R}^3) \times ((\mathbb{R}^3)^{1\times 3} \times \mathbb{R}^1)\end{aligned} \tag{4.6}$$

satisfy

$$W_{1,1,1} = (1,-1), \quad W_{1,2,1} = (2,-2), \quad W_{1,3,1} = (-3,3), \quad (B_{1,n})_{n\in\{1,2,3\}} = (1,2,3), \tag{4.7}$$

$$W_{2,1,1} = (1,-1,1), \quad W_{2,1,2} = (2,-2,2), \quad W_{2,1,3} = (-3,3,-3), \quad \text{and} \quad B_{2,1} = -2 \tag{4.8}$$

and let $v \in \mathbb{R}^9$ *satisfy* $v = (1,2,3,4,5,4,3,2,1)$*. Specify*

$$(\mathcal{R}^{\mathbf{C}}_{\mathfrak{r}}(\Phi))(v) \tag{4.9}$$

explicitly and prove that your result is correct (cf. Definitions 2.1.6 and 4.1.4)!

Exercise 4.1.2. *Prove or disprove the following statement: For every* $a \in C(\mathbb{R},\mathbb{R})$, $\Phi \in \mathbf{N}$ *there exists* $\Psi \in \mathbf{C}$ *such that for all* $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ *it holds that* $\mathbb{R}^{\mathcal{I}(\Phi)} \subseteq \operatorname{Domain}(\mathcal{R}^{\mathbf{C}}_a(\Psi))$ *and*

$$(\mathcal{R}^{\mathbf{C}}_a(\Psi))(x) = (\mathcal{R}_a(\Phi))(x) \tag{4.10}$$

(cf. Definitions 2.2.1, 2.2.3, 4.1.3, and 4.1.4).

Definition 4.1.5. *We denote by* $\langle\cdot,\cdot\rangle\colon \big[\bigcup_{d\in\mathbb{N}}(\mathbb{R}^d\times\mathbb{R}^d)\big] \to \mathbb{R}$ *the function which satisfies for all* $d \in \mathbb{N}$, $x = (x_1, x_2, \dots, x_d)$, $y = (y_1, y_2, \dots, y_d) \in \mathbb{R}^d$ *that*

$$\langle x, y\rangle = \sum_{i=1}^{d} x_i y_i. \tag{4.11}$$

Exercise 4.1.3. *For every $d \in \mathbb{N}$ let $\mathbf{e}_1^{(d)}, \mathbf{e}_2^{(d)}, \dots, \mathbf{e}_d^{(d)} \in \mathbb{R}^d$ satisfy $\mathbf{e}_1^{(d)} = (1, 0, \dots, 0)$, $\mathbf{e}_2^{(d)} = (0, 1, 0, \dots, 0)$, $\dots$, $\mathbf{e}_d^{(d)} = (0, \dots, 0, 1)$. Prove or disprove the following statement: For all $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \mathbf{N}$, $D \in \mathbb{N}$, $x = ((x_{i,j})_{j \in \{1,2,\dots,D\}})_{i \in \{1,2,\dots,\mathcal{I}(\Phi)\}} \in (\mathbb{R}^D)^{\mathcal{I}(\Phi)}$ it holds that*

$$(\mathcal{R}_a^{\mathbf{C}}(\Phi))(x) = \big((\langle \mathbf{e}_k^{(\mathcal{O}(\Phi))}, (\mathcal{R}_a(\Phi))((x_{i,j})_{i \in \{1,2,\dots,\mathcal{I}(\Phi)\}}) \rangle)_{j \in \{1,2,\dots,D\}} \big)_{k \in \{1,2,\dots,\mathcal{O}(\Phi)\}} \tag{4.12}$$

(cf. Definitions 2.2.1, 2.2.3, 4.1.4, and 4.1.5).

4.2 Recurrent ANNs

Definition 4.2.1 (Function unrolling)**.** *Let X, Y, I be sets, let $\mathbb{I} \in I$, $T \in \mathbb{N}$, and let $f \colon X \times I \to Y \times I$ be a function. Then we denote by $\mathfrak{R}_{f,T,\mathfrak{i}} \colon X^T \to Y^T$ the function which satisfies for all $x_1, x_2, \dots, x_T \in X$, $y_1, y_2, \dots, y_T \in Y$, $i_0, i_1, \dots, i_T \in I$ with $i_0 = \mathbb{I}$ and $\forall\, t \in \{1, 2, \dots, T\} \colon (y_t, i_t) = f(x_t, i_{t-1})$ that*

$$\mathfrak{R}_{f,T,\mathfrak{i}}(x_1, x_2, \dots, x_T) = (y_1, y_2, \dots, y_T) \tag{4.13}$$

and we call $\mathfrak{R}_{f,T,\mathfrak{i}}$ the T-times unrolled function f with initial information $\mathfrak{i}$.

Definition 4.2.2 (Simple recurrent ANN nodes)**.** *Let $\mathfrak{x}, \mathfrak{y}, \mathfrak{i} \in \mathbb{N}$, $\theta \in \mathbb{R}^{(\mathfrak{x}+\mathfrak{i}+1)\mathfrak{i}+(\mathfrak{i}+1)\mathfrak{y}}$ and let $\Psi_1 \colon \mathbb{R}^{\mathfrak{i}} \to \mathbb{R}^{\mathfrak{i}}$ and $\Psi_2 \colon \mathbb{R}^{\mathfrak{y}} \to \mathbb{R}^{\mathfrak{y}}$ be functions. Then we call r the realization function of the simple recurrent ANN node with parameter vector θ and activation functions Ψ_1 and Ψ_2 if and only if it holds that $r \colon \mathbb{R}^{\mathfrak{x}} \times \mathbb{R}^{\mathfrak{i}} \to \mathbb{R}^{\mathfrak{y}} \times \mathbb{R}^{\mathfrak{i}}$ is the function from $\mathbb{R}^{\mathfrak{x}} \times \mathbb{R}^{\mathfrak{i}}$ to $\mathbb{R}^{\mathfrak{y}} \times \mathbb{R}^{\mathfrak{i}}$ which satisfies for all $x \in \mathbb{R}^{\mathfrak{x}}$, $i \in \mathbb{R}^{\mathfrak{i}}$ that*

$$r(x, i) = \Big(\big(\Psi_2 \circ \mathcal{A}_{\mathfrak{y},\mathfrak{i}}^{\theta,(\mathfrak{x}+\mathfrak{i}+1)\mathfrak{i}} \circ \Psi_1 \circ \mathcal{A}_{\mathfrak{i},\mathfrak{x}+\mathfrak{i}}^{\theta,0} \big)(x, i), \big(\Psi_1 \circ \mathcal{A}_{\mathfrak{i},\mathfrak{x}+\mathfrak{i}}^{\theta,0} \big)(x, i) \Big) \tag{4.14}$$

(cf. Definition 2.1.1).

Definition 4.2.3 (Unrolled simple recurrent ANNs)**.** *Let $\mathfrak{x}, \mathfrak{y}, \mathfrak{i}, T \in \mathbb{N}$, $\theta \in \mathbb{R}^{(\mathfrak{x}+\mathfrak{i}+1)\mathfrak{i}+(\mathfrak{i}+1)\mathfrak{y}}$, $\mathbb{I} \in \mathbb{R}^{\mathfrak{i}}$ and let $\Psi_1 \colon \mathbb{R}^{\mathfrak{i}} \to \mathbb{R}^{\mathfrak{i}}$ and $\Psi_2 \colon \mathbb{R}^{\mathfrak{y}} \to \mathbb{R}^{\mathfrak{y}}$ be functions. Then we call R the realization function of the T-step unrolled simple recurrent ANN with parameter vector θ, activation functions Ψ_1 and Ψ_2, and initial information $\mathbb{I}$ if and only if there exists $r \colon \mathbb{R}^{\mathfrak{x}} \times \mathbb{R}^{\mathfrak{i}} \to \mathbb{R}^{\mathfrak{y}} \times \mathbb{R}^{\mathfrak{i}}$ such that*

- *(i) it holds that r is the realization of the simple recurrent ANN node with parameters θ and activation functions Ψ_1 and Ψ_2 and*
- *(ii) it holds that $R = \mathfrak{R}_{r,T,\mathbb{I}}$*

(cf. Definitions 4.2.1 and 4.2.2).

Exercise 4.2.1. *For every $T \in \mathbb{N}$, $\alpha \in (0, 1)$ let $R_{T,\alpha}$ be the realization function of the T-step unrolled simple recurrent ANN with parameter vector $(1, 0, 0, \alpha, 0, 1-\alpha, 0, 0, -1, 1, 0)$, activation functions $\mathfrak{M}_{\mathfrak{r},2}$ and $\mathrm{id}_{\mathbb{R}}$, and initial information $(0, 0)$ (cf. Definitions 2.1.4, 2.1.6, and 4.2.3). For every $T \in \mathbb{N}$, $\alpha \in (0, 1)$ specify $R_{T,\alpha}(1, 1, \dots, 1)$ explicitly and prove that your result is correct!*

4.3 Residual ANNs

Definition 4.3.1 (Structured description of residual ANNs)**.** *We denote by* $\mathbf{R}$ *the set given by*

$$\mathbf{R} = \textstyle\bigcup_{L\in\mathbb{N}} \bigcup_{l_0,l_1,\dots,l_L\in\mathbb{N}} \bigcup_{C\subseteq\{(r,k)\in(\mathbb{N}_0)^2\colon r<k\leq L\}} \Big(\big(\bigtimes_{k=1}^{L} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big) \times \big(\bigtimes_{(r,k)\in C} \mathbb{R}^{l_k\times l_r}\big)\Big) \tag{4.15}$$

and we call $\mathbf{R}$ *the set of residual ANNs.*

Definition 4.3.2 (Realization associated to a residual ANN)**.** *Let* $a \in C(\mathbb{R},\mathbb{R})$*. Then we denote by* $\mathcal{R}^{\mathbf{R}}_a \colon \mathbf{R} \to \big(\bigcup_{k,l\in\mathbb{N}} C(\mathbb{R}^k,\mathbb{R}^l)\big)$ *the function which satisfies for all* $L\in\mathbb{N}$, $l_0,l_1,\dots,l_L\in\mathbb{N}$, $C\subseteq\{(r,k)\in(\mathbb{N}_0)^2\colon r<k\leq L\}$, $\Phi = ((W_k,B_k)_{k\in\{1,2,\dots,L\}},(V_{r,k})_{(r,k)\in C}) \in \big(\big(\bigtimes_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big) \times \big(\bigtimes_{(r,k)\in C}\mathbb{R}^{l_k\times l_r}\big)\big)$, $x_0\in\mathbb{R}^{l_0}, x_1\in\mathbb{R}^{l_1},\dots,x_L\in\mathbb{R}^{l_L}$ *with* $\forall\, k\in\{1,2,\dots,L\}\colon x_k = \mathfrak{M}_{a\mathbb{1}_{(0,L)}(k)+\mathrm{id}_{\mathbb{R}}\mathbb{1}_{\{L\}}(k),l_k}(W_k x_{k-1} + B_k + \sum_{r\in\mathbb{N}_0,(r,k)\in C} V_{r,k}x_r)$ *that*

$$\mathcal{R}^{\mathbf{R}}_a(\Phi) \in C(\mathbb{R}^{l_0},\mathbb{R}^{l_L}) \qquad \text{and} \qquad (\mathcal{R}^{\mathbf{R}}_a(\Phi))(x_0) = x_L \tag{4.16}$$

and for every $\Phi\in\mathbf{R}$ *we call* $\mathcal{R}^{\mathbf{R}}_a(\Phi)$ *the realization function of the residual ANN* Φ *with activation function* a *(for every* $\Phi\in\mathbf{R}$ *we call* $\mathcal{R}^{\mathbf{R}}_a(\Phi)$ *the realization of the residual ANN* Φ *with activation* a*) (cf. Definitions 2.1.4 and 4.3.1).*

Exercise 4.3.1. *Let* $d=9$, $C=\{(1,3),(3,5)\}$, $V=(V_{r,k})_{(r,k)\in C}\in\bigtimes_{(r,k)\in C}\mathbb{R}^{d\times d}$ *satisfy* $V_{1,3} = V_{3,5} = \mathrm{I}_d$ *and let* $\Phi = (\mathfrak{I}_d \bullet \mathbf{P}_d(\mathbb{L}_1,\mathbb{L}_1,\dots,\mathbb{L}_1) \bullet \mathfrak{I}_d \bullet \mathbf{P}_d(\mathbb{L}_1,\mathbb{L}_1,\dots,\mathbb{L}_1), (V_{r,k})_{(r,k)\in C}) \in \mathbf{R}$ *(cf. Definitions 2.2.5, 2.2.9, 2.2.11, 2.2.16, 3.2.3, and 4.3.1). For every* $x\in\mathbb{R}^d$ *specify* $(\mathcal{R}^{\mathbf{R}}_{\mathfrak{r}}(\Phi))(x)$ *explicitly and prove that your result is correct (cf. Definitions 2.1.6 and 4.3.2)!*

Chapter 5

Optimization through flows of ordinary differential equations

In Chapter 6 and Chapter 7 below we study deterministic and stochastic gradient descent (GD) type optimization methods. Such methods are widely used in machine learning problems to approximatively compute minima of suitable objective functions. The stochastic GD type optimization methods in Chapter 7 can be viewed as suitable Monte Carlo approximations of the deterministic GD type optimization methods in Chapter 6 and the deterministic GD type optimization methods in Chapter 6 can, roughly speaking, be viewed as time-discrete approximations of solutions of suitable gradient flow ordinary differential equations (ODEs). To develop intuitions for GD type optimization methods and some of the tools which we employ to analyze such methods, we study in this chapter such gradient flow ODEs. In particular, we show in this chapter how such gradient flow ODEs can be used to approximatively solve appropriate optimization problems.

5.1 Introductory comments for the training of artificial neural networks

In this section we briefly sketch how gradient descent type optimization methods factor into machine learning problems. To do this, we now recall the deep supervised learning framework sketched in Section 1.1 above. Let $d, M \in \mathbb{N}$, $\mathcal{E} \in C(\mathbb{R}^d, \mathbb{R})$, $x_1, x_2, \dots, x_{M+1} \in \mathbb{R}^d$, $y_1, y_2, \dots, y_M \in \mathbb{R}$ satisfy for all $m \in \{1, 2, \dots, M\}$ that

$$y_m = \mathcal{E}(x_m), \tag{5.1}$$

and let $\Phi \colon C(\mathbb{R}^d, \mathbb{R}) \to [0, \infty)$ satisfy for all $\phi \in C(\mathbb{R}^d, \mathbb{R})$ that

$$\Phi(\phi) = \sum_{m=1}^{M} |\phi(x_m) - y_m|^2. \tag{5.2}$$

As in Section 1.1 we think of $M \in \mathbb{N}$ as the number of available input-output data pairs, we think of $d \in \mathbb{N}$ as the dimension of the input data, we think of $\mathcal{E}\colon \mathbb{R}^d \to \mathbb{R}$ as an unknown function which relates input and output data through (5.1), we think of $x_1, x_2, \dots, x_{M+1} \in \mathbb{R}^d$ as the available known input data, we think of $y_1, y_2, \dots, y_M \in \mathbb{R}$ as the available known output data, and the function $\Phi\colon C(\mathbb{R}^d, \mathbb{R}) \to [0, \infty)$ is the objective function in the optimization problem associated to the supervised learning problem in (5.2) above (cf. (1.2) in Section 1.1 above). In particular, observe that $\Phi(\mathcal{E}) = 0$ and we are trying to approximate the function $\mathcal{E}$ by approximatively computing a global minimizer of the function $\Phi\colon C(\mathbb{R}^d, \mathbb{R}) \to [0, \infty)$. In order to make this problem amenable to discrete numerical computations, we consider a spatially discretized version of the problem, where we compute minimizers of the function Φ restricted to a set of realization functions of neural networks. To do this, let $h \in \mathbb{N}$, $l_1, l_2, \dots, l_h, \mathbf{d} \in \mathbb{N}$ satisfy $\mathbf{d} = l_1(d+1) + \big[\sum_{k=2}^{h} l_k(l_{k-1}+1)\big] + l_h + 1$, and let

$$\mathfrak{N} = \big\{ \big(\mathbb{R}^d \ni x \mapsto \mathcal{N}^{\theta,d}_{\mathfrak{S}_{l_1},\mathfrak{S}_{l_2},\dots,\mathfrak{S}_{l_h},\mathrm{id}_{\mathbb{R}}}(x) \in \mathbb{R}\big) \colon \theta \in \mathbb{R}^{\mathbf{d}} \big\} \subseteq C(\mathbb{R}^d, \mathbb{R}) \tag{5.3}$$

(cf. Definitions 2.1.2 and 2.1.15). We think of h as the number of hidden layers of the neural networks we use as approximators, for every $i \in \{1, 2, \dots, h\}$ we think of $l_i \in \mathbb{N}$ as the number of neurons in the i-th hidden layer of the neural networks we use as approximators, we think of $\mathbf{d}$ as the number of real parameters necessary to describe the neural networks we use as approximators, and we think of $\mathfrak{N}$ as the set of realization functions of the neural networks we use as approximators.

We can now reformulate the optimization problem as the problem of approximately computing minima of the function $f\colon \mathbb{R}^{\mathbf{d}} \to [0, \infty)$ which satisfies for all $\theta \in \mathbb{R}^{\mathbf{d}}$ that

$$f(\theta) = \left[\sum_{m=1}^{M} \left| \big(\mathcal{N}^{\theta,d}_{\mathfrak{S}_{l_1},\mathfrak{S}_{l_2},\dots,\mathfrak{S}_{l_h},\mathrm{id}_{\mathbb{R}}}\big)(x_m) - y_m \right|^2 \right] \tag{5.4}$$

and this optimization is now accessible to discrete numerical computations.

Let $\xi \in \mathbb{R}^{\mathbf{d}}$ and let $\Theta = (\Theta_t)_{t \in [0,\infty)} \colon [0, \infty) \to \mathbb{R}^{\mathbf{d}}$ be a continuously differentiable function which satisfies for all $t \in [0, \infty)$ that

$$\Theta_0 = \xi \qquad \text{and} \qquad \dot{\Theta}_t = -(\nabla f)(\Theta_t). \tag{5.5}$$

Let $(\gamma_n)_{n \in \mathbb{N}} \subseteq [0, \infty)$ and let $\theta = (\theta_n)_{n \in \mathbb{N}_0} \colon \mathbb{N}_0 \to \mathbb{R}^{\mathbf{d}}$ satisfy for all $n \in \mathbb{N}$ that

$$\theta_0 = \xi \qquad \text{and} \qquad \theta_n = \theta_{n-1} - \gamma_n (\nabla f)(\theta_{n-1}). \tag{5.6}$$

Chapter 6

Deterministic gradient descent type optimization methods

This chapter reviews and studies deterministic gradient descent (GD) type optimization methods such as the classical plain vanilla gradient descent optimization method (see Section 6.1 below) as well as more sophisticated gradient descent type optimization methods including gradient descent optimization methods with momenta (cf. Sections 6.2, 6.3, and 6.7 below) and gradient descent optimization methods with adaptive modifications of the learning rates (cf. Sections 6.4–6.7 below).

6.1 The gradient descent optimization method

In this section we review and study the classical plain vanilla GD optimization method (cf., for example, Nesterov [9, Section 1.2.3], Boyd & Vandenberghe [1, Section 9.3], and Bubeck [2, Chapter 3]). A simple intuition behind the GD optimization method is the idea to solve a minimization problem by performing successive steps in direction of the steepest descents of the objective function, that is, by performing successive steps in the opposite direction of the gradients of the objective function.

Definition 6.1.1 (Gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $\xi \in \mathbb{R}^d$ and let $f\colon \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d \colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta). \tag{6.1}$$

Then we say that Θ is the gradient descent process for the objective function f with generalized gradient g, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, and initial value ξ (we say that Θ is the gradient descent process for the objective function f with learning rates $(\gamma_n)_{n\in\mathbb{N}}$ and initial value ξ) if and only if it holds that $\Theta\colon \mathbb{N}_0 \to \mathbb{R}^d$ is the function from $\mathbb{N}_0$ to $\mathbb{R}^d$ which satisfies for all $n \in \mathbb{N}$ that

$$\Theta_0 = \xi \qquad \textit{and} \qquad \Theta_n = \Theta_{n-1} - \gamma_n g(\Theta_{n-1}). \tag{6.2}$$

6.2 The gradient descent optimization method with classical momentum

In Section 6.1 above we have introduced and analyzed the classical plain vanilla GD optimization method. In the literature (see, for example, Ruder [11] for an overview) there are a number of somehow more sophisticated GD type optimization methods which aim to improve the convergence speed of the classical plain vanilla GD optimization method such as GD optimization methods with momenta (cf., for example, Sections 6.2, 6.3, and 6.7) or GD optimization methods with adaptive modifications of the learning rates (cf., for example, Sections 6.4–6.7). In this section we introduce one of such more sophisticated GD type optimization methods, that is, we introduce the so-called momentum gradient descent (GD) optimization method (see Definition 6.2.1 below).

Definition 6.2.1 (Momentum gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(\alpha_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $\xi \in \mathbb{R}^d$ and let $f\colon \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d\colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta). \tag{6.3}$$

Then we say that Θ is the momentum gradient descent process for the objective function f with generalized gradient g, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, and initial value ξ (we say that Θ is the momentum gradient descent process for the objective function f with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, and initial value ξ) if and only if it holds that $\Theta\colon \mathbb{N}_0 \to \mathbb{R}^d$ is the function from $\mathbb{N}_0$ to $\mathbb{R}^d$ which satisfies that there exists $\mathbf{m}\colon \mathbb{N}_0 \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbf{m}_0 = 0, \tag{6.4}$$

$$\mathbf{m}_n = \alpha_n \mathbf{m}_{n-1} + (1-\alpha_n) g(\Theta_{n-1}), \tag{6.5}$$

$$\text{and} \qquad \Theta_n = \Theta_{n-1} - \gamma_n \mathbf{m}_n. \tag{6.6}$$

6.2.1 A representation of the momentum GD optimization method

In (6.4)–(6.6) the momentum GD optimization method is formulated by means of a one-step recursion. This one-step recursion can efficiently be exploited in an implementation. The following elementary lemma, Lemma 6.2.2 below, provides a suitable full-history recursive representation for the momentum GD optimization method, which enables us to develop a better intuition for the momentum GD optimization method.

Lemma 6.2.2 (A representation of the momentum GD optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $\alpha \in [0,1]$, $\xi \in \mathbb{R}^d$, let $f\colon \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d\colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta), \tag{6.7}$$

and let $\Theta\colon \mathbb{N}_0 \to \mathbb{R}^d$ and $\mathbf{m}\colon \mathbb{N}_0 \to \mathbb{R}^d$ satisfy for all $n \in \mathbb{N}$ that

$$\Theta_0 = \xi, \qquad \mathbf{m}_0 = 0, \qquad \Theta_n = \Theta_{n-1} - \gamma_n \mathbf{m}_n, \tag{6.8}$$

$$\text{and} \qquad \mathbf{m}_n = \alpha\mathbf{m}_{n-1} + (1-\alpha) g(\Theta_{n-1}). \tag{6.9}$$

Then

(i) it holds for all $n \in \mathbb{N}_0$ that

$$\mathbf{m}_n = (1-\alpha)\left[\sum_{k=0}^{n-1} \alpha^k g(\Theta_{n-1-k})\right] \tag{6.10}$$

and

(ii) it holds for all $n \in \mathbb{N}$ that

$$\Theta_n = \Theta_{n-1} - \gamma_n (1-\alpha)\left[\sum_{k=0}^{n-1} \alpha^k g(\Theta_{n-1-k})\right]. \tag{6.11}$$

Proof of Lemma 6.2.2. We prove (6.10) by induction on $n \in \mathbb{N}_0$. For the base case $n = 0$ observe that (6.8) ensures that $\mathbf{m}_0 = (1-\alpha)0$. This establishes (6.10) in the base case $n = 0$. For the induction step observe that (6.9) assures that for all $n \in \mathbb{N}_0$ with

$$\mathbf{m}_n = (1-\alpha)\left[\sum_{k=0}^{n-1} \alpha^k g(\Theta_{n-1-k})\right] \tag{6.12}$$

it holds that

$$\begin{aligned}
\mathbf{m}_{n+1} &= \alpha \mathbf{m}_n + (1-\alpha) g(\Theta_n) \\
&= \alpha\left[(1-\alpha)\left[\sum_{k=0}^{n-1} \alpha^k g(\Theta_{n-1-k})\right]\right] + (1-\alpha) g(\Theta_n) \\
&= (1-\alpha)\left[\sum_{k=1}^{n} \alpha^k g(\Theta_{n-k})\right] + (1-\alpha)\alpha^0 g(\Theta_{n-0}) \\
&= (1-\alpha)\left[\sum_{k=0}^{n} \alpha^k g(\Theta_{n-k})\right] = (1-\alpha)\left[\sum_{k=0}^{(n+1)-1} \alpha^k g(\Theta_{(n+1)-1-k})\right].
\end{aligned} \tag{6.13}$$

Induction thus establishes (6.10). The proof of Lemma 6.2.2 is thus complete. □

6.2.2 Comparison of the GD optimization method with and without momentum in the case of a numerical example

In this subsection we provide a numerical comparison of the plain vanilla GD optimization method and the momentum GD optimization method in the case of the specific quadratic optimization problem in (6.14)–(6.15) below; see Illustration 6.2.3 below, Python code 6.1, and Figure 6.1 below.

Illustration 6.2.3. *Let $\mathcal{K} = 10$, $\kappa = 1$, $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2$, $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ satisfy*

$$\vartheta = \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \text{and} \qquad \xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \tag{6.14}$$

let $f \colon \mathbb{R}^2 \to \mathbb{R}$ satisfy for all $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ that

$$f(\theta) = \big(\tfrac{\kappa}{2}\big)|\theta_1 - \vartheta_1|^2 + \big(\tfrac{\mathcal{K}}{2}\big)|\theta_2 - \vartheta_2|^2, \tag{6.15}$$

let $\Theta \colon \mathbb{N}_0 \to \mathbb{R}^d$ satisfy for all $n \in \mathbb{N}$ that $\Theta_0 = \xi$ and

$$\begin{aligned} \Theta_n &= \Theta_{n-1} - \tfrac{2}{(\mathcal{K}+\kappa)}(\nabla f)(\Theta_{n-1}) = \Theta_{n-1} - \tfrac{2}{11}(\nabla f)(\Theta_{n-1}) \\ &= \Theta_{n-1} - 0.\overline{18}\,(\nabla f)(\Theta_{n-1}) \approx \Theta_{n-1} - 0.18\,(\nabla f)(\Theta_{n-1}), \end{aligned} \tag{6.16}$$

and let $\mathcal{M}, \mathbf{m} \colon \mathbb{N}_0 \to \mathbb{R}^d$ satisfy for all $n \in \mathbb{N}$ that $\mathcal{M}_0 = \xi$, $\mathbf{m}_0 = 0$, $\mathcal{M}_n = \mathcal{M}_{n-1} - 0.3\,\mathbf{m}_n$, and

$$\begin{aligned} \mathbf{m}_n &= 0.5\,\mathbf{m}_{n-1} + (1 - 0.5)\,(\nabla f)(\mathcal{M}_{n-1}) \\ &= 0.5\,(\mathbf{m}_{n-1} + (\nabla f)(\mathcal{M}_{n-1})). \end{aligned} \tag{6.17}$$

Then

(i) it holds for all $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ that

$$(\nabla f)(\theta) = \begin{pmatrix} \kappa(\theta_1 - \vartheta_1) \\ \mathcal{K}(\theta_2 - \vartheta_2) \end{pmatrix} = \begin{pmatrix} \theta_1 - 1 \\ 10\,(\theta_2 - 1) \end{pmatrix}, \tag{6.18}$$

(ii) it holds that

$$\Theta_0 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \tag{6.19}$$

$$\begin{aligned} \Theta_1 &= \Theta_0 - \tfrac{2}{11}(\nabla f)(\Theta_0) \approx \Theta_0 - 0.18(\nabla f)(\Theta_0) \\ &= \begin{pmatrix} 5 \\ 3 \end{pmatrix} - 0.18 \begin{pmatrix} 5 - 1 \\ 10(3 - 1) \end{pmatrix} = \begin{pmatrix} 5 - 0.18 \cdot 4 \\ 3 - 0.18 \cdot 10 \cdot 2 \end{pmatrix} \\ &= \begin{pmatrix} 5 - 0.72 \\ 3 - 3.6 \end{pmatrix} = \begin{pmatrix} 4.28 \\ -0.6 \end{pmatrix}, \end{aligned} \tag{6.20}$$

$$
\begin{aligned}
\Theta_2 \approx \Theta_1 - 0.18(\nabla f)(\Theta_1) &= \begin{pmatrix} 4.28 \\ -0.6 \end{pmatrix} - 0.18 \begin{pmatrix} 4.28 - 1 \\ 10(-0.6 - 1) \end{pmatrix} \\
&= \begin{pmatrix} 4.28 - 0.18 \cdot 3.28 \\ -0.6 - 0.18 \cdot 10 \cdot (-1.6) \end{pmatrix} = \begin{pmatrix} 4.10 - 0.18 \cdot 2 - 0.18 \cdot 0.28 \\ -0.6 + 1.8 \cdot 1.6 \end{pmatrix} \\
&= \begin{pmatrix} 4.10 - 0.36 - 2 \cdot 9 \cdot 4 \cdot 7 \cdot 10^{-4} \\ -0.6 + 1.6 \cdot 1.6 + 0.2 \cdot 1.6 \end{pmatrix} = \begin{pmatrix} 3.74 - 9 \cdot 56 \cdot 10^{-4} \\ -0.6 + 2.56 + 0.32 \end{pmatrix} \\
&= \begin{pmatrix} 3.74 - 504 \cdot 10^{-4} \\ 2.88 - 0.6 \end{pmatrix} = \begin{pmatrix} 3.6896 \\ 2.28 \end{pmatrix} \approx \begin{pmatrix} 3.69 \\ 2.28 \end{pmatrix},
\end{aligned} \tag{6.21}
$$

$$
\begin{aligned}
\Theta_3 \approx \Theta_2 - 0.18(\nabla f)(\Theta_2) &\approx \begin{pmatrix} 3.69 \\ 2.28 \end{pmatrix} - 0.18 \begin{pmatrix} 3.69 - 1 \\ 10(2.28 - 1) \end{pmatrix} \\
&= \begin{pmatrix} 3.69 - 0.18 \cdot 2.69 \\ 2.28 - 0.18 \cdot 10 \cdot 1.28 \end{pmatrix} = \begin{pmatrix} 3.69 - 0.2 \cdot 2.69 + 0.02 \cdot 2.69 \\ 2.28 - 1.8 \cdot 1.28 \end{pmatrix} \\
&= \begin{pmatrix} 3.69 - 0.538 + 0.0538 \\ 2.28 - 1.28 - 0.8 \cdot 1.28 \end{pmatrix} = \begin{pmatrix} 3.7438 - 0.538 \\ 1 - 1.28 + 0.2 \cdot 1.28 \end{pmatrix} \\
&= \begin{pmatrix} 3.2058 \\ 0.256 - 0.280 \end{pmatrix} = \begin{pmatrix} 3.2058 \\ -0.024 \end{pmatrix} \approx \begin{pmatrix} 3.21 \\ -0.02 \end{pmatrix},
\end{aligned} \tag{6.22}
$$

$$\vdots$$

and

(iii) it holds that

$$\mathcal{M}_0 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \tag{6.23}$$

$$
\begin{aligned}
\mathbf{m}_1 = 0.5\,(\mathbf{m}_0 + (\nabla f)(\mathcal{M}_0)) &= 0.5 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 5 - 1 \\ 10(3 - 1) \end{pmatrix} \right) \\
&= \begin{pmatrix} 0.5\,(0 + 4) \\ 0.5\,(0 + 10 \cdot 2) \end{pmatrix} = \begin{pmatrix} 2 \\ 10 \end{pmatrix},
\end{aligned} \tag{6.24}
$$

$$\mathcal{M}_1 = \mathcal{M}_0 - 0.3\,\mathbf{m}_1 = \begin{pmatrix} 5 \\ 3 \end{pmatrix} - 0.3 \begin{pmatrix} 2 \\ 10 \end{pmatrix} = \begin{pmatrix} 4.4 \\ 0 \end{pmatrix}, \tag{6.25}$$

$$\begin{aligned}\mathbf{m}_2 &= 0.5\,(\mathbf{m}_1 + (\nabla f)(\mathcal{M}_1)) = 0.5\left(\begin{pmatrix}2\\10\end{pmatrix} + \begin{pmatrix}4.4-1\\10(0-1)\end{pmatrix}\right)\\ &= \begin{pmatrix}0.5\,(2+3.4)\\0.5\,(10-10)\end{pmatrix} = \begin{pmatrix}2.7\\0\end{pmatrix},\end{aligned} \tag{6.26}$$

$$\mathcal{M}_2 = \mathcal{M}_1 - 0.3\,\mathbf{m}_2 = \begin{pmatrix}4.4\\0\end{pmatrix} - 0.3\begin{pmatrix}2.7\\0\end{pmatrix} = \begin{pmatrix}4.4-0.81\\0\end{pmatrix} = \begin{pmatrix}3.59\\0\end{pmatrix}, \tag{6.27}$$

$$\begin{aligned}\mathbf{m}_3 &= 0.5\,(\mathbf{m}_2 + (\nabla f)(\mathcal{M}_2)) = 0.5\left(\begin{pmatrix}2.7\\0\end{pmatrix} + \begin{pmatrix}3.59-1\\10(0-1)\end{pmatrix}\right)\\ &= \begin{pmatrix}0.5\,(2.7+2.59)\\0.5\,(0-10)\end{pmatrix} = \begin{pmatrix}0.5\cdot 5.29\\0.5(-10)\end{pmatrix}\\ &= \begin{pmatrix}2.5+0.145\\-5\end{pmatrix} = \begin{pmatrix}2.645\\-5\end{pmatrix} \approx \begin{pmatrix}2.65\\-5\end{pmatrix},\end{aligned} \tag{6.28}$$

$$\begin{aligned}\mathcal{M}_3 &= \mathcal{M}_2 - 0.3\,\mathbf{m}_3 \approx \begin{pmatrix}3.59\\0\end{pmatrix} - 0.3\begin{pmatrix}2.65\\-5\end{pmatrix}\\ &= \begin{pmatrix}3.59-0.795\\1.5\end{pmatrix} = \begin{pmatrix}3-0.205\\1.5\end{pmatrix} = \begin{pmatrix}2.795\\1.5\end{pmatrix} \approx \begin{pmatrix}2.8\\1.5\end{pmatrix},\end{aligned} \tag{6.29}$$

$$\vdots$$

.

```python
# Example for GD and momentum GD

import numpy as np
import matplotlib.pyplot as plt

# Number of steps for the schemes
N = 8

# Problem setting
d = 2
K = [1., 10.]

vartheta = np.array([1., 1.])
xi = np.array([5., 3.])

def f(x, y):
```

```
    result = K[0] / 2. * np.abs(x - vartheta[0]) ** 2 \
    + K[1] / 2. * np.abs(y - vartheta[1]) ** 2
    return result

def nabla_f(x):
    return K * (x - vartheta)

# Coefficients for GD
gamma_GD = 2 /(K[0] + K[1])

# Coefficients for momentum
gamma_momentum = 0.3
alpha = 0.5

# Placeholder for processes
Theta = np.zeros((N+1, d))
M = np.zeros((N+1, d))
m = np.zeros((N+1, d))

Theta[0] = xi
M[0] = xi

# Perform gradient descent
for i in range(N):
    Theta[i+1] = Theta[i] - gamma_GD * nabla_f(Theta[i])

# Perform momentum GD
for i in range(N):
    m[i+1] = alpha * m[i] + (1 - alpha) * nabla_f(M[i])
    M[i+1] = M[i] - gamma_momentum * m[i+1]


### Plot ###
plt.figure()

# Plot the gradient descent process
plt.plot(Theta[:, 0], Theta[:, 1],
         label = "GD", color = "c",
         linestyle = "--", marker = "*")

# Plot the momentum gradient descent process
plt.plot(M[:, 0], M[:, 1],
         label = "Momentum", color = "orange", marker = "*")

# Target value
plt.scatter(vartheta[0],vartheta[1],
            label = "vartheta", color = "red", marker = "x")

# Plot contour lines of f
```

```
x = np.linspace(-3., 7., 100)
y = np.linspace(-2., 4., 100)
X, Y = np.meshgrid(x, y)
Z = f(X, Y)
cp = plt.contour(X, Y, Z, colors="black",
                 levels = [0.5,2,4,8,16],
                 linestyles=":")

plt.legend()
plt.savefig("GD_momentum_plots.pdf")
plt.show()
```

Source code 6.1: PYTHON code for Figure 6.1

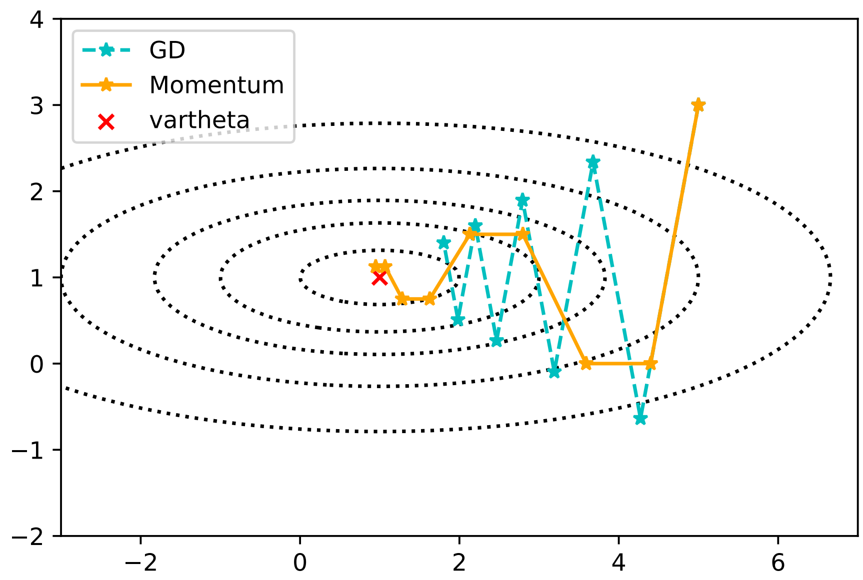

Figure 6.1: Result of a call of PYTHON code 6.1

Exercise 6.2.1. *Let $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(\alpha_n)_{n\in\mathbb{N}} \subseteq [0,1]$ satisfy for all $n \in \mathbb{N}$ that $\gamma_n = \frac{1}{n}$ and $\alpha_n = \frac{1}{2}$, let $f\colon \mathbb{R} \to \mathbb{R}$ satisfy for all $\theta \in \mathbb{R}$ that $f(\theta) = \theta^2$, and let Θ be the momentum gradient descent process for the objective function f with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, and initial value 1 (cf. Definition 6.2.1). Specify Θ_1, Θ_2, Θ_3, and Θ_4 explicitly and prove that your results are correct!*

6.3 The gradient descent optimization method with Nesterov momentum

Definition 6.3.1 (Nesterov accelerated gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(\alpha_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $\xi \in \mathbb{R}^d$ and let $f\colon \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d \colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta). \tag{6.30}$$

Then we say that Θ is the Nesterov accelerated gradient descent process for the objective function f with generalized gradient g, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, and initial value ξ (we say that Θ is the Nesterov accelerated gradient descent process for the objective function f with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, and initial value ξ) if and only if it holds that $\Theta\colon \mathbb{N}_0 \to \mathbb{R}^d$ is the function from $\mathbb{N}_0$ to $\mathbb{R}^d$ which satisfies that there exists $\mathbf{m}\colon \mathbb{N}_0 \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbf{m}_0 = 0, \tag{6.31}$$

$$\mathbf{m}_n = \alpha_n \mathbf{m}_{n-1} + (1-\alpha_n)\, g(\Theta_{n-1} - \gamma_n \alpha_n \mathbf{m}_{n-1}), \tag{6.32}$$

$$\text{and} \qquad \Theta_n = \Theta_{n-1} - \gamma_n \mathbf{m}_n. \tag{6.33}$$

6.4 The adaptive gradient descent optimization method (Adagrad optimization method)

Definition 6.4.1 (Adagrad optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $\varepsilon \in (0,\infty)$, $\xi \in \mathbb{R}^d$ and let $f\colon \mathbb{R}^d \to \mathbb{R}$ and $g = (g_1,\dots,g_d)\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d \colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta). \tag{6.34}$$

Then we say that Θ is the Adagrad gradient descent process for the objective function f with generalized gradient g, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, regularizing factor ε, and initial value ξ (we say that Θ is the Adagrad gradient descent process for the objective function f with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, regularizing factor ε, and initial value ξ) if and only if it holds that $\Theta = (\Theta^{(1)},\dots,\Theta^{(d)})\colon \mathbb{N}_0 \to \mathbb{R}^d$ is the function from $\mathbb{N}_0$ to $\mathbb{R}^d$ which satisfies for all $n \in \mathbb{N}$, $i \in \{1,2,\dots,d\}$ that

$$\Theta_0 = \xi \qquad \text{and} \qquad \Theta_n^{(i)} = \Theta_{n-1}^{(i)} - \gamma_n \Bigg[\varepsilon + \sum_{k=0}^{n-1} |g_i(\Theta_k)|^2\Bigg]^{-1/2} g_i(\Theta_{n-1}). \tag{6.35}$$

6.5 The root mean square propagation gradient descent optimization method (RMSprop gradient descent optimization method)

Definition 6.5.1 (RMSprop gradient descent optimization method). *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(\beta_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $\varepsilon \in (0,\infty)$, $\xi \in \mathbb{R}^d$ and let $f \colon \mathbb{R}^d \to \mathbb{R}$ and $g = (g_1,\dots,g_d) \colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d \colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta). \tag{6.36}$$

Then we say that Θ is the RMSprop gradient descent process for the objective function f with generalized gradient g, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, regularizing factor ε, and initial value ξ (we say that Θ is the RMSprop gradient descent process for the objective function f with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, regularizing factor ε, and initial value ξ) if and only if it holds that $\Theta = (\Theta^{(1)},\dots,\Theta^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$ is the function from $\mathbb{N}_0$ to $\mathbb{R}^d$ which satisfies that there exists $\mathbb{M} = (\mathbb{M}^{(1)},\dots,\mathbb{M}^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$, $i \in \{1,2,\dots,d\}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbb{M}_0 = 0, \qquad \mathbb{M}_n^{(i)} = \beta_n\,\mathbb{M}_{n-1}^{(i)} + (1-\beta_n)\left|g_i(\Theta_{n-1})\right|^2, \tag{6.37}$$

$$\text{and} \qquad \Theta_n^{(i)} = \Theta_{n-1}^{(i)} - \gamma_n\left[\varepsilon + \mathbb{M}_n^{(i)}\right]^{-1/2} g_i(\Theta_{n-1}). \tag{6.38}$$

6.6 The Adadelta gradient descent optimization method

Definition 6.6.1 (Adadelta gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\beta_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $(\delta_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $\varepsilon \in (0,\infty)$, $\xi \in \mathbb{R}^d$ and let $f \colon \mathbb{R}^d \to \mathbb{R}$ and $g = (g_1,\ldots,g_d) \colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d \colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta). \tag{6.39}$$

Then we say that Θ is the Adadelta gradient descent process for the objective function f with generalized gradient g, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, delta decay factors $(\delta_n)_{n\in\mathbb{N}}$, regularizing factor ε, and initial value ξ (we say that Θ is the Adadelta gradient descent process for the objective function f with second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, delta decay factors $(\delta_n)_{n\in\mathbb{N}}$, regularizing factor ε, and initial value ξ) if and only if it holds that $\Theta = (\Theta^{(1)},\ldots,\Theta^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$ is the function from $\mathbb{N}_0$ to $\mathbb{R}^d$ which satisfies that there exist $\mathbb{M} = (\mathbb{M}^{(1)},\ldots,\mathbb{M}^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$ and $\Delta = (\Delta^{(1)},\ldots,\Delta^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$, $i \in \{1,2,\ldots,d\}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbb{M}_0 = 0, \qquad \Delta_0 = 0, \tag{6.40}$$

$$\mathbb{M}^{(i)}_n = \beta_n \, \mathbb{M}^{(i)}_{n-1} + (1-\beta_n)|g_i(\Theta_{n-1})|^2, \tag{6.41}$$

$$\Theta^{(i)}_n = \Theta^{(i)}_{n-1} - \left[\frac{\varepsilon + \Delta^{(i)}_{n-1}}{\varepsilon + \mathbb{M}^{(i)}_n}\right]^{1/2} g_i(\Theta_{n-1}), \tag{6.42}$$

$$\text{and} \qquad \Delta^{(i)}_n = \delta_n \, \Delta^{(i)}_{n-1} + (1-\delta_n)\,|\Theta^{(i)}_n - \Theta^{(i)}_{n-1}|^2. \tag{6.43}$$

6.7 The adaptive moment estimation gradient descent optimization method (Adam gradient descent optimization method)

Definition 6.7.1 (Adam gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(\alpha_n)_{n\in\mathbb{N}} \subseteq [0,1)$, $(\beta_n)_{n\in\mathbb{N}} \subseteq [0,1)$, $\xi \in \mathbb{R}^d$ and let $f \colon \mathbb{R}^d \to \mathbb{R}$ and $g = (g_1,\ldots,g_d) \colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $\theta \in \{v \in \mathbb{R}^d \colon (f \text{ is differentiable at } v)\}$ that*

$$g(\theta) = (\nabla f)(\theta). \tag{6.44}$$

Then we say that Θ is the Adam gradient descent process for the objective function f with generalized gradient g, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, and initial value ξ (we say that Θ is the Adam gradient descent process for the objective function f with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, and initial value ξ) if and only if it holds that $\Theta = (\Theta^{(1)},\ldots,\Theta^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$ is the function from $\mathbb{N}_0$ to $\mathbb{R}^d$ which satisfies that there exist $\mathbf{m} = (\mathbf{m}^{(1)},\ldots,\mathbf{m}^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$, $\mathbb{M} = (\mathbb{M}^{(1)},\ldots,\mathbb{M}^{(d)}) \colon \mathbb{N}_0 \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$, $i \in \{1,2,\ldots,d\}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbf{m}_0 = 0, \qquad \mathbb{M}_0 = 0, \tag{6.45}$$

$$\mathbf{m}_n = \alpha_n \, \mathbf{m}_{n-1} + (1 - \alpha_n) \, g(\Theta_{n-1}), \tag{6.46}$$

$$\mathbb{M}_n^{(i)} = \beta_n \, \mathbb{M}_{n-1}^{(i)} + (1 - \beta_n) |g_i(\Theta_{n-1})|^2, \tag{6.47}$$

$$\text{and} \qquad \Theta_n^{(i)} = \Theta_{n-1}^{(i)} - \gamma_n \left[\varepsilon + \left[\tfrac{\mathbb{M}_n^{(i)}}{(1 - \prod_{l=1}^{n} \beta_l)} \right]^{1/2} \right]^{-1} \left[\frac{\mathbf{m}_n^{(i)}}{(1 - \prod_{l=1}^{n} \alpha_l)} \right]. \tag{6.48}$$

Chapter 7

Stochastic gradient descent type optimization methods

This chapter reviews and studies stochastic gradient descent (SGD) type optimization methods such as the classical plain vanilla SGD optimization method (see Section 7.1) as well as more sophisticated SGD type optimization methods including SGD type optimization methods with momenta (cf. Sections 7.2, 7.3, and 7.7 below) and SGD type optimization methods with adaptive modifications of the learning rate (cf. Sections 7.4–7.7 below). We also refer to the overview article Ruder [11] and the reference list in [4] for further references on SGD type optimization methods.

7.1 The stochastic gradient descent optimization method

Definition 7.1.1 (Stochastic gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(J_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $(S,\mathcal{S})$ be a measurable space, let $\xi\colon \Omega \to \mathbb{R}^d$ and $X_{n,j}\colon \Omega \to S$, $j \in \{1,2,\dots,J_n\}$, $n \in \mathbb{N}$, be random variables, and let $F = (F(\theta,x))_{(\theta,x)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}$ and $G\colon \mathbb{R}^d \times S \to \mathbb{R}^d$ satisfy for all $x \in S$, $\theta \in \{v \in \mathbb{R}^d\colon F(\cdot,x)$ is differentiable at $v\}$ that*

$$G(\theta,x) = (\nabla_\theta F)(\theta,x). \tag{7.1}$$

Then we say that Θ is the stochastic gradient descent process on $((\Omega,\mathcal{F},\mathbb{P}),(S,\mathcal{S}))$ for the loss function F with generalized gradient G, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$ (we say that Θ is the stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$) if and only if it holds that $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ is the function from $\mathbb{N}_0 \times \Omega$ to $\mathbb{R}^d$ which satisfies for all $n \in \mathbb{N}$ that

$$\Theta_0 = \xi \qquad \textit{and} \qquad \Theta_n = \Theta_{n-1} - \gamma_n \left[\frac{1}{J_n}\sum_{j=1}^{J_n} G(\Theta_{n-1}, X_{n,j})\right]. \tag{7.2}$$

7.1.1 Examples for stochastic optimization problems

Example 7.1.2 (Sums of optimization problems). *Let $d, N \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $\xi \in \mathbb{R}^d$, let $f_k\colon \mathbb{R}^d \to \mathbb{R}$, $k \in \{1,2,\dots,N\}$, be differentiable functions, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\mathsf{k}_n\colon \Omega \to \{1,2,\dots,N\}$, $n \in \mathbb{N}$, be independent $\mathcal{U}_{\{1,2,\dots,N\}}$-distributed random variables, let $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $n \in \mathbb{N}$ that*

$$\Theta_0 = \xi \qquad \text{and} \qquad \Theta_n = \Theta_{n-1} - \gamma_n(\nabla f_{\mathsf{k}_n})(\Theta_{n-1}), \tag{7.3}$$

and let $F\colon \mathbb{R}^d \times \{1,2,\dots,N\} \to \mathbb{R}$ satisfy for all $\theta \in \mathbb{R}^d$, $k \in \{1,2,\dots,N\}$ that

$$F(\theta,k) = f_k(\theta). \tag{7.4}$$

Then

(i) it holds that Θ is the stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $\mathbb{N} \ni n \mapsto 1 \in \mathbb{N}$, initial value ξ, and data $(\mathsf{k}_n)_{(n,j)\in\mathbb{N}\times\{1\}}$ (cf. Definition 7.1.1) and

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

$$\mathbb{E}\big[F(\theta,\mathsf{k}_1)\big] = \frac{1}{N}\left[\sum_{k=1}^N f_k(\theta)\right]. \tag{7.5}$$

Proof of Example 7.1.2. First, note that (7.4) ensures that for all $n \in \mathbb{N}$ it holds that

$$\Theta_n = \Theta_{n-1} - \gamma_n(\nabla f_{\mathsf{k}_n})(\Theta_{n-1}) = \Theta_{n-1} - \gamma_n(\nabla_\theta F)(\Theta_{n-1},\mathsf{k}_n). \tag{7.6}$$

Combining this with the assumption that $\Theta_0 = \xi$ proves item (i). Moreover, observe that (7.4) and the assumption that k_1 is a $\mathcal{U}_{\{1,2,\dots,N\}}$-distributed random variable demonstrate that

$$\mathbb{E}\big[F(\theta,\mathsf{k}_1)\big] = \frac{1}{N}\left[\sum_{k=1}^N F(\theta,k)\right] = \frac{1}{N}\left[\sum_{k=1}^N f_k(\theta)\right]. \tag{7.7}$$

This establishes item (ii). The proof of Example 7.1.2 is thus complete. □

Example 7.1.3 (Objective functions induced by data). *Let $d, N, \mathcal{I}, \mathcal{O} \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $\xi \in \mathbb{R}^d$, $x_1, x_2, \dots, x_N \in \mathbb{R}^{\mathcal{I}}$, let $\Phi\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{O}}$ be a function, let $u = (u_\theta(x))_{(\theta,x)\in\mathbb{R}^d\times\mathbb{R}^{\mathcal{I}}}\colon \mathbb{R}^d \times \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{O}}$ be a function which satisfies for every $x \in \mathbb{R}^{\mathcal{I}}$ that the function $\mathbb{R}^d \ni \theta \mapsto u_\theta(x) \in \mathbb{R}^{\mathcal{O}}$ is differentiable, let $F\colon \mathbb{R}^d \times \{1,2,\dots,N\} \to \mathbb{R}$ satisfy for all $\theta \in \mathbb{R}^d$, $k \in \{1,2,\dots,N\}$ that*

$$F(\theta,k) = \|u_\theta(x_k) - \Phi(x_k)\|_2^2, \tag{7.8}$$

let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\mathsf{k}_n\colon \Omega \to \{1,2,\dots,N\}$, $n \in \mathbb{N}$, be independent $\mathcal{U}_{\{1,2,\dots,N\}}$-distributed random variables, and let $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $n \in \mathbb{N}$ that $\Theta_0 = \xi$ and

$$\Theta_n = \Theta_{n-1} - \gamma_n(\nabla_\theta F)(\Theta_{n-1},\mathsf{k}_n) \tag{7.9}$$

Then

(i) *it holds that Θ is the stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $\mathbb{N} \ni n \mapsto 1 \in \mathbb{N}$, initial value ξ, and data $(\mathsf{k}_n)_{(n,j)\in\mathbb{N}\times\{1\}}$ (cf. Definition 7.1.1) and*

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

$$\mathbb{E}\big[F(\theta, \mathsf{k}_1)\big] = \frac{1}{N}\left[\sum_{k=1}^{N} \|u_\theta(x_k) - \Phi(x_k)\|_2^2\right]. \tag{7.10}$$

Proof of Example 7.1.3. Throughout this proof let $f_k\colon \mathbb{R}^d \to \mathbb{R}$, $k \in \{1, 2, \dots, N\}$, satisfy for all $\theta \in \mathbb{R}^d$, $k \in \{1, 2, \dots, N\}$ that

$$f_k(\theta) = \|u_\theta(x_k) - \Phi(x_k)\|_2^2. \tag{7.11}$$

Note that Example 7.1.2 (applied with $f_k \curvearrowleft f_k$ for $k \in \{1, 2, \dots, N\}$ in the notation of Example 7.1.2) establishes items (i) and (ii). The proof of Example 7.1.3 is thus complete. $\square$

7.2 The stochastic gradient descent optimization method with classical momentum

In this section we present the SGD optimization method with classical momentum. The idea for classical momentum was first introduced by Polyak for the (deterministic) GD optimization method (see Polyak [10] and Section 6.2 above).

Definition 7.2.1 (Momentum stochastic gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0, \infty)$, $(J_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $(\alpha_n)_{n\in\mathbb{N}} \subseteq [0, 1]$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, let $\xi\colon \Omega \to \mathbb{R}^d$ and $X_{n,j}\colon \Omega \to S$, $j \in \{1, 2, \dots, J_n\}$, $n \in \mathbb{N}$, be random variables, and let $F = (F(\theta, x))_{(\theta,x)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}$ and $G\colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be functions which satisfy for all $x \in S$, $\theta \in \{v \in \mathbb{R}^d\colon F(\cdot, x) \text{ is differentiable at } v\}$ that*

$$G(\theta, x) = (\nabla_\theta F)(\theta, x). \tag{7.12}$$

Then we say that Θ is the momentum stochastic gradient descent process on $((\Omega, \mathcal{F}, \mathbb{P}), (S, \mathcal{S}))$ for the loss function F with generalized gradient G, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\le J_k\}}$ (we say that Θ is the momentum stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\le J_k\}}$) if and only if $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ is the function from $\mathbb{N}_0 \times \Omega$ to $\mathbb{R}^d$ which satisfies that there exists a function $\mathbf{m}\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbf{m}_0 = 0, \tag{7.13}$$

$$\mathbf{m}_n = \alpha_n \mathbf{m}_{n-1} + (1-\alpha_n)\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G(\Theta_{n-1}, X_{n,j})\right], \tag{7.14}$$

$$\text{and} \qquad \Theta_n = \Theta_{n-1} - \gamma_n \mathbf{m}_n. \tag{7.15}$$

7.3 The stochastic gradient descent optimization method with Nesterov momentum

Nesterov accelerated stochastic gradient descent (NAG) builds on the idea of classical momentum and attemps to provide some kind of foresight to the scheme. This idea was first introduced by Nesterov as an adaption of the deterministic momentum GD optimization method (see Nesterov [8]).

Definition 7.3.1 (Nesterov accelerated stochastic gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(J_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $(\alpha_n)_{n\in\mathbb{N}} \subseteq [0,1]$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(S, \mathcal{S})$ be a measurable space, let $\xi \colon \Omega \to \mathbb{R}^d$ and $X_{n,j} \colon \Omega \to S$, $j \in \{1, 2, \dots, J_n\}$, $n \in \mathbb{N}$, be random variables, and let $F = (F(\theta, x))_{(\theta,x)\in\mathbb{R}^d\times S} \colon \mathbb{R}^d \times S \to \mathbb{R}$ and $G \colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be functions which satisfy for all $x \in S$, $\theta \in \{v \in \mathbb{R}^d \colon F(\cdot, x) \text{ is differentiable at } v\}$ that*

$$G(\theta, x) = (\nabla_\theta F)(\theta, x). \tag{7.16}$$

Then we say that Θ is the Nesterov accelerated stochastic gradient descent process on $((\Omega, \mathcal{F}, \mathbb{P}), (S, \mathcal{S}))$ for the loss function F with generalized gradient G, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2 \colon l\leq J_k\}}$ (we say that Θ is the Nesterov accelerated stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, momentum decay rates $(\alpha_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2 \colon l\leq J_k\}}$) if and only if $\Theta \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ is the function from $\mathbb{N}_0 \times \Omega$ to $\mathbb{R}^d$ which satisfies that there exists a function $\mathbf{m} \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbf{m}_0 = 0, \tag{7.17}$$

$$\mathbf{m}_n = \alpha_n \mathbf{m}_{n-1} + (1-\alpha_n)\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G\big(\Theta_{n-1} - \gamma_n \alpha_n \mathbf{m}_{n-1},\, X_{n,j}\big)\right], \tag{7.18}$$

$$\text{and} \qquad \Theta_n = \Theta_{n-1} - \gamma_n \mathbf{m}_n. \tag{7.19}$$

7.4 The adaptive stochastic gradient descent optimization method (Adagrad)

Definition 7.4.1 (Adagrad stochastic gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(J_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $\varepsilon \in (0,\infty)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $(S,\mathcal{S})$ be a measurable space, let $\xi\colon \Omega \to \mathbb{R}^d$ and $X_{n,j}\colon \Omega \to S$, $j \in \{1,2,\dots,J_n\}$, $n \in \mathbb{N}$, be random variables, and let $F = (F(\theta,x))_{(\theta,x)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}$ and $G = (G_1,G_2,\dots,G_d)\colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be functions which satisfy for all $x \in S$, $\theta \in \{v \in \mathbb{R}^d\colon F(\cdot,x) \text{ is differentiable at } v\}$ that*

$$G(\theta,x) = (\nabla_\theta F)(\theta,x). \tag{7.20}$$

Then we say that Θ is the Adagrad stochastic gradient descent process on $((\Omega,\mathcal{F},\mathbb{P}),(S,\mathcal{S}))$ for the loss function F with generalized gradient G, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, regularizing factor ε, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$ (we say that Θ is the Adagrad stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, regularizing factor ε, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$) if and only if it holds that $\Theta = (\Theta^{(1)},\Theta^{(2)},\dots,\Theta^{(d)})\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ is the function from $\mathbb{N}_0 \times \Omega$ to $\mathbb{R}^d$ which satisfies for all $n \in \mathbb{N}$, $i \in \{1,2,\dots,d\}$ that $\Theta_0 = \xi$ and

$$\Theta_n^{(i)} = \Theta_{n-1}^{(i)} - \gamma_n\biggl(\varepsilon + \sum_{k=1}^{n}\Bigl[\tfrac{1}{J_k}\textstyle\sum_{j=1}^{J_k} G_i(\Theta_{k-1},X_{k,j})\Bigr]^2\biggr)^{-1/2}\Biggl[\frac{1}{J_n}\sum_{j=1}^{J_n} G_i(\Theta_{n-1},X_{n,j})\Biggr]. \tag{7.21}$$

7.5 The root mean square propagation stochastic gradient descent optimization method (RMSprop)

Definition 7.5.1 (RMSprop stochastic gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(J_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $(\beta_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $\varepsilon \in (0,\infty)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $(S,\mathcal{S})$ be a measurable space, let $\xi\colon \Omega \to \mathbb{R}^d$ and $X_{n,j}\colon \Omega \to S$, $j \in \{1,2,\dots,J_n\}$, $n \in \mathbb{N}$, be random variables, and let $F = (F(\theta,x))_{(\theta,x)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}$ and $G = (G_1,G_2,\dots,G_d)\colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be functions which satisfy for all $x \in S$, $\theta \in \{v \in \mathbb{R}^d \colon F(\cdot,x)$ is differentiable at $v\}$ that*

$$G(\theta,x) = (\nabla_\theta F)(\theta,x). \tag{7.22}$$

Then we say that Θ is the RMSprop stochastic gradient descent process on $((\Omega,\mathcal{F},\mathbb{P}),(S,\mathcal{S}))$ for the loss function F with generalized gradient G, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, regularizing factor ε, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$ (we say that Θ is the RMSprop stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, regularizing factor ε, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$) if and only if it holds that $\Theta = (\Theta^{(1)},\dots,\Theta^{(d)})\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ is the function from $\mathbb{N}_0 \times \Omega$ to $\mathbb{R}^d$ which satisfies that there exists a function $\mathbb{M} = (\mathbb{M}^{(1)},\mathbb{M}^{(2)},\dots,\mathbb{M}^{(d)})\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$, $i \in \{1,2,\dots,d\}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbb{M}_0 = 0, \tag{7.23}$$

$$\mathbb{M}_n^{(i)} = \beta_n\,\mathbb{M}_{n-1}^{(i)} + (1-\beta_n)\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G_i(\Theta_{n-1},X_{n,j})\right]^2, \tag{7.24}$$

$$\text{and} \qquad \Theta_n^{(i)} = \Theta_{n-1}^{(i)} - \frac{\gamma_n}{\left[\varepsilon + \mathbb{M}_n^{(i)}\right]^{1/2}}\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G_i(\Theta_{n-1},X_{n,j})\right]. \tag{7.25}$$

Hinton et al. [3] suggests the choice that for all $n \in \mathbb{N}$ it holds that

$$\beta_n = 0.9 \tag{7.26}$$

as default values for the second moment decay factors $(\beta_n)_{n\in\mathbb{N}} \subseteq [0,1]$ in Definition 7.5.1. This default value in used several machine learning libraries that implement RMSprop (see, e.g., Tensorflow [12] and Lasagne [6]).

7.6 The Adadelta stochastic gradient descent optimization method

The Adadelta SGD optimization method was proposed in Zeiler [13]. It is a extension of RMSprop SGD optimization method. Like the RMSprop SGD optimization method, the Adadelta SGD optimization method adapts the learning rate for every component separately. To do this, the Adadelta SGD optimization method uses two exponentially decaying averages: one over the squares of the past partial derivatives and another one over the squares of the past increments (cf. Definition 7.6.1 below).

Definition 7.6.1 (Adadelta stochastic gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(J_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $(\beta_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $(\delta_n)_{n\in\mathbb{N}} \subseteq [0,1]$, $\varepsilon \in (0,\infty)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $(S,\mathcal{S})$ be a measurable space, let $\xi\colon \Omega \to \mathbb{R}^d$ and $X_{n,j}\colon \Omega \to S$, $j \in \{1,2,\dots,J_n\}$, $n \in \mathbb{N}$, be random variables, and let $F = (F(\theta,x))_{(\theta,x)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}$ and $G = (G_1,G_2,\dots,G_d)\colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be functions which satisfy for all $x \in S$, $\theta \in \{v \in \mathbb{R}^d \colon F(\cdot,x)$ is differentiable at $v\}$ that*

$$G(\theta,x) = (\nabla_\theta F)(\theta,x). \tag{7.27}$$

Then we say that Θ is the Adadelta stochastic gradient descent process on $((\Omega,\mathcal{F},\mathbb{P}),(S,\mathcal{S}))$ for the loss function F with generalized gradient G, batch sizes $(J_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, delta decay factors $(\delta_n)_{n\in\mathbb{N}}$, regularizing factor ε, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$ (we say that Θ is the Adadelta stochastic gradient descent process for the loss function F with batch sizes $(J_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, delta decay factors $(\delta_n)_{n\in\mathbb{N}}$, regularizing factor ε, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$) if and only if it holds that $\Theta = (\Theta^{(1)},\Theta^{(2)},\dots,\Theta^{(d)})\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ is the function from $\mathbb{N}_0 \times \Omega$ to $\mathbb{R}^d$ which satisfies that there exist $\mathbb{M} = (\mathbb{M}^{(1)},\mathbb{M}^{(2)},\dots,\mathbb{M}^{(d)})\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$, $\Delta = (\Delta^{(1)},\Delta^{(2)},\dots,\Delta^{(d)})\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$, $i \in \{1,2,\dots,d\}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbb{M}_0 = 0, \qquad \Delta_0 = 0, \tag{7.28}$$

$$\mathbb{M}_n^{(i)} = \beta_n\,\mathbb{M}_{n-1}^{(i)} + (1-\beta_n)\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G_i(\Theta_{n-1},X_{n,j})\right]^2, \tag{7.29}$$

$$\Theta_n^{(i)} = \Theta_{n-1}^{(i)} - \left(\frac{\varepsilon + \Delta_{n-1}^{(i)}}{\varepsilon + \mathbb{M}_n^{(i)}}\right)^{1/2}\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G_i(\Theta_{n-1},X_{n,j})\right], \tag{7.30}$$

$$\textit{and} \qquad \Delta_n^{(i)} = \delta_n\Delta_{n-1}^{(i)} + (1-\delta_n)\big|\Theta_n^{(i)} - \Theta_{n-1}^{(i)}\big|^2. \tag{7.31}$$

7.7 The adaptive moment estimation stochastic gradient descent optimization method (Adam stochastic gradient descent optimization method)

Definition 7.7.1 (Adam stochastic gradient descent optimization method)**.** *Let $d \in \mathbb{N}$, $(\gamma_n)_{n\in\mathbb{N}} \subseteq [0,\infty)$, $(J_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $(\alpha_n)_{n\in\mathbb{N}} \subseteq [0,1)$, $(\beta_n)_{n\in\mathbb{N}} \subseteq [0,1)$, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $(S,\mathcal{S})$ be a measurable space, let $\xi\colon \Omega \to \mathbb{R}^d$ and $X_{n,j}\colon \Omega \to S$, $j \in \{1,2,\dots,J_n\}$, $n \in \mathbb{N}$, be random variables, and let $F = (F(\theta,x))_{(\theta,x)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d\times S \to \mathbb{R}$ and $G = (G_1,G_2,\dots,G_d)\colon \mathbb{R}^d\times S \to \mathbb{R}^d$ be functions which satisfy for all $x \in S$, $\theta \in \{v \in \mathbb{R}^d\colon F(\cdot,x)$ is differentiable at $v\}$ that*

$$G(\theta,x) = (\nabla_\theta F)(\theta,x). \tag{7.32}$$

Then we say that Θ is the Adam stochastic gradient descent process on $((\Omega,\mathcal{F},\mathbb{P}),(S,\mathcal{S}))$ for the loss function F with generalized gradient G, learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$ (we say that Θ is the Adam stochastic gradient descent process for the loss function F with learning rates $(\gamma_n)_{n\in\mathbb{N}}$, batch sizes $(J_n)_{n\in\mathbb{N}}$, momentum decay factors $(\alpha_n)_{n\in\mathbb{N}}$, second moment decay factors $(\beta_n)_{n\in\mathbb{N}}$, initial value ξ, and data $(X_{n,j})_{(n,j)\in\{(k,l)\in\mathbb{N}^2\colon l\leq J_k\}}$) if and only if it holds that $\Theta = (\Theta^{(1)},\Theta^{(2)},\dots,\Theta^{(d)})\colon \mathbb{N}_0\times\Omega \to \mathbb{R}^d$ is the function from $\mathbb{N}_0\times\Omega$ to $\mathbb{R}^d$ which satisfies that there exist $\mathbf{m} = (\mathbf{m}^{(1)},\mathbf{m}^{(2)},\dots,\mathbf{m}^{(d)})\colon \mathbb{N}_0\times\Omega \to \mathbb{R}^d$, $\mathbb{M} = (\mathbb{M}^{(1)},\mathbb{M}^{(2)},\dots,\mathbb{M}^{(d)})\colon \mathbb{N}_0\times\Omega \to \mathbb{R}^d$ such that for all $n \in \mathbb{N}$, $i \in \{1,2,\dots,d\}$ it holds that

$$\Theta_0 = \xi, \qquad \mathbf{m}_0 = 0, \qquad \mathbb{M}_0 = 0, \tag{7.33}$$

$$\mathbf{m}_n = \alpha_n\,\mathbf{m}_{n-1} + (1-\alpha_n)\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G(\Theta_{n-1},X_{n,j})\right], \tag{7.34}$$

$$\mathbb{M}_n^{(i)} = \beta_n\,\mathbb{M}_{n-1}^{(i)} + (1-\beta_n)\left[\frac{1}{J_n}\sum_{j=1}^{J_n} G_i(\Theta_{n-1},X_{n,j})\right]^2, \tag{7.35}$$

$$\text{and} \qquad \Theta_n^{(i)} = \Theta_{n-1}^{(i)} - \gamma_n\left[\varepsilon + \left[\tfrac{\mathbb{M}_n^{(i)}}{(1-\prod_{l=1}^n \beta_l)}\right]^{1/2}\right]^{-1}\left[\frac{\mathbf{m}_n^{(i)}}{(1-\prod_{l=1}^n \alpha_l)}\right]. \tag{7.36}$$

Kingma & Ba [5] suggests the choice that for all $n \in \mathbb{N}$ it holds that that

$$\gamma_n = 0.001, \qquad \alpha_n = 0.9, \qquad \beta_n = 0.999, \qquad \text{and} \qquad \varepsilon = 10^{-8} \tag{7.37}$$

as default values for $(\gamma_n)_{n\in\mathbb{N}}$, $(\alpha_n)_{n\in\mathbb{N}}$, $(\beta_n)_{n\in\mathbb{N}}$, and ε in Definition 7.7.1.

Chapter 8

Backpropagation

8.1 Backpropagation for parametric functions

Proposition 8.1.1 (Backpropagation for parametric functions)**.** *Let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L, n_1, n_2, \ldots, n_L \in \mathbb{N}$, for every $k \in \{1, 2, \ldots, L\}$ let $F_k = (F_k(\theta_k, x_{k-1}))_{(\theta_k, x_{k-1}) \in \mathbb{R}^{n_k} \times \mathbb{R}^{l_{k-1}}} \colon \mathbb{R}^{n_k} \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_k}$ be differentiable, for every $k \in \{1, 2, \ldots, L\}$ let $J_k = (J_k(\theta_k, \theta_{k+1}, \ldots, \theta_L, x_{k-1}))_{(\theta_k, \theta_{k+1}, \ldots, \theta_L, x_{k-1}) \in \mathbb{R}^{n_k} \times \mathbb{R}^{n_{k+1}} \times \ldots \times \mathbb{R}^{n_L} \times \mathbb{R}^{l_{k-1}}} \colon \mathbb{R}^{n_k} \times \mathbb{R}^{n_{k+1}} \times \ldots \times \mathbb{R}^{n_L} \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_L}$ satisfy for all $\theta = (\theta_k, \theta_{k+1}, \ldots, \theta_L) \in \mathbb{R}^{n_k} \times \mathbb{R}^{n_{k+1}} \times \ldots \times \mathbb{R}^{n_L}$, $x_{k-1} \in \mathbb{R}^{l_{k-1}}$ that*

$$J_k(\theta, x_{k-1}) = \big(F_L(\theta_L, \cdot) \circ F_{L-1}(\theta_{L-1}, \cdot) \circ \ldots \circ F_k(\theta_k, \cdot)\big)(x_{k-1}), \tag{8.1}$$

let $\vartheta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_L) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L}$, $\mathfrak{x}_0 \in \mathbb{R}^{l_0}$, $\mathfrak{x}_1 \in \mathbb{R}^{l_1}$, $\ldots$, $\mathfrak{x}_L \in \mathbb{R}^{l_L}$ satisfy for all $k \in \{1, 2, \ldots, L\}$ that

$$\mathfrak{x}_k = F_k(\vartheta_k, \mathfrak{x}_{k-1}), \tag{8.2}$$

and let $D_k \in \mathbb{R}^{l_L \times l_{k-1}}$, $k \in \{1, 2, \ldots, L+1\}$, satisfy for all $k \in \{1, 2, \ldots, L\}$ that $D_{L+1} = \mathrm{I}_{l_L}$ and

$$D_k = D_{k+1}\left[\left(\frac{\partial F_k}{\partial x_{k-1}}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right] \tag{8.3}$$

(cf. Definition 2.2.9). Then

(i) it holds for all $k \in \{1, 2, \ldots, L\}$ that $J_k \colon \mathbb{R}^{n_k} \times \mathbb{R}^{n_{k+1}} \times \ldots \times \mathbb{R}^{n_L} \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_L}$ is differentiable,

(ii) it holds for all $k \in \{1, 2, \ldots, L\}$ that

$$D_k = \left(\frac{\partial J_k}{\partial x_{k-1}}\right)((\vartheta_k, \vartheta_{k+1}, \ldots, \vartheta_L), \mathfrak{x}_{k-1}), \tag{8.4}$$

and

(iii) it holds for all $k \in \{1, 2, \ldots, L\}$ that

$$\left(\frac{\partial J_1}{\partial \theta_k}\right)(\vartheta, \mathfrak{x}_0) = D_{k+1}\left[\left(\frac{\partial F_k}{\partial \theta_k}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right]. \tag{8.5}$$

Proof of Proposition 8.1.1. Note that (8.1) and the assumption that for all $k \in \{1, 2, \ldots, L\}$ it holds that $F_k \colon \mathbb{R}^{n_k} \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_k}$ is differentiable imply that for all $k \in \{1, 2, \ldots, L\}$ it holds that

$$J_k \colon \mathbb{R}^{n_k} \times \mathbb{R}^{n_{k+1}} \times \ldots \times \mathbb{R}^{n_L} \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_L} \tag{8.6}$$

is differentiable. This proves item (i). Next we prove (8.4) by induction on $k \in \{L, L-1, \ldots, 1\}$. Note that (8.1), (8.3), the assumption that $D_{L+1} = \mathrm{I}_{l_L}$, and the fact that $J_L = F_L$ assure that

$$D_L = D_{L+1}\left[\left(\frac{\partial F_L}{\partial x_{L-1}}\right)(\vartheta_L, \mathfrak{x}_{L-1})\right] = \left(\frac{\partial J_L}{\partial x_{L-1}}\right)(\vartheta_L, \mathfrak{x}_{L-1}). \tag{8.7}$$

This establishes (8.4) in the base case $k = L$. For the induction step note that the fact that for all $x_{k-1} \in \mathbb{R}^{l_{k-1}}$ it holds that $J_k((\vartheta_k, \vartheta_{k+1}, \ldots, \vartheta_L), x_{k-1}) = J_{k+1}((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), F_k(\vartheta_k, x_{k-1}))$ and the chain rule imply that for all $k \in \{L-1, L-2, \ldots, 1\}$ with $D_{k+1} = \big(\frac{\partial J_{k+1}}{\partial x_k}\big)((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), \mathfrak{x}_k)$ it holds that

$$\begin{aligned}
&\left(\frac{\partial J_k}{\partial x_{k-1}}\right)((\vartheta_k, \vartheta_{k+1}, \ldots, \vartheta_L), \mathfrak{x}_{k-1}) \\
&= \left(\frac{\mathrm{d}\,(\mathbb{R}^{l_{k-1}} \ni x_{k-1} \mapsto J_k((\vartheta_k, \vartheta_{k+1}, \ldots, \vartheta_L), x_{k-1}) \in \mathbb{R}^{l_L})}{\mathrm{d}x_{k-1}}\right)(\mathfrak{x}_{k-1}) \\
&= \left(\frac{\mathrm{d}\,(\mathbb{R}^{l_{k-1}} \ni x_{k-1} \mapsto J_{k+1}((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), F_k(\vartheta_k, x_{k-1})) \in \mathbb{R}^{l_L})}{\mathrm{d}x_{k-1}}\right)(\mathfrak{x}_{k-1}) \\
&= \left[\left(\frac{\mathrm{d}\,(\mathbb{R}^{l_{k-1}} \ni x_k \mapsto J_{k+1}((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), x_k)) \in \mathbb{R}^{l_L})}{\mathrm{d}x_k}\right)(F_k(\vartheta_k, \mathfrak{x}_{k-1}))\right] \\
&\quad \left[\left(\frac{\mathrm{d}\,(\mathbb{R}^{l_{k-1}} \ni x_{k-1} \mapsto F_k(\vartheta_k, x_{k-1})) \in \mathbb{R}^{l_k})}{\mathrm{d}x_{k-1}}\right)(\mathfrak{x}_{k-1})\right] \\
&= \left[\left(\frac{\partial J_{k+1}}{\partial x_k}\right)((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), \mathfrak{x}_k)\right]\left[\left(\frac{\partial F_k}{\partial x_{k-1}}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right] \\
&= D_{k+1}\left[\left(\frac{\partial F_k}{\partial x_{k-1}}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right] = D_k.
\end{aligned} \tag{8.8}$$

Induction thus proves (8.4). This establishes item (ii). Moreover, observe that (8.1) and (8.2) assure that for all $k \in \{1, 2, \ldots, L\}$, $\theta_k \in \mathbb{R}^{l_k}$ it holds that

$$\begin{aligned}
&J_1((\vartheta_1, \ldots, \vartheta_{k-1}, \theta_k, \vartheta_{k+1}, \ldots, \vartheta_L), \mathfrak{x}_0) \\
&= \big(F_L(\vartheta_L, \cdot) \circ \ldots \circ F_{k+1}(\vartheta_{k+1}, \cdot) \circ F_k(\theta_k, \cdot) \circ F_{k-1}(\vartheta_{k-1}, \cdot) \circ \circ F_1(\vartheta_1, \cdot)\big)(\mathfrak{x}_0) \\
&= \big(J_{k+1}((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), F_k(\theta_k, \cdot))\big)\big((F_{k-1}(\vartheta_{k-1}, \cdot) \circ \circ F_1(\vartheta_1, \cdot))(\mathfrak{x}_0)\big) \\
&= J_{k+1}((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), F_k(\theta_k, \mathfrak{x}_{k-1})).
\end{aligned} \tag{8.9}$$

Combining this with the chain rule, (8.2), and (8.4) demonstrates that for all $k \in \{1, 2, \ldots, L\}$ it holds that

$$\begin{aligned}
\left(\frac{\partial J_1}{\partial \theta_k}\right)(\vartheta, \mathfrak{x}_0) &= \left(\frac{\mathrm{d}\left(\mathbb{R}^{n_k} \ni \theta_k \mapsto J_{k+1}((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), F_k(\theta_k, \mathfrak{x}_{k-1})) \in \mathbb{R}^{l_L}\right)}{\mathrm{d}\theta_k}\right)(\vartheta_k) \\
&= \left[\left(\frac{\mathrm{d}\left(\mathbb{R}^{l_k} \ni x_k \mapsto J_{k+1}((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), x_k) \in \mathbb{R}^{l_L}\right)}{\mathrm{d}x_k}\right)(F_k(\vartheta_k, \mathfrak{x}_{k-1}))\right] \\
&\quad \left[\left(\frac{\mathrm{d}\left(\mathbb{R}^{n_k} \ni \theta_k \mapsto F_k(\theta_k, \mathfrak{x}_{k-1}) \in \mathbb{R}^{l_k}\right)}{\mathrm{d}\theta_k}\right)(\vartheta_k)\right] \\
&= \left[\left(\frac{\partial J_{k+1}}{\partial x_k}\right)((\vartheta_{k+1}, \vartheta_{k+2}, \ldots, \vartheta_L), \mathfrak{x}_k)\right]\left[\left(\frac{\partial F_k}{\partial \theta_k}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right] \\
&= D_{k+1}\left[\left(\frac{\partial F_k}{\partial \theta_k}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right].
\end{aligned} \tag{8.10}$$

This establishes item (iii). The proof of Proposition 8.1.1 is thus complete. ☐

Corollary 8.1.2 (Backpropagation for parametric functions with loss). *Let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L$, $n_1, n_2, \ldots, n_L \in \mathbb{N}$, $\vartheta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_L) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L}$, $\mathfrak{x}_0 \in \mathbb{R}^{l_0}$, $\mathfrak{x}_1 \in \mathbb{R}^{l_1}$, $\ldots$, $\mathfrak{x}_L \in \mathbb{R}^{l_L}$, $\mathfrak{y} \in \mathbb{R}^{l_L}$, let $\mathfrak{C} = (\mathfrak{C}(x, y))_{(x,y) \in \mathbb{R}^{l_L} \times \mathbb{R}^{l_L}} \colon \mathbb{R}^{l_L} \times \mathbb{R}^{l_L} \to \mathbb{R}$ be differentiable, for every $k \in \{1, 2, \ldots, L\}$ let $F_k = (F_k(\theta_k, x_{k-1}))_{(\theta_k, x_{k-1}) \in \mathbb{R}^{n_k} \times \mathbb{R}^{l_{k-1}}} \colon \mathbb{R}^{n_k} \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_k}$ be differentiable, let $J = (J(\theta_1, \theta_2, \ldots, \theta_L))_{(\theta_1, \theta_2, \ldots, \theta_L) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L}} \colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L} \to \mathbb{R}$ satisfy for all $\theta = (\theta_1, \theta_2, \ldots, \theta_L) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L}$ that*

$$J(\theta) = \big(\mathfrak{C}(\cdot, \mathfrak{y}) \circ F_L(\theta_L, \cdot) \circ F_{L-1}(\theta_{L-1}, \cdot) \circ \ldots \circ F_1(\theta_1, \cdot)\big)(\mathfrak{x}_0), \tag{8.11}$$

assume for all $k \in \{1, 2, \ldots, L\}$ that

$$\mathfrak{x}_k = F_k(\vartheta_k, \mathfrak{x}_{k-1}), \tag{8.12}$$

and let $D_k \in \mathbb{R}^{l_{k-1}}$, $k \in \{1, 2, \ldots, L+1\}$, satisfy for all $k \in \{1, 2, \ldots, L\}$ that

$$D_{L+1} = (\nabla_x \mathfrak{C})(\mathfrak{x}_L, \mathfrak{y}) \qquad \text{and} \qquad D_k = \left[\left(\frac{\partial F_k}{\partial x_{k-1}}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right]^* D_{k+1}. \tag{8.13}$$

Then

(i) it holds that $J \colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L} \to \mathbb{R}$ is differentiable and

(ii) it holds for all $k \in \{1, 2, \ldots, L\}$ that

$$\left(\nabla_{\theta_k} J\right)(\vartheta) = \left[\left(\frac{\partial F_k}{\partial \theta_k}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right]^* D_{k+1}. \tag{8.14}$$

Proof of Corollary 8.1.2. Throughout this proof let $\mathbf{J} = (\mathbf{J}(\theta_1, \theta_2, \ldots, \theta_L))_{(\theta_1,\theta_2,\ldots,\theta_L)\in\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}\times\ldots\times\mathbb{R}^{n_L}} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L} \to \mathbb{R}^{l_L}$ satisfy for all $\theta = (\theta_1, \theta_2, \ldots, \theta_L) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L}$ that

$$\mathbf{J}(\theta) = \big(F_L(\theta_L, \cdot) \circ F_{L-1}(\theta_{L-1}, \cdot) \circ \ldots \circ F_1(\theta_1, \cdot)\big)(\mathfrak{x}_0) \tag{8.15}$$

and let $\mathbf{D}_k \in \mathbb{R}^{l_L \times l_{k-1}}$, $k \in \{1, 2, \ldots, L+1\}$, satisfy for all $k \in \{1, 2, \ldots, L\}$ that $\mathbf{D}_{L+1} = \mathrm{I}_{l_L}$ and

$$\mathbf{D}_k = \mathbf{D}_{k+1}\bigg[\bigg(\frac{\partial F_k}{\partial x_{k-1}}\bigg)(\vartheta_k, \mathfrak{x}_{k-1})\bigg] \tag{8.16}$$

(cf. Definition 2.2.9). Observe that the assumption that $\mathfrak{C} \colon \mathbb{R}^{l_L} \times \mathbb{R}^{l_L} \to \mathbb{R}$ is differentiable, the assumption that for all $k \in \{1, 2, \ldots, L\}$ it holds that $F_k \colon \mathbb{R}^{n_k} \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_k}$ is differentiable, and (8.11) ensure that $J \colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_L} \to \mathbb{R}$ is differentiable. This establishes item (i). Next we claim that for all $k \in \{1, 2, \ldots, L+1\}$ it holds that

$$[D_k]^* = \bigg[\bigg(\frac{\partial \mathfrak{C}}{\partial x}\bigg)(\mathfrak{x}_L, \mathfrak{y})\bigg]\mathbf{D}_k. \tag{8.17}$$

We now prove (8.17) by induction on $k \in \{L+1, L, \ldots, 1\}$. For the base case $k = L+1$ note that (8.13) and (8.16) assure that

$$\begin{aligned}[D_{L+1}]^* &= [(\nabla_x \mathfrak{C})(\mathfrak{x}_L, \mathfrak{y})]^* = \bigg(\frac{\partial \mathfrak{C}}{\partial x}\bigg)(\mathfrak{x}_L, \mathfrak{y}) \\ &= \bigg[\bigg(\frac{\partial \mathfrak{C}}{\partial x}\bigg)(\mathfrak{x}_L, \mathfrak{y})\bigg]\mathrm{I}_{l_L} = \bigg[\bigg(\frac{\partial \mathfrak{C}}{\partial x}\bigg)(\mathfrak{x}_L, \mathfrak{y})\bigg]\mathbf{D}_{L+1}.\end{aligned} \tag{8.18}$$

This establishes (8.17) in the base case $k = L+1$. For the induction step observe (8.13) and (8.16) demonstrate that for all $k \in \{L, L-1, \ldots, 1\}$ with $[D_{k+1}]^* = \big[\big(\frac{\partial \mathfrak{C}}{\partial x}\big)(\mathfrak{x}_L, \mathfrak{y})\big]\mathbf{D}_{k+1}$ it holds that

$$\begin{aligned}[D_k]^* &= [D_{k+1}]^*\bigg[\bigg(\frac{\partial F_k}{\partial x_{k-1}}\bigg)(\vartheta_k, \mathfrak{x}_{k-1})\bigg] \\ &= \bigg[\bigg(\frac{\partial \mathfrak{C}}{\partial x}\bigg)(\mathfrak{x}_L, \mathfrak{y})\bigg]\mathbf{D}_{k+1}\bigg[\bigg(\frac{\partial F_k}{\partial x_{k-1}}\bigg)(\vartheta_k, \mathfrak{x}_{k-1})\bigg] = \bigg[\bigg(\frac{\partial \mathfrak{C}}{\partial x}\bigg)(\mathfrak{x}_L, \mathfrak{y})\bigg]\mathbf{D}_k.\end{aligned} \tag{8.19}$$

Induction thus establishes (8.17). Furthermore, note that item (iii) in Proposition 8.1.1 assures that for all $k \in \{1, 2, \ldots, L\}$ it holds that

$$\bigg(\frac{\partial \mathbf{J}}{\partial \theta_k}\bigg)(\vartheta) = \mathbf{D}_{k+1}\bigg[\bigg(\frac{\partial F_k}{\partial \theta_k}\bigg)(\vartheta_k, \mathfrak{x}_{k-1})\bigg]. \tag{8.20}$$

Combining this with chain rule, the fact that $J = \mathfrak{C}(\cdot, \mathfrak{y}) \circ \mathbf{J}$, and (8.17) ensures that for all $k \in \{1, 2, \dots, L\}$ it holds that

$$\begin{aligned}\left(\frac{\partial J}{\partial \theta_k}\right)(\vartheta) &= \left[\left(\frac{\partial \mathfrak{C}}{\partial x}\right)(\mathbf{J}(\vartheta), \mathfrak{y})\right]\left[\left(\frac{\partial \mathbf{J}}{\partial \theta_k}\right)(\vartheta)\right] \\ &= \left[\left(\frac{\partial \mathfrak{C}}{\partial x}\right)(\mathfrak{x}_0, \mathfrak{y})\right]\mathbf{D}_{k+1}\left[\left(\frac{\partial F_k}{\partial \theta_k}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right] \\ &= [D_{k+1}]^*\left[\left(\frac{\partial F_k}{\partial \theta_k}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right].\end{aligned} \tag{8.21}$$

Hence we obtain that for all $k \in \{1, 2, \dots, L\}$ it holds that

$$(\nabla_{\theta_k} J)(\vartheta) = \left[\left(\frac{\partial J}{\partial \theta_k}\right)(\vartheta)\right]^* = \left[\left(\frac{\partial F_k}{\partial \theta_k}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right]^* D_{k+1}. \tag{8.22}$$

This establishes item (ii). The proof of Corollary 8.1.2 is thus complete. □

8.2 Backpropagation for ANNs

Definition 8.2.1. *We denote by* $\operatorname{diag}\colon (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to (\cup_{d\in\mathbb{N}}\mathbb{R}^{d\times d})$ *the function which satisfies for all* $d \in \mathbb{N}$, $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ *that*

$$\operatorname{diag}(x) = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_d \end{pmatrix} \in \mathbb{R}^{d\times d}. \tag{8.23}$$

Corollary 8.2.2 (Backpropagation for ANNs)**.** *Let* $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((\mathbf{W}_1, \mathbf{B}_1), \dots, (\mathbf{W}_L, \mathbf{B}_L)) \in \bigtimes_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})$, *let* $\mathfrak{C} = (\mathfrak{C}(x, y))_{(x,y)\in\mathbb{R}^{l_L}\times\mathbb{R}^{l_L}} \colon \mathbb{R}^{l_L} \times \mathbb{R}^{l_L} \to \mathbb{R}$ *and* $a\colon \mathbb{R} \to \mathbb{R}$ *be differentiable, let* $\mathfrak{x}_0 \in \mathbb{R}^{l_0}$, $\mathfrak{x}_1 \in \mathbb{R}^{l_1}$, $\dots$, $\mathfrak{x}_L \in \mathbb{R}^{l_L}$, $\mathfrak{y} \in \mathbb{R}^{l_L}$ *satisfy for all* $k \in \{1, 2, \dots, L\}$ *that*

$$\mathfrak{x}_k = \mathfrak{M}_{a\mathbb{1}_{[0,L)}(k)+\mathrm{id}_{\mathbb{R}}\,\mathbb{1}_{\{L\}}(k), l_k}(\mathbf{W}_k \mathfrak{x}_{k-1} + \mathbf{B}_k), \tag{8.24}$$

let $J = \big(J((W_1, B_1), \dots, (W_L, B_L))\big)_{((W_1,B_1),\dots,(W_L,B_L))\in\bigtimes_{k=1}^L(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k})} \colon \bigtimes_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_k}) \to \mathbb{R}$ *satisfy for all* $\Psi \in \bigtimes_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})$ *that*

$$J(\Psi) = \mathfrak{C}((\mathcal{R}_a(\Psi))(\mathfrak{x}_0), \mathfrak{y}), \tag{8.25}$$

and let $D_k \in \mathbb{R}^{l_{k-1}}$, $k \in \{1, 2, \dots, L+1\}$, *satisfy for all* $k \in \{1, 2, \dots, L-1\}$ *that*

$$D_{L+1} = (\nabla_x \mathfrak{C})(\mathfrak{x}_L, \mathfrak{y}), \qquad D_L = [\mathbf{W}_L]^* D_{L+1}, \qquad \textit{and} \tag{8.26}$$

$$D_k = [\mathbf{W}_k]^*[\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k\mathfrak{x}_{k-1} + \mathbf{B}_k))]D_{k+1} \tag{8.27}$$

(cf. Definitions 2.1.4, 2.2.3, and 8.2.1). Then

(i) it holds that $J \colon \bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \to \mathbb{R}$ *is differentiable,*

(ii) it holds that $(\nabla_{B_L} J)(\Phi) = D_{L+1}$*,*

(iii) it holds for all $k \in \{1, 2, \ldots, L-1\}$ *that*

$$(\nabla_{B_k} J)(\Phi) = [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k\mathfrak{x}_{k-1} + \mathbf{B}_k))]D_{k+1}, \tag{8.28}$$

(iv) it holds that $(\nabla_{W_L} J)(\Phi) = D_{L+1}[\mathfrak{x}_{L-1}]^*$*, and*

(v) it holds for all $k \in \{1, 2, \ldots, L-1\}$ *that*

$$(\nabla_{W_k} J)(\Phi) = [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k\mathfrak{x}_{k-1} + \mathbf{B}_k))]D_{k+1}[\mathfrak{x}_{k-1}]^*. \tag{8.29}$$

Proof of Corollary 8.2.2. Throughout this proof for every $k \in \{1, 2, \ldots, L\}$ let

$$\begin{aligned} F_k &= (F_k^{(m)})_{m \in \{1,2,\ldots,l_k\}} \\ &= \big(F_k\big(((W_{k,i,j})_{(i,j)\in\{1,2,\ldots,l_k\}\times\{1,2,\ldots,l_{k-1}\}}, B_k), \\ &\quad x_{k-1})\big)_{(((W_{k,i,j})_{(i,j)\in\{1,2,\ldots,l_k\}\times\{1,2,\ldots,l_{k-1}\}},B_k),x_{k-1})\in(\mathbb{R}^{l_k\times l_{k-1}}\times\mathbb{R}^{l_{k-1}})\times\mathbb{R}^{l_{k-1}}} \\ &: (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_{k-1}}) \times \mathbb{R}^{l_{k-1}} \to \mathbb{R}^{l_k} \end{aligned} \tag{8.30}$$

satisfy for all $(W_k, B_k) \in \mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_{k-1}}$, $x_{k-1} \in \mathbb{R}^{l_{k-1}}$ that

$$F_k((W_k, B_k), x_{k-1}) = \mathfrak{M}_{a \mathbb{1}_{[0,L)}(k) + \mathrm{id}_{\mathbb{R}} \mathbb{1}_{\{L\}}(k), l_k}(W_k x_{k-1} + B_k) \tag{8.31}$$

and for every $d \in \mathbb{N}$ let $\mathbf{e}_1^{(d)}, \mathbf{e}_2^{(d)}, \ldots, \mathbf{e}_d^{(d)} \in \mathbb{R}^d$ satisfy $\mathbf{e}_1^{(d)} = (1, 0, \ldots, 0)$, $\mathbf{e}_2^{(d)} = (0, 1, 0, \ldots, 0)$, $\ldots$, $\mathbf{e}_d^{(d)} = (0, \ldots, 0, 1)$. Observe that the assumption that a is differentiable and (8.24) imply that $J \colon \bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \to \mathbb{R}$ is differentiable. This establishes item (i). Next note that (2.53), (8.25), and (8.31) ensure that for all $\Psi = ((W_1, B_1), \ldots, (W_L, B_L)) \in \bigtimes_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})$ it holds that

$$J(\Psi) = \big(\mathfrak{C}(\cdot, \mathfrak{y}) \circ F_L((W_L, B_L), \cdot) \circ F_{L-1}((W_{L-1}, B_{L-1}), \cdot) \circ \ldots \circ F_1((W_1, B_1), \cdot)\big)(\mathfrak{x}_0). \tag{8.32}$$

Moreover, observe that (8.24) and (8.31) imply that for all $k \in \{1, 2, \ldots, L\}$ it holds that

$$\mathfrak{x}_k = F_k((\mathbf{W}_k, \mathbf{B}_k), \mathfrak{x}_{k-1}). \tag{8.33}$$

In addition, observe that (8.31) assures that

$$\left(\frac{\partial F_L}{\partial x_{L-1}}\right)((\mathbf{W}_L, \mathbf{B}_L), \mathfrak{x}_{L-1}) = \mathbf{W}_L. \tag{8.34}$$

Moreover, note that (8.31) implies that for all $k \in \{1, 2, \dots, L-1\}$ it holds that

$$\left(\frac{\partial F_k}{\partial x_{k-1}}\right)((\mathbf{W}_k, \mathbf{B}_k), \mathfrak{x}_{k-1}) = [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k \mathfrak{x}_{k-1} + \mathbf{B}_k))]\mathbf{W}_k. \tag{8.35}$$

Combining this and (8.34) with (8.26) and (8.27) demonstrates that for all $k \in \{1, 2, \dots, L\}$ it holds that

$$D_{L+1} = (\nabla_x \mathfrak{C})(\mathfrak{x}_L, \mathfrak{y}) \qquad \text{and} \qquad D_k = \left[\left(\frac{\partial F_k}{\partial x_{k-1}}\right)(\vartheta_k, \mathfrak{x}_{k-1})\right]^* D_{k+1}. \tag{8.36}$$

Next note that this, (8.32), (8.33), and Corollary 8.1.2 prove that for all $k \in \{1, 2, \dots, L\}$ it holds that

$$(\nabla_{B_k} J)(\Phi) = \left[\left(\frac{\partial F_k}{\partial B_k}\right)((\mathbf{W}_k, \mathbf{B}_k), \mathfrak{x}_{k-1})\right]^* D_{k+1} \qquad \text{and} \tag{8.37}$$

$$(\nabla_{W_k} J)(\Phi) = \left[\left(\frac{\partial F_k}{\partial W_k}\right)((\mathbf{W}_k, \mathbf{B}_k), \mathfrak{x}_{k-1})\right]^* D_{k+1}. \tag{8.38}$$

Moreover, observe that (8.31) implies that

$$\left(\frac{\partial F_L}{\partial B_L}\right)((\mathbf{W}_L, \mathbf{B}_L), \mathfrak{x}_{L-1}) = \mathrm{I}_{l_L} \tag{8.39}$$

(cf. Definition 2.2.9). Combining this with (8.37) demonstrates that

$$(\nabla_{B_L} J)(\Phi) = [\mathrm{I}_{l_L}]^* D_{L+1} = D_{L+1}. \tag{8.40}$$

This establishes item (ii). Furthermore, note that (8.31) assures that for all $k \in \{1, 2, \dots, L-1\}$ it holds that

$$\left(\frac{\partial F_k}{\partial B_k}\right)((\mathbf{W}_k, \mathbf{B}_k), \mathfrak{x}_{k-1}) = \operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k \mathfrak{x}_{k-1} + \mathbf{B}_k)). \tag{8.41}$$

Combining this with (8.37) implies that for all $k \in \{1, 2, \dots, L-1\}$ it holds that

$$\begin{aligned} (\nabla_{B_k} J)(\Phi) &= [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k \mathfrak{x}_{k-1} + \mathbf{B}_k))]^* D_{k+1} \\ &= [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k \mathfrak{x}_{k-1} + \mathbf{B}_k))] D_{k+1}. \end{aligned} \tag{8.42}$$

This establishes item (iii). In addition, observe that (8.31) ensures that for all $m, i \in \{1, 2, \ldots, l_L\}$, $j \in \{1, 2, \ldots, l_{L-1}\}$ it holds that

$$\left(\frac{\partial F_L^{(m)}}{\partial W_{L,i,j}}\right)((\mathbf{W}_L, \mathbf{B}_L), \mathfrak{x}_{L-1}) = \mathbb{1}_{\{m\}}(i)\langle \mathfrak{x}_{L-1}, \mathbf{e}_j^{(l_{L-1})}\rangle \tag{8.43}$$

(cf. Definition 4.1.5). Combining this with (8.38) demonstrates that

$$\begin{aligned}
&(\nabla_{W_L} J)(\Phi) \\
&= \left(\sum_{m=1}^{l_L}\left[\left(\frac{\partial F_L^{(m)}}{\partial W_{L,i,j}}\right)((\mathbf{W}_L, \mathbf{B}_L), \mathfrak{x}_{L-1})\right]\langle D_{L+1}, \mathbf{e}_m^{(l_L)}\rangle\right)_{(i,j)\in\{1,2,\ldots,l_L\}\times\{1,2,\ldots,l_{L-1}\}} \\
&= \left(\textstyle\sum_{m=1}^{l_L} \mathbb{1}_{\{m\}}(i)\langle \mathbf{e}_j^{(l_{L-1})}, \mathfrak{x}_{L-1}\rangle\langle \mathbf{e}_m^{(l_L)}, D_{L+1}\rangle\right)_{(i,j)\in\{1,2,\ldots,l_L\}\times\{1,2,\ldots,l_{L-1}\}} \\
&= \left(\langle \mathbf{e}_j^{(l_{L-1})}, \mathfrak{x}_{L-1}\rangle\langle \mathbf{e}_i^{(l_L)}, D_{L+1}\rangle\right)_{(i,j)\in\{1,2,\ldots,l_L\}\times\{1,2,\ldots,l_{L-1}\}} \\
&= D_{L+1}[\mathfrak{x}_{L-1}]^*.
\end{aligned} \tag{8.44}$$

This establishes item (iv). Moreover, note that (8.31) implies that for all $k \in \{1, 2, \ldots, L-1\}$, $m, i \in \{1, 2, \ldots, l_k\}$, $j \in \{1, 2, \ldots, l_{k-1}\}$ it holds that

$$\left(\frac{\partial F_k^{(m)}}{\partial W_{k,i,j}}\right)((\mathbf{W}_k, \mathbf{B}_k), \mathfrak{x}_{k-1}) = \mathbb{1}_{\{m\}}(i)a'(\langle \mathbf{e}_i^{(l_k)}, \mathbf{W}_k\mathfrak{x}_{k-1} + \mathbf{B}_k\rangle)\langle \mathbf{e}_j^{(l_{k-1})}, \mathfrak{x}_{k-1}\rangle. \tag{8.45}$$

Combining this with (8.38) demonstrates that for all $k \in \{1, 2, \ldots, L-1\}$ it holds that

$$\begin{aligned}
&(\nabla_{W_k} J)(\Phi) \\
&= \left(\sum_{m=1}^{l_k}\left[\left(\frac{\partial F_k^{(m)}}{\partial W_{k,i,j}}\right)((\mathbf{W}_k, \mathbf{B}_k), \mathfrak{x}_{k-1})\right]\langle \mathbf{e}_m^{(l_k)}, D_{k+1}\rangle\right)_{(i,j)\in\{1,2,\ldots,l_k\}\times\{1,2,\ldots,l_{k-1}\}} \\
&= \left(\textstyle\sum_{m=1}^{l_k} \mathbb{1}_{\{m\}}(i)a'(\langle \mathbf{e}_i^{(l_k)}, \mathbf{W}_k\mathfrak{x}_{k-1} + \mathbf{B}_k\rangle)\langle \mathbf{e}_j^{(l_{k-1})}, \mathfrak{x}_{k-1}\rangle\langle \mathbf{e}_m^{(l_k)}, D_{k+1}\rangle\right)_{(i,j)\in\{1,2,\ldots,l_k\}\times\{1,2,\ldots,l_{k-1}\}} \\
&= \left(a'(\langle \mathbf{e}_i^{(l_k)}, \mathbf{W}_k\mathfrak{x}_{k-1} + \mathbf{B}_k\rangle)\langle \mathbf{e}_j^{(l_{k-1})}, \mathfrak{x}_{k-1}\rangle\langle \mathbf{e}_i^{(l_k)}, D_{k+1}\rangle\right)_{(i,j)\in\{1,2,\ldots,l_k\}\times\{1,2,\ldots,l_{k-1}\}} \\
&= [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k\mathfrak{x}_{k-1} + \mathbf{B}_k))]D_{k+1}[\mathfrak{x}_{k-1}]^*.
\end{aligned} \tag{8.46}$$

This establishes item (v). The proof of Corollary 8.2.2 is thus complete. □

Corollary 8.2.3 (Backpropagation for ANNs with minibatches)**.** *Let* $L, M \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((\mathbf{W}_1, \mathbf{B}_1), \ldots, (\mathbf{W}_L, \mathbf{B}_L)) \in \bigtimes_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})$, *let* $\mathfrak{C} = (\mathfrak{C}(x, y))_{(x,y)\in\mathbb{R}^{l_L}\times\mathbb{R}^{l_L}} \colon \mathbb{R}^{l_L} \times$

$\mathbb{R}^{l_L} \to \mathbb{R}$ and $a\colon \mathbb{R} \to \mathbb{R}$ be differentiable, for every $m \in \{1, 2, \ldots, M\}$ let $\mathfrak{x}_0^{(m)} \in \mathbb{R}^{l_0}$, $\mathfrak{x}_1^{(m)} \in \mathbb{R}^{l_1}$, $\ldots$, $\mathfrak{x}_L^{(m)} \in \mathbb{R}^{l_L}$, $\mathfrak{y}^{(m)} \in \mathbb{R}^{l_L}$ satisfy for all $k \in \{1, 2, \ldots, L\}$ that

$$\mathfrak{x}_k^{(m)} = \mathfrak{M}_{a \mathbb{1}_{[0,L)}(k) + \mathrm{id}_{\mathbb{R}} \mathbb{1}_{\{L\}}(k), l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k), \tag{8.47}$$

let $J = \big(J((W_1, B_1), \ldots, (W_L, B_L))\big)_{((W_1,B_1),\ldots,(W_L,B_L)) \in \times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})} \colon \times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \to \mathbb{R}$ satisfy for all $\Psi \in \times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})$ that

$$J(\Psi) = \frac{1}{M} \left[\sum_{m=1}^{M} \mathfrak{C}((\mathcal{R}_a(\Psi))(\mathfrak{x}_0^{(m)}), \mathfrak{y}^{(m)}) \right], \tag{8.48}$$

and for every $m \in \{1, 2, \ldots, M\}$ let $D_k^{(m)} \in \mathbb{R}^{l_{k-1}}$, $k \in \{1, 2, \ldots, L+1\}$, satisfy for all $k \in \{1, 2, \ldots, L-1\}$ that

$$D_{L+1}^{(m)} = (\nabla_x \mathfrak{C})(\mathfrak{x}_L^{(m)}, \mathfrak{y}^{(m)}), \qquad D_L^{(m)} = [\mathbf{W}_L]^* D_{L+1}^{(m)}, \qquad \textit{and} \tag{8.49}$$

$$D_k^{(m)} = [\mathbf{W}_k]^* [\mathrm{diag}(\mathfrak{M}_{a', l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k))] D_{k+1}^{(m)} \tag{8.50}$$

(cf. Definitions 2.1.4, 2.2.3, and 8.2.1). Then

(i) it holds that $J\colon \times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \to \mathbb{R}$ is differentiable,

(ii) it holds that $(\nabla_{B_L} J)(\Phi) = \frac{1}{M} \big[\sum_{m=1}^{M} D_{L+1}^{(m)} \big]$,

(iii) it holds for all $k \in \{1, 2, \ldots, L-1\}$ that

$$(\nabla_{B_k} J)(\Phi) = \frac{1}{M} \left[\sum_{m=1}^{M} [\mathrm{diag}(\mathfrak{M}_{a', l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k))] D_{k+1}^{(m)} \right], \tag{8.51}$$

(iv) it holds that $(\nabla_{W_L} J)(\Phi) = \frac{1}{M} \big[\sum_{m=1}^{M} D_{L+1}^{(m)} [\mathfrak{x}_{L-1}^{(m)}]^* \big]$, and

(v) it holds for all $k \in \{1, 2, \ldots, L-1\}$ that

$$(\nabla_{W_k} J)(\Phi) = \frac{1}{M} \left[\sum_{m=1}^{M} [\mathrm{diag}(\mathfrak{M}_{a', l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k))] D_{k+1}^{(m)} [\mathfrak{x}_{k-1}^{(m)}]^* \right]. \tag{8.52}$$

Proof of Corollary 8.2.3. Throughout this proof let $\mathcal{J}^{(m)}\colon \times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \to \mathbb{R}$, $m \in \{1, 2, \ldots, M\}$, satisfy for all $m \in \{1, 2, \ldots, M\}$, $\Psi \in \times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})$ that

$$\mathcal{J}^{(m)}(\Psi) = \mathfrak{C}((\mathcal{R}_a(\Psi))(\mathfrak{x}_0^{(m)}), \mathfrak{y}^{(m)}). \tag{8.53}$$

Observe that (8.53) and (8.48) ensure that for all $\Psi \in \times_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})$ it holds that

$$J(\Psi) = \frac{1}{M}\left[\sum_{m=1}^{M} \mathcal{J}^{(m)}(\Psi)\right]. \tag{8.54}$$

Corollary 8.2.2 hence establishes items (i), (ii), (iii), (iv), and (v). The proof of Corollary 8.2.3 is thus complete. □

Corollary 8.2.4 (Backpropagation for ANNs with quadratic loss and minibatches)**.** *Let $L, M \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((\mathbf{W}_1, \mathbf{B}_1), \dots, (\mathbf{W}_L, \mathbf{B}_L)) \in \times_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})$, let $a\colon \mathbb{R} \to \mathbb{R}$ be differentiable, for every $m \in \{1, 2, \dots, M\}$ let $\mathfrak{x}_0^{(m)} \in \mathbb{R}^{l_0}$, $\mathfrak{x}_1^{(m)} \in \mathbb{R}^{l_1}$, $\dots$, $\mathfrak{x}_L^{(m)} \in \mathbb{R}^{l_L}$, $\mathfrak{y}^{(m)} \in \mathbb{R}^{l_L}$ satisfy for all $k \in \{1, 2, \dots, L\}$ that*

$$\mathfrak{x}_k^{(m)} = \mathfrak{M}_{a\mathbb{1}_{[0,L)}(k)+\mathrm{id}_{\mathbb{R}}\mathbb{1}_{\{L\}}(k), l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k), \tag{8.55}$$

let $J = \big(J((W_1, B_1), \dots, (W_L, B_L))\big)_{((W_1,B_1),\dots,(W_L,B_L)) \in \times_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})}\colon \times_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \to \mathbb{R}$ satisfy for all $\Psi \in \big(\times_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})\big)$ that

$$J(\Psi) = \frac{1}{M}\left[\sum_{m=1}^{M} \|(\mathcal{R}_a(\Psi))(\mathfrak{x}_0^{(m)}) - \mathfrak{y}^{(m)}\|_2^2\right], \tag{8.56}$$

and for every $m \in \{1, 2, \dots, M\}$ let $D_k^{(m)} \in \mathbb{R}^{l_{k-1}}$, $k \in \{1, 2, \dots, L+1\}$, satisfy for all $k \in \{1, 2, \dots, L-1\}$ that

$$D_{L+1}^{(m)} = 2(\mathfrak{x}_L^{(m)} - \mathfrak{y}^{(m)}), \qquad D_L^{(m)} = [\mathbf{W}_L]^* D_{L+1}^{(m)}, \qquad \textit{and} \tag{8.57}$$

$$D_k^{(m)} = [\mathbf{W}_k]^*[\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k))] D_{k+1}^{(m)} \tag{8.58}$$

(cf. Definitions 2.1.4, 2.2.3, 3.1.16, and 8.2.1). Then

(i) it holds that $J\colon \times_{k=1}^{L}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \to \mathbb{R}$ is differentiable,

(ii) it holds that $(\nabla_{B_L} J)(\Phi) = \frac{1}{M}\big[\sum_{m=1}^{M} D_{L+1}^{(m)}\big]$,

(iii) it holds for all $k \in \{1, 2, \dots, L-1\}$ that

$$(\nabla_{B_k} J)(\Phi) = \frac{1}{M}\left[\sum_{m=1}^{M} [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k))] D_{k+1}^{(m)}\right], \tag{8.59}$$

(iv) it holds that $(\nabla_{W_L} J)(\Phi) = \frac{1}{M}\big[\sum_{m=1}^{M} D_{L+1}^{(m)} [\mathfrak{x}_{L-1}^{(m)}]^*\big]$, and

(v) it holds for all $k \in \{1, 2, \dots, L-1\}$ that

$$(\nabla_{W_k} J)(\Phi) = \frac{1}{M}\left[\sum_{m=1}^{M} [\operatorname{diag}(\mathfrak{M}_{a',l_k}(\mathbf{W}_k \mathfrak{x}_{k-1}^{(m)} + \mathbf{B}_k))] D_{k+1}^{(m)} [\mathfrak{x}_{k-1}^{(m)}]^*\right]. \tag{8.60}$$

Proof of Corollary 8.2.4. Throughout this proof let $\mathfrak{C} = (\mathfrak{C}(x,y))_{(x,y)\in\mathbb{R}^{l_L}\times\mathbb{R}^{l_L}} \colon \mathbb{R}^{l_L}\times\mathbb{R}^{l_L}\to\mathbb{R}$ satisfy for all $x, y \in \mathbb{R}^{l_L}$ that

$$\mathfrak{C}(x,y) = \|x-y\|_2^2, \tag{8.61}$$

Observe that (8.61) ensures that for all $m \in \{1,2,\dots,M\}$ it holds that

$$(\nabla_x\mathfrak{C})(\mathfrak{x}_L^{(m)}, \mathfrak{y}^{(m)}) = 2(\mathfrak{x}_L^{(m)} - \mathfrak{y}^{(m)}) = D_{L+1}^{(m)}. \tag{8.62}$$

Combining this, (8.55), (8.56), (8.57), and (8.58) with Corollary 8.2.3 establishes items (i), (ii), (iii), (iv), and (v). The proof of Corollary 8.2.4 is thus complete. $\square$

Bibliography

[1] Boyd, S., and Vandenberghe, L. *Convex optimization.* Cambridge University Press, 2004.

[2] Bubeck, S. Convex Optimization: Algorithms and Complexity. *Foundations and Trends® in Machine Learning* 8, 3-4 (2015), 231–357. ISSN: 1935-8237. URL: http://dx.doi.org/10.1561/2200000050.

[3] Hinton, G., Srivastava, N., and Swersky, K. *Lecture 6e: Rmsprop: Divide the gradient by a running average of its recent magnitude.* http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. 26-31, Accessed: 01.12.2017.

[4] Jentzen, A., Kuckuck, B., Neufeld, A., and von Wurstemberger, P. Strong error analysis for stochastic gradient descent optimization algorithms. *IMA Journal of Numerical Analysis* 41, 1 (May 2020), 455–492. URL: https://doi.org/10.1093/imanum/drz055.

[5] Kingma, D. P., and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980* (2014), 15 pages. arXiv: 1412.6980. URL: http://arxiv.org/abs/1412.6980.

[6] *Lasagne: Updates.* http://lasagne.readthedocs.io/en/latest/modules/updates.html#lasagne.updates.rmsprop. [Accessed 6-December-2017].

[7] LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature* 521 (2015), 436–444.

[8] Nesterov, Y. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In: *Soviet Mathematics Doklady.* Vol. 27. 1983, 372–376.

[9] Nesterov, Y. *Introductory lectures on convex optimization: A basic course.* Vol. 87. Springer Science & Business Media, 2013.

[10] Polyak, B. T. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics* 4, 5 (1964), 1–17.

[11] Ruder, S. An overview of gradient descent optimization algorithms. *arXiv:1609.04747* (2016), 12 pages.

[12] *Tensorflow: RMSProp.* https://www.tensorflow.org/api_docs/python/tf/train/RMSPropOptimizer. [Accessed 6-December-2017].

[13] Zeiler, M. D. ADADELTA: An Adaptive Learning Rate Method. *arXiv:1212.5701* (2012), 6 pages. arXiv: 1212.5701. url: http://arxiv.org/abs/1212.5701.