Adv Comput Math (2023) 49:4
https://doi.org/10.1007/s10444-022-09970-2

Space-time error estimates for deep neural network approximations for differential equations

Philipp Grohs[1] · Fabian Hornung[2,3] · Arnulf Jentzen[2,4,5] · Philipp Zimmermann[2]

Received: 10 August 2019 / Accepted: 27 June 2022

Abstract

Over the last few years deep artificial neural networks (ANNs) have very successfully been used in numerical simulations for a wide variety of computational problems including computer vision, image classification, speech recognition, natural language processing, as well as computational advertisement. In addition, it has recently been proposed to approximate solutions of high-dimensional partial differential equations (PDEs) by means of stochastic learning problems involving deep ANNs. There are now also a few rigorous mathematical results in the scientific literature which provide error estimates for such deep learning based approximation methods for PDEs. All of these articles provide spatial error estimates for ANN approximations for PDEs but do not provide error estimates for the entire space-time error for the considered ANN approximations. It is the subject of the main result of this article to provide space-time error estimates for deep ANN approximations of Euler approximations of certain perturbed differential equations. Our proof of this result is based (i) on a certain ANN calculus and (ii) on ANN approximation results for products of the form $[0, T] \times \mathbb{R}^d \ni (t, x) \mapsto tx \in \mathbb{R}^d$ where $T \in (0, \infty)$, $d \in \mathbb{N}$, which we both develop within this article.

Keywords ANNs · PDEs · Space-time error

1 Introduction

Over the last few years deep artificial neural networks (ANNs) have very successfully been used in numerical simulations for a wide variety of computational problems

---

Communicated by: Holger Rauhut

✉ Philipp Zimmermann
philipp.zimmermann@math.ethz.ch

Extended author information available on the last page of the article.

including computer vision, image classification, speech recognition, natural language processing, as well as computational advertisement (cf., e.g., the references mentioned in [14, 17, 27]). In addition, the articles [9, 19] suggest to approximate solutions of high-dimensional partial differential equations (PDEs) by means of stochastic learning problems involving deep ANNs. We also refer to [1–6, 8, 10, 12, 13, 15, 20, 21, 23, 26, 30–32, 35, 36, 39] for extensions and improvements of such deep learning based approximation methods for PDEs.

There are now also a few rigorous mathematical results in the scientific literature which provide error estimates for such deep learning based approximation methods for PDEs; see, e.g., [7, 11, 16, 20, 24, 27, 28, 37, 39]. The articles in this reference list all provide spatial error estimates for ANN approximations for PDEs but do not provide error estimates for the entire space-time error for the considered ANN approximations.

It is the subject of Theorem 3.12 in this article, which is the main result of this article, to provide space-time error estimates for ANN approximations of Euler approximations of certain perturbed ordinary differential equations (ODEs). To illustrate the findings of the main result of this article in more details, we formulate in Theorem 1.1 below a special case of Theorem 3.12. In the following we briefly illuminate the statement of Theorem 1.1 in words and also add some explanatory comments regarding the mathematical objects appearing in Theorem 1.1, thereafter, we present the precise statement of Theorem 1.1, and, thereafter, we remark on the proofs of Theorem 1.1 and Theorem 3.12, respectively, and we also add some comments on the benefits of Theorem 1.1 and Theorem 3.12, respectively.

In Theorem 1.1 we study ANN approximations of Euler approximations of suitable ODEs and the real number $T \in (0, \infty)$ in Theorem 1.1 specifies the time horizon of the ODE under consideration. To precisely specify the ANN approximations in Theorem 1.1, we need to mathematically formulate the considered ANNs and their realization functions. In Theorem 1.1 we employ fully-connected feedforward ANNs and, roughly speaking, we can think of such ANNs as nested tuples of real matrices and vectors and the realization function of such a ANN is then an iterated finite composition of appropriate affine linear functions (uniquely described through pairs of real matrices and vectors in the nested tuples) and certain fixed nonlinear functions.

The fixed nonlinear functions are often appropriate multidimensional versions of a one-dimensional function from $\mathbb{R}$ to $\mathbb{R}$ and this one-dimensional function is usually referred to as activation function of the ANN. In Theorem 1.1 we use the rectifier function $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$ as the activation function and the functions

$$A_d \colon \mathbb{R}^d \to \mathbb{R}^d \tag{1}$$

for $d \in \mathbb{N}$ satisfying that for all $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ we have that $A_d(x) = (\max\{x_1, 0\}, \ldots, \max\{x_d, 0\})$ in Theorem 1.1 serve as the appropriate multidimensional versions of the one-dimensional rectifier function $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$.

The set $\mathbf{N}$ in Theorem 1.1 represents the set of all fully-connected feedforward ANNs and the function

$$R \colon \mathbf{N} \to \cup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l) \tag{2}$$

in (6) in Theorem 1.1 assigns to each ANN its realization function. More formally, for every ANN $\Phi \in \mathbf{N}$ we have that the continuous function $R(\Phi) \in \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ is the realization function associated to Φ.

The function $P\colon \mathbf{N} \to \mathbb{N}$ in Theorem 1.1 counts the number of real weight and bias parameters used to describe the considered ANN. Specifically, for every ANN $\Phi \in \mathbf{N}$ we have that $P(\Phi)$ is the number of real numbers employed to describe Φ. In that sense we have for every $\Phi \in \mathbf{N}$ that $P(\Phi)$ corresponds to the amount of memory needed to store Φ on a computer. We also refer to Fig. 1 for a graphical illustration for an example ANN within the class of fully-connected feedforward ANNs used in Theorem 1.1 below.

In Theorem 1.1 we study ANN approximations of Euler approximations of appropriate ODEs in which the vector field of the considered ODE is the realization function of a ANN and the ANNs $\Phi_d \in \mathbf{N}$, $d \in \mathbb{N}$, in Theorem 1.1 serve as the ANNs whose realization functions are the vector fields of the considered ODEs. More formally, for every $d \in \mathbb{N}$ let $X^d = (X^d_{t,x})_{(t,x)\in[0,T]\times\mathbb{R}^d}\colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$

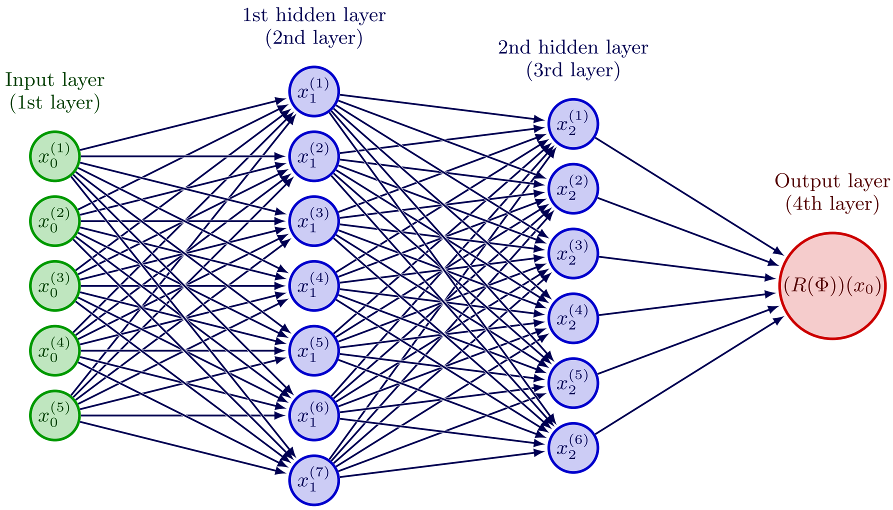

Fig. 1 Graphical illustration of a fully-connected feedforward example ANN $\Phi \in \mathbf{N}$ consisting of 4 layers (corresponding to $L = 3$ affine linear transformations: one affine linear transformation between the 1st layer [input layer] and the 2nd layer [1st hidden layer], one affine linear transformation between the 2nd layer [1st hidden layer] and the 3rd layer [2nd hidden layer], and one affine linear transformation between the 3rd layer [2nd hidden layer] and the 4th layer [output layer]) with $l_0 = 5$ neurons on the input layer (with a 5-dimensional 1st layer), with $l_1 = 7$ neurons on the 1st hidden layer (with a 7-dimensional 2nd layer), with $l_2 = 6$ neurons on the 2nd hidden layer (with a 6-dimensional 3rd layer), and with $l_3 = 1$ neuron on the output layer (with an 1-dimensional output layer). In the situation of (6) in Theorem 1.1, we have that $L = 3$, $l_0 = 5$, $l_1 = 7$, $l_2 = 6$, $l_3 = 1$, we have that $x_0 = (x_0^{(1)}, \dots, x_0^{(5)}) \in \mathbb{R}^5$, $x_1 = (x_1^{(1)}, \dots, x_1^{(7)}) = A_7(W_1 x_0 + B_1) \in \mathbb{R}^7$, $x_2 = (x_2^{(1)}, \dots, x_2^{(6)}) = A_6(W_2 x_1 + B_2) \in \mathbb{R}^6$, we have that the ANN $\Phi = ((W_1, B_1), (W_2, B_2), (W_3, B_3))$ is an element of the set $\times_{k=1}^{3}(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) = ((\mathbb{R}^{7\times 5} \times \mathbb{R}^7) \times (\mathbb{R}^{6\times 7} \times \mathbb{R}^6) \times (\mathbb{R}^{1\times 6} \times \mathbb{R}^1)) \subsetneq \mathbf{N}$, and we have that the number of ANN parameters of Φ satisfies $P(\Phi) = l_1(l_0+1)+l_2(l_1+1)+l_3(l_2+1) = 7(5+1)+6(7+1)+1(6+1) = 42+48+7 = 97$.

be the unique continuous function which satisfies for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that $([0, T] \ni s \mapsto X^d_{s,x} \in \mathbb{R}^d) \in C^1([0, T], \mathbb{R}^d)$, $X^d_{t,x} = x$, and

$$\tfrac{d}{dt} X^d_{t,x} = \big(R(\Phi_d)\big)(X^d_{t,x}) \tag{3}$$

and note that for all $d, N \in \mathbb{N}$ we have that the function

$$Y^{d,N} = (Y^{d,N}_{t,x})_{(t,x)\in[0,T]\times\mathbb{R}^d} \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \tag{4}$$

in (8) in Theorem 1.1 is nothing else but the time-continuous Euler approximation of the ODE in (3) with $N \in \mathbb{N}$ equidistant time steps.

Roughly speaking, Theorem 1.1 then demonstrates that there exist ANNs $(\Psi_{\varepsilon,d,N})_{(\varepsilon,d,N)\in(0,1]\times\mathbb{N}\times\mathbb{N}} \subseteq \mathbf{N}$ whose realization functions $R(\Psi_{\varepsilon,d,N}) \colon \mathbb{R}^{d+1} \to \mathbb{R}^d$, $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, approximate the time-continuous Euler approximations $Y^{d,N} = (Y^{d,N}_{t,x})_{(t,x)\in[0,T]\times\mathbb{R}^d} \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $d, N \in \mathbb{N}$, according to (9) in Theorem 1.1 and whose parameters $P(\Psi_{\varepsilon,d,N}) \in \mathbb{N}$, $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, grow at most polynomially in the dimension $d \in \mathbb{N}$, at most polynomially in the number of time steps $N \in \mathbb{N}$, and at most logarithmically in the accuracy parameter $\varepsilon > 0$. We now present the precise statement of Theorem 1.1 and, thereafter, comment on the proof and the use of Theorem 1.1.

Theorem 1.1 *Let $\mathfrak{C}, T, \mathfrak{d} \in (0, \infty)$, let $A_d \in C(\mathbb{R}^d, \mathbb{R}^d)$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that*

$$A_d(x) = (\max\{x_1, 0\}, \ldots, \max\{x_d, 0\}), \tag{5}$$

let $\mathbf{N} = \cup_{L\in\mathbb{N}} \cup_{l_0,l_1,\ldots,l_L\in\mathbb{N}} \big(\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\big)$, let $R \colon \mathbf{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ and $P \colon \mathbf{N} \to \mathbb{N}$ satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi \in (\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $\Psi = ((W_1, B_1), \ldots, (W_L, B_L)) \in (\times_{k=1}^L (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_L \in \mathbb{R}^{l_L}$ with $\forall\, k \in \mathbb{N} \cap (0, L) \colon x_k = A_{l_k}(W_k x_{k-1} + B_k)$ that

$$R(\Psi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}), \qquad (R(\Psi))(x_0) = W_L x_{L-1} + B_L, \tag{6}$$

and $P(\Phi) = \sum_{k=1}^L l_k(l_{k-1} + 1)$, let $\Phi_d \in \mathbf{N}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that

$$R(\Phi_d) \in C(\mathbb{R}^d, \mathbb{R}^d), \quad \|(R(\Phi_d))(x)\| \leqslant \mathfrak{C}(1 + \|x\|), \quad \text{and} \quad P(\Phi_d) \leqslant \mathfrak{C} d^{\mathfrak{d}}, \tag{7}$$

and let $Y^{d,N} = (Y^{d,N}_{t,x})_{(t,x)\in[0,T]\times\mathbb{R}^d} \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $N, d \in \mathbb{N}$, satisfy for all $d, N \in \mathbb{N}$, $n \in \{0, 1, \ldots, N-1\}$, $t \in \big[\frac{nT}{N}, \frac{(n+1)T}{N}\big]$, $x \in \mathbb{R}^d$ that $Y^{d,N}_{0,x} = x$ and

$$Y^{d,N}_{t,x} = Y^{d,N}_{\frac{nT}{N},x} + \big(t - \tfrac{nT}{N}\big)\, (R(\Phi_d))(Y^{d,N}_{\frac{nT}{N},x}). \tag{8}$$

Then there exist $C \in \mathbb{R}$ and $\Psi_{\varepsilon,d,N} \in \mathbf{N}$, $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, such that

(i) *it holds for all* $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$ *that* $R(\Psi_{\varepsilon,d,N}) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,

(ii) *it holds for all* $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ *that*

$$\|Y_{t,x}^{d,N} - (R(\Psi_{\varepsilon,d,N}))(t,x)\| \leqslant C d^{1/2} N^{3/2} \varepsilon (1 + \|x\|^3), \tag{9}$$

(iii) *it holds for all* $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ *that*

$$\|(R(\Psi_{\varepsilon,d,N}))(t,x)\| \leqslant C d^{1/2} N (1 + \|x\|^2), \tag{10}$$

and

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

$$P(\Psi_{\varepsilon,d,N}) \leqslant C d^{16+8\mathfrak{d}} N^6 \big[1 + |\ln(\varepsilon)|^2\big]. \tag{11}$$

Theorem 1.1 is an immediate consequence of Corollary 3.13 in Section 3.3.5 below. Corollary 3.13, in turn, follows from Theorem 3.12 in Section 3.3.5, which is the main result of this article.

Theorem 1.1 and Theorem 3.12, respectively, can be used to establish that ANNs can approximate solutions of certain second-order Kolmogorov PDEs on entire space-time regions without the curse of dimensionality and this is precisely the subject of our follow-up article [22]. To illustrate this issue, let $f_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $g_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, be Lipschitz continuous functions, let $u_d \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, $d \in \mathbb{N}$, satisfy for all $t \in [0, T]$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that

$$u_d(0, x) = g_d(x) \qquad \text{and} \qquad (\tfrac{\partial}{\partial t} u_d)(t, x) = (\tfrac{\partial}{\partial x} u_d)(t, x)\, f_d(x), \tag{12}$$

and let $\mathcal{X}^d = (\mathcal{X}^d_{t,x})_{(t,x)\in[0,T]\times\mathbb{R}^d} \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, satisfy for all $t \in [0, T]$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that

$$\mathcal{X}^d_{0,x} = x \qquad \text{and} \qquad \tfrac{d}{dt} \mathcal{X}^d_{t,x} = f_d(\mathcal{X}^d_{t,x}). \tag{13}$$

In the outline of this introductory section we restrict ourselves to first-order Kolmogorov PDEs of the form (12) and deterministic ODEs of the form (13), respectively, while the later results in this article are also applicable in the case of certain second-order Kolmogorov PDEs and stochastic ODEs, respectively.

Our goal is to verify under reasonable assumptions that ANNs can approximate the solutions $u_d \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, of the PDEs in (12) on entire space-time regions without the curse of dimensionality. For this we observe that (12), (13), and the method of characteristics (or, in the context of second-order PDEs, the Feynman-Kac formula) show that for all $t \in [0, T]$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that

$$u_d(t, x) = g_d(\mathcal{X}^d_{t,x}). \tag{14}$$

In the next step we assume that the functions $f_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, and $g_d \colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, themselves can be approximated by ANNs without the curse of dimensionality in a suitable sense.

More specifically, we assume that there exist ANNs $F_{\varepsilon,d} \in \mathbf{N}$, $(\varepsilon, d) \in (0, 1] \times \mathbb{N}$, whose ANN parameters grow at most polynomially, both, in $\varepsilon^{-1} \in [1, \infty)$ and $d \in \mathbb{N}$ and whose realization functions converge in a suitable sense to the functions $f_d \colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, as ε converges to zero and we assume that there exist ANNs $G_{\varepsilon,d} \in \mathbf{N}$, $(\varepsilon, d) \in (0, 1] \times \mathbb{N}$, whose ANN parameters grow at most polynomially, both, in $\varepsilon^{-1} \in [1,\infty)$ and $d \in \mathbb{N}$ and whose realization functions converge in a suitable sense to the functions $g_d\colon \mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, as ε converges to zero.

Beside approximations for the functions f_d, $d \in \mathbb{N}$, and g_d, $d \in \mathbb{N}$, in (12), we also employ time-continuous Euler approximations. More formally, for every $d, N \in \mathbb{N}$ and every continuous $\mathscr{F}\colon \mathbb{R}^d \to \mathbb{R}^d$ let $\mathbb{Y}^{\mathscr{F},N} = (\mathbb{Y}^{\mathscr{F},N}_{t,x})_{(t,x)\in[0,T]\times\mathbb{R}^d}\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $n \in \{0,1,\ldots,N-1\}$, $t \in [\frac{nT}{N},\frac{(n+1)T}{N}]$, $x \in \mathbb{R}^d$ that

$$\mathbb{Y}^{\mathscr{F},N}_{0,x} = x \qquad \text{and} \qquad \mathbb{Y}^{\mathscr{F},N}_{t,x} = \mathbb{Y}^{\mathscr{F},N}_{\frac{nT}{N},x} + \big(t - \tfrac{nT}{N}\big)\,\mathscr{F}\big(\mathbb{Y}^{\mathscr{F},N}_{\frac{nT}{N},x}\big). \tag{15}$$

Observe that for all $d, N \in \mathbb{N}$, $\mathscr{F} \in C(\mathbb{R}^d,\mathbb{R}^d)$, $x \in \mathbb{R}^d$ we have that

$$[0,T] \ni t \mapsto \mathbb{Y}^{\mathscr{F},N}_{t,x} \in \mathbb{R}^d \tag{16}$$

is the time-continuous Euler approximation for the ODE with initial value x, with the vector field function $\mathscr{F}\colon \mathbb{R}^d \to \mathbb{R}^d$ and with the equidistant time step size $\frac{T}{N}$. In particular, we observe that (8) in Theorem 1.1 ensures that for all $d, N \in \mathbb{N}$ we have that $\mathbb{Y}^{R(\Phi_d),N} = Y^{d,N}$.

In the next step we combine (14) with the triangle inequality to obtain that for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ we have that

$$\begin{aligned}
&\big|u_d(t,x) - \big(R(G_{\epsilon,d})\big)(\mathbb{Y}^{R(F_{\varepsilon,d}),N}_{t,x})\big| = \big|g_d(\mathcal{X}^d_{t,x}) - \big(R(G_{\epsilon,d})\big)(\mathbb{Y}^{R(F_{\varepsilon,d}),N}_{t,x})\big| \\
&\leqslant \big|g_d(\mathcal{X}^d_{t,x}) - \big(R(G_{\epsilon,d})\big)(\mathcal{X}^d_{t,x})\big| + \big|\big(R(G_{\epsilon,d})\big)(\mathcal{X}^d_{t,x}) - \big(R(G_{\epsilon,d})\big)(\mathbb{Y}^{f_d,N}_{t,x})\big| \\
&+\big|\big(R(G_{\epsilon,d})\big)(\mathbb{Y}^{f_d,N}_{t,x}) - \big(R(G_{\epsilon,d})\big)(\mathbb{Y}^{R(F_{\varepsilon,d}),N}_{t,x})\big|.
\end{aligned} \tag{17}$$

The first summand on the right hand side of (17) can be controlled through the assumption that the functions $R(G_{\varepsilon,d})$, $\varepsilon \in (0,1]$, $d \in \mathbb{N}$, converge in a suitable sense to the functions g_d, $d \in \mathbb{N}$, as ε converges to zero, the second summand on the right hand side of (17) can be bounded from above by employing standard error analysis for Euler approximations from the literature (cf., e.g., [18, Section II.3] and [22, Section 2.3]), and the third summand on the right hand side of (17) can be estimated from above by using suitable elementary perturbation estimates for Euler approximations of ODEs. The estimate in (17) thus illustrates that it is sufficient to verify that

$$[0,T]\times\mathbb{R}^d \ni (t,x) \mapsto \big(R(G_{\epsilon,d})\big)(\mathbb{Y}^{R(F_{\varepsilon,d}),N}_{t,x}) \in \mathbb{R} \tag{18}$$

for $d \in \mathbb{N}$ can be approximated by ANNs without the curse of dimensionality in order to prove that $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ for $d \in \mathbb{N}$ can be approximated by ANNs without the curse of dimensionality. This is precisely where Theorem 1.1 above and the more general result in Theorem 3.12, respectively, can be brought into play. Indeed we observe that Theorem 1.1 above and Theorem 3.12, respectively, reveal that time-continuous Euler approximations of the form

$$[0,T]\times\mathbb{R}^d \ni (t,x) \mapsto \mathbb{Y}^{R(F_{\varepsilon,d}),N}_{t,x} \in \mathbb{R}^d \tag{19}$$

for $d \in \mathbb{N}$ can be approximated by ANNs without the curse of dimensionality. Combining this with (17) then allows us to conclude the solution functions $u_d\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$, $d \in \mathbb{N}$, can in a suitable sense be approximated by ANNs without the curse of dimensionality.

This finishes our sketch on how Theorem 1.1 above and Theorem 3.12, respectively, can be used to verify that solutions of first-order Kolmogorov PDEs of the form (12) can be approximated by ANNs without the curse of dimensionality. In the same spirit as above, Theorem 3.12 can actually also be employed to verify that certain second-order Kolmogorov PDEs can be approximated by ANNs without the curse of dimensionality. We refer to the precise statement of Theorem 3.12 in Section 3.3.5 in this article as well as to our follow-up article [22] for the details, also in this more general situation.

Our proofs of Theorem 1.1 and Theorem 3.12, respectively, are based on a certain ANN calculus, which we develop in Section 2. Section 2 is in parts based on several well-known concepts and results in the scientific literature (cf., e.g., [11, 27, 34, 40]). We refer to the beginning of Section 2 for a more detailed comparison of the content of Section 2 with the material in related articles in the scientific literature. Our proof of Theorem 1.1 and Theorem 3.12, respectively, is mainly inspired by [27], [11, Section 6], and [40, Section 3.1].

Theorem 1.1 and Theorem 3.12, respectively, provide error estimates for rectified ANN approximations of Euler approximations of certain perturbed ODEs. Many of the ANN approximation and representation results of this work, however, apply to ANNs with more general activation functions than only the rectifier function (cf., e.g., Li et al. [29, Section 1] and Petersen et al. [33, Section 2] for further activation functions).

The error estimates for rectified ANN approximations of Euler approximations of perturbed ODEs, which we establish in Theorem 1.1 and Theorem 3.12, respectively, can then be used to establish space-time error estimates for ANN approximations for PDEs. This will be the subject of a future research article, which will be based on this article.

The remainder of this article is organized as follows. In Section 2 we develop the above mentioned ANN calculus and, in particular, we establish in Section 2.5 ANN representation results for Euler approximations. In Section 3.1 we develop ANN approximation results for the square function $\mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$. These ANN approximation results for the square function are then used in Section 3.2 to develop ANN approximation results for products of the form $[0, T] \times \mathbb{R}^d \ni (t, x) \mapsto tx \in \mathbb{R}^d$ where $T \in (0, \infty)$, $d \in \mathbb{N}$. In Section 3.3 we then combine the ANN representation results in Section 2.5 with the ANN approximation results for products in Section 3.2 to establish in Theorem 3.12 the main result of this article.

2 Artificial neural network (ANN) calculus

This section develops a certain calculus for ANNs. Some of the notions and results which we present here are rather elementary, but for convenience of the reader we present here all details and we include the proof of every result. The material in this section is also in parts based on several well-known concepts and results in the scientific literature. In particular, Definition 2.1, Definition 2.2, and Definition 2.3 are slight reformulations of Petersen & Voigtlaender [34, Definition 2.1].

Moreover, Lemma 2.4 is elementary and well-known in the scientific literature. Furthermore, Definition 2.5 is also a slight reformulation of Petersen & Voigtlaender [34, Definition 2.2]. In addition, Proposition 2.6, Corollary 2.7, and Lemma 2.8 are elementary and essentially well-known in the scientific literature (cf., e.g., Petersen & Voigtlaender [34]). Moreover, Definition 2.11 is an extension of Elbrächter et al. [11, Setting 5.2] and Proposition 2.16 is in parts an extension of Elbrächter et al. [11, Lemma 5.3]. Furthermore, Definition 2.17 and Definition 2.22 extend Elbrächter et al. [11, Setting 5.2] (cf., e.g., Petersen & Voigtlaender [34, Definition 2.7]). In addition, Proposition 2.25 is a reformulation of [27, Lemma 5.1]. Moreover, Lemma 2.27 and Proposition 2.28 are significantly inspired by [27, Proposition 5.3]. Furthermore, item (iv) in Lemma 2.27 and item (iv) in Proposition 2.28, respectively, improve the parameter estimates in [27, Proposition 5.3]. In addition, Corollary 2.31 in Section 2.5.2 below is also in parts inspired by [27, Proposition 6.1].

2.1 ANNs and their realization functions

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

$$\mathbf{N} = \cup_{L\in\mathbb{N}} \cup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} \left(\times_{k=1}^{L} (\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}) \right) \tag{20}$$

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

Definition 2.2 (Multidimensional versions) Let $d \in \mathbb{N}$ and let $\psi \colon \mathbb{R} \to \mathbb{R}$ be a function. Then we denote by $\mathfrak{M}_{\psi,d} \colon \mathbb{R}^d \to \mathbb{R}^d$ the function which satisfies for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ that

$$\mathfrak{M}_{\psi,d}(x) = (\psi(x_1), \ldots, \psi(x_d)). \tag{21}$$

Definition 2.3 (Realizations associated to ANNs) Let $a \in C(\mathbb{R},\mathbb{R})$. Then we denote by $\mathcal{R}_a \colon \mathbf{N} \to \cup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ the function which satisfies for all $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, \ldots, x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall\, k \in \mathbb{N} \cap (0, L) \colon x_k = \mathfrak{M}_{a,l_k}(W_k x_{k-1} + B_k)$ that

$$\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}) \qquad \text{and} \qquad (\mathcal{R}_a(\Phi))(x_0) = W_L x_{L-1} + B_L \tag{22}$$

(cf. Definition 2.2 and Definition 2.1).

Lemma 2.4 *Let* $\Phi \in \mathbf{N}$ *(cf. Definition 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 2.3).

Proof of Lemma 2.4 Note that the assumption that $\Phi \in \mathbf{N} = \cup_{L\in\mathbb{N}} \cup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$ ensures that there exist $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$ such that

$$\Phi \in \left(\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k})\right). \tag{23}$$

Observe that (23) assures that

$$\mathcal{L}(\Phi) = L, \qquad \mathcal{I}(\Phi) = l_0, \qquad \mathcal{O}(\Phi) = l_L, \tag{24}$$

$$\text{and} \qquad \mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L) \in \mathbb{N}^{L+1} = \mathbb{N}^{\mathcal{L}(\Phi)+1}. \tag{25}$$

This establishes item (i). Moreover, note that (24) and (22) 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 2.4 is thus completed. □

2.2 Compositions of ANNs

2.2.1 Standard compositions of ANNs

Definition 2.5 (Standard compositions of ANNs) We denote by $(\cdot) \bullet (\cdot) \colon \{(\Phi_1, \Phi_2) \in \mathbf{N} \times \mathbf{N} \colon \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)\} \to \mathbf{N}$ the function which satisfies for all $L, \mathfrak{L} \in \mathbb{N}$, $l_0, l_1, \ldots, l_L, \mathfrak{l}_0, \mathfrak{l}_1, \ldots, \mathfrak{l}_{\mathfrak{L}} \in \mathbb{N}$, $\Phi_1 = ((W_1, B_1), (W_2, B_2), \ldots, (W_L, B_L)) \in (\times_{k=1}^{L}(\mathbb{R}^{l_k\times l_{k-1}} \times \mathbb{R}^{l_k}))$, $\Phi_2 = ((\mathcal{W}_1, \mathfrak{B}_1), (\mathcal{W}_2, \mathfrak{B}_2), \ldots, (\mathcal{W}_{\mathfrak{L}}, \mathfrak{B}_{\mathfrak{L}})) \in (\times_{k=1}^{\mathfrak{L}} (\mathbb{R}^{\mathfrak{l}_k\times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k}))$ with $l_0 = \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2) = \mathfrak{l}_{\mathfrak{L}}$ that

$$\Phi_1 \bullet \Phi_2 = \begin{cases} \begin{array}{r}((\mathcal{W}_1, \mathfrak{B}_1), (\mathcal{W}_2, \mathfrak{B}_2), \ldots, (\mathcal{W}_{\mathfrak{L}-1}, \mathfrak{B}_{\mathfrak{L}-1}), (W_1\mathcal{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1), \\ (W_2, B_2), (W_3, B_3), \ldots, (W_L, B_L))\end{array} & : L > 1 < \mathfrak{L} \\ ((W_1\mathcal{W}_1, W_1\mathfrak{B}_1 + B_1), (W_2, B_2), (W_3, B_3), \ldots, (W_L, B_L)) & : L > 1 = \mathfrak{L} \\ ((\mathcal{W}_1, \mathfrak{B}_1), (\mathcal{W}_2, \mathfrak{B}_2), \ldots, (\mathcal{W}_{\mathfrak{L}-1}, \mathfrak{B}_{\mathfrak{L}-1}), (W_1\mathcal{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1)) & : L = 1 < \mathfrak{L} \\ (W_1\mathcal{W}_1, W_1\mathfrak{B}_1 + B_1) & : L = 1 = \mathfrak{L} \end{cases} \tag{26}$$

(cf. Definition 2.1).

Proposition 2.6 *Let* $\Phi_1, \Phi_2 \in \mathbf{N}$, $l_{1,0}, l_{1,1}, \ldots, l_{1,\mathcal{L}(\Phi_1)}, l_{2,0}, l_{2,1}, \ldots, l_{2,\mathcal{L}(\Phi_2)} \in \mathbb{N}$ *satisfy for all* $k \in \{1, 2\}$ *that* $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ *and* $\mathcal{D}(\Phi_k) = (l_{k,0}, l_{k,1}, \ldots, l_{k,\mathcal{L}(\Phi_k)})$ *(cf. Definition 2.1). Then*

(i) *it holds that*

$$\mathcal{D}(\Phi_1 \bullet \Phi_2) = (l_{2,0}, l_{2,1}, \ldots, l_{2,\mathcal{L}(\Phi_2)-1}, l_{1,1}, l_{1,2}, \ldots, l_{1,\mathcal{L}(\Phi_1)}), \tag{27}$$

(ii) *it holds that*

$$[\mathcal{L}(\Phi_1 \bullet \Phi_2) - 1] = [\mathcal{L}(\Phi_1) - 1] + [\mathcal{L}(\Phi_2) - 1], \tag{28}$$

(iv) *it holds that*

$$\mathcal{H}(\Phi_1 \bullet \Phi_2) = \mathcal{H}(\Phi_1) + \mathcal{H}(\Phi_2), \tag{29}$$

(v) *it holds that*

$$\begin{aligned}\mathcal{P}(\Phi_1 \bullet \Phi_2) &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1}(l_{2,\mathcal{L}(\Phi_2)-1} + 1) \\ &\quad - l_{1,1}(l_{1,0} + 1) - l_{2,\mathcal{L}(\Phi_2)}(l_{2,\mathcal{L}(\Phi_2)-1} + 1) \\ &\leqslant \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1} l_{2,\mathcal{L}(\Phi_2)-1},\end{aligned} \tag{30}$$

and

(vi) *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*

$$\mathcal{R}_a(\Phi_1 \bullet \Phi_2) = [\mathcal{R}_a(\Phi_1)] \circ [\mathcal{R}_a(\Phi_2)] \tag{31}$$

(cf. Definition 2.3 and Definition 2.5).

Proof of Proposition 2.6 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 $\big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big) \in (\times_{j=1}^{L_k}(\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}}))$, $k \in \{1, 2\}$, satisfy for all $k \in \{1, 2\}$ that

$$\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big), \tag{32}$$

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

$$\forall\, j \in \mathbb{N} \cap (0, L_2)\colon\ x_j = \mathfrak{M}_{a,l_{2,j}}(W_{2,j} x_{j-1} + B_{2,j}) \tag{33}$$

(cf. Definition 2.2), let $y_0 \in \mathbb{R}^{l_{1,0}}, y_1 \in \mathbb{R}^{l_{1,1}}, \ldots, y_{L_1-1} \in \mathbb{R}^{l_{1,L_1-1}}$ satisfy that $y_0 = W_{2,L_2} x_{L_2-1} + B_{2,L_2}$ and

$$\forall\, j \in \mathbb{N} \cap (0, L_1)\colon\ y_j = \mathfrak{M}_{a,l_{1,j}}(W_{1,j} y_{j-1} + B_{1,j}), \tag{34}$$

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

$$\forall\, j \in \mathbb{N} \cap (0, L_3)\colon\ z_j = \mathfrak{M}_{a,l_{3,j}}(W_{3,j} z_{j-1} + B_{3,j}). \tag{35}$$

Note that (26) ensures that

$$\Phi_3 = \Phi_1 \bullet \Phi_2 = \begin{cases} \begin{aligned}&\big((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \ldots, (W_{2,L_2-1}, B_{2,L_2-1}),\\ &\quad (W_{1,1} W_{2,L_2}, W_{1,1} B_{2,L_2} + B_{1,1}), (W_{1,2}, B_{1,2}),\\ &\quad (W_{1,3}, B_{1,3}), \ldots, (W_{1,L_1}, B_{1,L_1})\big)\end{aligned} & : L_1 > 1 < L_2 \\ \begin{aligned}&\big((W_{1,1} W_{2,1}, W_{1,1} B_{2,1} + B_{1,1}), (W_{1,2}, B_{1,2}),\\ &\quad (W_{1,3}, B_{1,3}), \ldots, (W_{1,L_1}, B_{1,L_1})\big)\end{aligned} & : L_1 > 1 = L_2 \\ \begin{aligned}&\big((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \ldots, (W_{2,L_2-1}, B_{2,L_2-1}),\\ &\quad (W_{1,1} W_{2,L_2}, W_{1,1} B_{2,L_2} + B_{1,1})\big)\end{aligned} & : L_1 = 1 < L_2 \\ (W_{1,1} W_{2,1}, W_{1,1} B_{2,1} + B_{1,1}) & : L_1 = 1 = L_2 \end{cases}. \tag{36}$$

Hence, we obtain that

$$\begin{aligned}[\mathcal{L}(\Phi_1 \bullet \Phi_2) - 1] &= [(L_2 - 1) + 1 + (L_1 - 1)] - 1 \\ &= [L_1 - 1] + [L_2 - 1] = [\mathcal{L}(\Phi_1) - 1] + [\mathcal{L}(\Phi_2) - 1]\end{aligned} \tag{37}$$

$$\text{and} \qquad \mathcal{D}(\Phi_1 \bullet \Phi_2) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1}, l_{1,2}, \ldots, l_{1,L_1}). \tag{38}$$

This establishes items (i)–(iii). In addition, observe that (38) demonstrates that

$$\begin{aligned}&\mathcal{P}(\Phi_1 \bullet \Phi_2) = \sum_{j=1}^{L_3} l_{3,j}(l_{3,j-1}+1) \\ &= \left[\sum_{j=1}^{L_2-1} l_{3,j}(l_{3,j-1}+1)\right] + l_{3,L_2}(l_{3,L_2-1}+1) + \left[\sum_{j=L_2+1}^{L_3} l_{3,j}(l_{3,j-1}+1)\right] \\ &= \left[\sum_{j=1}^{L_2-1} l_{2,j}(l_{2,j-1}+1)\right] + l_{1,1}(l_{2,L_2-1}+1) + \left[\sum_{j=L_2+1}^{L_3} l_{1,j-L_2+1}(l_{1,j-L_2}+1)\right] \\ &= \left[\sum_{j=1}^{L_2-1} l_{2,j}(l_{2,j-1}+1)\right] + \left[\sum_{j=2}^{L_1} l_{1,j}(l_{1,j-1}+1)\right] + l_{1,1}\big(l_{2,L_2-1}+1\big) \\ &= \left[\sum_{j=1}^{L_2} l_{2,j}(l_{2,j-1}+1)\right] + \left[\sum_{j=1}^{L_1} l_{1,j}(l_{1,j-1}+1)\right] + l_{1,1}(l_{2,L_2-1}+1) \\ &\quad - l_{2,L_2}(l_{2,L_2-1}+1) - l_{1,1}(l_{1,0}+1) \\ &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1}(l_{2,L_2-1}+1) - l_{2,L_2}(l_{2,L_2-1}+1) \\ &\quad - l_{1,1}(l_{1,0}+1) \\ &\leqslant \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1} l_{2,L_2-1}.\end{aligned} \tag{39}$$

This establishes item (iv). Moreover, observe that (36) and the fact that $a \in C(\mathbb{R}, \mathbb{R})$ ensure that

$$\mathcal{R}_a(\Phi_1 \bullet \Phi_2) \in C(\mathbb{R}^{l_{2,0}}, \mathbb{R}^{l_{1,L_1}}) = C(\mathbb{R}^{\mathcal{I}(\Phi_2)}, \mathbb{R}^{\mathcal{O}(\Phi_1)}). \tag{40}$$

Next note that (37) implies that $L_3 = L_1 + L_2 - 1$. This, (36), and (38) ensure that

$$(l_{3,0}, l_{3,1}, \ldots, l_{3,L_1+L_2-1}) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1}, l_{1,2}, \ldots, l_{1,L_1}), \tag{41}$$

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_2) \colon\ (W_{3,j}, B_{3,j}) = (W_{2,j}, B_{2,j})\big], \tag{42}$$

$$(W_{3,L_2}, B_{3,L_2}) = (W_{1,1} W_{2,L_2}, W_{1,1} B_{2,L_2} + B_{1,1}), \tag{43}$$

$$\text{and} \qquad \big[\forall\, j \in \mathbb{N} \cap (L_2, L_1 + L_2) \colon\ (W_{3,j}, B_{3,j}) = (W_{1,j+1-L_2}, B_{1,j+1-L_2})\big]. \tag{44}$$

This, (33), (35), 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 (43) and the fact that $y_0 = W_{2,L_2} x_{L_2-1} + B_{2,L_2}$ ensures that

$$\begin{aligned}W_{3,L_2} z_{L_2-1} + B_{3,L_2} &= W_{3,L_2} x_{L_2-1} + B_{3,L_2} \\ &= W_{1,1} W_{2,L_2} x_{L_2-1} + W_{1,1} B_{2,L_2} + B_{1,1} \\ &= W_{1,1}(W_{2,L_2} x_{L_2-1} + B_{2,L_2}) + B_{1,1} = W_{1,1} y_0 + B_{1,1}.\end{aligned} \tag{45}$$

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

$$W_{3,j} z_{j-1} + B_{3,j} = W_{1,j+1-L_2} y_{j-L_2} + B_{1,j+1-L_2}. \tag{46}$$

We prove (46) by induction on $j \in \mathbb{N} \cap [L_2, L_1 + L_2)$. Note that (45) establishes (46) in the base case $j = L_2$. For the induction step note that the fact that $L_3 = L_1 + L_2 - 1$, (34), (35), (41), and (44) imply that for all $j \in \mathbb{N} \cap [L_2, \infty) \cap (0, L_1 + L_2 - 1)$ with

$$W_{3,j} z_{j-1} + B_{3,j} = W_{1,j+1-L_2} y_{j-L_2} + B_{1,j+1-L_2} \tag{47}$$

it holds that

$$\begin{aligned} & W_{3,j+1} z_j + B_{3,j+1} = W_{3,j+1} \mathfrak{M}_{a,l_{3,j}}(W_{3,j} z_{j-1} + B_{3,j}) + B_{3,j+1} \\ & = W_{1,j+2-L_2} \mathfrak{M}_{a,l_{1,j+1-L_2}}(W_{1,j+1-L_2} y_{j-L_2} + B_{1,j+1-L_2}) + B_{1,j+2-L_2} \\ & = W_{1,j+2-L_2} y_{j+1-L_2} + B_{1,j+2-L_2}. \end{aligned} \tag{48}$$

Induction hence proves (46). Next observe that (46) and the fact that $L_3 = L_1 + L_2 - 1$ assure that

$$W_{3,L_3} z_{L_3-1} + B_{3,L_3} = W_{3,L_1+L_2-1} z_{L_1+L_2-2} + B_{3,L_1+L_2-1} = W_{1,L_1} y_{L_1-1} + B_{1,L_1}. \tag{49}$$

The fact that $\Phi_3 = \Phi_1 \bullet \Phi_2$, (33), (34), and (35) therefore prove that

$$\begin{aligned} [\mathcal{R}_a(\Phi_1 \bullet \Phi_2)](x_0) &= [\mathcal{R}_a(\Phi_3)](x_0) = [\mathcal{R}_a(\Phi_3)](z_0) = W_{3,L_3} z_{L_3-1} + B_{3,L_3} \\ &= W_{1,L_1} y_{L_1-1} + B_{1,L_1} = [\mathcal{R}_a(\Phi_1)](y_0) \\ &= [\mathcal{R}_a(\Phi_1)]\big(W_{2,L_2} x_{L_2-1} + B_{2,L_2}\big) \\ &= [\mathcal{R}_a(\Phi_1)]\big([\mathcal{R}_a(\Phi_2)](x_0)\big) = [(\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))](x_0). \end{aligned} \tag{50}$$

Combining this with (40) establishes item (v). The proof of Proposition 2.6 is thus completed. □

Corollary 2.7 *Let $L_1, L_2, L_3 \in \mathbb{N}$, $l_{1,0}, l_{1,1}, \ldots, l_{1,L_1}, l_{2,0}, l_{2,1}, \ldots, l_{2,L_2}, l_{3,0}, l_{3,1}, \ldots, l_{3,L_3} \in \mathbb{N}$ satisfy that $l_{1,0} = l_{2,L_2}$ and let $\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \ldots, (W_{k,L_k}, B_{k,L_k})\big) \in \big(\times_{j=1}^{L_k} (\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}})\big)$, $k \in \{1, 2, 3\}$, satisfy that $\Phi_3 = \Phi_1 \bullet \Phi_2$ (cf. Definition 2.1 and Definition 2.5). Then*

(i) *it holds that*

$$L_3 = \mathcal{L}(\Phi_3) = \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2) - 1 = L_1 + L_2 - 1 \geqslant \max\{L_1, L_2\}, \tag{51}$$

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

$$(W_{3,j}, B_{3,j}) = (W_{2,j}, B_{2,j}), \tag{52}$$

(iii) *it holds that*

$$(W_{3,L_2}, B_{3,L_2}) = (W_{1,1} W_{2,L_2}, W_{1,1} B_{2,L_2} + B_{1,1}), \tag{53}$$

and

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

$$(W_{3,j}, B_{3,j}) = (W_{1,j-L_2+1}, B_{1,j-L_2+1}). \tag{54}$$

Proof of Corollary 2.7 Observe that item (ii) in Proposition 2.6 proves item (i). Moreover, note that (26) establishes items (ii)–(iv). The proof of Corollary 2.7 is thus completed. □

2.2.2 Associativity of standard compositions of ANNs

Lemma 2.8 *Let $\Phi_1, \Phi_2, \Phi_3 \in \mathbf{N}$ satisfy that $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ and $\mathcal{I}(\Phi_2) = \mathcal{O}(\Phi_3)$ (cf. Definition 2.1). Then it holds that*

$$(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3) \tag{55}$$

(cf. Definition 2.5).

Proof of Lemma 2.8 Throughout this proof let $\Phi_4, \Phi_5, \Phi_6, \Phi_7 \in \mathbf{N}$ satisfy that $\Phi_4 = \Phi_1 \bullet \Phi_2$, $\Phi_5 = \Phi_2 \bullet \Phi_3$, $\Phi_6 = \Phi_4 \bullet \Phi_3$, and $\Phi_7 = \Phi_1 \bullet \Phi_5$, let $L_k \in \mathbb{N}$, $k \in \{1, 2, \dots, 7\}$, 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 $\big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \dots, (W_{k,L_k}, B_{k,L_k})\big) \in (\times_{j=1}^{L_k}(\mathbb{R}^{l_{k,j} \times l_{k,j-1}} \times \mathbb{R}^{l_{k,j}}))$, $k \in \{1, 2, \dots, 7\}$, satisfy for all $k \in \{1, 2, \dots, 7\}$ that

$$\Phi_k = \big((W_{k,1}, B_{k,1}), (W_{k,2}, B_{k,2}), \dots, (W_{k,L_k}, B_{k,L_k})\big). \tag{56}$$

Observe that item (ii) in Proposition 2.6 and the fact that for all $k \in \{1, 2, 3\}$ it holds that $\mathcal{L}(\Phi_k) = L_k$ proves that

$$\begin{aligned} \mathcal{L}(\Phi_6) &= \mathcal{L}((\Phi_1 \bullet \Phi_2) \bullet \Phi_3) = \mathcal{L}(\Phi_1 \bullet \Phi_2) + \mathcal{L}(\Phi_3) - 1 \\ &= \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2) + \mathcal{L}(\Phi_3) - 2 = L_1 + L_2 + L_3 - 2 \\ &= \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2 \bullet \Phi_3) - 1 = \mathcal{L}(\Phi_1 \bullet (\Phi_2 \bullet \Phi_3)) = \mathcal{L}(\Phi_7). \end{aligned} \tag{57}$$

Next note that Corollary 2.7, (56), and the fact that $\Phi_4 = \Phi_1 \bullet \Phi_2$ imply that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_2)\colon\ (W_{4,j}, B_{4,j}) = (W_{2,j}, B_{2,j})\big], \tag{58}$$

$$(W_{4,L_2}, B_{4,L_2}) = (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), \tag{59}$$

and $$\big[\forall\, j \in \mathbb{N} \cap (L_2, L_1 + L_2)\colon\ (W_{4,j}, B_{4,j}) = (W_{1,j+1-L_2}, B_{1,j+1-L_2})\big]. \tag{60}$$

Hence, we obtain that

$$\begin{aligned} &\big[\forall\, j \in \mathbb{N} \cap (L_3 - 1, L_2 + L_3 - 1)\colon \\ &\qquad (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3})\big], \end{aligned} \tag{61}$$

$$(W_{4,L_2}, B_{4,L_2}) = (W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}), \tag{62}$$

and

$$\begin{aligned} &\big[\forall\, j \in \mathbb{N} \cap (L_2 + L_3 - 1, L_1 + L_2 + L_3 - 1)\colon \\ &\qquad (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{1,j+2-L_2-L_3}, B_{1,j+2-L_2-L_3})\big]. \end{aligned} \tag{63}$$

In addition, observe that Corollary 2.7, (56), and the fact that $\Phi_5 = \Phi_2 \bullet \Phi_3$ demonstrate that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_3)\colon\ (W_{5,j}, B_{5,j}) = (W_{3,j}, B_{3,j})\big], \tag{64}$$

$$(W_{5,L_3}, B_{5,L_3}) = (W_{2,1}W_{3,L_3}, W_{2,1}B_{3,L_3} + B_{2,1}), \tag{65}$$

and $$\big[\forall\, j \in \mathbb{N} \cap (L_3, L_2 + L_3)\colon\ (W_{5,j}, B_{5,j}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3})\big]. \tag{66}$$

Moreover, note that Corollary 2.7, (56), and the fact that $\Phi_6 = \Phi_4 \bullet \Phi_3$ ensure that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_3)\colon\ (W_{6,j}, B_{6,j}) = (W_{3,j}, B_{3,j})\big], \tag{67}$$

$$(W_{6,L_3}, B_{6,L_3}) = (W_{4,1} W_{3,L_3}, W_{4,1} B_{3,L_3} + B_{4,1}), \tag{68}$$

and $$\big[\forall\, j \in \mathbb{N} \cap (L_3, L_4 + L_3)\colon\ (W_{6,j}, B_{6,j}) = (W_{4,j+1-L_3}, B_{4,j+1-L_3})\big]. \tag{69}$$

Furthermore, observe that Corollary 2.7, (56), and the fact that $\Phi_7 = \Phi_1 \bullet \Phi_5$ show that

$$\big[\forall\, j \in \mathbb{N} \cap (0, L_5)\colon\ (W_{7,j}, B_{7,j}) = (W_{5,j}, B_{5,j})\big], \tag{70}$$

$$(W_{7,L_5}, B_{7,L_5}) = (W_{1,1} W_{5,L_5}, W_{1,1} B_{5,L_5} + B_{1,1}), \tag{71}$$

and $$\big[\forall\, j \in \mathbb{N} \cap (L_5, L_1 + L_5)\colon\ (W_{7,j}, B_{7,j}) = (W_{1,j+1-L_5}, B_{1,j+1-L_5})\big]. \tag{72}$$

This, the fact that $L_3 \leqslant L_2 + L_3 - 1 = L_5$, (64), and (67) imply that for all $j \in \mathbb{N} \cap (0, L_3)$ it holds that

$$(W_{6,j}, B_{6,j}) = (W_{3,j}, B_{3,j}) = (W_{5,j}, B_{5,j}) = (W_{7,j}, B_{7,j}). \tag{73}$$

In addition, observe that (58), (59), (64), (65), (68), (70), (71), and the fact that $L_5 = L_2 + L_3 - 1$ demonstrate that

$$\begin{aligned}
&(W_{6,L_3}, B_{6,L_3}) = (W_{4,1} W_{3,L_3}, W_{4,1} B_{3,L_3} + B_{4,1}) \\
&= \begin{cases} (W_{2,1} W_{3,L_3}, W_{2,1} B_{3,L_3} + B_{2,1}) & : L_2 > 1 \\ (W_{1,1} W_{2,1} W_{3,L_3}, W_{1,1} W_{2,1} B_{3,L_3} + W_{1,1} B_{2,1} + B_{1,1}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{2,1} W_{3,L_3}, W_{2,1} B_{3,L_3} + B_{2,1}) & : L_2 > 1 \\ (W_{1,1} (W_{2,1} W_{3,L_3}), W_{1,1} (W_{2,1} B_{3,L_3} + B_{2,1}) + B_{1,1}) & : L_2 = 1 \end{cases} \\
&= \begin{cases} (W_{5,L_3}, B_{5,L_3}) & : L_2 > 1 \\ (W_{1,1} W_{5,L_3}, W_{1,1} B_{5,L_3} + B_{1,1}) & : L_2 = 1 \end{cases} \\
&= (W_{7,L_3}, B_{7,L_3}).
\end{aligned} \tag{74}$$

Next note that the fact that $L_5 = L_2 + L_3 - 1 < L_1 + L_2 + L_3 - 1 = L_3 + L_4$, (69), (61), (66), and (70) ensure that for all $j \in \mathbb{N}$ with $L_3 < j < L_5$ it holds that

$$\begin{aligned}
(W_{6,j}, B_{6,j}) &= (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{2,j+1-L_3}, B_{2,j+1-L_3}) \\
&= (W_{5,j}, B_{5,j}) = (W_{7,j}, B_{7,j}).
\end{aligned} \tag{75}$$

Moreover, observe that the fact that $L_5 = L_2+L_3-1 < L_1+L_2+L_3-1 = L_3+L_4$, (69), (74), (59), (66), and (71) prove that

$$\begin{aligned}(W_{6,L_5}, B_{6,L_5}) &= \begin{cases}(W_{4,L_5+1-L_3}, B_{4,L_5+1-L_3}) & : L_2 > 1\\ (W_{6,L_3}, B_{6,L_3}) & : L_2 = 1\end{cases}\\ &= \begin{cases}(W_{4,L_2}, B_{4,L_2}) & : L_2 > 1\\ (W_{7,L_3}, B_{7,L_3}) & : L_2 = 1\end{cases}\\ &= \begin{cases}(W_{1,1}W_{2,L_2}, W_{1,1}B_{2,L_2} + B_{1,1}) & : L_2 > 1\\ (W_{7,L_5}, B_{7,L_5}) & : L_2 = 1\end{cases}\\ &= \begin{cases}(W_{1,1}W_{5,L_5}, W_{1,1}B_{5,L_5} + B_{1,1}) & : L_2 > 1\\ (W_{7,L_5}, B_{7,L_5}) & : L_2 = 1\end{cases}\\ &= (W_{7,L_5}, B_{7,L_5}).\end{aligned} \tag{76}$$

Furthermore, note that (69), (63), (72), and the fact that $L_5 = L_2 + L_3 - 1 \geqslant L_3$ assure that for all $j \in \mathbb{N}$ with $L_5 < j \leqslant L_6$ it holds that

$$\begin{aligned}(W_{6,j}, B_{6,j}) &= (W_{4,j+1-L_3}, B_{4,j+1-L_3}) = (W_{1,j+2-L_2-L_3}, B_{1,j+2-L_2-L_3})\\ &= (W_{1,j+1-L_5}, B_{1,j+1-L_5}) = (W_{7,j}, B_{7,j}).\end{aligned} \tag{77}$$

Combining this with (57), (73), (74), (75), and (76) establishes that

$$(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_4 \bullet \Phi_3 = \Phi_6 = \Phi_7 = \Phi_1 \bullet \Phi_5 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3). \tag{78}$$

The proof of Lemma 2.8 is thus completed. □

2.2.3 Compositions of ANNs and affine linear transformations

Corollary 2.9 *Let $\Phi \in \mathbf{N}$ (cf. Definition 2.1). Then*

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

$$\mathcal{P}(\mathbb{A} \bullet \Phi) \leqslant \left[\max\left\{1, \frac{\mathcal{O}(\mathbb{A})}{\mathcal{O}(\Phi)}\right\}\right] \mathcal{P}(\Phi) \tag{79}$$

and

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

$$\mathcal{P}(\Phi \bullet \mathbb{A}) \leqslant \left[\max\left\{1, \frac{\mathcal{I}(\mathbb{A})+1}{\mathcal{I}(\Phi)+1}\right\}\right] \mathcal{P}(\Phi) \tag{80}$$

(cf. Definition 2.5).

Proof of Corollary 2.9 Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\mathbb{A}_1, \mathbb{A}_2 \in \mathbf{N}$ satisfy that $\mathcal{L}(\mathbb{A}_1) = \mathcal{L}(\mathbb{A}_2) = 1$, $\mathcal{I}(\mathbb{A}_1) = \mathcal{O}(\Phi)$, $\mathcal{I}(\Phi) = \mathcal{O}(\mathbb{A}_2)$, and $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$. Observe that item (iv) in Proposition 2.6, the fact that $\mathcal{O}(\Phi) = l_L$, the fact that $\mathcal{I}(\Phi) = l_0$, and the fact that for all $k \in \{1, 2\}$ it holds that $\mathcal{D}(\mathbb{A}_k) = (\mathcal{I}(\mathbb{A}_k), \mathcal{O}(\mathbb{A}_k))$ ensure that

$$\begin{aligned}
\mathcal{P}(\mathbb{A}_1 \bullet \Phi) &= \left[\sum_{m=1}^{L-1} l_m(l_{m-1}+1)\right] + \big[\mathcal{O}(\mathbb{A}_1)\big](l_{L-1}+1) \\
&= \left[\sum_{m=1}^{L-1} l_m(l_{m-1}+1)\right] + \left[\tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right] l_L(l_{L-1}+1) \\
&\leqslant \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right\}\right] \left[\sum_{m=1}^{L-1} l_m(l_{m-1}+1)\right] + \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right\}\right] l_L(l_{L-1}+1) \\
&= \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{l_L}\right\}\right] \left[\sum_{m=1}^{L} l_m(l_{m-1}+1)\right] = \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{\mathcal{O}(\Phi)}\right\}\right] \mathcal{P}(\Phi)
\end{aligned} \tag{81}$$

and

$$\begin{aligned}
\mathcal{P}(\Phi \bullet \mathbb{A}_2) &= \left[\sum_{m=2}^{L} l_m(l_{m-1}+1)\right] + l_1\big[\mathcal{I}(\mathbb{A}_2)+1\big] \\
&= \left[\sum_{m=2}^{L} l_m(l_{m-1}+1)\right] + \left[\tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right] l_1(l_0+1) \\
&\leqslant \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right\}\right] \left[\sum_{m=2}^{L} l_m(l_{m-1}+1)\right] + \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right\}\right] l_1(l_0+1) \\
&= \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{l_0+1}\right\}\right] \left[\sum_{m=1}^{L} l_m(l_{m-1}+1)\right] = \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{\mathcal{I}(\Phi)+1}\right\}\right] \mathcal{P}(\Phi).
\end{aligned} \tag{82}$$

This establishes items (i)–(ii). The proof of Corollary 2.9 is thus completed. □

2.2.4 Powers and extensions of ANNs

Definition 2.10 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 2.11 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

$$\Phi^{\bullet n} = \begin{cases} \big(\mathrm{I}_{\mathcal{O}(\Phi)}, (0, 0, \dots, 0)\big) \in \mathbb{R}^{\mathcal{O}(\Phi)\times\mathcal{O}(\Phi)} \times \mathbb{R}^{\mathcal{O}(\Phi)} & : n = 0 \\ \Phi \bullet (\Phi^{\bullet(n-1)}) & : n \in \mathbb{N} \end{cases} \tag{83}$$

(cf. Definition 2.1, Definition 2.5, and Definition 2.10).

Definition 2.12 (Extension of ANNs) Let $L \in \mathbb{N}$, $\Psi \in \mathbf{N}$ satisfy that $\mathcal{I}(\Psi) = \mathcal{O}(\Psi)$. Then we denote by $\mathcal{E}_{L,\Psi}\colon \{\Phi \in \mathbf{N}\colon (\mathcal{L}(\Phi) \leqslant L \text{ and } \mathcal{O}(\Phi) = \mathcal{I}(\Psi))\} \to \mathbf{N}$ the function which satisfies for all $\Phi \in \mathbf{N}$ with $\mathcal{L}(\Phi) \leqslant L$ and $\mathcal{O}(\Phi) = \mathcal{I}(\Psi)$ that

$$\mathcal{E}_{L,\Psi}(\Phi) = \left(\Psi^{\bullet(L-\mathcal{L}(\Phi))}\right) \bullet \Phi \tag{84}$$

(cf. Definition 2.1, Definition 2.5, and Definition 2.11).

Lemma 2.13 *Let $d, \mathfrak{i} \in \mathbb{N}$, $\Psi \in \mathbf{N}$ satisfy that $\mathcal{D}(\Psi) = (d, \mathfrak{i}, d)$ (cf. Definition 2.1). Then*

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

$$\mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d, d) & : n = 0 \\ (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & : n \in \mathbb{N} \end{cases} \tag{85}$$

and

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

$$\begin{aligned} &\mathcal{P}(\mathcal{E}_{L,\Psi}(\Phi)) \\ &\leqslant \begin{cases} \mathcal{P}(\Phi) & : \mathcal{L}(\Phi) = L \\ \Big[\big(\max\{1, \tfrac{\mathfrak{i}}{d}\}\big)\mathcal{P}(\Phi) + \big((L - \mathcal{L}(\Phi) - 1)\,\mathfrak{i} + d\big)(\mathfrak{i} + 1)\Big] & : \mathcal{L}(\Phi) < L \end{cases} \end{aligned} \tag{86}$$

(cf. Definition 2.11 and Definition 2.12).

Proof of Lemma 2.13 Throughout this proof let $\Phi \in \mathbf{N}$, $l_0, l_1, \ldots, l_{\mathcal{L}(\Phi)} \in \mathbb{N}$ satisfy that $\mathcal{O}(\Phi) = d$ and $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_{\mathcal{L}(\Phi)}) \in \mathbb{N}^{\mathcal{L}(\Phi)+1}$ and let $a_{L,k} \in \mathbb{N}$, $k \in \mathbb{N}_0 \cap [0, L]$, $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$, satisfy for all $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$, $k \in \mathbb{N}_0 \cap [0, L]$ that

$$a_{L,k} = \begin{cases} l_k & : k < \mathcal{L}(\Phi) \\ \mathfrak{i} & : \mathcal{L}(\Phi) \leqslant k < L \\ d & : k = L \end{cases}. \tag{87}$$

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

$$\mathcal{L}(\Psi^{\bullet n}) = n + 1 \qquad \text{and} \qquad \mathbb{N}^{n+2} \ni \mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d, d) & : n = 0 \\ (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & : n \in \mathbb{N} \end{cases}. \tag{88}$$

We now prove (88) by induction on $n \in \mathbb{N}_0$. Note that the fact that $\Psi^{\bullet 0} = (\mathrm{I}_d, 0) \in \mathbb{R}^{d\times d} \times \mathbb{R}^d$ (cf. Definition 2.10) establishes (85) in the base case $n = 0$. For the induction step $\mathbb{N}_0 \ni n \to n+1 \in \mathbb{N}$ assume that there exists $n \in \mathbb{N}_0$ such that

$$\mathcal{L}(\Psi^{\bullet n}) = n+1 \qquad \text{and} \qquad \mathbb{N}^{n+2} \ni \mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d, d) & : n = 0 \\ (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) & : n \in \mathbb{N} \end{cases}. \tag{89}$$

Observe that Lemma 2.4, (83), items (i)–(ii) in Proposition 2.6, (89), and the hypothesis that $\mathcal{D}(\Psi) = (d, \mathfrak{i}, d)$ imply that

$$\begin{aligned} \mathcal{L}(\Psi^{\bullet(n+1)}) &= \mathcal{L}(\Psi \bullet (\Psi^{\bullet n})) = \mathcal{L}(\Psi) + \mathcal{L}(\Psi^{\bullet n}) - 1 = 2 + (n+1) - 1 = (n+1) + 1 \\ \text{and} \qquad & \mathcal{D}(\Psi^{\bullet(n+1)}) = \mathcal{D}(\Psi \bullet (\Psi^{\bullet n})) = (d, \mathfrak{i}, \mathfrak{i}, \ldots, \mathfrak{i}, d) \in \mathbb{N}^{n+3}. \end{aligned} \tag{90}$$

Induction thus proves (88). Next note that (88) establishes item (i). In addition, observe that items (i)–(ii) in Proposition 2.6, item (i), (84), and (87) ensure that for all $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$ it holds that

$$\begin{aligned} \mathcal{L}\big(\mathcal{E}_{L,\Psi}(\Phi)\big) &= \mathcal{L}\big((\Psi^{\bullet(L-\mathcal{L}(\Phi))}) \bullet \Phi\big) = \mathcal{L}\big(\Psi^{\bullet(L-\mathcal{L}(\Phi))}\big) + \mathcal{L}(\Phi) - 1 \\ &= (L - \mathcal{L}(\Phi) + 1) + \mathcal{L}(\Phi) - 1 = L \end{aligned} \tag{91}$$

and

$$\mathcal{D}\big(\mathcal{E}_{L,\Psi}(\Phi)\big) = \mathcal{D}\big((\Psi^{\bullet(L-\mathcal{L}(\Phi))}) \bullet \Phi\big) = (a_{L,0}, a_{L,1}, \ldots, a_{L,L}). \tag{92}$$

Combining this with (87) demonstrates that

$$\mathcal{L}\big(\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi)\big) = \mathcal{L}(\Phi) \tag{93}$$

and

$$\begin{aligned} \mathcal{D}\big(\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi)\big) &= (a_{\mathcal{L}(\Phi),0}, a_{\mathcal{L}(\Phi),1}, \ldots, a_{\mathcal{L}(\Phi),\mathcal{L}(\Phi)}) \\ &= (l_0, l_1, \ldots, l_{\mathcal{L}(\Phi)}) = \mathcal{D}(\Phi). \end{aligned} \tag{94}$$

Hence, we obtain that

$$\mathcal{P}\big(\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi)\big) = \mathcal{P}(\Phi). \tag{95}$$

Next note that (87), (92), and the fact that $l_{\mathcal{L}(\Phi)} = \mathcal{O}(\Phi) = d$ imply that for all $L \in \mathbb{N} \cap (\mathcal{L}(\Phi), \infty)$ it holds that

$$
\begin{aligned}
\mathcal{P}\big(\mathcal{E}_{L,\Psi}(\Phi)\big) &= \sum_{k=1}^{L} a_{L,k}(a_{L,k-1}+1) \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} a_{L,k}(a_{L,k-1}+1)\right] + \left[\sum_{k=\mathcal{L}(\Phi)}^{L} a_{L,k}(a_{L,k-1}+1)\right] \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + \left[\sum_{k=\mathcal{L}(\Phi)}^{\mathcal{L}(\Phi)} a_{L,k}(a_{L,k-1}+1)\right] \\
&\quad + \left[\sum_{k=\mathcal{L}(\Phi)+1}^{L} a_{L,k}(a_{L,k-1}+1)\right] \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + a_{L,\mathcal{L}(\Phi)}(a_{L,\mathcal{L}(\Phi)-1}+1) \\
&\quad + \left[\sum_{k=\mathcal{L}(\Phi)+1}^{L-1} a_{L,k}(a_{L,k-1}+1)\right] + \left[\sum_{k=L}^{L} a_{L,k}(a_{L,k-1}+1)\right] \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + \mathfrak{i}(l_{\mathcal{L}(\Phi)-1}+1) \\
&\quad + \big(L-1-(\mathcal{L}(\Phi)+1)+1\big)\mathfrak{i}(\mathfrak{i}+1) + a_{L,L}(a_{L,L-1}+1) \\
&= \left[\sum_{k=1}^{\mathcal{L}(\Phi)-1} l_k(l_{k-1}+1)\right] + \tfrac{\mathfrak{i}}{d}\big[l_{\mathcal{L}(\Phi)}(l_{\mathcal{L}(\Phi)-1}+1)\big] \\
&\quad + \big(L-\mathcal{L}(\Phi)-1\big)\mathfrak{i}(\mathfrak{i}+1) + d(\mathfrak{i}+1) \\
&\leqslant \left[\max\{1, \tfrac{\mathfrak{i}}{d}\}\right] \left[\sum_{k=1}^{\mathcal{L}(\Phi)} l_k(l_{k-1}+1)\right] + \big(L-\mathcal{L}(\Phi)-1\big)\mathfrak{i}(\mathfrak{i}+1) + d(\mathfrak{i}+1) \\
&= \left[\max\{1, \tfrac{\mathfrak{i}}{d}\}\right] \mathcal{P}(\Phi) + \big(L-\mathcal{L}(\Phi)-1\big)\mathfrak{i}(\mathfrak{i}+1) + d(\mathfrak{i}+1).
\end{aligned} \tag{96}
$$

Combining this with (95) establishes (86). The proof of Lemma 2.13 is thus completed. □

Lemma 2.14 *Let $a \in C(\mathbb{R}, \mathbb{R})$, $\mathbb{I} \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}^{\mathcal{I}(\mathbb{I})}$ that $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\mathbb{I})$ and $(\mathcal{R}_a(\mathbb{I}))(x) = x$ (cf. Definition 2.1 and Definition 2.3). Then*

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

$$\mathcal{R}_a(\mathbb{I}^{\bullet n}) \in C(\mathbb{R}^{\mathcal{I}(\mathbb{I})}, \mathbb{R}^{\mathcal{I}(\mathbb{I})}) \qquad \text{and} \qquad (\mathcal{R}_a(\mathbb{I}^{\bullet n}))(x) = x \tag{97}$$

and

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

$$\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi)) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)}) \quad \text{and} \quad \big(\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi))\big)(x) = \big(\mathcal{R}_a(\Phi)\big)(x) \tag{98}$$

(cf. Definition 2.11 and Definition 2.12).

Proof of Lemma 2.14 Throughout this proof let $\Phi \in \mathbf{N}$, $L, d \in \mathbb{N}$ satisfy that $\mathcal{L}(\Phi) \leqslant L$ and $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\Phi) = d$. We claim that for all $n \in \mathbb{N}_0$ it holds that

$$\mathcal{R}_a(\mathbb{I}^{\bullet n}) \in C(\mathbb{R}^d, \mathbb{R}^d) \qquad \text{and} \qquad \forall\, x \in \mathbb{R}^d\colon\ (\mathcal{R}_a(\mathbb{I}^{\bullet n}))(x) = x. \tag{99}$$

We now prove (99) by induction on $n \in \mathbb{N}_0$. Note that (83) and the fact that $\mathcal{O}(\mathbb{I}) = d$ demonstrate that $\mathcal{R}_a(\mathbb{I}^{\bullet 0}) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and $\forall\, x \in \mathbb{R}^d\colon (\mathcal{R}_a(\mathbb{I}^{\bullet 0}))(x) = x$. This establishes (99) in the base case $n = 0$. For the induction step observe that for all $n \in \mathbb{N}_0$ with $\mathcal{R}_a(\mathbb{I}^{\bullet n}) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and $\forall\, x \in \mathbb{R}^d\colon (\mathcal{R}_a(\mathbb{I}^{\bullet n}))(x) = x$ it holds that

$$\mathcal{R}_a(\mathbb{I}^{\bullet (n+1)}) = \mathcal{R}_a(\mathbb{I} \bullet (\mathbb{I}^{\bullet n})) = (\mathcal{R}_a(\mathbb{I})) \circ (\mathcal{R}_a(\mathbb{I}^{\bullet n})) \in C(\mathbb{R}^d, \mathbb{R}^d) \tag{100}$$

and

$$\begin{aligned} \forall\, x \in \mathbb{R}^d\colon\ \big(\mathcal{R}_a(\mathbb{I}^{\bullet (n+1)})\big)(x) &= \big([\mathcal{R}_a(\mathbb{I})] \circ [\mathcal{R}_a(\mathbb{I}^{\bullet n})]\big)(x) \\ &= (\mathcal{R}_a(\mathbb{I}))\big(\big(\mathcal{R}_a(\mathbb{I}^{\bullet n})\big)(x)\big) = (\mathcal{R}_a(\mathbb{I}))(x) = x. \end{aligned} \tag{101}$$

Induction thus proves (99). Next observe that (99) establishes item (i). Moreover, note that (84), item (v) in Proposition 2.6, item (i), and the fact that $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\Phi)$ ensure that

$$\begin{aligned} &\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi)) = \mathcal{R}_a((\mathbb{I}^{\bullet (L - \mathcal{L}(\Phi))}) \bullet \Phi) \\ &\in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\mathbb{I})}) = C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{I}(\mathbb{I})}) = C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)}) \end{aligned} \tag{102}$$

and

$$\begin{aligned} \forall\, x \in \mathbb{R}^{\mathcal{I}(\Phi)}\colon\ \big(\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}}(\Phi))\big)(x) &= \big(\mathcal{R}_a(\mathbb{I}^{\bullet (L - \mathcal{L}(\Phi))})\big)\big((\mathcal{R}_a(\Phi))(x)\big) \\ &= (\mathcal{R}_a(\Phi))(x). \end{aligned} \tag{103}$$

This establishes item (ii). The proof of Lemma 2.14 is thus completed. □

2.2.5 Compositions of ANNs involving artificial identities

Definition 2.15 (Composition of ANNs involving artificial identities) Let $\Psi \in \mathbf{N}$. Then we denote by

$$(\cdot) \odot_\Psi (\cdot)\colon \{(\Phi_1, \Phi_2) \in \mathbf{N} \times \mathbf{N}\colon \mathcal{I}(\Phi_1) = \mathcal{O}(\Psi) \text{ and } \mathcal{O}(\Phi_2) = \mathcal{I}(\Psi)\} \to \mathbf{N} \tag{104}$$

the function which satisfies for all $\Phi_1, \Phi_2 \in \mathbf{N}$ with $\mathcal{I}(\Phi_1) = \mathcal{O}(\Psi)$ and $\mathcal{O}(\Phi_2) = \mathcal{I}(\Psi)$ that

$$\Phi_1 \odot_\Psi \Phi_2 = \Phi_1 \bullet (\Psi \bullet \Phi_2) = (\Phi_1 \bullet \Psi) \bullet \Phi_2 \tag{105}$$

(cf. Definition 2.1, Definition 2.5, and Lemma 2.8).

Proposition 2.16 *Let* $\Psi, \Phi_1, \Phi_2 \in \mathbf{N}$, $\mathfrak{i}, l_{1,0}, l_{1,1}, \ldots, l_{1,\mathcal{L}(\Phi_1)}, l_{2,0}, l_{2,1}, \ldots, l_{2,\mathcal{L}(\Phi_2)} \in \mathbb{N}$ *satisfy for all* $k \in \{1, 2\}$ *that* $\mathcal{D}(\Psi) = (\mathcal{I}(\Psi), \mathfrak{i}, \mathcal{O}(\Psi))$, $\mathcal{I}(\Phi_1) = \mathcal{O}(\Psi)$, $\mathcal{O}(\Phi_2) = \mathcal{I}(\Psi)$, *and* $\mathcal{D}(\Phi_k) = (l_{k,0}, l_{k,1}, \ldots, l_{k,\mathcal{L}(\Phi_k)})$ *(cf. Definition 2.1). Then*

(i) *it holds that*

$$\mathcal{D}(\Phi_1 \odot_\Psi \Phi_2) = (l_{2,0}, l_{2,1}, \ldots, l_{2,\mathcal{L}(\Phi_2)-1}, \mathfrak{i}, l_{1,1}, l_{1,2}, \ldots, l_{1,\mathcal{L}(\Phi_1)}), \tag{106}$$

(ii) *it holds that*

$$\mathcal{L}(\Phi_1 \odot_\Psi \Phi_2) = \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2), \tag{107}$$

(iii) *it holds that*

$$\mathcal{P}(\Phi_1 \odot_\Psi \Phi_2) \leqslant \left[\max\left\{1, \tfrac{\mathfrak{i}}{\mathcal{I}(\Psi)}, \tfrac{\mathfrak{i}}{\mathcal{O}(\Psi)}\right\}\right] \big(\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2)\big), \tag{108}$$

and

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

$$\mathcal{R}_a(\Phi_1 \odot_\Psi \Phi_2) = [\mathcal{R}_a(\Phi_1)] \circ [\mathcal{R}_a(\Psi)] \circ [\mathcal{R}_a(\Phi_2)] \tag{109}$$

(cf. Definition 2.3 and Definition 2.15).

Proof of Propositions 2.16 Throughout this proof let $a \in C(\mathbb{R}, \mathbb{R})$, $L_1, L_2 \in \mathbb{N}$ satisfy that $L_1 = \mathcal{L}(\Phi_1)$ and $L_2 = \mathcal{L}(\Phi_2)$. Note that item (i) in Proposition 2.6, the hypothesis that $\mathcal{D}(\Phi_2) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2})$, the hypothesis that $\mathcal{D}(\Psi) = (\mathcal{I}(\Psi), \mathfrak{i}, \mathcal{O}(\Psi))$, and the hypothesis that $\mathcal{I}(\Psi) = \mathcal{O}(\Phi_2)$ show that

$$\mathcal{D}(\Psi \bullet \Phi_2) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, \mathfrak{i}, \mathcal{O}(\Psi)) \tag{110}$$

(cf. Definition 2.5). Combining this with item (i) in Proposition 2.6, the hypothesis that $\mathcal{D}(\Phi_1) = (l_{1,0}, l_{1,1}, \ldots, l_{1,L_1})$, and the hypothesis that $\mathcal{I}(\Phi_1) = \mathcal{O}(\Psi)$ proves that

$$\mathcal{D}(\Phi_1 \odot_\Psi \Phi_2) = \mathcal{D}\big(\Phi_1 \bullet (\Psi \bullet \Phi_2)\big) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, \mathfrak{i}, l_{1,1}, l_{1,2}, \ldots, l_{1,L_1}). \tag{111}$$

This establishes item (i). Moreover, observe that item (ii) in Proposition 2.6 and the fact that $\mathcal{L}(\Psi) = 2$ ensure that

$$\begin{aligned} \mathcal{L}(\Phi_1 \odot_\Psi \Phi_2) &= \mathcal{L}\big(\Phi_1 \bullet (\Psi \bullet \Phi_2)\big) = \mathcal{L}(\Phi_1) + \mathcal{L}(\Psi \bullet \Phi_2) - 1 \\ &= \mathcal{L}(\Phi_1) + \mathcal{L}(\Psi) + \mathcal{L}(\Phi_2) - 2 = \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2). \end{aligned} \tag{112}$$

This establishes item (ii). In addition, observe that (111), the fact that $\mathcal{I}(\Psi) = \mathcal{O}(\Phi_2) = l_{2,L_2}$, and the fact that $\mathcal{O}(\Psi) = \mathcal{I}(\Phi_1) = l_{1,0}$ demonstrate that

$$\begin{aligned}
\mathcal{P}(\Phi_1 \odot_\Psi \Phi_2) &= \left[\sum_{m=1}^{L_2-1} l_{2,m}(l_{2,m-1}+1)\right] + \left[\sum_{m=2}^{L_1} l_{1,m}(l_{1,m-1}+1)\right] \\
&\quad + \mathfrak{i}\big(l_{2,L_2-1}+1\big) + l_{1,1}(\mathfrak{i}+1) \\
&= \left[\sum_{m=1}^{L_2-1} l_{2,m}(l_{2,m-1}+1)\right] + \left[\sum_{m=2}^{L_1} l_{1,m}(l_{1,m-1}+1)\right] \\
&\quad + \tfrac{\mathfrak{i}}{\mathcal{I}(\Psi)}\, l_{2,L_2}\big(l_{2,L_2-1}+1\big) + l_{1,1}\big(\tfrac{\mathfrak{i}}{\mathcal{O}(\Psi)}\, l_{1,0}+1\big) \\
&\leqslant \left[\max\left\{1, \tfrac{\mathfrak{i}}{\mathcal{I}(\Psi)}\right\}\right] \left[\sum_{m=1}^{L_2} l_{2,m}(l_{2,m-1}+1)\right] \\
&\quad + \left[\max\left\{1, \tfrac{\mathfrak{i}}{\mathcal{O}(\Psi)}\right\}\right] \left[\sum_{m=1}^{L_1} l_{1,m}(l_{1,m-1}+1)\right] \\
&\leqslant \left[\max\left\{1, \tfrac{\mathfrak{i}}{\mathcal{I}(\Psi)}, \tfrac{\mathfrak{i}}{\mathcal{O}(\Psi)}\right\}\right] \big(\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2)\big).
\end{aligned} \tag{113}$$

This establishes item (iii). Next note that item (v) in Proposition 2.6 implies that

$$\begin{aligned}
\mathcal{R}_a(\Phi_1 \odot_\Psi \Phi_2) &= \mathcal{R}_a\big(\Phi_1 \bullet (\Psi \bullet \Phi_2)\big) \\
&= \big[\mathcal{R}_a(\Phi_1)\big] \circ \big[\mathcal{R}_a(\Psi \bullet \Phi_2)\big] \\
&= \big(\big[\mathcal{R}_a(\Phi_1)\big] \circ \big[\mathcal{R}_a(\Psi)\big] \circ \big[\mathcal{R}_a(\Phi_2)\big]\big) \in C(\mathbb{R}^{\mathcal{I}(\Phi_2)}, \mathbb{R}^{\mathcal{O}(\Phi_1)}).
\end{aligned} \tag{114}$$

This establishes item (iv). The proof of Proposition 2.16 is thus completed. □

2.3 Parallelizations of ANNs

2.3.1 Parallelizations of ANNs with the same length

Definition 2.17 (Parallelization of ANNs with the same length) Let $n \in \mathbb{N}$. Then we denote by

$$\mathbf{P}_n \colon \big\{(\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n \colon \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)\big\} \to \mathbf{N} \tag{115}$$

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

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

$$\mathbf{P}_n(\Phi_1, \Phi_2, \ldots, \Phi_n) = \left(\left(\begin{pmatrix} W_{1,1} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,1} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,1} \end{pmatrix}, \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ B_{3,1} \\ \vdots \\ B_{n,1} \end{pmatrix} \right), \left(\begin{pmatrix} W_{1,2} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,2} \end{pmatrix}, \begin{pmatrix} B_{1,2} \\ B_{2,2} \\ B_{3,2} \\ \vdots \\ B_{n,2} \end{pmatrix} \right), \ldots, \left(\begin{pmatrix} W_{1,L} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,L} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,L} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,L} \end{pmatrix}, \begin{pmatrix} B_{1,L} \\ B_{2,L} \\ B_{3,L} \\ \vdots \\ B_{n,L} \end{pmatrix} \right) \right) \tag{116}$$

(cf. Definition 2.1).

Lemma 2.18 *Let $n, L \in \mathbb{N}$, $(l_{1,0}, l_{1,1}, \ldots, l_{1,L})$, $(l_{2,0}, l_{2,1}, \ldots, l_{2,L})$, $\ldots$, $(l_{n,0}, l_{n,1}, \ldots, l_{n,L}) \in \mathbb{N}^{L+1}$, $\Phi_1 = ((W_{1,1}, B_{1,1}), (W_{1,2}, B_{1,2}), \ldots, (W_{1,L}, B_{1,L})) \in (\times_{k=1}^{L}(\mathbb{R}^{l_{1,k}\times l_{1,k-1}} \times \mathbb{R}^{l_{1,k}}))$, $\Phi_2 = ((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \ldots, (W_{2,L}, B_{2,L})) \in (\times_{k=1}^{L}(\mathbb{R}^{l_{2,k}\times l_{2,k-1}} \times \mathbb{R}^{l_{2,k}}))$, $\ldots$, $\Phi_n = ((W_{n,1}, B_{n,1}), (W_{n,2}, B_{n,2}), \ldots, (W_{n,L}, B_{n,L})) \in (\times_{k=1}^{L}(\mathbb{R}^{l_{n,k}\times l_{n,k-1}} \times \mathbb{R}^{l_{n,k}}))$. Then it holds that*

$$\mathbf{P}_n(\Phi_1, \Phi_2, \ldots, \Phi_n) \in \left(\times_{k=1}^{L} \left(\mathbb{R}^{(\sum_{j=1}^{n} l_{j,k})\times(\sum_{j=1}^{n} l_{j,k-1})} \times \mathbb{R}^{(\sum_{j=1}^{n} l_{j,k})}\right)\right) \tag{117}$$

(cf. Definition 2.17).

Proof of Lemma 2.18 Note that (116) establishes (117). The proof of Lemma 2.18 is thus completed. □

Proposition 2.19 *Let $a \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n$ satisfy that $\mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \ldots = \mathcal{L}(\Phi_n)$ (cf. Definition 2.1). Then*

(i) *it holds that*

$$\mathcal{R}_a(\mathbf{P}_n(\Phi)) \in C\left(\mathbb{R}^{[\sum_{j=1}^{n} \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^{n} \mathcal{O}(\Phi_j)]}\right) \tag{118}$$

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*

$$\begin{aligned} &\big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(x_1, x_2, \ldots, x_n) \\ &= \big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \ldots, (\mathcal{R}_a(\Phi_n))(x_n)\big) \in \mathbb{R}^{[\sum_{j=1}^{n} \mathcal{O}(\Phi_j)]} \end{aligned} \tag{119}$$

(cf. Definition 2.3 and Definition 2.17).

Proof of Proposition 2.19 Throughout this proof let $L \in \mathbb{N}$ satisfy that $L = \mathcal{L}(\Phi_1)$, let $l_{j,0}, l_{j,1}, \ldots, l_{j,L} \in \mathbb{N}$, $j \in \{1, 2, \ldots, n\}$, satisfy for all $j \in \{1, 2, \ldots, n\}$ that $\mathcal{D}(\Phi_j) = (l_{j,0}, l_{j,1}, \ldots, l_{j,L})$, let $\big((W_{j,1}, B_{j,1}), (W_{j,2}, B_{j,2}), \ldots, (W_{j,L}, B_{j,L})\big) \in (\times_{k=1}^{L}(\mathbb{R}^{l_{j,k} \times l_{j,k-1}} \times \mathbb{R}^{l_{j,k}}))$, $j \in \{1, 2, \ldots, n\}$, satisfy for all $j \in \{1, 2, \ldots, n\}$ that

$$\Phi_j = \big((W_{j,1}, B_{j,1}), (W_{j,2}, B_{j,2}), \ldots, (W_{j,L}, B_{j,L})\big), \tag{120}$$

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

$$\mathbf{P}_n(\Phi) = \big((A_1, b_1), (A_2, b_2), \ldots, (A_L, b_L)\big) \tag{121}$$

(cf. Lemma 2.18), 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

$$x_{j,k} = \mathfrak{M}_{a,l_{j,k}}(W_{j,k} x_{j,k-1} + B_{j,k}) \tag{122}$$

(cf. Definition 2.2), 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 (121) 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 2.4, 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

$$\begin{aligned}\mathcal{R}_a(\mathbf{P}_n(\Phi)) \in C(\mathbb{R}^{\alpha_0}, \mathbb{R}^{\alpha_L}) &= C\big(\mathbb{R}^{[\sum_{j=1}^{n} l_{j,0}]}, \mathbb{R}^{[\sum_{j=1}^{n} l_{j,L}]}\big) \\ &= C\big(\mathbb{R}^{[\sum_{j=1}^{n} \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^{n} \mathcal{O}(\Phi_j)]}\big).\end{aligned} \tag{123}$$

This proves item (i). Moreover, observe that (116) and (121) demonstrate that for all $k \in \{1, 2, \ldots, L\}$ it holds that

$$A_k = \begin{pmatrix} W_{1,k} & 0 & 0 & \cdots & 0 \\ 0 & W_{2,k} & 0 & \cdots & 0 \\ 0 & 0 & W_{3,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & W_{n,k} \end{pmatrix} \quad \text{and} \quad b_k = \begin{pmatrix} B_{1,k} \\ B_{2,k} \\ B_{3,k} \\ \vdots \\ B_{n,k} \end{pmatrix}. \tag{124}$$

Combining this with (21), (122), and the fact that for all $k \in \mathbb{N} \cap [0, L)$ it holds that $\mathfrak{x}_k = (x_{1,k}, x_{2,k}, \ldots, x_{n,k})$ implies that for all $k \in \mathbb{N} \cap (0, L)$ it holds that

$$\mathfrak{M}_{a,\alpha_k}(A_k \mathfrak{x}_{k-1} + b_k) = \begin{pmatrix} \mathfrak{M}_{a,l_{1,k}}(W_{1,k}x_{1,k-1} + B_{1,k}) \\ \mathfrak{M}_{a,l_{2,k}}(W_{2,k}x_{2,k-1} + B_{2,k}) \\ \vdots \\ \mathfrak{M}_{a,l_{n,k}}(W_{n,k}x_{n,k-1} + B_{n,k}) \end{pmatrix} = \begin{pmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{pmatrix} = \mathfrak{x}_k. \tag{125}$$

This, (22), (120), (121), (122), (124), the fact that $\mathfrak{x}_0 = (x_{1,0}, x_{2,0}, \ldots, x_{n,0})$, and the fact that $\mathfrak{x}_{L-1} = (x_{1,L-1}, x_{2,L-1}, \ldots, x_{n,L-1})$ ensure that

$$\begin{aligned} &\big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(x_{1,0}, x_{2,0}, \ldots, x_{n,0}) = \big(\mathcal{R}_a\big(\mathbf{P}_n(\Phi)\big)\big)(\mathfrak{x}_0) \\ &= A_L \mathfrak{x}_{L-1} + b_L = \begin{pmatrix} W_{1,L}x_{1,L-1} + B_{1,L} \\ W_{2,L}x_{2,L-1} + B_{2,L} \\ \vdots \\ W_{n,L}x_{n,L-1} + B_{n,L} \end{pmatrix} = \begin{pmatrix} (\mathcal{R}_a(\Phi_1))(x_{1,0}) \\ (\mathcal{R}_a(\Phi_2))(x_{2,0}) \\ \vdots \\ (\mathcal{R}_a(\Phi_n))(x_{n,0}) \end{pmatrix}. \end{aligned} \tag{126}$$

This establishes item (ii). The proof of Proposition 2.19 is thus completed. □

Proposition 2.20 *Let $n, L \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n$, $(l_{1,0}, l_{1,1}, \ldots, l_{1,L})$, $(l_{2,0}, l_{2,1}, \ldots, l_{2,L}), \ldots, (l_{n,0}, l_{n,1}, \ldots, l_{n,L}) \in \mathbb{N}^{L+1}$ satisfy for all $j \in \{1, 2, \ldots, n\}$ that $\mathcal{D}(\Phi_j) = (l_{j,0}, l_{j,1}, \ldots, l_{j,L})$ (cf. Definition 2.1). Then*

(i) *it holds that*

$$\mathcal{D}\big(\mathbf{P}_n(\Phi)\big) = \big(\textstyle\sum_{j=1}^n l_{j,0}, \sum_{j=1}^n l_{j,1}, \ldots, \sum_{j=1}^n l_{j,L}\big) \tag{127}$$

and

(ii) *it holds that*

$$\mathcal{P}(\mathbf{P}_n(\Phi)) \leqslant \tfrac{1}{2}\big[\textstyle\sum_{j=1}^n \mathcal{P}(\Phi_j)\big]^2 \tag{128}$$

(cf. Definition 2.17).

Proof of Proposition 2.20 Note that the hypothesis that $\forall\, j \in \{1, 2, \ldots, n\} \colon \mathcal{D}(\Phi_j) = (l_{j,0}, l_{j,1}, \ldots, l_{j,L})$ and Lemma 2.18 assure that

$$\mathcal{D}\big(\mathbf{P}_n(\Phi)\big) = \big(\textstyle\sum_{j=1}^n l_{j,0}, \sum_{j=1}^n l_{j,1}, \ldots, \sum_{j=1}^n l_{j,L}\big). \tag{129}$$

This establishes item (i). Moreover, observe that (129) demonstrates that

$$
\begin{aligned}
\mathcal{P}(\mathbf{P}_n(\Phi)) &= \sum_{k=1}^{L} \Big[\textstyle\sum_{i=1}^{n} l_{i,k}\Big]\Big[\big(\textstyle\sum_{i=1}^{n} l_{i,k-1}\big) + 1\Big] \\
&= \sum_{k=1}^{L} \Big[\textstyle\sum_{i=1}^{n} l_{i,k}\Big]\Big[\big(\textstyle\sum_{j=1}^{n} l_{j,k-1}\big) + 1\Big] \\
&\leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{L} l_{i,k}(l_{j,k-1}+1) \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k,\ell=1}^{L} l_{i,k}(l_{j,\ell-1}+1) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \Big[\textstyle\sum_{k=1}^{L} l_{i,k}\Big]\Big[\textstyle\sum_{\ell=1}^{L} (l_{j,\ell-1}+1)\Big] \\
&\leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} \Big[\textstyle\sum_{k=1}^{L} \tfrac{1}{2} l_{i,k}(l_{i,k-1}+1)\Big]\Big[\textstyle\sum_{\ell=1}^{L} l_{j,\ell}(l_{j,\ell-1}+1)\Big] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \tfrac{1}{2}\mathcal{P}(\Phi_i)\mathcal{P}(\Phi_j) = \tfrac{1}{2}\Big[\textstyle\sum_{i=1}^{n} \mathcal{P}(\Phi_i)\Big]^2.
\end{aligned} \tag{130}
$$

The proof of Proposition 2.20 is thus completed. □

Corollary 2.21 *Let $n \in \mathbb{N}$, $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n$ satisfy that $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_n)$ (cf. Definition 2.1). Then it holds that $\mathcal{P}(\mathbf{P}_n(\Phi)) \leqslant n^2\mathcal{P}(\Phi_1)$ (cf. Definition 2.17).*

Proof of Corollary 2.21 Throughout this proof let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$ satisfy that $\mathcal{D}(\Phi_1) = (l_0, l_1, \ldots, l_L)$. Note that item (i) in Proposition 2.20 and the fact that $\forall\, j \in \{1, 2, \ldots, n\} \colon \mathcal{D}(\Phi_j) = (l_0, l_1, \ldots, l_L)$ demonstrate that

$$
\begin{aligned}
\mathcal{P}(\mathbf{P}_n(\Phi_1, \Phi_2, \ldots, \Phi_n)) &= \sum_{j=1}^{L} (nl_j)\big((nl_{j-1}) + 1\big) \leqslant \sum_{j=1}^{L} (nl_j)\big((nl_{j-1}) + n\big) \\
&= n^2\Bigg[\sum_{j=1}^{L} l_j(l_{j-1}+1)\Bigg] = n^2\mathcal{P}(\Phi_1).
\end{aligned} \tag{131}
$$

The proof of Corollary 2.21 is thus completed. □

2.3.2 Parallelizations of ANNs with different lengths

Definition 2.22 (Parallelization of ANNs with different length) Let $n \in \mathbb{N}$, $\Psi = (\Psi_1, \Psi_2, \ldots, \Psi_n) \in \mathbf{N}^n$ satisfy for all $j \in \{1, 2, \ldots, n\}$ that $\mathcal{H}(\Psi_j) = 1$ and $\mathcal{I}(\Psi_j) = \mathcal{O}(\Psi_j)$. Then we denote by

$$
\mathrm{P}_{n,\Psi} \colon \{(\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n \colon (\forall\, j \in \{1, 2, \ldots, n\} \colon \mathcal{O}(\Phi_j) = \mathcal{I}(\Psi_j))\} \to \mathbf{N} \tag{132}
$$

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

$$\mathrm{P}_{n,\Psi}(\Phi) = \mathbf{P}_n\big(\mathcal{E}_{\max_{k\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_k),\Psi_1}(\Phi_1), \dots, \mathcal{E}_{\max_{k\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_k),\Psi_n}(\Phi_n)\big) \tag{133}$$

(cf. Definition 2.1, Definition 2.12, Lemma 2.13, and Definition 2.17).

Corollary 2.23 *Let $a \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $\mathbb{I} = (\mathbb{I}_1, \mathbb{I}_2, \dots, \mathbb{I}_n)$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy for all $j \in \{1, 2, \dots, n\}$, $x \in \mathbb{R}^{\mathcal{O}(\Phi_j)}$ that $\mathcal{H}(\mathbb{I}_j) = 1$, $\mathcal{I}(\mathbb{I}_j) = \mathcal{O}(\mathbb{I}_j) = \mathcal{O}(\Phi_j)$, and $(\mathcal{R}_a(\mathbb{I}_j))(x) = x$ (cf. Definition 2.1 and Definition 2.3). Then*

(i) *it holds that*

$$\mathcal{R}_a\big(\mathrm{P}_{n,\mathbb{I}}(\Phi)\big) \in C\big(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\big) \tag{134}$$

and

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

$$\begin{aligned}&\big(\mathcal{R}_a(\mathrm{P}_{n,\mathbb{I}}(\Phi))\big)(x_1, x_2, \dots, x_n)\\ &= \big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \dots, (\mathcal{R}_a(\Phi_n))(x_n)\big) \in \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\end{aligned} \tag{135}$$

(cf. Definition 2.22).

Proof of Corollary 2.23 Throughout this proof let $L \in \mathbb{N}$ satisfy that $L = \max_{j\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_j)$. Note that item (ii) in Lemma 2.13, the hypothesis that for all $j \in \{1, 2, \dots, n\}$ it holds that $\mathcal{H}(\mathbb{I}_j) = 1$, (84), (28), and item (ii) in Lemma 2.14 demonstrate

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

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

$$\big(\mathcal{R}_a(\mathcal{E}_{L,\mathbb{I}_j}(\Phi_j))\big)(x) = (\mathcal{R}_a(\Phi_j))(x) \tag{136}$$

(cf. Definition 2.12). Items (i)–(ii) in Proposition 2.19 therefore imply

(A) that

$$\mathcal{R}_a\big(\mathbf{P}_n\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1), \mathcal{E}_{L,\mathbb{I}_2}(\Phi_2), \dots, \mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big) \in C\big(\mathbb{R}^{[\sum_{j=1}^n \mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^n \mathcal{O}(\Phi_j)]}\big) \tag{137}$$

and

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

$$\begin{aligned}&\big(\mathcal{R}_a\big(\mathbf{P}_n\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1), \mathcal{E}_{L,\mathbb{I}_2}(\Phi_2), \dots, \mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big)\big)\big)(x_1, x_2, \dots, x_n)\\ &= \Big(\big(\mathcal{R}_a\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1)\big)\big)(x_1), \big(\mathcal{R}_a\big(\mathcal{E}_{L,\mathbb{I}_2}(\Phi_2)\big)\big)(x_2), \dots, \big(\mathcal{R}_a\big(\mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big)\big)(x_n)\Big)\\ &= \Big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \dots, (\mathcal{R}_a(\Phi_n))(x_n)\Big)\end{aligned} \tag{138}$$

(cf. Definition 2.17). Combining this with (133) and the fact that $L = \max_{j\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_j)$ ensures

(C) that

$$\mathcal{R}_a\big(\mathrm{P}_{n,\mathbb{I}}(\Phi)\big) \in C\big(\mathbb{R}^{[\sum_{j=1}^{n}\mathcal{I}(\Phi_j)]}, \mathbb{R}^{[\sum_{j=1}^{n}\mathcal{O}(\Phi_j)]}\big) \tag{139}$$

and

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

$$\begin{aligned}&\big(\mathcal{R}_a\big(\mathrm{P}_{n,\mathbb{I}}(\Phi)\big)\big)(x_1, x_2, \dots, x_n)\\ &= \big(\mathcal{R}_a\big(\mathbf{P}_n\big(\mathcal{E}_{L,\mathbb{I}_1}(\Phi_1), \mathcal{E}_{L,\mathbb{I}_2}(\Phi_2), \dots, \mathcal{E}_{L,\mathbb{I}_n}(\Phi_n)\big)\big)\big)(x_1, x_2, \dots, x_n)\\ &= \Big((\mathcal{R}_a(\Phi_1))(x_1), (\mathcal{R}_a(\Phi_2))(x_2), \dots, (\mathcal{R}_a(\Phi_n))(x_n)\Big).\end{aligned} \tag{140}$$

This establishes items (i)–(ii). The proof of Corollary 2.23 is thus completed. □

Corollary 2.24 *Let $n, L \in \mathbb{N}$, $\mathfrak{i}_1, \mathfrak{i}_2, \dots, \mathfrak{i}_n \in \mathbb{N}$, $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_n)$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy for all $j \in \{1, 2, \dots, n\}$ that $\mathcal{D}(\Psi_j) = (\mathcal{O}(\Phi_j), \mathfrak{i}_j, \mathcal{O}(\Phi_j))$ and $L = \max_{k\in\{1,2,\dots,n\}} \mathcal{L}(\Phi_k)$ (cf. Definition 2.1). Then it holds that*

$$\begin{aligned}&\mathcal{P}\big(\mathrm{P}_{n,\Psi}(\Phi)\big)\\ &\leqslant \tfrac{1}{2}\bigg(\Big[\textstyle\sum_{j=1}^{n}\big[\max\big\{1, \tfrac{\mathfrak{i}_j}{\mathcal{O}(\Phi_j)}\big\}\big]\,\mathcal{P}(\Phi_j)\,\mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big]\\ &\quad + \Big[\textstyle\sum_{j=1}^{n}\big((L - \mathcal{L}(\Phi_j) - 1)\,\mathfrak{i}_j\,(\mathfrak{i}_j + 1) + \mathcal{O}(\Phi_j)\,(\mathfrak{i}_j + 1)\big)\,\mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big]\\ &\quad + \Big[\textstyle\sum_{j=1}^{n}\mathcal{P}(\Phi_j)\,\mathbb{1}_{\{\mathcal{L}(\Phi_j)\}}(L)\Big]\bigg)^2\end{aligned} \tag{141}$$

(cf. Definition 2.22).

Proof of Corollary 2.24 Observe that (133), item (ii) in Proposition 2.20, and item (ii) in Lemma 2.13 assure that

$$\begin{aligned}&\mathcal{P}\big(\mathrm{P}_{n,\Psi}(\Phi)\big)\\ &= \mathcal{P}\big(\mathbf{P}_n\big(\mathcal{E}_{L,\Psi_1}(\Phi_1), \mathcal{E}_{L,\Psi_2}(\Phi_2), \dots, \mathcal{E}_{L,\Psi_n}(\Phi_n)\big)\big)\\ &\leqslant \tfrac{1}{2}\Big[\textstyle\sum_{j=1}^{n}\mathcal{P}(\mathcal{E}_{L,\Psi_j}(\Phi_j))\Big]^2\\ &\leqslant \tfrac{1}{2}\bigg(\Big[\textstyle\sum_{j=1}^{n}\big[\max\big\{1, \tfrac{\mathfrak{i}_j}{\mathcal{O}(\Phi_j)}\big\}\big]\,\mathcal{P}(\Phi_j)\,\mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big]\\ &\quad + \Big[\textstyle\sum_{j=1}^{n}\big((L - \mathcal{L}(\Phi_j) - 1)\,\mathfrak{i}_j\,(\mathfrak{i}_j + 1) + \mathcal{O}(\Phi_j)\,(\mathfrak{i}_j + 1)\big)\,\mathbb{1}_{(\mathcal{L}(\Phi_j),\infty)}(L)\Big]\\ &\quad + \Big[\textstyle\sum_{j=1}^{n}\mathcal{P}(\Phi_j)\,\mathbb{1}_{\{\mathcal{L}(\Phi_j)\}}(L)\Big]\bigg)^2\end{aligned} \tag{142}$$

(cf. Definition 2.12 and Definition 2.17). The proof of Corollary 2.24 is thus completed. □

2.4 Sums of ANNs

2.4.1 Sums of ANNs with the same length

Proposition 2.25 *Let $a \in C(\mathbb{R}, \mathbb{R})$, $M \in \mathbb{N}$, $h_1, h_2, \ldots, h_M \in \mathbb{R}$, $\Phi_1, \Phi_2, \ldots, \Phi_M \in \mathbf{N}$ satisfy that $\mathcal{D}(\Phi_1) = \mathcal{D}(\Phi_2) = \ldots = \mathcal{D}(\Phi_M)$ (cf. Definition 2.1). Then there exists $\Psi \in \mathbf{N}$ such that*

(i) *it holds that $\mathcal{P}(\Psi) \leqslant M^2 \mathcal{P}(\Phi_1)$,*

(ii) *it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^{\mathcal{I}(\Phi_1)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$, and*

(iii) *it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi_1)}$ that*

$$(\mathcal{R}_a(\Psi))(x) = \sum_{m=1}^{M} h_m \, (\mathcal{R}_a(\Phi_m))(x) \tag{143}$$

(cf. Definition 2.3).

Proof of Proposition 2.25 Throughout this proof let $d, \mathfrak{d} \in \mathbb{N}$ satisfy that $\mathcal{I}(\Phi_1) = d$ and $\mathcal{O}(\Phi_1) = \mathfrak{d}$, let $(A_1, b_1) \in \mathbb{R}^{\mathfrak{d} \times (M\mathfrak{d})} \times \mathbb{R}^{\mathfrak{d}}$, $(A_2, b_2) \in \mathbb{R}^{(Md) \times d} \times \mathbb{R}^{Md}$ satisfy that

$$A_1 = \begin{pmatrix} h_1 \, \mathrm{I}_{\mathfrak{d}} & h_2 \, \mathrm{I}_{\mathfrak{d}} & \ldots & h_M \, \mathrm{I}_{\mathfrak{d}} \end{pmatrix}, \quad A_2 = \begin{pmatrix} \mathrm{I}_d \\ \mathrm{I}_d \\ \vdots \\ \mathrm{I}_d \end{pmatrix}, \quad b_1 = 0, \quad \text{and} \quad b_2 = 0 \tag{144}$$

(cf. Definition 2.10), let $\mathbb{A}_1, \mathbb{A}_2 \in \mathbf{N}$ satisfy that $\mathbb{A}_1 = (A_1, b_1)$ and $\mathbb{A}_2 = (A_2, b_2)$, and let $\Psi \in \mathbf{N}$ satisfy that

$$\Psi = \mathbb{A}_1 \bullet \big[\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M)\big] \bullet \mathbb{A}_2 \tag{145}$$

(cf. Definition 2.5, Definition 2.17, Lemma 2.8, and Proposition 2.19). Note that (145) and items (i)–(ii) in Corollary 2.9 demonstrate that

$$\begin{aligned}
\mathcal{P}(\Psi) &\leqslant \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{\mathcal{O}(\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M))}\right\}\right] \mathcal{P}\big(\big[\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M)\big] \bullet \mathbb{A}_2\big) \\
&\leqslant \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{\mathcal{O}(\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M))}\right\}\right] \left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{\mathcal{I}(\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M))+1}\right\}\right] \\
&\quad \cdot \mathcal{P}\big(\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M)\big) \\
&= \left[\max\left\{1, \tfrac{\mathfrak{d}}{M\mathfrak{d}}\right\}\right] \left[\max\left\{1, \tfrac{d+1}{Md+1}\right\}\right] \mathcal{P}\big(\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M)\big) \\
&= \mathcal{P}\big(\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M)\big).
\end{aligned} \tag{146}$$

Corollary 2.21 and the hypothesis that for all $m \in \{1, 2, \ldots, M\}$ it holds that $\mathcal{D}(\Phi_m) = \mathcal{D}(\Phi_1)$ hence prove that

$$\mathcal{P}(\Psi) \leqslant \mathcal{P}\big(\mathbf{P}_M(\Phi_1, \Phi_2, \ldots, \Phi_M)\big) \leqslant M^2 \mathcal{P}(\Phi_1). \tag{147}$$

Next note that (144) and the fact that $\mathbb{A}_2 = (A_2, b_2)$ prove that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\mathbb{A}_2) \in C(\mathbb{R}^d, \mathbb{R}^{Md})$ and $(\mathcal{R}_a(\mathbb{A}_2))(x) = (x, x, \ldots, x) \in \mathbb{R}^{Md}$.

Proposition 2.19 and item (v) in Proposition 2.6 therefore ensure that

$$\begin{aligned}\mathcal{R}_a\big((\mathbf{P}_M(\Phi_1,\Phi_2,\ldots,\Phi_M))\bullet\mathbb{A}_2\big) &= \big(\mathcal{R}_a(\mathbf{P}_M(\Phi_1,\Phi_2,\ldots,\Phi_M))\big)\circ(\mathcal{R}_a(\mathbb{A}_2))\\ &\in C\big(\mathbb{R}^{\mathcal{I}(\mathbb{A}_2)},\mathbb{R}^{\mathcal{O}(\mathbf{P}_M(\Phi_1,\Phi_2,\ldots,\Phi_M))}\big)\\ &= C\big(\mathbb{R}^d,\mathbb{R}^{\mathcal{O}(\Phi_1)+\mathcal{O}(\Phi_2)+\ldots+\mathcal{O}(\Phi_M)}\big)\\ &= C(\mathbb{R}^d,\mathbb{R}^{M\mathfrak{d}})\end{aligned} \tag{148}$$

and

$$\begin{aligned}\forall\, x\in\mathbb{R}^d\colon\ &\big(\mathcal{R}_a\big((\mathbf{P}_M(\Phi_1,\Phi_2,\ldots,\Phi_M))\bullet\mathbb{A}_2\big)\big)(x)\\ &= \big(\big[\mathcal{R}_a(\mathbf{P}_M(\Phi_1,\Phi_2,\ldots,\Phi_M))\big]\circ[\mathcal{R}_a(\mathbb{A}_2)]\big)(x)\\ &= \big(\mathcal{R}_a(\mathbf{P}_M(\Phi_1,\Phi_2,\ldots,\Phi_M))\big)(x,x,\ldots,x)\\ &= \big((\mathcal{R}_a(\Phi_1))(x),(\mathcal{R}_a(\Phi_2))(x),\ldots,(\mathcal{R}_a(\Phi_M))(x)\big).\end{aligned} \tag{149}$$

Furthermore, observe that (144) and the fact that $\mathbb{A}_1 = (A_1, b_1)$ assure that for all $y_1, y_2, \ldots, y_M \in \mathbb{R}^{\mathfrak{d}}$ it holds that $\mathcal{R}_a(\mathbb{A}_1) \in C(\mathbb{R}^{M\mathfrak{d}}, \mathbb{R}^{\mathfrak{d}})$ and

$$(\mathcal{R}_a(\mathbb{A}_1))(y_1,y_2,\ldots,y_M) = \sum_{m=1}^{M} h_m y_m. \tag{150}$$

Combining this and item (v) in Proposition 2.6 with (145), (148), and (149) demonstrates that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^{\mathfrak{d}})$ and

$$(\mathcal{R}_a(\Psi))(x) = \sum_{m=1}^{M} h_m(\mathcal{R}_a(\Phi_m))(x). \tag{151}$$

This and (147) establish items (i)–(iii). The proof of Proposition 2.25 is thus completed. □

2.4.2 Sums of ANNs with different lengths

Proposition 2.26 *Let $a \in C(\mathbb{R},\mathbb{R})$, $M, d, \mathfrak{d}, \mathfrak{i}, L \in \mathbb{N}$, $h_1, h_2, \ldots, h_M \in \mathbb{R}$, $\mathbb{I}, \Phi_1, \Phi_2, \ldots, \Phi_M \in \mathbf{N}$ satisfy for all $m \in \{1, 2, \ldots, M\}$, $x \in \mathbb{R}^{\mathfrak{d}}$ that $\mathcal{D}(\mathbb{I}) = (\mathfrak{d}, \mathfrak{i}, \mathfrak{d})$, $(\mathcal{R}_a(\mathbb{I}))(x) = x$, $\mathcal{I}(\Phi_m) = d$, $\mathcal{O}(\Phi_m) = \mathfrak{d}$, and $L = \max_{m\in\{1,2,\ldots,M\}} \mathcal{L}(\Phi_m)$ (cf. Definition 2.1 and Definition 2.3). Then there exists $\Psi \in \mathbf{N}$ such that*

(i) *it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^{\mathfrak{d}})$,*
(ii) *it holds for all $x \in \mathbb{R}^d$ that*

$$(\mathcal{R}_a(\Psi))(x) = \sum_{m=1}^{M} h_m\,(\mathcal{R}_a(\Phi_m))(x), \tag{152}$$

and

(iii) *it holds that*

$$\begin{aligned} \mathcal{P}(\Psi) \leqslant \tfrac{1}{2}\bigg(& \Big[\textstyle\sum_{m=1}^{M} \big[\max\big\{1, \tfrac{\mathfrak{i}}{\mathfrak{d}}\big\}\big] \mathcal{P}(\Phi_m)\, \mathbb{1}_{(\mathcal{L}(\Phi_m),\infty)}(L)\Big] \\ & + \Big[\textstyle\sum_{m=1}^{M} \big((L - \mathcal{L}(\Phi_m) - 1)\, \mathfrak{i}\,(\mathfrak{i}+1) \\ & + \mathfrak{d}\,(\mathfrak{i}+1)\big)\, \mathbb{1}_{(\mathcal{L}(\Phi_m),\infty)}(L)\Big] \\ & + \Big[\textstyle\sum_{m=1}^{M} \mathcal{P}(\Phi_m)\, \mathbb{1}_{\{\mathcal{L}(\Phi_m)\}}(L)\Big]\bigg)^2. \end{aligned} \tag{153}$$

Proof of Proposition 2.26 Throughout this proof let $\mathfrak{I} = (\mathfrak{I}_1, \mathfrak{I}_2, \ldots, \mathfrak{I}_M) \in \mathbf{N}^M$ satisfy for all $m \in \{1, 2, \ldots, M\}$ that $\mathfrak{I}_m = \mathbb{I}$, let $(A_1, b_1) \in \mathbb{R}^{\mathfrak{d}\times(M\mathfrak{d})} \times \mathbb{R}^{\mathfrak{d}}$, $(A_2, b_2) \in \mathbb{R}^{(Md)\times d} \times \mathbb{R}^{Md}$ satisfy that

$$A_1 = \big(h_1\, \mathrm{I}_{\mathfrak{d}}\;\; h_2\, \mathrm{I}_{\mathfrak{d}}\;\; \ldots\;\; h_M\, \mathrm{I}_{\mathfrak{d}}\big), \quad A_2 = \begin{pmatrix} \mathrm{I}_d \\ \mathrm{I}_d \\ \vdots \\ \mathrm{I}_d \end{pmatrix}, \quad b_1 = 0, \quad \text{and} \quad b_2 = 0 \tag{154}$$

(cf. Definition 2.10), let $\mathbb{A}_1, \mathbb{A}_2 \in \mathbf{N}$ satisfy that $\mathbb{A}_1 = (A_1, b_1)$ and $\mathbb{A}_2 = (A_2, b_2)$, and let $\Psi \in \mathbf{N}$ satisfy that

$$\Psi = \mathbb{A}_1 \bullet \big(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1, \Phi_2, \ldots, \Phi_M)\big) \bullet \mathbb{A}_2 \tag{155}$$

(cf. Definition 2.5, Definition 2.22, Lemma 2.8, and Corollary 2.23). Note that (155) and items (i)–(ii) in Corollary 2.9 demonstrate that

$$\begin{aligned} \mathcal{P}(\Psi) & \leqslant \Big[\max\Big\{1, \tfrac{\mathcal{O}(\mathbb{A}_1)}{\mathcal{O}(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1,\Phi_2,\ldots,\Phi_M))}\Big\}\Big] \Big[\max\Big\{1, \tfrac{\mathcal{I}(\mathbb{A}_2)+1}{\mathcal{I}(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1,\Phi_2,\ldots,\Phi_M))+1}\Big\}\Big] \\ & \quad \cdot \mathcal{P}\big(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1, \Phi_2, \ldots, \Phi_M)\big) \\ & = \big[\max\{1, \tfrac{\mathfrak{d}}{M\mathfrak{d}}\}\big] \big[\max\{1, \tfrac{d+1}{Md+1}\}\big]\, \mathcal{P}\big(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1, \Phi_2, \ldots, \Phi_M)\big) \\ & = \mathcal{P}\big(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1, \Phi_2, \ldots, \Phi_M)\big). \end{aligned} \tag{156}$$

Corollary 2.24 hence proves that

$$\begin{aligned} & \mathcal{P}(\Psi) \leqslant \mathcal{P}\big(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1, \Phi_2, \ldots, \Phi_M)\big) \\ & \leqslant \tfrac{1}{2}\bigg(\Big[\textstyle\sum_{m=1}^{M} \big[\max\big\{1, \tfrac{\mathfrak{i}}{\mathfrak{d}}\big\}\big] \mathcal{P}(\Phi_m)\, \mathbb{1}_{(\mathcal{L}(\Phi_m),\infty)}(L)\Big] \\ & \quad + \Big[\textstyle\sum_{m=1}^{M} \big((L - \mathcal{L}(\Phi_m) - 1)\, \mathfrak{i}\,(\mathfrak{i}+1) + \mathfrak{d}\,(\mathfrak{i}+1)\big)\, \mathbb{1}_{(\mathcal{L}(\Phi_m),\infty)}(L)\Big] \\ & \quad + \Big[\textstyle\sum_{m=1}^{M} \mathcal{P}(\Phi_m)\, \mathbb{1}_{\{\mathcal{L}(\Phi_m)\}}(L)\Big]\bigg)^2. \end{aligned} \tag{157}$$

Next note that (154) and the fact that $\mathbb{A}_2 = (A_2, b_2)$ prove that $\mathcal{R}_a(\mathbb{A}_2) \in C(\mathbb{R}^d, \mathbb{R}^{Md})$ and

$$\forall\, x \in \mathbb{R}^d\colon\; (\mathcal{R}_a(\mathbb{A}_2))(x) = (x, x, \ldots, x) \in \mathbb{R}^{Md}. \tag{158}$$

Corollary 2.23 and item (v) in Proposition 2.6 therefore ensure that

$$\mathcal{R}_a\big(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1,\Phi_2,\ldots,\Phi_M)\bullet\mathbb{A}_2\big) \in C\big(\mathbb{R}^{\mathcal{I}(\mathbb{A}_2)},\mathbb{R}^{\mathcal{O}(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1,\Phi_2,\ldots,\Phi_M))}\big) = C(\mathbb{R}^d,\mathbb{R}^{M\mathfrak{d}}) \tag{159}$$

and

$$\forall\, x\in\mathbb{R}^d\colon\ \big(\mathcal{R}_a\big(\mathrm{P}_{M,\mathfrak{I}}(\Phi_1,\Phi_2,\ldots,\Phi_M)\bullet\mathbb{A}_2\big)\big)(x) = \big((\mathcal{R}_a(\Phi_1))(x),(\mathcal{R}_a(\Phi_2))(x),\ldots,(\mathcal{R}_a(\Phi_M))(x)\big). \tag{160}$$

In addition, observe that (154) and the fact that $\mathbb{A}_1=(A_1,b_1)$ assure that $\mathcal{R}_a(\mathbb{A}_1)\in C(\mathbb{R}^{M\mathfrak{d}},\mathbb{R}^{\mathfrak{d}})$ and

$$\forall\, y_1,y_2,\ldots,y_M\in\mathbb{R}^{\mathfrak{d}}\colon\ (\mathcal{R}_a(\mathbb{A}_1))(y_1,y_2,\ldots,y_M) = \sum_{m=1}^{M} h_m y_m. \tag{161}$$

Combining this, (159), (160), and (155) with item (v) in Proposition 2.6 demonstrates that

$$\mathcal{R}_a(\Psi)\in C\big(\mathbb{R}^{\mathcal{I}((\mathrm{P}_{M,\mathfrak{I}}(\Phi_1,\Phi_2,\ldots,\Phi_M))\bullet\mathbb{A}_2)},\mathbb{R}^{\mathcal{O}(\mathbb{A}_1)}\big) = C(\mathbb{R}^d,\mathbb{R}^{\mathfrak{d}}) \tag{162}$$

and

$$\forall\, x\in\mathbb{R}^d\colon\ (\mathcal{R}_a(\Psi))(x) = \sum_{m=1}^{M} h_m(\mathcal{R}_a(\Phi_m))(x). \tag{163}$$

This and (157) establish items (i)–(iii). The proof of Proposition 2.26 is thus completed. □

2.5 ANN representations for Euler approximations

2.5.1 ANN representations for one Euler step

Lemma 2.27 *Let $a\in C(\mathbb{R},\mathbb{R})$, $L_1\in\mathbb{N}\cap[2,\infty)$, $L_2\in\mathbb{N}$, $d,\mathfrak{i},l_{1,0},l_{1,1},\ldots,l_{1,L_1},l_{2,0},l_{2,1},\ldots,l_{2,L_2}\in\mathbb{N}$, $\mathbb{I},\Phi_1,\Phi_2\in\mathbf{N}$ satisfy for all $k\in\{1,2\}$, $x\in\mathbb{R}^d$ that $\mathcal{D}(\mathbb{I})=(d,\mathfrak{i},d)$, $(\mathcal{R}_a(\mathbb{I}))(x)=x$, $\mathcal{I}(\Phi_k)=\mathcal{O}(\Phi_k)=d$, and $\mathcal{D}(\Phi_k)=(l_{k,0},l_{k,1},\ldots,l_{k,L_k})$ (cf. Definition 2.1 and Definition 2.3). Then there exists $\Psi\in\mathbf{N}$ such that*

(i) *it holds that $\mathcal{R}_a(\Psi)\in C(\mathbb{R}^d,\mathbb{R}^d)$,*

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

$$(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\Phi_2))(x) + \big((\mathcal{R}_a(\Phi_1))\circ(\mathcal{R}_a(\Phi_2))\big)(x), \tag{164}$$

(iii) *it holds that*

$$\mathcal{D}(\Psi) = (l_{2,0},l_{2,1},\ldots,l_{2,L_2-1},l_{1,1}+\mathfrak{i},l_{1,2}+\mathfrak{i},\ldots,l_{1,L_1-1}+\mathfrak{i},l_{1,L_1}), \tag{165}$$

and

(iv) *it holds that*

$$\mathcal{P}(\Psi) = \mathcal{P}(\Phi_1)+\mathcal{P}(\Phi_2)+(\mathfrak{i}-d)(l_{2,L_2-1}+1)+l_{1,1}(l_{2,L_2-1}-d) +(L_1-2)\mathfrak{i}(\mathfrak{i}+1)+\mathfrak{i}\left[\sum_{m=2}^{L_1} l_{1,m}\right]+\mathfrak{i}\left[\sum_{m=1}^{L_1-2} l_{1,m}\right]. \tag{166}$$

Proof of Lemma 2.27 Throughout this proof let $A_1 \in \mathbb{R}^{d \times 2d}$, $A_2 \in \mathbb{R}^{2d \times d}$, $b_1 \in \mathbb{R}^d$, $b_2 \in \mathbb{R}^{2d}$ satisfy that

$$A_1 = \begin{pmatrix} \mathrm{I}_d & \mathrm{I}_d \end{pmatrix}, \qquad A_2 = \begin{pmatrix} \mathrm{I}_d \\ \mathrm{I}_d \end{pmatrix}, \qquad b_1 = 0, \qquad \text{and} \qquad b_2 = 0 \tag{167}$$

(cf. Definition 2.10) and let $\mathbb{A}_1 \in (\mathbb{R}^{d \times 2d} \times \mathbb{R}^d) \subseteq \mathbf{N}$, $\mathbb{A}_2 \in (\mathbb{R}^{2d \times d} \times \mathbb{R}^{2d}) \subseteq \mathbf{N}$, $\Psi \in \mathbf{N}$ satisfy that $\mathbb{A}_1 = (A_1, b_1)$, $\mathbb{A}_2 = (A_2, b_2)$, and

$$\Psi = \mathbb{A}_1 \bullet \big[\mathbf{P}_2\big(\Phi_1, \mathbb{I}^{\bullet(L_1-1)}\big)\big] \bullet \mathbb{A}_2 \bullet \Phi_2 \tag{168}$$

(cf. Definition 2.5, Definition 2.11, Definition 2.17, Lemma 2.8, and item (i) in Lemma 2.13). Observe that (167) and the fact that $\mathbb{A}_2 = (A_2, b_2)$ ensure that for all $x \in \mathbb{R}^d$ it holds that

$$\mathcal{R}_a(\mathbb{A}_2) \in C(\mathbb{R}^d, \mathbb{R}^{2d}) \qquad \text{and} \qquad (\mathcal{R}_a(\mathbb{A}_2))(x) = (x, x). \tag{169}$$

Item (v) in Proposition 2.6, item (i) in Lemma 2.14, and Proposition 2.19 hence imply that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a([\mathbf{P}_2(\Phi_1, \mathbb{I}^{\bullet(L_1-1)})] \bullet \mathbb{A}_2) \in C(\mathbb{R}^d, \mathbb{R}^{2d})$ and

$$\begin{aligned} \big(\mathcal{R}_a\big(\big[\mathbf{P}_2\big(\Phi_1, \mathbb{I}^{\bullet(L_1-1)}\big)\big] \bullet \mathbb{A}_2\big)\big)(x) &= \big(\mathcal{R}_a\big(\mathbf{P}_2\big(\Phi_1, \mathbb{I}^{\bullet(L_1-1)}\big)\big)\big)(x, x) \\ &= \big((\mathcal{R}_a(\Phi_1))(x), (\mathcal{R}_a(\mathbb{I}^{\bullet(L_1-1)}))(x)\big) = \big((\mathcal{R}_a(\Phi_1))(x), x\big). \end{aligned} \tag{170}$$

Item (v) in Proposition 2.6 therefore demonstrates that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a\big(\big[\mathbf{P}_2\big(\Phi_1, \mathbb{I}^{\bullet(L_1-1)}\big)\big] \bullet \mathbb{A}_2 \bullet \Phi_2\big) \in C(\mathbb{R}^d, \mathbb{R}^{2d})$ and

$$\big(\mathcal{R}_a\big(\big[\mathbf{P}_2\big(\Phi_1, \mathbb{I}^{\bullet(L_1-1)}\big)\big] \bullet \mathbb{A}_2 \bullet \Phi_2\big)\big)(x) = \Big((\mathcal{R}_a(\Phi_1))\big((\mathcal{R}_a(\Phi_2))(x)\big), (\mathcal{R}_a(\Phi_2))(x)\Big). \tag{171}$$

In addition, note that (167) and the fact that $\mathbb{A}_1 = (A_1, b_1)$ ensure that for all $y = (y_1, y_2) \in \mathbb{R}^d \times \mathbb{R}^d$ it holds that

$$\mathcal{R}_a(\mathbb{A}_1) \in C(\mathbb{R}^{2d}, \mathbb{R}^d) \qquad \text{and} \qquad (\mathcal{R}_a(\mathbb{A}_1))(y) = y_1 + y_2. \tag{172}$$

Item (v) in Proposition 2.6, (168), and (171) hence prove that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and

$$(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\Phi_1))\big((\mathcal{R}_a(\Phi_2))(x)\big) + (\mathcal{R}_a(\Phi_2))(x). \tag{173}$$

Next note that item (i) in Lemma 2.13 and item (i) in Proposition 2.20 demonstrate that

$$\mathcal{D}\big(\mathbf{P}_2\big(\Phi_1, \mathbb{I}^{\bullet(L_1-1)}\big)\big) = (2d, l_{1,1} + \mathfrak{i}, l_{1,2} + \mathfrak{i}, \ldots, l_{1,L_1-1} + \mathfrak{i}, 2d). \tag{174}$$

Item (i) in Proposition 2.6 therefore ensures that

$$\mathcal{D}\big(\mathbb{A}_1 \bullet \big[\mathbf{P}_2\big(\Phi_1, \mathbb{I}^{\bullet(L_1-1)}\big)\big] \bullet \mathbb{A}_2\big) = (d, l_{1,1} + \mathfrak{i}, l_{1,2} + \mathfrak{i}, \ldots, l_{1,L_1-1} + \mathfrak{i}, d). \tag{175}$$

Combining this with item (i) in Proposition 2.6, (168), and the fact that $\mathcal{O}(\Phi_2) = l_{2,L_2} = d$ shows that

$$\mathcal{D}(\Psi) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1} + \mathfrak{i}, l_{1,2} + \mathfrak{i}, \ldots, l_{1,L_1-1} + \mathfrak{i}, d). \tag{176}$$

The fact that $l_{1,L_1} = \mathcal{O}(\Phi_1) = d$ hence ensures that

$$\begin{aligned}
\mathcal{P}(\Psi) &= \left[\sum_{m=1}^{L_2-1} l_{2,m}(l_{2,m-1}+1)\right] + (l_{1,1}+\mathfrak{i})(l_{2,L_2-1}+1) \\
&\quad + \left[\sum_{m=2}^{L_1-1} (l_{1,m}+\mathfrak{i})(l_{1,m-1}+\mathfrak{i}+1)\right] + d(l_{1,L_1-1}+\mathfrak{i}+1) \\
&= \mathcal{P}(\Phi_2) - l_{2,L_2}(l_{2,L_2-1}+1) + (l_{1,1}+\mathfrak{i})(l_{2,L_2-1}+1) \\
&\quad + \mathfrak{i}\left[\sum_{m=2}^{L_1-1} l_{1,m}\right] + \mathfrak{i}\left[\sum_{m=2}^{L_1-1} l_{1,m-1}\right] + \left[\sum_{m=2}^{L_1-1} l_{1,m}(l_{1,m-1}+1)\right] \\
&\quad + (L_1-2)\mathfrak{i}(\mathfrak{i}+1) + l_{1,L_1}(l_{1,L_1-1}+1) + l_{1,L_1}\mathfrak{i}.
\end{aligned} \tag{177}$$

This, the fact that $l_{2,L_2} = \mathcal{O}(\Phi_2) = d$, and the fact that $l_{1,0} = \mathcal{I}(\Phi_1) = d$ demonstrate that

$$\begin{aligned}
\mathcal{P}(\Psi) &= \mathcal{P}(\Phi_2) + (l_{1,1}-d+\mathfrak{i})(l_{2,L_2-1}+1) + (L_1-2)\mathfrak{i}(\mathfrak{i}+1) \\
&\quad + \mathfrak{i}\left[\sum_{m=2}^{L_1} l_{1,m}\right] + \mathfrak{i}\left[\sum_{m=1}^{L_1-2} l_{1,m}\right] + \mathcal{P}(\Phi_1) - l_{1,1}(l_{1,0}+1) \\
&= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + (\mathfrak{i}-d)(l_{2,L_2-1}+1) + l_{1,1}(l_{2,L_2-1}-d) \\
&\quad + (L_1-2)\mathfrak{i}(\mathfrak{i}+1) + \mathfrak{i}\left[\sum_{m=2}^{L_1} l_{1,m}\right] + \mathfrak{i}\left[\sum_{m=1}^{L_1-2} l_{1,m}\right].
\end{aligned} \tag{178}$$

Combining this with (173) and (176) establishes items (i)–(iv). The proof of Lemma 2.27 is thus completed. □

Proposition 2.28 *Let $a \in C(\mathbb{R}, \mathbb{R})$, $L_1 \in \mathbb{N} \cap [2, \infty)$, $L_2 \in \mathbb{N}$, $\mathbb{I}, \Phi_1, \Phi_2 \in \mathbf{N}$, $d, \mathfrak{i}, l_{1,0}, l_{1,1}, \ldots, l_{1,L_1}, l_{2,0}, l_{2,1}, \ldots, l_{2,L_2} \in \mathbb{N}$ satisfy for all $k \in \{1, 2\}$, $x \in \mathbb{R}^d$ that $2 \leqslant \mathfrak{i} \leqslant 2d$, $l_{2,L_2-1} \leqslant l_{1,L_1-1} + \mathfrak{i}$, $\mathcal{D}(\mathbb{I}) = (d, \mathfrak{i}, d)$, $(\mathcal{R}_a(\mathbb{I}))(x) = x$, $\mathcal{I}(\Phi_k) = \mathcal{O}(\Phi_k) = d$, and $\mathcal{D}(\Phi_k) = (l_{k,0}, l_{k,1}, \ldots, l_{k,L_k})$ (cf. Definition 2.1 and Definition 2.3). Then there exists $\Psi \in \mathbf{N}$ such that*

(i) *it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,*
(ii) *it holds for all $x \in \mathbb{R}^d$ that*

$$(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\Phi_2))(x) + \big((\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))\big)(x), \tag{179}$$

(iii) *it holds that*

$$\mathcal{D}(\Psi) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1}+\mathfrak{i}, l_{1,2}+\mathfrak{i}, \ldots, l_{1,L_1-1}+\mathfrak{i}, l_{1,L_1}), \tag{180}$$

and

(iv) *it holds that*

$$\begin{aligned}
\mathcal{P}(\Psi) &\leqslant \mathcal{P}(\Phi_2) + \mathcal{P}(\Phi_1)\left[\tfrac{1}{4}\mathcal{P}(\Phi_1) + \mathcal{P}(\mathbb{I}) - 1\right] \\
&\leqslant \mathcal{P}(\Phi_2) + \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\Phi_1)\right]^2.
\end{aligned} \tag{181}$$

Proof of Proposition 2.28 Throughout this proof let $\Psi \in \mathbf{N}$ satisfy that

(I) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(II) it holds for all $x \in \mathbb{R}^d$ that

$$(\mathcal{R}_a(\Psi))(x) = (\mathcal{R}_a(\Phi_2))(x) + \big((\mathcal{R}_a(\Phi_1)) \circ (\mathcal{R}_a(\Phi_2))\big)(x), \tag{182}$$

(III) it holds that

$$\mathcal{D}(\Psi) = (l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1}+\mathfrak{i}, l_{1,2}+\mathfrak{i}, \ldots, l_{1,L_1-1}+\mathfrak{i}, l_{1,L_1}), \quad (183)$$

and

(IV) it holds that

$$\begin{aligned}\mathcal{P}(\Psi) &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + (\mathfrak{i} - d)(l_{2,L_2-1} + 1) + l_{1,1}(l_{2,L_2-1} - d) \\ &\quad + (L_1 - 2)\mathfrak{i}(\mathfrak{i} + 1) + \mathfrak{i}\left[\sum_{m=2}^{L_1} l_{1,m}\right] + \mathfrak{i}\left[\sum_{m=1}^{L_1-2} l_{1,m}\right]\end{aligned} \quad (184)$$

(cf. Lemma 2.27). Note that the fact that $l_{1,0} = \mathcal{I}(\Phi_1) = d = \mathcal{O}(\Phi_1) = l_{1,L_1}$ implies that

$$\mathfrak{i}\left[\sum_{m=2}^{L_1} l_{1,m}\right] \leqslant \tfrac{1}{2}\mathfrak{i}\left[\sum_{m=2}^{L_1} l_{1,m}(l_{1,m-1} + 1)\right] = \tfrac{1}{2}\mathfrak{i}\big[\mathcal{P}(\Phi_1) - l_{1,1}(d+1)\big] \quad (185)$$

and

$$\begin{aligned}\mathfrak{i}\left[\sum_{m=1}^{L_1-2} l_{1,m}\right] &\leqslant \tfrac{1}{2}\mathfrak{i}\left[\sum_{m=1}^{L_1-2} l_{1,m}(l_{1,m-1} + 1)\right] \\ &= \tfrac{1}{2}\mathfrak{i}\big[\mathcal{P}(\Phi_1) - d(l_{1,L_1-1} + 1) - l_{1,L_1-1}(l_{1,L_1-2} + 1)\big].\end{aligned} \quad (186)$$

Combining this with (IV) and the hypothesis that $l_{2,L_2-1} \leqslant l_{1,L_1-1} + \mathfrak{i}$ ensures that

$$\begin{aligned}\mathcal{P}(\Psi) &\leqslant [1 + \mathfrak{i}]\,\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + (\mathfrak{i} - d)(l_{2,L_2-1} + 1) + l_{1,1}(l_{2,L_2-1} - d) \\ &\quad + (L_1 - 2)\mathfrak{i}(\mathfrak{i} + 1) - \tfrac{1}{2}\mathfrak{i}\,l_{1,1}(d + 1) \\ &\quad - \tfrac{1}{2}\mathfrak{i}\,d(l_{1,L_1-1} + 1) - \tfrac{1}{2}\mathfrak{i}\,l_{1,L_1-1}(l_{1,L_1-2} + 1) \\ &\leqslant [1 + \mathfrak{i}]\,\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + [\max\{\mathfrak{i} - d, 0\}](l_{1,L_1-1} + \mathfrak{i} + 1) \\ &\quad + l_{1,1}\big[l_{1,L_1-1} + \mathfrak{i} - d - \tfrac{1}{2}\mathfrak{i}\,(d + 1)\big] + (L_1 - 2)\mathfrak{i}(\mathfrak{i} + 1) \\ &\quad - \tfrac{1}{2}\mathfrak{i}\,d(l_{1,L_1-1} + 1) - \tfrac{1}{2}\mathfrak{i}\,l_{1,L_1-1}(l_{1,L_1-2} + 1).\end{aligned} \quad (187)$$

Moreover, observe that the hypothesis that $2 \leqslant \mathfrak{i} \leqslant 2d$ shows that

$$\begin{aligned}&l_{1,1}\big[\mathfrak{i} - d - \tfrac{1}{2}\mathfrak{i}\,(d + 1)\big] - \tfrac{1}{2}\mathfrak{i}\,d(l_{1,L_1-1} + 1) - \tfrac{1}{2}\mathfrak{i}\,l_{1,L_1-1}(l_{1,L_1-2} + 1) \\ &\leqslant l_{1,1}\big[2d - d - (d + 1)\big] - \tfrac{1}{2}\mathfrak{i}\,l_{1,L_1-1} - \mathfrak{i}\,l_{1,L_1-1} \leqslant -\tfrac{3}{2}\mathfrak{i}\,l_{1,L_1-1}.\end{aligned} \quad (188)$$

This and (187) prove that

$$\begin{aligned}\mathcal{P}(\Psi) &\leqslant [1 + \mathfrak{i}]\,\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + [\max\{\mathfrak{i} - d, 0\}]\,l_{1,L_1-1} \\ &\quad + [\max\{\mathfrak{i} - d, 0\}]\,(\mathfrak{i} + 1) + l_{1,1}l_{1,L_1-1} - \tfrac{3}{2}\mathfrak{i}\,l_{1,L_1-1} + (L_1 - 2)\mathfrak{i}(\mathfrak{i} + 1) \\ &\leqslant [1 + \mathfrak{i}]\,\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + [\max\{\mathfrak{i} - d, 0\}]\,l_{1,L_1-1} \\ &\quad + \mathfrak{i}(\mathfrak{i} + 1) + l_{1,1}l_{1,L_1-1} + (L_1 - 2)\mathfrak{i}(\mathfrak{i} + 1) - \tfrac{3}{2}\mathfrak{i}\,l_{1,L_1-1} \\ &\leqslant [1 + \mathfrak{i}]\,\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + [\max\{\mathfrak{i} - d, 0\}]\,l_{1,L_1-1} + l_{1,1}l_{1,L_1-1} \\ &\quad + (L_1 - 1)\mathfrak{i}(\mathfrak{i} + 1) - \tfrac{3}{2}\mathfrak{i}\,l_{1,L_1-1} \\ &\leqslant [1 + \mathfrak{i}]\,\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1}l_{1,L_1-1} + (L_1 - 1)\mathfrak{i}(\mathfrak{i} + 1).\end{aligned} \quad (189)$$

Moreover, observe that

$$L_1 - 1 \leqslant \left[\sum_{m=1}^{L_1} l_{1,m} \right] - 1 \leqslant \tfrac{1}{2} \left[\sum_{m=1}^{L_1} l_{1,m}(l_{1,m-1}+1) \right] - 1 \\ \leqslant \tfrac{1}{2}\mathcal{P}(\Phi_1) - 1 \leqslant \tfrac{1}{2}\mathcal{P}(\Phi_1). \tag{190}$$

Combining this and (189) with the fact that $\forall\, k \in \mathbb{N} \cap [1, L_1] \colon l_{1,k} \leqslant \frac{1}{2} l_{1,k}(l_{1,k-1} + 1) \leqslant \frac{1}{2}\mathcal{P}(\Phi_1)$ demonstrates that

$$\begin{aligned} \mathcal{P}(\Psi) &\leqslant [1+\mathfrak{i}]\mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + l_{1,1} l_{1,L_1-1} + \tfrac{1}{2}\mathcal{P}(\Phi_1)\mathfrak{i}(\mathfrak{i}+1) \\ &= \mathcal{P}(\Phi_2) + \big[1 + \mathfrak{i} + \tfrac{1}{2}\mathfrak{i}(\mathfrak{i}+1)\big]\mathcal{P}(\Phi_1) + l_{1,1} l_{1,L_1-1} \\ &\leqslant \mathcal{P}(\Phi_2) + \big[1 + \mathfrak{i} + \tfrac{1}{2}\mathfrak{i}(\mathfrak{i}+1)\big]\mathcal{P}(\Phi_1) + \tfrac{1}{4}[\mathcal{P}(\Phi_1)]^2. \end{aligned} \tag{191}$$

Furthermore, note that the hypothesis that $2 \leqslant \mathfrak{i} \leqslant 2d$ and the hypothesis that $\mathcal{D}(\mathbb{I}) = (d, \mathfrak{i}, d)$ prove that

$$\begin{aligned} \mathfrak{i} + \tfrac{1}{2}\mathfrak{i}(\mathfrak{i}+1) &= \mathfrak{i}^2 + \mathfrak{i} + \tfrac{1}{2}\mathfrak{i} - \tfrac{1}{2}\mathfrak{i}^2 \leqslant 2d\mathfrak{i} + \mathfrak{i} + d - \tfrac{1}{2}\mathfrak{i}^2 \\ &= \mathfrak{i}(d+1) + d(\mathfrak{i}+1) - \tfrac{1}{2}\mathfrak{i}^2 = \mathcal{P}(\mathbb{I}) - \tfrac{1}{2}\mathfrak{i}^2 \leqslant \mathcal{P}(\mathbb{I}) - 2. \end{aligned} \tag{192}$$

Combining this and (191) implies that

$$\begin{aligned} \mathcal{P}(\Psi) &\leqslant \mathcal{P}(\Phi_2) + [1 + \mathcal{P}(\mathbb{I}) - 2]\,\mathcal{P}(\Phi_1) + \tfrac{1}{4}[\mathcal{P}(\Phi_1)]^2 \\ &= \mathcal{P}(\Phi_2) + \big[\tfrac{1}{4}\mathcal{P}(\Phi_1) + \mathcal{P}(\mathbb{I}) - 1\big]\mathcal{P}(\Phi_1) \\ &\leqslant \mathcal{P}(\Phi_2) + \mathcal{P}(\mathbb{I})\,\mathcal{P}(\Phi_1) + \tfrac{1}{4}[\mathcal{P}(\mathbb{I})]^2 + [\mathcal{P}(\Phi_1)]^2 \\ &= \mathcal{P}(\Phi_2) + \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\Phi_1)\big]^2. \end{aligned} \tag{193}$$

This, (I), (II), and (III) establish items (i)–(iv). The proof of Proposition 2.28 is thus completed. □

2.5.2 ANN representations for multiple nested Euler steps

Corollary 2.29 *Let* $a \in C(\mathbb{R}, \mathbb{R})$, $d, \mathfrak{i}, \mathfrak{L} \in \mathbb{N}$, $\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}} \in \mathbb{N}$, $(L_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{N} \cap [2, \infty)$, $\mathbb{I}, \psi \in \mathbf{N}$, $(\phi_n)_{n \in \mathbb{N}_0} \subseteq \mathbf{N}$, *let* $l_{n,k} \in \mathbb{N}$, $k \in \{0, 1, \ldots, L_n\}$, $n \in \mathbb{N}_0$, *assume for all* $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ *that* $2 \leqslant \mathfrak{i} \leqslant 2d$, $\ell_{\mathfrak{L}-1} \leqslant l_{0,L_0-1} + \mathfrak{i}$, $l_{n,L_n-1} \leqslant l_{n+1,L_{n+1}-1}$, $\mathcal{D}(\mathbb{I}) = (d, \mathfrak{i}, d)$, $(\mathcal{R}_a(\mathbb{I}))(x) = x$, $\mathcal{I}(\phi_n) = \mathcal{O}(\phi_n) = \mathcal{I}(\psi) = \mathcal{O}(\psi) = d$, $\mathcal{D}(\phi_n) = (l_{n,0}, l_{n,1}, \ldots, l_{n,L_n})$, *and* $\mathcal{D}(\psi) = (\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}})$, *and let* $f_n \colon \mathbb{R}^d \to \mathbb{R}^d$, $n \in \mathbb{N}_0$, *be the functions which satisfy for all* $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ *that*

$$f_0(x) = (\mathcal{R}_a(\psi))(x) \qquad \text{and} \qquad f_{n+1}(x) = f_n(x) + \big([\mathcal{R}_a(\phi_n)] \circ f_n\big)(x) \tag{194}$$

(cf. Definition 2.1 and Definition 2.3). Then for every $n \in \mathbb{N}$ *there exists* $\Psi \in \mathbf{N}$ *such that*

(i) *it holds that* $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(ii) *it holds for all* $x \in \mathbb{R}^d$ *that* $(\mathcal{R}_a(\Psi))(x) = f_n(x)$,

(iii) *it holds that* $\mathcal{H}(\Psi) = \mathcal{H}(\psi) + \sum_{k=0}^{n-1} \mathcal{H}(\phi_k)$,

(iv) *it holds that*

$$\mathcal{D}(\Psi) = \big(\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}-1}, l_{0,1} + \mathfrak{i}, l_{0,2} + \mathfrak{i}, \ldots, l_{0,L_0-1} + \mathfrak{i}, \\ l_{1,1} + \mathfrak{i}, l_{1,2} + \mathfrak{i}, \ldots, l_{1,L_1-1} + \mathfrak{i}, \ldots, l_{n-1,1} + \mathfrak{i}, l_{n-1,2} + \mathfrak{i}, \ldots, l_{n-1,L_{n-1}-1} + \mathfrak{i}, d\big), \tag{195}$$

and

(v) *it holds that* $\mathcal{P}(\Psi) \leqslant \mathcal{P}(\psi) + \sum_{k=0}^{n-1} \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\big]^2$.

Proof of Corollary 2.29 We prove items (i)–(v) by induction on $n \in \mathbb{N}$. Note that the hypothesis that $\mathcal{D}(\psi) = (\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}})$, the fact that $\ell_0 = \mathcal{I}(\psi) = \ell_{\mathfrak{L}} = \mathcal{O}(\psi) = d$, the hypothesis that $\mathcal{D}(\phi_0) = (l_{0,0}, l_{0,1}, \ldots, l_{0,L_0})$, the hypothesis that $\ell_{\mathfrak{L}-1} \leqslant l_{0,L_0-1} + \mathfrak{i}$, the hypothesis that $L_0 \in \mathbb{N} \cap [2, \infty)$, Proposition 2.28 (with $a = a$, $L_1 = L_0$, $L_2 = \mathfrak{L}$, $\mathbb{I} = \mathbb{I}$, $\Phi_1 = \phi_0$, $\Phi_2 = \psi$, $d = d$, $\mathfrak{i} = \mathfrak{i}$, $l_{1,v} = l_{0,v}$, $l_{2,w} = \ell_w$ for $v \in \{0, 1, \ldots, L_0\}$, $w \in \{0, 1, \ldots, \mathfrak{L}\}$ in the notation of Proposition 2.28), and (194) imply that there exists $\Upsilon \in \mathbf{N}$ which satisfies that

(I) it holds that $\mathcal{R}_a(\Upsilon) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

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

$$\begin{aligned}(\mathcal{R}_a(\Upsilon))(x) &= (\mathcal{R}_a(\psi))(x) + \big([\mathcal{R}_a(\phi_0)] \circ [\mathcal{R}_a(\psi)]\big)(x) \\ &= f_0(x) + ([\mathcal{R}_a(\phi_0)] \circ f_0)(x) = f_1(x),\end{aligned} \tag{196}$$

(III) it holds that

$$\mathcal{D}(\Upsilon) = \big(\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}-1}, l_{0,1} + \mathfrak{i}, l_{0,2} + \mathfrak{i}, \ldots, l_{0,L_0-1} + \mathfrak{i}, l_{0,L_0}\big), \tag{197}$$

and

(IV) it holds that $\mathcal{P}(\Upsilon) \leqslant \mathcal{P}(\psi) + \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_0)\big]^2$.

Observe that (III) shows that $\mathcal{L}(\Upsilon) = \mathfrak{L} + L_0 - 1$. Hence, we obtain that

$$\mathcal{H}(\Upsilon) = \mathcal{L}(\Upsilon) - 1 = (\mathfrak{L} - 1) + (L_0 - 1) = \mathcal{H}(\psi) + \mathcal{H}(\phi_0). \tag{198}$$

Combining this with (I)–(IV) establishes items (i)–(v) in the base case $n = 1$. For the induction step $\mathbb{N} \ni n \to n + 1 \in \mathbb{N} \cap [2, \infty)$ let $n \in \mathbb{N}$, $\Psi \in \mathbf{N}$, $\mathfrak{l}_0, \mathfrak{l}_1, \ldots, \mathfrak{l}_{\mathfrak{L}+\sum_{k=0}^{n-1}(L_k-1)} \in \mathbb{N}$ satisfy that

(a) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

(b) it holds for all $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Psi))(x) = f_n(x)$,

(c) it holds that $\mathcal{H}(\Psi) = \mathcal{H}(\psi) + \sum_{k=0}^{n-1} \mathcal{H}(\phi_k)$,

(d) it holds that

$$\begin{aligned}\mathcal{D}(\Psi) &= \big(\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}-1}, l_{0,1} + \mathfrak{i}, l_{0,2} + \mathfrak{i}, \ldots, l_{0,L_0-1} + \mathfrak{i}, l_{1,1} + \mathfrak{i}, l_{1,2} + \mathfrak{i}, \\ &\qquad \ldots, l_{1,L_1-1} + \mathfrak{i}, \ldots, l_{n-1,1} + \mathfrak{i}, l_{n-1,2} + \mathfrak{i}, \ldots, l_{n-1,L_{n-1}-1} + \mathfrak{i}, d\big) \\ &= \big(\mathfrak{l}_0, \mathfrak{l}_1, \ldots, \mathfrak{l}_{\mathfrak{L}+\sum_{k=0}^{n-1}(L_k-1)}\big),\end{aligned} \tag{199}$$

and

1. it holds that $\mathcal{P}(\Psi) \leqslant \mathcal{P}(\psi) + \sum_{k=0}^{n-1} \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\big]^2$.

Observe that (d) and the hypothesis that $\forall k \in \mathbb{N}_0 \colon l_{k,L_k-1} \leqslant l_{k+1,L_{k+1}-1}$ demonstrate that

$$\mathfrak{l}_{\mathcal{L}(\Psi)-1} = \mathfrak{l}_{\mathfrak{L}-1+\sum_{k=0}^{n-1}(L_k-1)} = l_{n-1,L_{n-1}-1} + \mathfrak{i} \leqslant l_{n,L_n-1} + \mathfrak{i}. \tag{200}$$

The hypothesis that $\mathcal{D}(\phi_n) = (l_{n,0}, l_{n,1}, \ldots, l_{n,L_n})$, (d), the hypothesis that $L_n \in \mathbb{N} \cap [2, \infty)$, and Proposition 2.28 (with $a = a$, $L_1 = L_n$, $L_2 = \mathfrak{L} + \sum_{k=0}^{n-1}(L_k - 1)$, $\mathbb{I} = \mathbb{I}$, $\Phi_1 = \phi_n$, $\Phi_2 = \Psi$, $d = d$, $\mathfrak{i} = \mathfrak{i}$, $l_{1,v} = l_{n,v}$, $l_{2,w} = \ell_w$ for $v \in \{0, 1, \ldots, L_n\}$, $w \in \{0, 1, \ldots, \mathfrak{L} + \sum_{k=0}^{n-1}(L_k - 1)\}$ in the notation of Proposition 2.28) hence prove that there exists $\Phi \in \mathbf{N}$ which satisfies that

(A) it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(B) it holds for all $x \in \mathbb{R}^d$ that

$$(\mathcal{R}_a(\Phi))(x) = (\mathcal{R}_a(\Psi))(x) + \big([\mathcal{R}_a(\phi_n)] \circ [\mathcal{R}_a(\Psi)]\big)(x), \tag{201}$$

(C) it holds that

$$\mathcal{D}(\Phi) = \big(\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}-1}, l_{0,1}+\mathfrak{i}, l_{0,2}+\mathfrak{i}, \ldots, l_{0,L_0-1}+\mathfrak{i}, l_{1,1}+\mathfrak{i}, l_{1,2}+\mathfrak{i}, \\ \ldots, l_{1,L_1-1}+\mathfrak{i}, \ldots, l_{n,1}+\mathfrak{i}, l_{n,2}+\mathfrak{i}, \ldots, l_{n,L_n-1}+\mathfrak{i}, l_{n,L_n}\big), \tag{202}$$

and
(D) it holds that $\mathcal{P}(\Phi) \leqslant \mathcal{P}(\Psi) + \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_n)\big]^2$.

Next note that (C) implies that $\mathcal{L}(\Phi) = \mathfrak{L} + \sum_{k=0}^{n}(L_k - 1)$. Hence, we obtain that

$$\mathcal{H}(\Phi) = \mathcal{L}(\Phi) - 1 = (\mathfrak{L} - 1) + \sum\nolimits_{k=0}^{n}(L_k - 1) = \mathcal{H}(\psi) + \sum\nolimits_{k=0}^{n}\mathcal{H}(\phi_k). \tag{203}$$

Moreover, observe that (B), (194), and (b) demonstrate that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Phi))(x) &= (\mathcal{R}_a(\Psi))(x) + \big([\mathcal{R}_a(\phi_n)] \circ [\mathcal{R}_a(\Psi)]\big)(x) \\ &= f_n(x) + ([\mathcal{R}_a(\phi_n)] \circ f_n)(x) = f_{n+1}(x).\end{aligned} \tag{204}$$

In addition, note that (D) and (e) ensure that

$$\begin{aligned}\mathcal{P}(\Phi) &\leqslant \mathcal{P}(\psi) + \left[\sum_{k=0}^{n-1}\big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\big]^2\right] + \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_n)\big]^2 \\ &= \mathcal{P}(\psi) + \sum_{k=0}^{n}\big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\big]^2.\end{aligned} \tag{205}$$

This, (A), (C), (203), and (204) prove items (i)–(v) in the case $n + 1$. Induction thus establishes items (i)–(v). The proof of Corollary 2.29 is thus completed. □

Proposition 2.30 *Let $a \in C(\mathbb{R}, \mathbb{R})$, $d, \mathfrak{L} \in \mathbb{N}$, $\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}} \in \mathbb{N}$, $\psi \in \mathbf{N}$, $(\phi_n)_{n \in \mathbb{N}_0} \subseteq \mathbf{N}$ satisfy for all $n \in \mathbb{N}_0$ that $\mathcal{I}(\phi_n) = \mathcal{O}(\phi_n) = \mathcal{I}(\psi) = \mathcal{O}(\psi) = d$, $\mathcal{L}(\phi_n) = 1$, and $\mathcal{D}(\psi) = (\ell_0, \ell_1, \ldots, \ell_{\mathfrak{L}})$ and let $f_n \colon \mathbb{R}^d \to \mathbb{R}^d$, $n \in \mathbb{N}_0$, be the functions which satisfy for all $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ that*

$$f_0(x) = (\mathcal{R}_a(\psi))(x) \qquad \textit{and} \qquad f_{n+1}(x) = f_n(x) + \big([\mathcal{R}_a(\phi_n)] \circ f_n\big)(x) \tag{206}$$

(cf. Definition 2.1 and Definition 2.3). Then for every $n \in \mathbb{N}_0$ there exists $\Psi \in \mathbf{N}$ such that

(i) *it holds that* $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(ii) *it holds for all* $x \in \mathbb{R}^d$ *that* $(\mathcal{R}_a(\Psi))(x) = f_n(x)$, *and*
(iii) *it holds that* $\mathcal{D}(\Psi) = \mathcal{D}(\psi)$.

Proof of Proposition 2.30 We prove items (i)–(iii) by induction on $n \in \mathbb{N}_0$. Note that (206) and the fact that $\mathcal{R}_a(\psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$ establish items (i)–(iii) in the base case $n = 0$. For the induction step $\mathbb{N}_0 \ni n \to n + 1 \in \mathbb{N}$ let $n \in \mathbb{N}_0$, $\Psi \in \mathbf{N}$ satisfy that

(I) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(II) it holds for all $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Psi))(x) = f_n(x)$, and
(III) it holds that $\mathcal{D}(\Psi) = \mathcal{D}(\psi)$,

and let $(A, b) \in (\mathbb{R}^{d\times d} \times \mathbb{R}^d) \subseteq \mathbf{N}$, $\mathbb{A} \in (\mathbb{R}^{d\times d} \times \mathbb{R}^d) \subseteq \mathbf{N}$, $\Phi \in \mathbf{N}$ satisfy that $\phi_n = (A, b)$, $\mathbb{A} = (A + \mathrm{I}_d, b)$, and $\Phi = \mathbb{A} \bullet \Psi$ (cf. Definition 2.5 and Definition 2.10). Observe that item (v) in Proposition 2.6 demonstrates that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and

$$\begin{aligned}(\mathcal{R}_a(\Phi))(x) &= (\mathcal{R}_a(\mathbb{A}))\big((\mathcal{R}_a(\Psi))(x)\big)\\ &= (A + \mathrm{I}_d)\big((\mathcal{R}_a(\Psi))(x)\big) + b\\ &= A\big((\mathcal{R}_a(\Psi))(x)\big) + b + (\mathcal{R}_a(\Psi))(x)\\ &= (\mathcal{R}_a(\phi_n))\big((\mathcal{R}_a(\Psi))(x)\big) + (\mathcal{R}_a(\Psi))(x).\end{aligned} \tag{207}$$

Combining this with (206) and (II) proves that for all $x \in \mathbb{R}^d$ it holds that

$$(\mathcal{R}_a(\Phi))(x) = (\mathcal{R}_a(\phi_n))\big(f_n(x)\big) + f_n(x) = f_{n+1}(x). \tag{208}$$

In addition, note that (III), the fact that $\Phi = \mathbb{A} \bullet \Psi$, the fact that $\mathcal{L}(\mathbb{A}) = 1$, the fact that $\mathcal{I}(\mathbb{A}) = \mathcal{O}(\mathbb{A}) = \mathcal{O}(\Psi) = d$, and item (i) in Proposition 2.6 imply that $\mathcal{D}(\Phi) = \mathcal{D}(\Psi) = \mathcal{D}(\psi)$. Combining this and the fact that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^d, \mathbb{R}^d)$ with (208) proves items (i)–(iii) in the case $n + 1$. Induction thus establishes items (i)–(iii). The proof of Proposition 2.30 is thus completed. □

Corollary 2.31 *Let* $a \in C(\mathbb{R}, \mathbb{R})$, $d, \mathfrak{i}, L, \mathfrak{L} \in \mathbb{N}$, $\ell_0, \ell_1, \dots, \ell_{\mathfrak{L}} \in \mathbb{N}$, $\mathbb{I}, \psi \in \mathbf{N}$, $(\phi_n)_{n\in\mathbb{N}_0} \subseteq \mathbf{N}$, *let* $l_{n,k} \in \mathbb{N}$, $k \in \{0, 1, \dots, L\}$, $n \in \mathbb{N}_0$, *assume for all* $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ *that* $2 \leqslant \mathfrak{i} \leqslant 2d$, $\ell_{\mathfrak{L}-1} \leqslant l_{0,L-1} + \mathfrak{i}$, $l_{n,L-1} \leqslant l_{n+1,L-1}$, $\mathcal{D}(\mathbb{I}) = (d, \mathfrak{i}, d)$, $(\mathcal{R}_a(\mathbb{I}))(x) = x$, $\mathcal{I}(\phi_n) = \mathcal{O}(\phi_n) = \mathcal{I}(\psi) = \mathcal{O}(\psi) = d$, $\mathcal{D}(\phi_n) = (l_{n,0}, l_{n,1}, \dots, l_{n,L})$, *and* $\mathcal{D}(\psi) = (\ell_0, \ell_1, \dots, \ell_{\mathfrak{L}})$, *and let* $f_n \colon \mathbb{R}^d \to \mathbb{R}^d$, $n \in \mathbb{N}_0$, *be the functions which satisfy for all* $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ *that*

$$f_0(x) = (\mathcal{R}_a(\psi))(x) \qquad \text{and} \qquad f_{n+1}(x) = f_n(x) + \big([\mathcal{R}_a(\phi_n)] \circ f_n\big)(x) \tag{209}$$

(cf. Definition 2.1 and Definition 2.3). Then for every $n \in \mathbb{N}_0$ *there exists* $\Psi \in \mathbf{N}$ *such that*

(i) *it holds that* $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(ii) *it holds for all* $x \in \mathbb{R}^d$ *that* $(\mathcal{R}_a(\Psi))(x) = f_n(x)$,
(iii) *it holds that* $\mathcal{H}(\Psi) = \mathcal{H}(\psi) + \sum_{k=0}^{n-1} \mathcal{H}(\phi_k) = \mathcal{H}(\psi) + n\,\mathcal{H}(\phi_0)$, *and*
(iv) *it holds that* $\mathcal{P}(\Psi) \leqslant \mathcal{P}(\psi) + \sum_{k=0}^{n-1} \big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\big]^2$.

Proof of Corollary 2.31 To prove items (i)–(iv) we distinguish between the case $L = 1$ and the case $L \in \mathbb{N} \cap [2, \infty)$. We first prove items (i)–(iv) in the case $L = 1$. Observe that Proposition 2.30 ensures that there exist $\Psi_n \in \mathbf{N}$, $n \in \mathbb{N}_0$, which satisfy that

(I) it holds for all $n \in \mathbb{N}_0$ that $\mathcal{R}_a(\Psi_n) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(II) it holds for all $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Psi_n))(x) = f_n(x)$, and
(III) it holds for all $n \in \mathbb{N}_0$ that $\mathcal{D}(\Psi_n) = \mathcal{D}(\psi)$.

Next note that the hypothesis that $L = 1$ demonstrates that for all $n \in \mathbb{N}_0$ it holds that $\mathcal{H}(\phi_n) = 0$. Combining this with (III) implies that for all $n \in \mathbb{N}_0$ it holds that

$$\mathcal{H}(\Psi_n) = \mathcal{H}(\psi) = \mathcal{H}(\psi) + \sum_{k=0}^{n-1} \mathcal{H}(\phi_k) = \mathcal{H}(\psi) + n\mathcal{H}(\phi_0). \tag{210}$$

In addition, observe that (III) shows that for all $n \in \mathbb{N}_0$ it holds that

$$\mathcal{P}(\Psi_n) = \mathcal{P}(\psi) \leqslant \mathcal{P}(\psi) + \sum_{k=0}^{n-1} \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\right]^2. \tag{211}$$

Combining this and (210) with (I)–(II) establishes items (i)–(iv) in the case $L = 1$. We now prove items (i)–(iv) in the case $L \in \mathbb{N} \cap [2, \infty)$. Note that (209), the fact that $\mathcal{R}_a(\psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$, the fact that $\mathcal{H}(\psi) = \mathcal{H}(\psi) + \sum_{k=0}^{-1} \mathcal{H}(\phi_k) = \mathcal{H}(\psi) + 0 \cdot \mathcal{H}(\phi_0)$, and the fact that $\mathcal{P}(\psi) = \mathcal{P}(\psi) + \sum_{k=0}^{-1} \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\right]^2$ prove that there exists $\Psi \in \mathbf{N}$ such that

(a) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(b) it holds for all $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Psi))(x) = f_0(x)$,
(c) it holds that $\mathcal{H}(\Psi) = \mathcal{H}(\psi) + \sum_{k=0}^{-1} \mathcal{H}(\phi_k) = \mathcal{H}(\psi) + 0 \cdot \mathcal{H}(\phi_0)$, and
(d) it holds that $\mathcal{P}(\Psi) \leqslant \mathcal{P}(\psi) + \sum_{k=0}^{-1} \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\right]^2$.

Moreover, observe that Corollary 2.29 and the fact that for all $k \in \mathbb{N}_0$ it holds that $\mathcal{H}(\phi_k) = L - 1 = \mathcal{H}(\phi_0)$ ensure that for every $n \in \mathbb{N}$ there exists $\Psi \in \mathbf{N}$ such that

(A) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^d)$,
(B) it holds for all $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Psi))(x) = f_n(x)$,
(C) it holds that $\mathcal{H}(\Psi) = \mathcal{H}(\psi) + \sum_{k=0}^{n-1} \mathcal{H}(\phi_k) = \mathcal{H}(\psi) + n\, \mathcal{H}(\phi_0)$, and
(D) it holds that $\mathcal{P}(\Psi) \leqslant \mathcal{P}(\psi) + \sum_{k=0}^{n-1} \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\phi_k)\right]^2$.

Combining this with (a)–(d) proves items (i)–(iv) in the case $L \in \mathbb{N} \cap [2, \infty)$. The proof of Corollary 2.31 is thus completed. □

2.5.3 ANN representations for multiple perturbed nested Euler steps

Proposition 2.32 *Let $a \in C(\mathbb{R}, \mathbb{R})$, $N, d, \mathfrak{i} \in \mathbb{N}$, $\mathbb{I}, \Phi \in \mathbf{N}$, $A_1, A_2, \ldots, A_N \in \mathbb{R}^{d \times d}$ satisfy for all $x \in \mathbb{R}^d$ that $2 \leqslant \mathfrak{i} \leqslant 2d$, $\mathcal{D}(\mathbb{I}) = (d, \mathfrak{i}, d)$, $(\mathcal{R}_a(\mathbb{I}))(x) = x$, and $\mathcal{I}(\Phi) = \mathcal{O}(\Phi) = d$ and let $Y_n = (Y_n^{x,y})_{(x,y) \in \mathbb{R}^d \times (\mathbb{R}^d)^N} \colon \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$, $n \in \{0, 1, \ldots, N\}$, be the functions which satisfy for all $n \in \{0, 1, \ldots, N-1\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that $Y_0^{x,y} = x$ and*

$$Y_{n+1}^{x,y} = Y_n^{x,y} + A_{n+1}\big((\mathcal{R}_a(\Phi))(Y_n^{x,y})\big) + y_{n+1} \tag{212}$$

(cf. Definition 2.1 and Definition 2.3). Then there exists $(\Psi_{n,y})_{(n,y)\in\{0,1,\dots,N\}\times(\mathbb{R}^d)^N} \subseteq \mathbf{N}$ *such that*

(i) *it holds for all* $n \in \{0, 1, \dots, N\}$, $y \in (\mathbb{R}^d)^N$ *that* $\mathcal{R}_a(\Psi_{n,y}) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

(ii) *it holds for all* $n \in \{0, 1, \dots, N\}$, $y \in (\mathbb{R}^d)^N$, $x \in \mathbb{R}^d$ *that* $(\mathcal{R}_a(\Psi_{n,y}))(x) = Y_n^{x,y}$,

(iii) *it holds for all* $n \in \{0, 1, \dots, N\}$, $y \in (\mathbb{R}^d)^N$ *that*

$$\mathcal{H}(\Psi_{n,y}) = \mathcal{H}(\mathbb{I}) + n\,\mathcal{H}(\Phi) = 1 + n\,\mathcal{H}(\Phi), \tag{213}$$

(iv) *it holds for all* $n \in \{0, 1, \dots, N\}$, $y \in (\mathbb{R}^d)^N$ *that*

$$\mathcal{P}(\Psi_{n,y}) \leqslant \mathcal{P}(\mathbb{I}) + n\big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\Phi)\big]^2, \tag{214}$$

(v) *it holds for all* $n \in \{0, 1, \dots, N\}$, $x \in \mathbb{R}^d$ *that*

$$\big[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Psi_{n,y}))(x) \in \mathbb{R}^d\big] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big), \tag{215}$$

and

(vi) *it holds for all* $n \in \{0, 1, \dots, N\}$, $m \in \mathbb{N}_0 \cap [0, n]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \dots, y_N)$, $z = (z_1, z_2, \dots, z_N) \in (\mathbb{R}^d)^N$ *with* $\forall\, k \in \mathbb{N}\cap[0, n]\colon y_k = z_k$ *that*

$$(\mathcal{R}_a(\Psi_{m,y}))(x) = (\mathcal{R}_a(\Psi_{m,z}))(x). \tag{216}$$

Proof of Proposition 2.32 Throughout this proof let $l_0, l_1, \dots, l_{\mathcal{L}(\Phi)} \in \mathbb{N}$ satisfy that $\mathcal{D}(\Phi) = (l_0, l_1, \dots, l_{\mathcal{L}(\Phi)})$, let $\mathbb{A}_{n,b} \in (\mathbb{R}^{d\times d} \times \mathbb{R}^d) \subseteq \mathbf{N}$, $n \in \{1, 2, \dots, N\}$, $b \in \mathbb{R}^d$, satisfy for all $n \in \{1, 2, \dots, N\}$, $b \in \mathbb{R}^d$ that

$$\mathbb{A}_{n,b} = (A_n, b) \in (\mathbb{R}^{d\times d} \times \mathbb{R}^d), \tag{217}$$

let $\rho_{n,y} \in \mathbf{N}$, $n \in \mathbb{N}$, $y \in (\mathbb{R}^d)^N$, satisfy for all $n \in \mathbb{N}$, $y = (y_1, y_2, \dots, y_N) \in (\mathbb{R}^d)^N$ that

$$\rho_{n,y} = \mathbb{A}_{\min\{n,N\},y_{\min\{n,N\}}} \bullet \Phi \tag{218}$$

(cf. Definition 2.5), and let $\mathcal{Y}_n = (\mathcal{Y}_n^{x,y})_{(x,y)\in\mathbb{R}^d\times(\mathbb{R}^d)^N} : \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$, $n \in \mathbb{N}_0$, be the functions which satisfy for all $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \dots, y_N) \in (\mathbb{R}^d)^N$ that $\mathcal{Y}_0^{x,y} = x$ and

$$\mathcal{Y}_{n+1}^{x,y} = \mathcal{Y}_n^{x,y} + \big(\mathcal{R}_a(\rho_{n+1,y})\big)(\mathcal{Y}_n^{x,y}). \tag{219}$$

Observe that item (i) in Proposition 2.6 and the fact that for all $n \in \{1, 2, \dots, N\}$, $y = (y_1, y_2, \dots, y_N) \in (\mathbb{R}^d)^N$ it holds that $\rho_{n,y} = \mathbb{A}_{n,y_n} \bullet \Phi$ prove that for all $n \in \{1, 2, \dots, N\}$, $y \in (\mathbb{R}^d)^N$ it holds that $\mathcal{D}(\rho_{n,y}) = \mathcal{D}(\Phi) = (l_0, l_1, \dots, l_{\mathcal{L}(\Phi)})$. Corollary 2.31 (with $a = a$, $d = d$, $\mathfrak{i} = \mathfrak{i}$, $L = \mathcal{L}(\Phi)$, $\mathfrak{L} = 2$, $\ell_0 = d$, $\ell_1 = \mathfrak{i}$, $\ell_2 = d$, $\mathbb{I} = \mathbb{I}$, $\psi = \mathbb{I}$, $(\mathbb{N}_0 \ni n \mapsto \phi_n \in \mathbf{N}) = (\mathbb{N}_0 \ni n \mapsto \rho_{n+1,y} \in \mathbf{N})$, $(\mathbb{N}_0 \times \{0, 1, \dots, \mathcal{L}(\Phi)\} \ni (n, k) \mapsto l_{n,k} \in \mathbb{N}) = (\mathbb{N}_0 \times \{0, 1, \dots, \mathcal{L}(\Phi)\} \ni (n, k) \mapsto l_k \in \mathbb{N})$, $(\mathbb{N}_0 \ni n \mapsto f_n \in C(\mathbb{R}^d, \mathbb{R}^d)) = (\mathbb{N}_0 \ni n \mapsto (\mathbb{R}^d \ni x \mapsto \mathcal{Y}_n^{x,y} \in \mathbb{R}^d) \in C(\mathbb{R}^d, \mathbb{R}^d))$ for $y \in (\mathbb{R}^d)^N$ in the notation of Corollary 2.31) and the fact that for all $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that $(\mathcal{R}_a(\mathbb{I}))(x) = x = \mathcal{Y}_0^{x,y} = Y_0^{x,y}$ hence prove that there exist $\Psi_{n,y} \in \mathbf{N}$, $(n, y) \in \{0, 1, \dots, N\} \times (\mathbb{R}^d)^N$, which satisfy that

(I) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Psi_{n,y}) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

(II) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Psi_{n,y}))(x) = \mathcal{Y}_n^{x,y} = Y_n^{x,y}$,

(III) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ that

$$\mathcal{H}(\Psi_{n,y}) = \mathcal{H}(\mathbb{I}) + \sum_{k=0}^{n-1} \mathcal{H}(\rho_{k+1,y}) = 1 + n\mathcal{H}(\Phi), \tag{220}$$

and

(IV) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ that

$$\mathcal{P}(\Psi_{n,y}) \leqslant \mathcal{P}(\mathbb{I}) + \sum_{k=0}^{n-1} \left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\rho_{k+1,y})\right]^2 = \mathcal{P}(\mathbb{I}) + n\left[\tfrac{1}{2}\mathcal{P}(\mathbb{I}) + \mathcal{P}(\Phi)\right]^2. \tag{221}$$

Next we claim that for all $n \in \{0, 1, \ldots, N\}$ it holds that

$$\forall\, x \in \mathbb{R}^d \colon \left[\left((\mathbb{R}^d)^N \ni y \mapsto Y_n^{x,y} \in \mathbb{R}^d\right) \in C\left((\mathbb{R}^d)^N, \mathbb{R}^d\right)\right]. \tag{222}$$

We now prove (222) by induction on $n \in \{0, 1, \ldots, N\}$. Note that the fact that for all $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that $Y_0^{x,y} = x$ proves (222) in the base case $n = 0$. For the induction step observe that (212) and the fact that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^d, \mathbb{R}^d)$ ensure that for all $n \in \{0, 1, \ldots, N-1\}$ with

$$\forall\, x \in \mathbb{R}^d \colon \left[\left((\mathbb{R}^d)^N \ni y \mapsto Y_n^{x,y} \in \mathbb{R}^d\right) \in C\left((\mathbb{R}^d)^N, \mathbb{R}^d\right)\right] \tag{223}$$

it holds that

$$\forall\, x \in \mathbb{R}^d \colon \left[\left((\mathbb{R}^d)^N \ni y \mapsto Y_{n+1}^{x,y} \in \mathbb{R}^d\right) \in C\left((\mathbb{R}^d)^N, \mathbb{R}^d\right)\right]. \tag{224}$$

Induction thus proves (222). In addition, observe that (222) and (II) imply that for all $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$ it holds that

$$\left((\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Psi_{n,y}))(x) \in \mathbb{R}^d\right) \in C\left((\mathbb{R}^d)^N, \mathbb{R}^d\right). \tag{225}$$

Next let $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ satisfy for all $k \in \mathbb{N} \cap [0, n]$ that $y_k = z_k$. We claim that for all $m \in \mathbb{N}_0 \cap [0, n]$ it holds that

$$Y_m^{x,y} = Y_m^{x,z}. \tag{226}$$

We now prove (226) by induction on $m \in \mathbb{N}_0 \cap [0, n]$. Note that the fact that $Y_0^{x,y} = x = Y_0^{x,z}$ implies (226) in the base case $m = 0$. For the induction step observe that (212) and the fact that for all $k \in \mathbb{N} \cap [0, n]$ it holds that $y_k = z_k$ ensure that for all $m \in \mathbb{N}_0 \cap (-\infty, n)$ with $Y_m^{x,y} = Y_m^{x,z}$ it holds that

$$\begin{aligned} Y_{m+1}^{x,y} &= Y_m^{x,y} + A_{m+1}\big((\mathcal{R}_a(\Phi))(Y_m^{x,y})\big) + y_{m+1} \\ &= Y_m^{x,z} + A_{m+1}\big((\mathcal{R}_a(\Phi))(Y_m^{x,z})\big) + z_{m+1} = Y_{m+1}^{x,z}. \end{aligned} \tag{227}$$

Induction thus proves (226). Note that (226) and (2.5.3) assure that for all $n \in \{0, 1, \ldots, N\}$, $m \in \mathbb{N}_0 \cap [0, n]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ with $\forall\, k \in \mathbb{N} \cap [0, n] \colon y_k = z_k$ it holds that

$$(\mathcal{R}_a(\Psi_{m,y}))(x) = (\mathcal{R}_a(\Psi_{m,z}))(x). \tag{228}$$

Combining this with (225) and (I)–(IV) establishes items (i)–(vi). The proof of Proposition 2.32 is thus completed. □

3 ANN approximation results

This section establishes in Theorem 3.12 in Section 3.3 below the main result of this article. Some of the material presented in Sections 3.1 and 3.2 are well-known concepts and results in the scientific literature. In particular, the material in Sections 3.1.1 and 3.1.2 consists mainly of reformulations of concepts and results in Elbrächter et al. [11, Appendix A.3 and Appendix A.4]. Moreover, our proof of Proposition 3.5 in Section 3.2.1 below is inspired by Elbrächter et al. [11, Section 6] and Yarotsky [40, Section 3.1] (cf., e.g., also [34, Lemma A.3] and [38, Lemma A.2]). Furthermore, Lemma 3.8 and Lemma 3.9 are elementary and essentially well-known in the scientific literature. In addition, our proof of Lemma 3.11 is based on a well-known Gronwall argument.

3.1 ANN approximations for the square function

3.1.1 Explicit approximations for the square function on $[0, 1]$

Lemma 3.1 *Let $g_n : \mathbb{R} \to [0, 1]$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$, $x \in \mathbb{R}$ that*

$$g_1(x) = \begin{cases} 2x & : x \in [0, \frac{1}{2}) \\ 2 - 2x & : x \in [\frac{1}{2}, 1] \\ 0 & : x \in \mathbb{R}\backslash[0, 1] \end{cases} \tag{229}$$

and $g_{n+1}(x) = g_1(g_n(x))$. Then

(i) *it holds for all $n \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\frac{k}{2^{n-1}}, \frac{k+1}{2^{n-1}}\right]$ that*

$$g_n(x) = \begin{cases} 2^n(x - \frac{2k}{2^n}) & : x \in \left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right] \\ 2^n(\frac{2k+2}{2^n} - x) & : x \in \left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right] \end{cases} \tag{230}$$

and

(ii) *it holds for all $n \in \mathbb{N}$, $x \in \mathbb{R}\backslash[0, 1]$ that $g_n(x) = 0$.*

Proof of Lemma 3.1 First, we claim that for all $n \in \mathbb{N}$ it holds that

$$\left(\forall\, k \in \{0, 1, \ldots, 2^{n-1} - 1\}, x \in \left[\frac{k}{2^{n-1}}, \frac{k+1}{2^{n-1}}\right] : g_n(x) = \begin{cases} 2^n(x - \frac{2k}{2^n}) & : x \in \left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right] \\ 2^n(\frac{2k+2}{2^n} - x) & : x \in \left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right] \end{cases}\right). \tag{231}$$

We now prove (231) by induction on $n \in \mathbb{N}$. Note that (229) establishes (231) in the base case $n = 1$. For the induction step $\mathbb{N} \ni n \to n+1 \in \mathbb{N} \cap [2, \infty)$ assume that there exists $n \in \mathbb{N}$ such that for all $k \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\frac{k}{2^{n-1}}, \frac{k+1}{2^{n-1}}\right]$ it holds that

$$g_n(x) = \begin{cases} 2^n\left(x - \frac{2k}{2^n}\right) & : x \in \left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right] \\ 2^n\left(\frac{2k+2}{2^n} - x\right) & : x \in \left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right] \end{cases}. \tag{232}$$

Observe that (229) and (232) imply that for all $l \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\frac{2l}{2^n}, \frac{2l+(1/2)}{2^n}\right]$ it holds that

$$g_{n+1}(x) = g_1(g_n(x)) = g_1\left(2^n\left(x - \tfrac{2l}{2^n}\right)\right) = 2\left[2^n\left(x - \tfrac{2l}{2^n}\right)\right] = 2^{n+1}\left(x - \tfrac{2l}{2^n}\right). \tag{233}$$

In addition, note that (229) and (232) ensure that for all $l \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\frac{2l+(1/2)}{2^n}, \frac{2l+1}{2^n}\right]$ it holds that

$$\begin{aligned} g_{n+1}(x) &= g_1(g_n(x)) = g_1\left(2^n\left(x - \tfrac{2l}{2^n}\right)\right) = 2 - 2\left[2^n\left(x - \tfrac{2l}{2^n}\right)\right] \\ &= 2 - 2^{n+1}x + 4l = 2^{n+1}\left(\tfrac{4l+2}{2^{n+1}} - x\right). \end{aligned} \tag{234}$$

Moreover, observe that (229) and (232) demonstrate that for all $l \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\frac{2l+1}{2^n}, \frac{2l+(3/2)}{2^n}\right]$ it holds that

$$\begin{aligned} g_{n+1}(x) &= g_1(g_n(x)) = g_1\left(2^n\left(\tfrac{2l+2}{2^n} - x\right)\right) = 2 - 2\left[2^n\left(\tfrac{2l+2}{2^n} - x\right)\right] \\ &= 2 - 2(2l+2) + 2^{n+1}x = 2^{n+1}x - 4l - 2 \\ &= 2^{n+1}\left(x - \tfrac{4l+2}{2^{n+1}}\right). \end{aligned} \tag{235}$$

Next note that (229) and (232) prove that for all $l \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\frac{2l+(3/2)}{2^n}, \frac{2l+2}{2^n}\right]$ it holds that

$$g_{n+1}(x) = g_1(g_n(x)) = g_1\left(2^n\left(\tfrac{2l+2}{2^n} - x\right)\right) = 2\left[2^n\left(\tfrac{2l+2}{2^n} - x\right)\right] = 2^{n+1}\left(\tfrac{2l+2}{2^n} - x\right). \tag{236}$$

Moreover, observe that for all $k \in \{0, 2, 4, 6, \ldots\} \cap [0, 2^n - 2]$ it holds that

$$\left[\tfrac{2k}{2^{n+1}}, \tfrac{2k+1}{2^{n+1}}\right] = \left[\tfrac{2(k/2)}{2^n}, \tfrac{2(k/2)+(1/2)}{2^n}\right], \qquad \left[\tfrac{2k+1}{2^{n+1}}, \tfrac{2k+2}{2^{n+1}}\right] = \left[\tfrac{2(k/2)+(1/2)}{2^n}, \tfrac{2(k/2)+1}{2^n}\right], \tag{237}$$

and $k/2 \in \{0, 1, \ldots, 2^{n-1} - 1\}$. This, (233), and (234) demonstrate that for all $k \in \{0, 2, 4, 6, \ldots\} \cap [0, 2^n - 2]$, $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]$ it holds that

$$\begin{aligned} g_{n+1}(x) &= \begin{cases} 2^{n+1}\left(x - \frac{2(k/2)}{2^n}\right) & : x \in \left[\frac{2(k/2)}{2^n}, \frac{2(k/2)+(1/2)}{2^n}\right] \\ 2^{n+1}\left(\frac{4(k/2)+2}{2^{n+1}} - x\right) & : x \in \left[\frac{2(k/2)+(1/2)}{2^n}, \frac{2(k/2)+1}{2^n}\right] \end{cases} \\ &= \begin{cases} 2^{n+1}\left(x - \frac{2k}{2^{n+1}}\right) & : x \in \left[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right] \\ 2^{n+1}\left(\frac{2k+2}{2^{n+1}} - x\right) & : x \in \left[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right] \end{cases}. \end{aligned} \tag{238}$$

In addition, observe that for all $k \in \{1, 3, 5, 7, \dots\} \cap [1, 2^n - 1]$ it holds that

$$\left[\tfrac{2k}{2^{n+1}}, \tfrac{2k+1}{2^{n+1}}\right] = \left[\tfrac{2((k-1)/2)+1}{2^n}, \tfrac{2((k-1)/2)+(3/2)}{2^n}\right],$$
$$\left[\tfrac{2k+1}{2^{n+1}}, \tfrac{2k+2}{2^{n+1}}\right] = \left[\tfrac{2((k-1)/2)+(3/2)}{2^n}, \tfrac{2((k-1)/2)+2}{2^n}\right], \tag{239}$$

and $(k-1)/2 \in \{0, 1, \dots, 2^{n-1} - 1\}$. This, (235), and (236) demonstrate that for all $k \in \{1, 3, 5, 7, \dots\} \cap [1, 2^n - 1]$, $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]$ it holds that

$$\begin{aligned} g_{n+1}(x) &= \begin{cases} 2^{n+1}\left(x - \frac{4((k-1)/2)+2}{2^{n+1}}\right) & : x \in \left[\frac{2((k-1)/2)+1}{2^n}, \frac{2((k-1)/2)+(3/2)}{2^n}\right] \\ 2^{n+1}\left(\frac{2((k-1)/2)+2}{2^n} - x\right) & : x \in \left[\frac{2((k-1)/2)+(3/2)}{2^n}, \frac{2((k-1)/2)+2}{2^n}\right] \end{cases} \\ &= \begin{cases} 2^{n+1}\left(x - \frac{2k}{2^{n+1}}\right) & : x \in \left[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right] \\ 2^{n+1}\left(\frac{2k+2}{2^{n+1}} - x\right) & : x \in \left[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right] \end{cases}. \end{aligned} \tag{240}$$

Combining this with (238) ensures that for all $k \in \{0, 1, \dots, 2^n - 1\}$, $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]$ it holds that

$$g_{n+1}(x) = \begin{cases} 2^{n+1}\left(x - \frac{2k}{2^{n+1}}\right) & : x \in \left[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right] \\ 2^{n+1}\left(\frac{2k+2}{2^{n+1}} - x\right) & : x \in \left[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right] \end{cases}. \tag{241}$$

Induction thus proves (231). Observe that (231) establishes item (i). Next we claim that for all $n \in \mathbb{N}$ it holds that

$$\forall\, x \in \mathbb{R}\backslash[0, 1]\colon\ g_n(x) = 0. \tag{242}$$

We now prove (242) by induction on $n \in \mathbb{N}$. Note that (229) establishes (242) in the base case $n = 1$. For the induction step observe that (229) ensures that for all $n \in \mathbb{N}$ with $(\forall x \in \mathbb{R}\backslash[0, 1]\colon g_n(x) = 0)$ it holds that

$$\big(\forall\, x \in \mathbb{R}\backslash[0, 1]\colon\ g_{n+1}(x) = g_1(g_n(x)) = g_1(0) = 0\big). \tag{243}$$

Induction thus proves (242). Note that (242) establishes item (ii). The proof of Lemma 3.1 is thus completed. □

Lemma 3.2 *Let $g_n\colon [0, 1] \to [0, 1]$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$, $x \in [0, 1]$ that*

$$g_1(x) = \begin{cases} 2x & : x \in [0, \frac{1}{2}) \\ 2 - 2x & : x \in [\frac{1}{2}, 1] \end{cases} \tag{244}$$

and $g_{n+1}(x) = g_1(g_n(x))$, and let $f_n\colon [0, 1] \to [0, 1]$, $n \in \mathbb{N}_0$, be the functions which satisfy for all $n \in \mathbb{N}_0$, $k \in \{0, 1, \dots, 2^n - 1\}$, $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$ that $f_n(1) = 1$ and

$$f_n(x) = \left[\tfrac{2k+1}{2^n}\right] x - \tfrac{(k^2+k)}{2^{2n}}. \tag{245}$$

Then it holds for all $n \in \mathbb{N}_0$, $x \in [0, 1]$ that

$$f_n(x) = x - \left[\sum_{m=1}^{n} \left(2^{-2m} g_m(x)\right)\right] \quad \textit{and} \quad \left|x^2 - f_n(x)\right| \le 2^{-2n-2}. \tag{246}$$

Proof of Lemma 3.2 Note that (245) proves that for all $n \in \mathbb{N}_0, l \in \{0, 1, \ldots, 2^n - 1\}$ it holds that

$$f_n(\tfrac{l}{2^n}) = \left[\tfrac{2l+1}{2^n}\right] \tfrac{l}{2^n} - \tfrac{(l^2+l)}{2^{2n}} = \tfrac{(2l+1)l-(l^2+l)}{2^{2n}} = \tfrac{l^2}{2^{2n}} = \left[\tfrac{l}{2^n}\right]^2. \tag{247}$$

The hypothesis that for all $n \in \mathbb{N}_0$ it holds that $f_n(1) = 1$ hence ensures that for all $n \in \mathbb{N}_0, l \in \{0, 1, \ldots, 2^n\}$ it holds that

$$f_n(\tfrac{l}{2^n}) = \left[\tfrac{l}{2^n}\right]^2. \tag{248}$$

This and Lemma 3.1 demonstrate that for all $n \in \mathbb{N}, k \in \{0, 1, \ldots, 2^{n-1}\}$ it holds that

$$\begin{aligned} f_{n-1}(\tfrac{2k}{2^n}) - f_n(\tfrac{2k}{2^n}) &= f_{n-1}(\tfrac{k}{2^{n-1}}) - f_n(\tfrac{2k}{2^n}) = \left[\tfrac{k}{2^{n-1}}\right]^2 - \left[\tfrac{2k}{2^n}\right]^2 \\ &= 0 = 2^{-2n} g_n(\tfrac{2k}{2^n}). \end{aligned} \tag{249}$$

In addition, note that (245) and (248) imply that for all $n \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^{n-1} - 1\}$ it holds that

$$\begin{aligned} f_{n-1}(\tfrac{2k+1}{2^n}) = f_{n-1}\left(\tfrac{k+(1/2)}{2^{n-1}}\right) &= \left[\tfrac{2k+1}{2^{n-1}}\right]\left[\tfrac{2k+1}{2^n}\right] - \tfrac{(k^2+k)}{2^{2(n-1)}} = \tfrac{(4k^2+4k+1)}{2^{2n-1}} - \tfrac{(2k^2+2k)}{2^{2n-1}} \\ &= \tfrac{2k^2+2k+1}{2^{2n-1}} = \tfrac{4k^2+4k+2}{2^{2n}} \end{aligned} \tag{250}$$

and

$$f_n(\tfrac{2k+1}{2^n}) = \left[\tfrac{2k+1}{2^n}\right]^2 = \tfrac{4k^2+4k+1}{2^{2n}}. \tag{251}$$

Lemma 3.1 hence assures that for all $n \in \mathbb{N}, k \in \{0, 1, \ldots, 2^{n-1} - 1\}$ it holds that

$$f_{n-1}(\tfrac{2k+1}{2^n}) - f_n(\tfrac{2k+1}{2^n}) = \tfrac{(4k^2+4k+2)}{2^{2n}} - \tfrac{(4k^2+4k+1)}{2^{2n}} = 2^{-2n} = 2^{-2n} g_n(\tfrac{2k+1}{2^n}). \tag{252}$$

Combining this with (249) shows that for all $n \in \mathbb{N}, l \in \{0, 1, \ldots, 2^n\}$ it holds that

$$f_{n-1}(\tfrac{l}{2^n}) - f_n(\tfrac{l}{2^n}) = 2^{-2n} g_n(\tfrac{l}{2^n}). \tag{253}$$

Furthermore, observe that (248) demonstrates that for all $n \in \mathbb{N}_0$, $l \in \{0, 1, \ldots, 2^n - 1\}$ it holds that

$$\left[\tfrac{2l+1}{2^n}\right]\left[\tfrac{l+1}{2^n}\right] - \tfrac{(l^2+l)}{2^{2n}} = \tfrac{(2l+1)(l+1)-l(l+1)}{2^{2n}} = \tfrac{(l+1)^2}{2^{2n}} = \left[\tfrac{l+1}{2^n}\right]^2 = f_n(\tfrac{l+1}{2^n}). \tag{254}$$

Combining this with (245) implies that for all $n \in \mathbb{N}_0$ it holds that $f_n \in C([0,1], \mathbb{R})$ and

$$\forall\, l \in \{0, 1, \ldots, 2^n - 1\}, x \in \left[\tfrac{l}{2^n}, \tfrac{l+1}{2^n}\right]\colon\ f_n(x) = \left[\tfrac{2l+1}{2^n}\right] x - \tfrac{(l^2+l)}{2^{2n}}. \tag{255}$$

The fact that for all $n \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^{n-1} - 1\}$ it holds that $\left[\tfrac{k}{2^{n-1}}, \tfrac{k+1}{2^{n-1}}\right] = \left[\tfrac{2k}{2^n}, \tfrac{2k+1}{2^n}\right] \cup \left[\tfrac{2k+1}{2^n}, \tfrac{2k+2}{2^n}\right]$ hence ensures that there exist $(a_{n,k}, b_{n,k}, c_{n,k}) \in \mathbb{R}^3$, $k \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $n \in \mathbb{N}$, such that for all $n \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\tfrac{k}{2^{n-1}}, \tfrac{k+1}{2^{n-1}}\right]$ it holds that

$$f_{n-1}(x) - f_n(x) = \begin{cases} a_{n,k}\left(x - \tfrac{(2k+1)}{2^n}\right) + b_{n,k} & : x \in \left[\tfrac{2k}{2^n}, \tfrac{2k+1}{2^n}\right] \\ c_{n,k}\left(x - \tfrac{(2k+1)}{2^n}\right) + b_{n,k} & : x \in \left[\tfrac{2k+1}{2^n}, \tfrac{2k+2}{2^n}\right] \end{cases}. \tag{256}$$

Lemma 3.1 and (253) therefore prove that for all $n \in \mathbb{N}$, $k \in \{0, 1, \ldots, 2^{n-1} - 1\}$, $x \in \left[\frac{k}{2^{n-1}}, \frac{k+1}{2^{n-1}}\right]$ it holds that

$$f_{n-1}(x) - f_n(x) = 2^{-2n} g_n(x). \tag{257}$$

Hence, we obtain that for all $n \in \mathbb{N}$, $x \in [0, 1]$ it holds that

$$f_{n-1}(x) - f_n(x) = 2^{-2n} g_n(x). \tag{258}$$

Next note that (245) ensures that for all $x \in [0, 1]$ it holds that $f_0(x) = x$. Combining this with (258) implies that for all $m \in \mathbb{N}_0$, $x \in [0, 1]$ it holds that

$$\begin{aligned} f_m(x) &= f_0(x) + \left[\sum_{n=1}^{m} (f_n(x) - f_{n-1}(x)) \right] \\ &= f_0(x) - \left[\sum_{n=1}^{m} (f_{n-1}(x) - f_n(x)) \right] = x - \left[\sum_{n=1}^{m} 2^{-2n} g_n(x) \right]. \end{aligned} \tag{259}$$

Moreover, observe that (255) demonstrates that for all $m \in \mathbb{N}_0$, $l \in \{0, 1, \ldots, 2^m - 1\}$, $x \in \left[\frac{l}{2^m}, \frac{l+1}{2^m}\right]$ it holds that

$$\begin{aligned} f_m(x) - x^2 &= \left[\frac{2l+1}{2^m}\right] x - \frac{(l^2+l)}{2^{2m}} - x^2 = \left[\frac{l+1}{2^m}\right] x + \left[\frac{l}{2^m}\right] x - \left[\frac{l+1}{2^m}\right]\left[\frac{l}{2^m}\right] - x^2 \\ &= \left(x - \frac{l}{2^m}\right)\left(\frac{l+1}{2^m} - x\right) \geqslant 0. \end{aligned} \tag{260}$$

The fact that for all $a \in \mathbb{R}$, $b \in (a, \infty)$, $r \in [a, b]$ it holds that $(r - a)(b - r) \leqslant \frac{1}{4}(b - a)^2$ hence proves that for all $m \in \mathbb{N}_0$, $l \in \{0, 1, \ldots, 2^m - 1\}$, $x \in \left[\frac{l}{2^m}, \frac{l+1}{2^m}\right]$ it holds that

$$\begin{aligned} \left|f_m(x) - x^2\right| &= f_m(x) - x^2 = \left(x - \frac{l}{2^m}\right)\left(\frac{l+1}{2^m} - x\right) \\ &\leqslant \frac{1}{4}\left(\frac{l+1}{2^m} - \frac{l}{2^m}\right)^2 = \frac{1}{4}\left(\frac{1}{2^m}\right)^2 = \frac{1}{2^2}\left(\frac{1}{2^{2m}}\right) = \frac{1}{2^{2m+2}} = 2^{-2m-2}. \end{aligned} \tag{261}$$

Therefore, we obtain that for all $m \in \mathbb{N}_0$, $x \in [0, 1]$ it holds that

$$\left|f_m(x) - x^2\right| \leqslant 2^{-2m-2}. \tag{262}$$

Combining this with (259) establishes (246). The proof of Lemma 3.2 is thus completed. □

3.1.2 ANN approximations for the square function on [0, 1]

Proposition 3.3 *Let* $\varepsilon \in (0, 1]$, $a \in C(\mathbb{R}, \mathbb{R})$ *satisfy for all* $x \in \mathbb{R}$ *that* $a(x) = \max\{x, 0\}$. *Then there exists* $\Phi \in \mathbf{N}$ *such that*

(i) *it holds that* $\mathcal{R}_a(\Phi) \in C(\mathbb{R}, \mathbb{R})$,
(ii) *it holds for all* $x \in \mathbb{R}\backslash[0, 1]$ *that* $(\mathcal{R}_a(\Phi))(x) = a(x)$,
(iii) *it holds for all* $x \in [0, 1]$ *that* $\left|x^2 - (\mathcal{R}_a(\Phi))(x)\right| \leqslant \varepsilon$,
(iv) *it holds that* $\mathcal{P}(\Phi) \leqslant \max\{10 \log_2(\varepsilon^{-1}) - 7, 13\}$, *and*
(v) *it holds that* $\mathcal{L}(\Phi) \leqslant \max\{\frac{1}{2} \log_2(\varepsilon^{-1}) + 1, 2\}$

(cf. Definition 2.1 and Definition 2.3).

Proof of Proposition 3.3 Throughout this proof let $M \in \mathbb{N}$ satisfy that

$$M = \min\Big(\mathbb{N} \cap [2, \infty) \cap \big[\tfrac{1}{2}\log_2(\varepsilon^{-1}), \infty\big)\Big), \tag{263}$$

let $g_n : \mathbb{R} \to [0, 1]$, $n \in \mathbb{N}$, be the functions which satisfy for all $n \in \mathbb{N}$, $x \in \mathbb{R}$ that

$$g_1(x) = \begin{cases} 2x & : x \in [0, \frac{1}{2}) \\ 2 - 2x & : x \in [\frac{1}{2}, 1] \\ 0 & : x \in \mathbb{R}\backslash[0, 1] \end{cases} \tag{264}$$

and $g_{n+1}(x) = g_1(g_n(x))$, let $f_n : [0, 1] \to [0, 1]$, $n \in \mathbb{N}_0$, be the functions which satisfy for all $n \in \mathbb{N}_0$, $k \in \{0, 1, \ldots, 2^n - 1\}$, $x \in \big[\frac{k}{2^n}, \frac{k+1}{2^n}\big)$ that $f_n(1) = 1$ and

$$f_n(x) = \left[\tfrac{2k+1}{2^n}\right] x - \tfrac{(k^2+k)}{2^{2n}}, \tag{265}$$

let $(A_k, b_k) \in \mathbb{R}^{4\times 4} \times \mathbb{R}^4$, $k \in \mathbb{N} \cap [2, \infty)$, satisfy for all $k \in \mathbb{N} \cap [2, \infty)$ that

$$A_k = \begin{pmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ (-2)^{3-2k} & 2^{4-2k} & (-2)^{3-2k} & 1 \end{pmatrix} \quad \text{and} \quad b_k = \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \tag{266}$$

let $\mathbb{A}_k \in \mathbb{R}^{1\times 4} \times \mathbb{R}$, $k \in \mathbb{N} \cap [2, \infty)$, satisfy for all $k \in \mathbb{N} \cap [2, \infty)$ that

$$\mathbb{A}_k = \big(\big((-2)^{3-2k} \;\; 2^{4-2k} \;\; (-2)^{3-2k} \;\; 1\big), 0\big), \tag{267}$$

let $\phi_k \in \mathbf{N}$, $k \in \mathbb{N} \cap [2, \infty)$, satisfy that

$$\phi_2 = \left(\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}\right), \mathbb{A}_2\right) \tag{268}$$

and

$$\forall\, k \in \mathbb{N} \cap [3, \infty) : \phi_k = \left(\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{pmatrix}\right), (A_2, b_2), \ldots, (A_{k-1}, b_{k-1}), \mathbb{A}_k\right), \tag{269}$$

and let $r_k = (r_{k,1}, r_{k,2}, r_{k,3}, r_{k,4}) : \mathbb{R} \to \mathbb{R}^4$, $k \in \mathbb{N}$, be the functions which satisfy for all $x \in \mathbb{R}$, $k \in \mathbb{N}$ that

$$r_1(x) = (r_{1,1}(x), r_{1,2}(x), r_{1,3}(x), r_{1,4}(x)) = \mathfrak{M}_{a,4}\big(x, x - \tfrac{1}{2}, x - 1, x\big) \tag{270}$$

and

$$r_{k+1}(x) = (r_{k+1,1}(x), r_{k+1,2}(x), r_{k+1,3}(x), r_{k+1,4}(x)) = \mathfrak{M}_{a,4}\big(A_{k+1} r_k(x) + b_{k+1}\big) \tag{271}$$

(cf. Definition 2.2). Note that (270), (21), (264), and the hypothesis that for all $x \in \mathbb{R}$ it holds that $a(x) = \max\{x, 0\}$ show that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned}&2r_{1,1}(x) - 4r_{1,2}(x) + 2r_{1,3}(x) = 2a(x) - 4a(x - \tfrac{1}{2}) + 2a(x - 1)\\&= 2\max\{x, 0\} - 4\max\{x - \tfrac{1}{2}, 0\} + 2\max\{x - 1, 0\} = g_1(x).\end{aligned} \tag{272}$$

Furthermore, observe that (270), (21), the hypothesis that for all $x \in \mathbb{R}$ it holds that $a(x) = \max\{x, 0\}$, and the fact that for all $x \in [0, 1]$ it holds that $f_0(x) = x = \max\{x, 0\}$ imply that for all $x \in \mathbb{R}$ it holds that

$$r_{1,4}(x) = \max\{x, 0\} = \begin{cases} f_0(x) & : x \in [0, 1] \\ \max\{x, 0\} & : x \in \mathbb{R}\backslash[0, 1] \end{cases}. \tag{273}$$

Next we claim that for all $k \in \mathbb{N}$ it holds that

$$\big(\forall\, x \in \mathbb{R}\colon\ 2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) = g_k(x)\big) \tag{274}$$

and

$$\left(\forall\, x \in \mathbb{R}\colon\ r_{k,4}(x) = \begin{cases} f_{k-1}(x) & : x \in [0, 1] \\ \max\{x, 0\} & : x \in \mathbb{R}\backslash[0, 1] \end{cases}\right). \tag{275}$$

We now prove (274)–(275) by induction on $k \in \mathbb{N}$. Note that (272) and (273) prove (274)–(275) in the base case $k = 1$. For the induction step $\mathbb{N} \ni k \to k+1 \in \mathbb{N} \cap [2, \infty)$ assume that there exists $k \in \mathbb{N}$ such that for all $x \in \mathbb{R}$ it holds that

$$2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) = g_k(x) \tag{276}$$

$$\text{and} \qquad r_{k,4}(x) = \begin{cases} f_{k-1}(x) & : x \in [0, 1] \\ \max\{x, 0\} & : x \in \mathbb{R}\backslash[0, 1] \end{cases}. \tag{277}$$

Observe that (276), (272), (266), (21), and (271) ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} g_{k+1}(x) &= g_1(g_k(x)) = g_1(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)) \\ &= 2\,a\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)\big) \\ &\quad -4\,a\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) - \tfrac{1}{2}\big) \\ &\quad +2\,a\big(2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x) - 1\big) \\ &= 2r_{k+1,1}(x) - 4r_{k+1,2}(x) + 2r_{k+1,3}(x). \end{aligned} \tag{278}$$

In addition, observe that (21), (266), (271), and (276) demonstrate that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} &r_{k+1,4}(x) \\ &= a\big((-2)^{3-2(k+1)} r_{k,1}(x) + 2^{4-2(k+1)} r_{k,2}(x) + (-2)^{3-2(k+1)} r_{k,3}(x) + r_{k,4}(x)\big) \\ &= a\big((-2)^{1-2k} r_{k,1}(x) + 2^{2-2k} r_{k,2}(x) + (-2)^{1-2k} r_{k,3}(x) + r_{k,4}(x)\big) \\ &= a\big(2^{-2k}\big[-2r_{k,1}(x) + 2^2 r_{k,2}(x) - 2r_{k,3}(x)\big] + r_{k,4}(x)\big) \\ &= a\big(-\big[2^{-2k}\big]\big[2r_{k,1}(x) - 4r_{k,2}(x) + 2r_{k,3}(x)\big] + r_{k,4}(x)\big) \\ &= a\big(-\big[2^{-2k}\big] g_k(x) + r_{k,4}(x)\big). \end{aligned} \tag{279}$$

Combining this with (277), Lemma 3.2, the hypothesis that for all $x \in \mathbb{R}$ it holds that $a(x) = \max\{x, 0\}$, and the fact that for all $x \in [0, 1]$ it holds that $f_k(x) \geqslant 0$ shows that for all $x \in [0, 1]$ it holds that

$$\begin{aligned} r_{k+1,4}(x) &= a\big(-\big[2^{-2k} g_k(x)\big] + f_{k-1}(x)\big) \\ &= a\Big(-\big(2^{-2k} g_k(x)\big) + x - \Big[\sum_{j=1}^{k-1} \big(2^{-2j} g_j(x)\big)\Big]\Big) \\ &= a\Big(x - \Big[\sum_{j=1}^{k} 2^{-2j} g_j(x)\Big]\Big) = a(f_k(x)) = f_k(x). \end{aligned} \tag{280}$$

Next note that (279), (277), item (ii) in Lemma 3.1, and the hypothesis that for all $x \in \mathbb{R}$ it holds that $a(x) = \max\{x, 0\}$ prove that for all $x \in \mathbb{R}\backslash[0, 1]$ it holds that

$$r_{k+1,4}(x) = a\Big(-\big(2^{-2k} g_k(x)\big) + r_{k,4}(x)\Big) = a(\max\{x, 0\}) = \max\{x, 0\}. \tag{281}$$

Combining (278) and (280) hence proves (274)–(275) in the case $k+1$. Induction thus establishes (274)–(275). Next note that (22), (266), (267), (274), (268), (269), (270), and (271) assure that for all $m \in \mathbb{N} \cap [2, \infty)$, $x \in \mathbb{R}$ it holds that $\mathcal{R}_a(\phi_m) \in C(\mathbb{R}, \mathbb{R})$ and

$$\begin{aligned} &(\mathcal{R}_a(\phi_m))(x) \\ &= (-2)^{3-2m} r_{m-1,1}(x) + 2^{4-2m} r_{m-1,2}(x) + (-2)^{3-2m} r_{m-1,3}(x) + r_{m-1,4}(x) \\ &= (-2)^{4-2m}\Big(\Big[\tfrac{r_{m-1,1}(x)+r_{m-1,3}(x)}{(-2)}\Big] + r_{m-1,2}(x)\Big) + r_{m-1,4}(x) \\ &= 2^{4-2m}\Big(\Big[\tfrac{r_{m-1,1}(x)+r_{m-1,3}(x)}{(-2)}\Big] + r_{m-1,2}(x)\Big) + r_{m-1,4}(x) \\ &= 2^{2-2m}\big(4r_{m-1,2}(x) - 2r_{m-1,1}(x) - 2r_{m-1,3}(x)\big) + r_{m-1,4}(x) \\ &= -\big[2^{-2(m-1)}\big]\big[2r_{m-1,1}(x) - 4r_{m-1,2}(x) + 2r_{m-1,3}(x)\big] + r_{m-1,4}(x) \\ &= -\big[2^{-2(m-1)}\big] g_{m-1}(x) + r_{m-1,4}(x). \end{aligned} \tag{282}$$

Combining this with (275) and Lemma 3.2 shows that for all $m \in \mathbb{N} \cap [2, \infty)$, $x \in [0, 1]$ it holds that

$$\begin{aligned} (\mathcal{R}_a(\phi_m))(x) &= -\big(2^{-2(m-1)} g_{m-1}(x)\big) + f_{m-2}(x) \\ &= -\big(2^{-2(m-1)} g_{m-1}(x)\big) + x - \Big[\sum_{j=1}^{m-2} 2^{-2j} g_j(x)\Big] \\ &= x - \Big[\sum_{j=1}^{m-1} 2^{-2j} g_j(x)\Big] = f_{m-1}(x). \end{aligned} \tag{283}$$

Lemma 3.2 therefore implies that for all $m \in \mathbb{N} \cap [2, \infty)$, $x \in [0, 1]$ it holds that

$$\Big|x^2 - (\mathcal{R}_a(\phi_m))(x)\Big| \leqslant 2^{-2(m-1)-2} = 2^{-2m}. \tag{284}$$

Next note that (263) assures that

$$\begin{aligned} M &= \min\Big(\mathbb{N} \cap \big[\max\{2, \tfrac{1}{2}\log_2(\varepsilon^{-1})\}, \infty)\Big) \\ &\geqslant \min\Big(\big[\max\{2, \tfrac{1}{2}\log_2(\varepsilon^{-1})\}, \infty)\Big) \\ &= \max\{2, \tfrac{1}{2}\log_2(\varepsilon^{-1})\} \geqslant \tfrac{1}{2}\log_2(\varepsilon^{-1}). \end{aligned} \tag{285}$$

This and (284) demonstrate that for all $x \in [0, 1]$ it holds that

$$\Big|x^2 - (\mathcal{R}_a(\phi_M))(x)\Big| \leqslant 2^{-2M} \leqslant 2^{-\log_2(\varepsilon^{-1})} = \varepsilon. \tag{286}$$

Moreover, observe that item (ii) in Lemma 3.1, (275), and (282) ensure that for all $m \in \mathbb{N} \cap [2, \infty)$, $x \in \mathbb{R}\backslash[0, 1]$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\phi_m))(x) &= -2^{-2(m-1)} g_{m-1}(x) + r_{m-1,4}(x)\\ &= r_{m-1,4}(x) = \max\{x, 0\} = a(x).\end{aligned} \tag{287}$$

Furthermore, observe that (263), (268), and (269) assure that

$$\mathcal{L}(\phi_M) = M \leqslant \max\{\tfrac{1}{2}\log_2(\varepsilon^{-1}) + 1, 2\}. \tag{288}$$

This, (263), (268), and (269) show that

$$\begin{aligned}\mathcal{P}(\phi_M) &= 4(1+1) + \left[\sum_{j=2}^{M-1} 4(4+1)\right] + (4+1)\\ &= 8 + 20(M-2) + 5 \leqslant 20\max\{\tfrac{1}{2}\log_2(\varepsilon^{-1}) - 1, 0\} + 13\\ &= \max\{10\log_2(\varepsilon^{-1}) - 20, 0\} + 13 = \max\{10\log_2(\varepsilon^{-1}) - 7, 13\}.\end{aligned} \tag{289}$$

Combining (286), (288), (287), and the fact that $\mathcal{R}_a(\phi_M) \in C(\mathbb{R}, \mathbb{R})$ hence establishes items (i)–(v). The proof of Proposition 3.3 is thus completed. □

3.1.3 ANN approximations for the square function on $\mathbb{R}$

Proposition 3.4 *Let $\varepsilon \in (0, 1]$, $q \in (2, \infty)$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$. Then there exists $\Phi \in \mathbf{N}$ such that*

(i) *it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}, \mathbb{R})$,*
(ii) *it holds that $(\mathcal{R}_a(\Phi))(0) = 0$,*
(iii) *it holds for all $x \in \mathbb{R}$ that $0 \leqslant (\mathcal{R}_a(\Phi))(x) \leqslant \varepsilon + |x|^2$,*
(iv) *it holds for all $x \in \mathbb{R}$ that $|x^2 - (\mathcal{R}_a(\Phi))(x)| \leqslant \varepsilon \max\{1, |x|^q\}$,*
(v) *it holds that $\mathcal{P}(\Phi) \leqslant \max\{\left[\frac{40q}{(q-2)}\right]\log_2(\varepsilon^{-1}) + \frac{80}{(q-2)} - 28, 52\}$, and*
(vi) *it holds that $\mathcal{L}(\Phi) \leqslant \max\{\frac{q}{2(q-2)}\log_2(\varepsilon^{-1}) + \frac{1}{(q-2)} + 1, 2\}$*

(cf. Definition 2.1 and Definition 2.3).

Proof of Proposition 3.4 Throughout this proof let $\delta \in (0, 1]$ satisfy that $\delta = 2^{-2/(q-2)}\varepsilon^{q/(q-2)}$, let $\mathbb{A}_1 \in (\mathbb{R}^{2\times 1} \times \mathbb{R}^2) \subseteq \mathbf{N}$, $\mathbb{A}_2 \in (\mathbb{R}^{1\times 2} \times \mathbb{R}) \subseteq \mathbf{N}$ satisfy that

$$\mathbb{A}_1 = \left(\begin{pmatrix}(\frac{\varepsilon}{2})^{1/(q-2)}\\ -(\frac{\varepsilon}{2})^{1/(q-2)}\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\right) \quad \text{and} \quad \mathbb{A}_2 = \left(\left((\tfrac{\varepsilon}{2})^{-2/(q-2)} \;\; (\tfrac{\varepsilon}{2})^{-2/(q-2)}\right), 0\right), \tag{290}$$

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

(I) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}, \mathbb{R})$,
(II) it holds for all $x \in \mathbb{R}\backslash[0, 1]$ that $(\mathcal{R}_a(\Psi))(x) = a(x)$,
(III) it holds for all $x \in [0, 1]$ that $\left|x^2 - (\mathcal{R}_a(\Psi))(x)\right| \leqslant \delta$,
(IV) it holds that $\mathcal{P}(\Psi) \leqslant \max\{10\log_2(\delta^{-1}) - 7, 13\}$, and
(V) it holds that $\mathcal{L}(\Psi) \leqslant \max\{\frac{1}{2}\log_2(\delta^{-1}) + 1, 2\}$

(cf. Proposition 3.3), and let $\Phi \in \mathbf{N}$ satisfy that

$$\Phi = \mathbb{A}_2 \bullet \big[\mathbf{P}_2(\Psi, \Psi)\big] \bullet \mathbb{A}_1 \tag{291}$$

(cf. Definition 2.5, Definition 2.17, and Lemma 2.8). Note that Proposition 2.19 and item (v) in Proposition 2.6 ensure that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} &\big(\mathcal{R}_a\big(\big(\mathbf{P}_2(\Psi, \Psi)\big) \bullet \mathbb{A}_1\big)\big)(x) = \big(\mathcal{R}_a\big(\mathbf{P}_2(\Psi, \Psi)\big)\big)\big(\big(\mathcal{R}_a(\mathbb{A}_1)\big)(x)\big) \\ &= \big(\mathcal{R}_a\big(\mathbf{P}_2(\Psi, \Psi)\big)\big)\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x, -(\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) \\ &= \begin{pmatrix} (\mathcal{R}_a(\Psi))\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) \\ (\mathcal{R}_a(\Psi))\big(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) \end{pmatrix}. \end{aligned} \tag{292}$$

Item (v) in Proposition 2.6 and (291) therefore demonstrate that for all $x \in \mathbb{R}$ it holds that

$$\begin{aligned} (\mathcal{R}_a(\Phi))(x) &= (\mathcal{R}_a(\mathbb{A}_2))\big(\mathcal{R}_a\big(\big[\mathbf{P}_2(\Psi, \Psi)\big] \bullet \mathbb{A}_1\big)(x)\big) \\ &= \big((\tfrac{\varepsilon}{2})^{-2/(q-2)} \;\; (\tfrac{\varepsilon}{2})^{-2/(q-2)}\big) \begin{pmatrix} [\mathcal{R}_a(\Psi)]\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) \\ [\mathcal{R}_a(\Psi)]\big(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) \end{pmatrix} \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)}\Big([\mathcal{R}_a(\Psi)]\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) + [\mathcal{R}_a(\Psi)]\big(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x\big)\Big). \end{aligned} \tag{293}$$

This, (I), (II), and the hypothesis that for all $x \in \mathbb{R}$ it holds that $a(x) = \max\{x, 0\}$ imply that

$$\begin{aligned} (\mathcal{R}_a(\Phi))(0) &= (\tfrac{\varepsilon}{2})^{-2/(q-2)}\big([\mathcal{R}_a(\Psi)](0) + [\mathcal{R}_a(\Psi)](0)\big) \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)}\big(a(0) + a(0)\big) = 0. \end{aligned} \tag{294}$$

Moreover, observe that (I) and (II) ensure that for all $x \in \mathbb{R}\backslash[-1, 1]$ it holds that

$$\begin{aligned} [\mathcal{R}_a(\Psi)](x) + [\mathcal{R}_a(\Psi)](-x) &= a(x) + a(-x) = \max\{x, 0\} + \max\{-x, 0\} \\ &= \max\{x, 0\} - \min\{x, 0\} = |x|. \end{aligned} \tag{295}$$

Furthermore, note that (II) and (III) show that

$$\begin{aligned} &\sup\nolimits_{x\in[-1,1]} \big|x^2 - \big([\mathcal{R}_a(\Psi)](x) + [\mathcal{R}_a(\Psi)](-x)\big)\big| \\ &= \max\Big\{ \sup\nolimits_{x\in[-1,0]} \big|x^2 - \big(a(x) + [\mathcal{R}_a(\Psi)](-x)\big)\big|, \sup\nolimits_{x\in[0,1]} \big|x^2 - \big([\mathcal{R}_a(\Psi)](x) + a(-x)\big)\big| \Big\} \\ &= \max\Big\{ \sup\nolimits_{x\in[-1,0]} \big|(-x)^2 - (\mathcal{R}_a(\Psi))(-x)\big|, \sup\nolimits_{x\in[0,1]} \big|x^2 - (\mathcal{R}_a(\Psi))(x)\big| \Big\} \\ &= \sup\nolimits_{x\in[0,1]} \big|x^2 - (\mathcal{R}_a(\Psi))(x)\big| \leqslant \delta. \end{aligned} \tag{296}$$

Next observe that (293) and (295) prove that for all $x \in \mathbb{R}\backslash[-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$\begin{aligned} 0 &\leqslant [\mathcal{R}_a(\Phi)](x) \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)}\Big([\mathcal{R}_a(\Psi)]\big((\tfrac{\varepsilon}{2})^{1/(q-2)}x\big) + [\mathcal{R}_a(\Psi)]\big(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x\big)\Big) \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)}\Big|(\tfrac{\varepsilon}{2})^{1/(q-2)}x\Big| = (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x| \leqslant |x|^2. \end{aligned} \tag{297}$$

The triangle inequality therefore ensures that for all $x \in \mathbb{R}\backslash[-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$\begin{aligned} \left|x^2 - (\mathcal{R}_a(\Phi))(x)\right| &= \left|x^2 - (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|\right| \leqslant \left(|x|^2 + (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|\right) \\ &= \left(|x|^q|x|^{-(q-2)} + (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|^q|x|^{-(q-1)}\right) \\ &\leqslant \left(|x|^q(\tfrac{\varepsilon}{2})^{(q-2)/(q-2)} + (\tfrac{\varepsilon}{2})^{-1/(q-2)}|x|^q(\tfrac{\varepsilon}{2})^{(q-1)/(q-2)}\right) \\ &= (\tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2})|x|^q = \varepsilon|x|^q \leqslant \varepsilon \max\{1, |x|^q\}. \end{aligned} \tag{298}$$

Next note that (293), (296), and the fact that $\delta = 2^{-2/(q-2)}\varepsilon^{q/(q-2)}$ demonstrate that for all $x \in [-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$\begin{aligned} &\left|x^2 - (\mathcal{R}_a(\Phi))(x)\right| \\ &= (\tfrac{\varepsilon}{2})^{-2/(q-2)}\left|\left((\tfrac{\varepsilon}{2})^{1/(q-2)}x\right)^2 - \left([\mathcal{R}_a(\Psi)]\left((\tfrac{\varepsilon}{2})^{1/(q-2)}x\right) + [\mathcal{R}_a(\Psi)]\left(-(\tfrac{\varepsilon}{2})^{1/(q-2)}x\right)\right)\right| \\ &\leqslant (\tfrac{\varepsilon}{2})^{-2/(q-2)}\left[\sup_{y\in[-1,1]}\left|y^2 - \left([\mathcal{R}_a(\Psi)](y) + [\mathcal{R}_a(\Psi)](-y)\right)\right|\right] \\ &\leqslant (\tfrac{\varepsilon}{2})^{-2/(q-2)}\delta = (\tfrac{\varepsilon}{2})^{-2/(q-2)}2^{-2/(q-2)}\varepsilon^{q/(q-2)} = \varepsilon \leqslant \varepsilon\max\{1, |x|^q\}. \end{aligned} \tag{299}$$

Combining this and (298) implies that for all $x \in \mathbb{R}$ it holds that

$$\left|x^2 - (\mathcal{R}_a(\Phi))(x)\right| \leqslant \varepsilon \max\{1, |x|^q\} \leqslant \varepsilon\left(1 + |x|^q\right). \tag{300}$$

In addition, note that (299) ensures that for all $x \in [-(\varepsilon/2)^{-1/(q-2)}, (\varepsilon/2)^{-1/(q-2)}]$ it holds that

$$|(\mathcal{R}_a(\Phi))(x)| \leqslant \left|x^2 - (\mathcal{R}_a(\Phi))(x)\right| + |x|^2 \leqslant \varepsilon + |x|^2. \tag{301}$$

This and (297) show for all $x \in \mathbb{R}$ that

$$|(\mathcal{R}_a(\Phi))(x)| \leqslant \varepsilon + |x|^2. \tag{302}$$

Furthermore, observe that the fact that $\delta = 2^{-2/(q-2)}\varepsilon^{q/(q-2)}$ ensures that

$$\log_2(\delta^{-1}) = \log_2(2^{2/(q-2)}\varepsilon^{-q/(q-2)}) = \tfrac{2}{(q-2)} + \left[\left[\tfrac{q}{(q-2)}\right]\log_2(\varepsilon^{-1})\right]. \tag{303}$$

Next note that Corollary 2.21 implies that $\mathcal{P}\big(\mathbf{P}_2(\Psi, \Psi)\big) \leqslant 4\,\mathcal{P}(\Psi)$. Corollary 2.9, (291), (IV), and (303) hence ensure that

$$\begin{aligned} \mathcal{P}(\Phi) &\leqslant \left[\max\left\{1, \tfrac{\mathcal{O}(\mathbb{A}_2)}{\mathcal{O}(\mathbf{P}_2(\Psi,\Psi))}\right\}\right]\left[\max\left\{1, \tfrac{\mathcal{I}(\mathbb{A}_1)+1}{\mathcal{I}(\mathbf{P}_2(\Psi,\Psi))+1}\right\}\right]\mathcal{P}\big(\mathbf{P}_2(\Psi, \Psi)\big) \\ &= \left[\max\{1, \tfrac{1}{2}\}\right]\left[\max\{1, \tfrac{2}{3}\}\right]\mathcal{P}\big(\mathbf{P}_2(\Psi, \Psi)\big) \\ &= \mathcal{P}\big(\mathbf{P}_2(\Psi, \Psi)\big) \leqslant 4\,\mathcal{P}(\Psi) \leqslant 4\max\{10\log_2(\delta^{-1}) - 7, 13\} \\ &= \max\left\{40\left[\tfrac{2}{(q-2)}\right] + 40\left[\tfrac{q}{(q-2)}\right]\log_2(\varepsilon^{-1}) - 28, 52\right\} \\ &= \max\left\{\left[\tfrac{40q}{(q-2)}\right]\log_2(\varepsilon^{-1}) + \tfrac{80}{(q-2)} - 28, 52\right\}. \end{aligned} \tag{304}$$

In addition, observe that item (ii) in Proposition 2.6, (291), (V), and (303) demonstrate that

$$\begin{aligned}\mathcal{L}(\Phi) &= \mathcal{L}\big(\mathbf{P}_2(\Psi, \Psi)\big) = \mathcal{L}(\Psi) \leqslant \max\left\{\tfrac{1}{2}\log_2(\delta^{-1}) + 1, 2\right\} \\ &= \max\left\{\left[\tfrac{q}{2(q-2)}\right]\log_2(\varepsilon^{-1}) + \tfrac{1}{(q-2)} + 1, 2\right\}.\end{aligned} \tag{305}$$

Combining this with (294), (297), (302), (300), and (304) establishes items (i)–(vi). The proof of Proposition 3.4 is thus completed. □

3.2 ANN approximations for products

3.2.1 ANN approximations for one-dimensional products

Proposition 3.5 *Let $\varepsilon \in (0, 1]$, $q \in (2, \infty)$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$. Then there exists $\Phi \in \mathbf{N}$ such that*

(i) *it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^2, \mathbb{R})$,*

(ii) *it holds for all $x \in \mathbb{R}$ that $(\mathcal{R}_a(\Phi))(x, 0) = (\mathcal{R}_a(\Phi))(0, x) = 0$,*

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

$$|xy - (\mathcal{R}_a(\Phi))(x, y)| \leqslant \varepsilon \max\{1, |x|^q, |y|^q\}, \tag{306}$$

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

$$|(\mathcal{R}_a(\Phi))(x, y)| \leqslant \tfrac{3}{2}\left(\tfrac{\varepsilon}{3} + x^2 + y^2\right) \leqslant 1 + 2x^2 + 2y^2, \tag{307}$$

(v) *it holds that*

$$\begin{aligned}\mathcal{P}(\Phi) &\leqslant \tfrac{360q}{(q-2)}\left[\log_2(\varepsilon^{-1}) + \log_2(2^{q-1} + 1)\right] + \tfrac{1}{(q-2)} - 252 \\ &\leqslant \tfrac{360q}{(q-2)}\left[\log_2(\varepsilon^{-1}) + q + 1\right] - 252,\end{aligned} \tag{308}$$

and

(vi) *it holds that*

$$\begin{aligned}\mathcal{L}(\Phi) &\leqslant \tfrac{q}{2(q-2)}\left[\log_2(\varepsilon^{-1}) + \log_2(2^{q-1} + 1)\right] + \tfrac{(q-1)}{(q-2)} \\ &\leqslant \tfrac{q}{(q-2)}\left[\log_2(\varepsilon^{-1}) + q\right]\end{aligned} \tag{309}$$

(cf. Definition 2.1 and Definition 2.3).

Proof of Proposition 3.5 Throughout this proof let $\delta \in (0, 1]$ satisfy that $\delta = \varepsilon(2^{q-1} + 1)^{-1}$, let $\mathbb{A}_1 \in (\mathbb{R}^{3\times 2} \times \mathbb{R}^3) \subseteq \mathbf{N}$, $\mathbb{A}_2 \in (\mathbb{R}^{1\times 3} \times \mathbb{R}) \subseteq \mathbf{N}$ satisfy that

$$\mathbb{A}_1 = \left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right) \qquad \text{and} \qquad \mathbb{A}_2 = \left(\begin{pmatrix} \tfrac{1}{2} & -\tfrac{1}{2} & -\tfrac{1}{2} \end{pmatrix}, 0\right), \tag{310}$$

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

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

(II) it holds that $[\mathcal{R}_a(\Psi)](0) = 0$,

(III) it holds for all $x \in \mathbb{R}$ that $0 \leqslant [\mathcal{R}_a(\Psi)](x) \leqslant \delta + |x|^2$,

(IV) it holds for all $x \in \mathbb{R}$ that $|x^2 - [\mathcal{R}_a(\Psi)](x)| \leqslant \delta \max\{1, |x|^q\}$,

(V) it holds that $\mathcal{P}(\Psi) \leqslant \max\{\lceil\frac{40q}{(q-2)}\rceil \log_2(\delta^{-1}) + \frac{80}{(q-2)} - 28, 52\}$, and

(VI) it holds that $\mathcal{L}(\Psi) \leqslant \max\{\lceil\frac{q}{2(q-2)}\rceil \log_2(\delta^{-1}) + \frac{1}{(q-2)} + 1, 2\}$

(cf. Proposition 3.4), and let $\Phi \in \mathbf{N}$ satisfy that

$$\Phi = \mathbb{A}_2 \bullet \big[\mathbf{P}_3(\Psi, \Psi, \Psi)\big] \bullet \mathbb{A}_1 \tag{311}$$

(cf. Definition 2.5, Definition 2.17, and Lemma 2.8). Note that item (v) in Proposition 2.6 and Proposition 2.19 ensure that for all $x, y \in \mathbb{R}$ it holds that $\mathcal{R}_a\big(\big[\mathbf{P}_3(\Psi, \Psi, \Psi)\big] \bullet \mathbb{A}_1\big) \in C(\mathbb{R}^2, \mathbb{R}^3)$ and

$$\begin{aligned} &\big[\mathcal{R}_a\big(\big[\mathbf{P}_3(\Psi, \Psi, \Psi)\big] \bullet \mathbb{A}_1\big)\big](x, y) = \big[\mathcal{R}_a\big(\mathbf{P}_3(\Psi, \Psi, \Psi)\big)\big]\big([\mathcal{R}_a(\mathbb{A}_1)](x, y)\big) \\ &= \big[\mathcal{R}_a\big(\mathbf{P}_3(\Psi, \Psi, \Psi)\big)\big](x + y, x, y) = \begin{pmatrix} [\mathcal{R}_a(\Psi)](x + y) \\ [\mathcal{R}_a(\Psi)](x) \\ [\mathcal{R}_a(\Psi)](y) \end{pmatrix}. \end{aligned} \tag{312}$$

Item (v) in Proposition 2.6 and (311) therefore demonstrate that for all $x, y \in \mathbb{R}$ it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^2, \mathbb{R})$ and

$$\begin{aligned} [\mathcal{R}_a(\Phi)](x, y) &= \big(\mathcal{R}_a\big(\mathbb{A}_2 \bullet \big[\mathbf{P}_3(\Psi, \Psi, \Psi)\big] \bullet \mathbb{A}_1\big)\big)(x, y) \\ &= [\mathcal{R}_a(\mathbb{A}_2)]\big(\mathcal{R}_a\big(\big[\mathbf{P}_3(\Psi, \Psi, \Psi)\big] \bullet \mathbb{A}_1\big)(x, y)\big) \\ &= \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} [\mathcal{R}_a(\Psi)](x + y) \\ [\mathcal{R}_a(\Psi)](x) \\ [\mathcal{R}_a(\Psi)](y) \end{pmatrix} \\ &= \tfrac{1}{2}[\mathcal{R}_a(\Psi)](x + y) - \tfrac{1}{2}[\mathcal{R}_a(\Psi)](x) - \tfrac{1}{2}[\mathcal{R}_a(\Psi)](y). \end{aligned} \tag{313}$$

The fact that for all $\alpha, \beta \in \mathbb{R}$ it holds that $\alpha\beta = \frac{1}{2}|\alpha+\beta|^2 - \frac{1}{2}|\alpha|^2 - \frac{1}{2}|\beta|^2$, the triangle inequality, and (IV) hence ensure that for all $x, y \in \mathbb{R}$ it holds that

$$\begin{aligned} &|[\mathcal{R}_a(\Phi)](x, y) - xy| \\ &= \big|\tfrac{1}{2}\big[[\mathcal{R}_a(\Psi)](x + y) - |x + y|^2\big] - \tfrac{1}{2}\big[[\mathcal{R}_a(\Psi)](x) - |x|^2\big] - \tfrac{1}{2}\big[[\mathcal{R}_a(\Psi)](y) - |y|^2\big]\big| \\ &\leqslant \tfrac{1}{2}\big|[\mathcal{R}_a(\Psi)](x + y) - |x + y|^2\big| + \tfrac{1}{2}\big|[\mathcal{R}_a(\Psi)](x) - |x|^2\big| + \tfrac{1}{2}\big|[\mathcal{R}_a(\Psi)](y) - |y|^2\big| \\ &\leqslant \tfrac{\delta}{2}\big[\max\{1, |x + y|^q\} + \max\{1, |x|^q\} + \max\{1, |y|^q\}\big]. \end{aligned} \tag{314}$$

This, the fact that for all $\alpha, \beta \in \mathbb{R}$, $p \in [1, \infty)$ it holds that $|\alpha+\beta|^p \leqslant 2^{p-1}(|\alpha|^p + |\beta|^p)$, and the fact that $\delta = \varepsilon(2^{q-1} + 1)^{-1}$ establish that for all $x, y \in \mathbb{R}$ it holds that

$$\begin{aligned} &|[\mathcal{R}_a(\Phi)](x, y) - xy| \\ &\leqslant \tfrac{\delta}{2}\big[\max\{1, 2^{q-1}|x|^q + 2^{q-1}|y|^q\} + \max\{1, |x|^q\} + \max\{1, |y|^q\}\big] \\ &\leqslant \tfrac{\delta}{2}\big[\max\{1, 2^{q-1}|x|^q\} + 2^{q-1}|y|^q + \max\{1, |x|^q\} + \max\{1, |y|^q\}\big] \\ &\leqslant \tfrac{\delta}{2}\big[2^q + 2\big]\max\{1, |x|^q, |y|^q\} = \varepsilon\max\{1, |x|^q, |y|^q\}. \end{aligned} \tag{315}$$

Moreover, observe that (III), (313), the triangle inequality, the fact that for all $\alpha, \beta \in \mathbb{R}$ it holds that $|\alpha+\beta|^2 \leqslant 2(|\alpha|^2+|\beta|^2)$, and the fact that $\delta = \varepsilon(2^{q-1}+1)^{-1}$ prove that for all $x, y \in \mathbb{R}$ it holds that

$$\begin{aligned}|[\mathcal{R}_a(\Phi)](x,y)| &\leqslant \tfrac{1}{2}|[\mathcal{R}_a(\Psi)](x+y)| + \tfrac{1}{2}|[\mathcal{R}_a(\Psi)](x)| + \tfrac{1}{2}|[\mathcal{R}_a(\Psi)](y)| \\ &\leqslant \tfrac{1}{2}\big(\delta+|x+y|^2\big) + \tfrac{1}{2}\big(\delta+|x|^2\big) + \tfrac{1}{2}\big(\delta+|y|^2\big) \\ &\leqslant \tfrac{3\delta}{2} + \tfrac{3}{2}\big(|x|^2+|y|^2\big) = \big[\tfrac{3\varepsilon}{2}\big][2^{q-1}+1]^{-1} + \tfrac{3}{2}\big(|x|^2+|y|^2\big) \\ &= \tfrac{3}{2}\Big[\tfrac{\varepsilon}{(2^{q-1}+1)} + |x|^2+|y|^2\Big] \leqslant \tfrac{3}{2}\big[\tfrac{\varepsilon}{3} + |x|^2+|y|^2\big].\end{aligned} \tag{316}$$

Next note that (I) and (313) prove that for all $x, y \in \mathbb{R}$ it holds that

$$\begin{aligned}[\mathcal{R}_a(\Phi)](x,0) &= \tfrac{1}{2}[\mathcal{R}_a(\Psi)](x) - \tfrac{1}{2}[\mathcal{R}_a(\Psi)](x) - \tfrac{1}{2}[\mathcal{R}_a(\Psi)](0) = 0 \\ &= \tfrac{1}{2}[\mathcal{R}_a(\Psi)](y) - \tfrac{1}{2}[\mathcal{R}_a(\Psi)](0) - \tfrac{1}{2}[\mathcal{R}_a(\Psi)](y) = [\mathcal{R}_a(\Phi)](0,y).\end{aligned} \tag{317}$$

Furthermore, observe that the fact that $\delta = \varepsilon(2^{q-1}+1)^{-1}$ shows that

$$\begin{aligned}&\big[\tfrac{q}{2(q-2)}\big]\log_2(\delta^{-1}) + \tfrac{1}{(q-2)} = \big[\tfrac{q}{2(q-2)}\big]\log_2\big(\varepsilon^{-1}(2^{q-1}+1)\big) + \tfrac{1}{(q-2)} \\ &= \tfrac{q}{2(q-2)}\big[\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+1)\big] + \tfrac{1}{(q-2)} \\ &= \big[\tfrac{q}{2(q-2)}\big]\log_2(\varepsilon^{-1}) + \big[\tfrac{q}{2(q-2)}\big]\log_2(2^{q-1}+1) + \tfrac{1}{(q-2)}.\end{aligned} \tag{318}$$

Moreover, observe that Corollary 2.21 implies that $\mathcal{P}\big(\mathbf{P}_3(\Psi,\Psi,\Psi)\big) \leqslant 9\,\mathcal{P}(\Psi)$. Items (i)–(ii) in Corollary 2.9, (V), (311), and (318) hence ensure that

$$\begin{aligned}\mathcal{P}(\Phi) &\leqslant \Big[\max\Big\{1, \tfrac{\mathcal{O}(\mathbb{A}_2)}{\mathcal{O}(\mathbf{P}_3(\Psi,\Psi,\Psi))}\Big\}\Big]\Big[\max\Big\{1, \tfrac{\mathcal{I}(\mathbb{A}_1)+1}{\mathcal{I}(\mathbf{P}_3(\Psi,\Psi,\Psi))+1}\Big\}\Big]\mathcal{P}\big(\mathbf{P}_3(\Psi,\Psi,\Psi)\big) \\ &= \Big[\max\{1, \tfrac{1}{3}\}\Big]\Big[\max\{1, \tfrac{3}{4}\}\Big]\mathcal{P}\big(\mathbf{P}_3(\Psi,\Psi,\Psi)\big) = \mathcal{P}\big(\mathbf{P}_3(\Psi,\Psi,\Psi)\big) \\ &\leqslant 9\,\mathcal{P}(\Psi) \leqslant 9\max\big\{\big[\tfrac{40q}{(q-2)}\big]\log_2(\delta^{-1}) + \tfrac{80}{(q-2)} - 28, 52\big\} \\ &= \max\big\{720\big(\big[\tfrac{q}{2(q-2)}\big]\log_2(\delta^{-1}) + \tfrac{1}{(q-2)}\big) - 252, 468\big\} \\ &= \max\big\{720\big(\big[\tfrac{q}{2(q-2)}\big]\log_2(\varepsilon^{-1}) + \big[\tfrac{q}{2(q-2)}\big]\log_2(2^{q-1}+1) + \tfrac{1}{(q-2)}\big) - 252, 468\big\} \\ &= \max\big\{\tfrac{360q}{(q-2)}\big(\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+1)\big) + \tfrac{720}{(q-2)} - 252, 468\big\}.\end{aligned} \tag{319}$$

Next note that the fact that for all $r \in (-\infty, 4]$ it holds that $r \geqslant 2r-4 = 2(r-2)$ ensures that for all $r \in (2,4]$ it holds that $\frac{r(r-1)}{(r-2)} \geqslant \frac{r}{(r-2)} \geqslant 2$. This and the fact that for all $r \in [3,\infty)$ it holds that $\frac{r(r-1)}{(r-2)} \geqslant r-1 \geqslant 2$ imply that for all $r \in (2,\infty)$ it holds that $\frac{r(r-1)}{(r-2)} \geqslant 2$. Hence, we obtain that for all $r \in (2,\infty)$ it holds that

$$\begin{aligned}\big[\tfrac{360r}{(r-2)}\big]\log_2(2^{r-1}+1) - 252 &\geqslant \big[\tfrac{360r}{(r-2)}\big]\log_2(2^{r-1}) - 252 \\ &= \tfrac{360r(r-1)}{(r-2)} - 252 \geqslant 720 - 252 = 468.\end{aligned} \tag{320}$$

Combining this with (319) shows that

$$\mathcal{P}(\Phi) \leqslant \tfrac{360q}{(q-2)}\big(\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+1)\big) + \tfrac{720}{(q-2)} - 252. \tag{321}$$

The fact that

$$\begin{aligned}\log_2(2^{q-1}+1) &= \log_2(2^{q-1}+1) - \log_2(2^q) + q = \log_2\!\left(\tfrac{2^{q-1}+1}{2^q}\right) + q \\ &= \log_2\!\left(2^{-1}+2^{-q}\right) + q \leqslant \log_2\!\left(2^{-1}+2^{-2}\right) + q \\ &= \log_2\!\left(\tfrac{3}{4}\right) + q = \log_2(3) - 2 + q\end{aligned} \tag{322}$$

hence proves that

$$\begin{aligned}\mathcal{P}(\Phi) &\leqslant \tfrac{360q}{(q-2)}\left(\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+1)\right) + \tfrac{720}{(q-2)} - 252 \\ &\leqslant \tfrac{360q}{(q-2)}\left(\log_2(\varepsilon^{-1}) + q + \log_2(3) - 2\right) + \tfrac{720}{(q-2)} - 252 \\ &= \tfrac{360q}{(q-2)}\left(\log_2(\varepsilon^{-1}) + q + \log_2(3) - 2 + \tfrac{2}{q}\right) - 252 \\ &\leqslant \tfrac{360q}{(q-2)}\left(\log_2(\varepsilon^{-1}) + q + \log_2(3) - 1\right) - 252.\end{aligned} \tag{323}$$

In addition, observe that item (ii) in Proposition 2.6, (311), (VI), the fact that $\delta = \varepsilon(2^{q-1}+1)^{-1}$, and (318) demonstrate that

$$\begin{aligned}\mathcal{L}(\Phi) &= \mathcal{L}\big(\mathbf{P}_3(\Psi,\Psi,\Psi)\big) = \mathcal{L}(\Psi) \\ &\leqslant \max\left\{\left[\tfrac{q}{2(q-2)}\right]\log_2(\delta^{-1}) + \tfrac{1}{(q-2)} + 1, 2\right\} \\ &\leqslant \max\left\{\tfrac{q}{2(q-2)}\left[\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+1)\right] + \tfrac{(q-1)}{(q-2)}, 2\right\}.\end{aligned} \tag{324}$$

Furthermore, note that the fact for all $r \in (2,\infty)$ it holds that $\frac{r(r-1)}{(r-2)} \geqslant 2$ assures that

$$\begin{aligned}&\tfrac{q}{2(q-2)}\left[\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+1)\right] + \tfrac{(q-1)}{(q-2)} \\ &\geqslant \left[\tfrac{q}{2(q-2)}\right]\log_2(2^{q-1}) + 1 = \tfrac{q(q-1)}{2(q-2)} + 1 \geqslant 2.\end{aligned} \tag{325}$$

Combining this with (324) proves that

$$\begin{aligned}\mathcal{L}(\Phi) &\leqslant \tfrac{q}{2(q-2)}\left[\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+1)\right] + \tfrac{(q-1)}{(q-2)} \\ &\leqslant \tfrac{q}{2(q-2)}\left[\log_2(\varepsilon^{-1}) + \log_2(2^{q-1}+2^{q-1})\right] + \tfrac{q}{(q-2)} \\ &= \tfrac{q}{(q-2)}\left[\tfrac{\log_2(\varepsilon^{-1})}{2} + \tfrac{q}{2} + 1\right] \leqslant \tfrac{q}{(q-2)}\left[\log_2(\varepsilon^{-1}) + \tfrac{q}{2} + \tfrac{q}{2}\right] \\ &= \tfrac{q}{(q-2)}\left[\log_2(\varepsilon^{-1}) + q\right].\end{aligned} \tag{326}$$

This, the fact that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^2,\mathbb{R})$, (315), (316), (317), and (323) establish items (i)–(vi). The proof of Proposition 3.5 is thus completed. □

3.2.2 ANN approximations for multi-dimensional products

Definition 3.6 (The Euclidean norm) We denote by $\|\cdot\| : (\cup_{d\in\mathbb{N}}\mathbb{R}^d) \to [0,\infty)$ the function which satisfies for all $d \in \mathbb{N}$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ that

$$\|x\| = \left[\textstyle\sum_{j=1}^d |x_j|^2\right]^{1/2}. \tag{327}$$

Proposition 3.7 *Let $\varepsilon \in (0,1]$, $q \in (2,\infty)$, $d \in \mathbb{N}$, $a \in C(\mathbb{R},\mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x,0\}$. Then there exists $\Phi \in \mathbf{N}$ such that*

(i) *it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{d+1},\mathbb{R}^d)$,*
(ii) *it holds for all $t \in \mathbb{R}$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Phi))(t,0) = (\mathcal{R}_a(\Phi))(0,x) = 0$,*

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

$$\|tx - (\mathcal{R}_a(\Phi))(t,x)\| \leqslant \varepsilon\big(\sqrt{d}\,\big[\max\{1, |t|^q\}\big] + \|x\|^q\big), \tag{328}$$

(iv) *it holds for all* $t \in \mathbb{R}$, $x \in \mathbb{R}^d$ *that*

$$\|(\mathcal{R}_a(\Phi))(t,x)\| \leqslant \sqrt{d}\big(1 + 2t^2\big) + 2\|x\|^2, \tag{329}$$

(v) *it holds that* $\mathcal{P}(\Phi) \leqslant d^2\big[\frac{360q}{(q-2)}\big]\big[\log_2(\varepsilon^{-1}) + q + 1\big] - 252d^2$, *and*

(vi) *it holds that* $\mathcal{L}(\Phi) \leqslant \frac{q}{(q-2)}[\log_2(\varepsilon^{-1}) + q]$

(cf. Definition 2.1, Definition 2.3, and Definition 3.6).

Proof of Proposition 3.7 Throughout this proof let $v, w \in \mathbb{R}^{2\times 1}$, $b \in \mathbb{R}^{2d}$, $A \in \mathbb{R}^{(2d)\times(d+1)}$ satisfy that

$$v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad b = 0, \tag{330}$$

and

$$A = \begin{pmatrix} w & v & 0 & 0 & \cdots & 0 \\ w & 0 & v & 0 & \cdots & 0 \\ w & 0 & 0 & v & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w & 0 & 0 & 0 & \cdots & v \end{pmatrix}, \tag{331}$$

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

(I) it holds that $\mathcal{R}_a(\Psi) \in C(\mathbb{R}^2, \mathbb{R})$,
(II) it holds for all $x \in \mathbb{R}$ that $[\mathcal{R}_a(\Psi)](x,0) = [\mathcal{R}_a(\Psi)](0,x) = 0$,
(III) it holds for all $x, y \in \mathbb{R}$ that $|xy - [\mathcal{R}_a(\Psi)](x,y)| \leqslant \varepsilon \max\{1, |x|^q, |y|^q\}$,
(IV) it holds for all $x, y \in \mathbb{R}$ that $|[\mathcal{R}_a(\Psi)](x,y)| \leqslant 1 + 2x^2 + 2y^2$,
(V) it holds that $\mathcal{P}(\Psi) \leqslant \frac{360q}{(q-2)}\big[\log_2(\varepsilon^{-1}) + q + 1\big] - 252$, and
(VI) it holds that $\mathcal{L}(\Psi) \leqslant \frac{q}{(q-2)}[\log_2(\varepsilon^{-1}) + q]$

(cf. Proposition 3.5), and let $\mathbb{A} \in (\mathbb{R}^{2d\times(d+1)} \times \mathbb{R}^{2d}) \subseteq \mathbf{N}$, $\Phi \in \mathbf{N}$ satisfy that

$$\mathbb{A} = (A, b) \qquad \text{and} \qquad \Phi = \big[\mathbf{P}_d(\Psi, \Psi, \ldots, \Psi)\big] \bullet \mathbb{A} \tag{332}$$

(cf. Definition 2.5 and Definition 2.17). Observe that (330) and (331) ensure that for all $y = (y_1, y_2, \ldots, y_{d+1}) \in \mathbb{R}^{d+1}$ it holds that

$$Ay = \begin{pmatrix} y_1 w + y_2 v \\ y_1 w + y_3 v \\ \vdots \\ y_1 w + y_{d+1} v \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_1 \\ y_3 \\ \vdots \\ y_1 \\ y_{d+1} \end{pmatrix}. \tag{333}$$

Combining this with (332) proves that for all $t \in \mathbb{R}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\mathcal{R}_a(\mathbb{A}) \in C(\mathbb{R}^{d+1}, \mathbb{R}^{2d}) \qquad \text{and} \qquad (\mathcal{R}_a(\mathbb{A}))(t, x) = (t, x_1, t, x_2, \dots, t, x_d). \tag{334}$$

Proposition 2.19, (332), and item (v) in Proposition 2.6 hence demonstrate that for all $t \in \mathbb{R}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$ and

$$\begin{aligned}(\mathcal{R}_a(\Phi))(t, x) &= \big(\big[\mathcal{R}_a\big(\mathbf{P}_d(\Psi, \Psi, \dots, \Psi)\big)\big] \circ \big[\mathcal{R}_a(\mathbb{A})\big]\big)(t, x) \\ &= \big[\mathcal{R}_a\big(\mathbf{P}_d(\Psi, \Psi, \dots, \Psi)\big)\big](t, x_1, t, x_2, \dots, t, x_d) \\ &= \big((\mathcal{R}_a(\Psi))(t, x_1), (\mathcal{R}_a(\Psi))(t, x_2), \dots, (\mathcal{R}_a(\Psi))(t, x_d)\big).\end{aligned} \tag{335}$$

Combining this with (II) proves that for all $t \in \mathbb{R}$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Phi))(t, 0, 0, \dots, 0) &= \big((\mathcal{R}_a(\Psi))(t, 0), (\mathcal{R}_a(\Psi))(t, 0), \dots, (\mathcal{R}_a(\Psi))(t, 0)\big) \\ &= (0, 0, \dots, 0) = 0.\end{aligned} \tag{336}$$

Next note that (II) and (335) imply that for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Phi))(0, x) &= \big((\mathcal{R}_a(\Psi))(0, x_1), (\mathcal{R}_a(\Psi))(0, x_2), \dots, (\mathcal{R}_a(\Psi))(0, x_d)\big) \\ &= (0, 0, \dots, 0) = 0.\end{aligned} \tag{337}$$

In addition, observe that the triangle inequality and the fact that for all $r \in [1, \infty)$, $(x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\Big[\textstyle\sum_{j=1}^d |x_j|^{2r}\Big]^{1/2} \leqslant \Big[\textstyle\sum_{j=1}^d |x_j|^2\Big]^{r/2} \tag{338}$$

prove that for all $b \in \mathbb{R}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, $r \in [1, \infty)$ it holds that

$$\begin{aligned}\Big[\textstyle\sum_{j=1}^d \big(|b| + |x_j|^r\big)^2\Big]^{1/2} &\leqslant \Big[\textstyle\sum_{j=1}^d b^2\Big]^{1/2} + \Big[\textstyle\sum_{j=1}^d |x_j|^{2r}\Big]^{1/2} \\ &\leqslant |b|\sqrt{d} + \Big[\textstyle\sum_{j=1}^d |x_j|^2\Big]^{r/2} = |b|\sqrt{d} + \|x\|^r.\end{aligned} \tag{339}$$

This, (III), and (335) assure that for all $t \in \mathbb{R}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}&\|tx - (\mathcal{R}_a(\Phi))(t, x)\| = \Big[\textstyle\sum_{j=1}^d |tx_j - (\mathcal{R}_a(\Psi))(t, x_j)|^2\Big]^{1/2} \\ &\leqslant \Big[\textstyle\sum_{j=1}^d \big[\varepsilon \max\{1, |t|^q, |x_j|^q\}\big]^2\Big]^{1/2} \leqslant \varepsilon\Big[\textstyle\sum_{j=1}^d \big[\max\{1, |t|^q\} + |x_j|^q\big]^2\Big]^{1/2} \\ &\leqslant \varepsilon\big(\sqrt{d}\,\big[\max\{1, |t|^q\}\big] + \|x\|^q\big).\end{aligned} \tag{340}$$

Furthermore, observe that (IV), (335), and (339) show that for all $t \in \mathbb{R}$, $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}\|(\mathcal{R}_a(\Phi))(t, x)\| &= \Big[\textstyle\sum_{j=1}^d |(\mathcal{R}_a(\Psi))(t, x_j)|^2\Big]^{1/2} \\ &\leqslant \Big[\textstyle\sum_{j=1}^d \big(1 + 2|t|^2 + 2|x_j|^2\big)^2\Big]^{1/2} \\ &= \Big[\textstyle\sum_{j=1}^d \big(1 + 2|t|^2 + |\sqrt{2}x_j|^2\big)^2\Big]^{1/2} \\ &\leqslant \sqrt{d}\big(1 + 2|t|^2\big) + \big\|\sqrt{2}x\big\|^2 = \sqrt{d}\big(1 + 2|t|^2\big) + 2\|x\|^2.\end{aligned} \tag{341}$$

In addition, note that Corollary 2.21 implies that

$$\mathcal{P}\big(\mathbf{P}_d(\Psi, \Psi, \ldots, \Psi)\big) \leqslant d^2\, \mathcal{P}(\Psi). \tag{342}$$

Item (ii) in Corollary 2.9, (V), and (332) hence ensure that

$$\begin{aligned} \mathcal{P}(\Phi) &\leqslant \Big[\max\Big\{1, \tfrac{\mathcal{I}(\mathbb{A})+1}{\mathcal{I}(\mathbf{P}_d(\Psi,\Psi,\ldots,\Psi))+1}\Big\}\Big]\, \mathcal{P}\big(\mathbf{P}_d(\Psi, \Psi, \ldots, \Psi)\big) \\ &= \Big[\max\{1, \tfrac{d+2}{2d+1}\}\Big]\, \mathcal{P}\big(\mathbf{P}_d(\Psi, \Psi, \ldots, \Psi)\big) = \mathcal{P}\big(\mathbf{P}_d(\Psi, \Psi, \ldots, \Psi)\big) \\ &\leqslant d^2\, \mathcal{P}(\Psi) \leqslant d^2\big[\tfrac{360q}{(q-2)}\big]\big[\log_2(\varepsilon^{-1}) + q + 1\big] - 252 d^2. \end{aligned} \tag{343}$$

Next note that item (ii) in Proposition 2.6, (VI), and (332) demonstrate that

$$\mathcal{L}(\Phi) = \mathcal{L}\big(\mathbf{P}_d(\Psi, \Psi, \ldots, \Psi)\big) = \mathcal{L}(\Psi) \leqslant \tfrac{q}{(q-2)}[\log_2(\varepsilon^{-1}) + q]. \tag{344}$$

This, the fact that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$, (336), (337), (340), (341), and (343) establish items (i)–(vi). The proof of Proposition 3.7 is thus completed. □

3.3 Space-time ANN approximations for Euler approximations

3.3.1 Space-time representations for Euler approximations

Lemma 3.8 *Let $N, d \in \mathbb{N}$, $\mu \in C(\mathbb{R}^d, \mathbb{R}^d)$, $T \in (0, \infty)$, $(t_n)_{n\in\{-1,0,1,\ldots,N+1\}} \subseteq \mathbb{R}$ satisfy that $t_{-1} < 0 = t_0 < t_1 < \ldots < t_N = T < t_{N+1}$, let $f_n\colon \mathbb{R} \to \mathbb{R}$, $n \in \{0, 1, \ldots, N\}$, be the functions which satisfy for all $n \in \{0, 1, \ldots, N\}$, $t \in \mathbb{R}$ that*

$$f_n(t) = \Big[\tfrac{(t-t_{n-1})}{(t_n-t_{n-1})}\Big]\, \mathbb{1}_{(t_{n-1},t_n]}(t) + \Big[\tfrac{(t_{n+1}-t)}{(t_{n+1}-t_n)}\Big]\, \mathbb{1}_{(t_n,t_{n+1})}(t), \tag{345}$$

and let $Y = (Y_t^{x,y})_{(t,x,y)\in[0,T]\times\mathbb{R}^d\times(\mathbb{R}^d)^N}\colon [0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$ be the function which satisfies for all $n \in \{0, 1, \ldots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that $Y_0^{x,y} = x$ and

$$Y_t^{x,y} = Y_{t_n}^{x,y} + \tfrac{(t-t_n)}{(t_{n+1}-t_n)}\, \big[(t_{n+1} - t_n)\, \mu\big(Y_{t_n}^{x,y}\big) + y_{n+1}\big] \tag{346}$$

(cf. Definition 2.1 and Definition 2.3). Then

(i) *it holds that*

$$\big([0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \ni (t, x, y) \mapsto Y_t^{x,y} \in \mathbb{R}^d\big) \in C([0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N, \mathbb{R}^d) \tag{347}$$

and

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

$$Y_t^{x,y} = \sum_{n=0}^{N} f_n(t)\, Y_{t_n}^{x,y}. \tag{348}$$

Proof of Lemma 3.8 Observe that (346) ensures that for all $n \in \{0, 1, \ldots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} & Y^{x,y}_{t_n}\big(\tfrac{t_{n+1}-t}{t_{n+1}-t_n}\big) + Y^{x,y}_{t_{n+1}}\big(\tfrac{t-t_n}{t_{n+1}-t_n}\big) \\ &= Y^{x,y}_{t_n}\big(1 - \tfrac{t-t_n}{t_{n+1}-t_n}\big) + Y^{x,y}_{t_{n+1}}\big(\tfrac{t-t_n}{t_{n+1}-t_n}\big) \\ &= Y^{x,y}_{t_n}\big(1 - \tfrac{t-t_n}{t_{n+1}-t_n}\big) \\ &\quad + \Big(Y^{x,y}_{t_n} + \tfrac{t_{n+1}-t_n}{t_{n+1}-t_n}\big[(t_{n+1}-t_n)\,\mu\big(Y^{x,y}_{t_n}\big) + y_{n+1}\big]\Big)\big(\tfrac{t-t_n}{t_{n+1}-t_n}\big) \\ &= Y^{x,y}_{t_n} + \big[(t_{n+1}-t_n)\,\mu\big(Y^{x,y}_{t_n}\big) + y_{n+1}\big]\big(\tfrac{t-t_n}{t_{n+1}-t_n}\big) = Y^{x,y}_t. \end{aligned} \tag{349}$$

Hence, we obtain that for all $t \in [0, T]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} Y^{x,y}_t &= Y^{x,y}_{t_0}\,\mathbb{1}_{\{t_0\}}(t) + \sum_{n=0}^{N-1}\big(Y^{x,y}_t\,\mathbb{1}_{(t_n,t_{n+1}]}(t)\big) \\ &= Y^{x,y}_{t_0}\,\mathbb{1}_{\{t_0\}}(t) + \sum_{n=0}^{N-1}\Big(\Big[Y^{x,y}_{t_n}\big(\tfrac{t_{n+1}-t}{t_{n+1}-t_n}\big) + Y^{x,y}_{t_{n+1}}\big(\tfrac{t-t_n}{t_{n+1}-t_n}\big)\Big]\mathbb{1}_{(t_n,t_{n+1}]}(t)\Big) \\ &= Y^{x,y}_{t_0}\,\mathbb{1}_{\{t_0\}}(t) + \left[\sum_{n=0}^{N-1} Y^{x,y}_{t_n}\big(\tfrac{t-t_{n+1}}{t_n-t_{n+1}}\big)\,\mathbb{1}_{(t_n,t_{n+1}]}(t)\right] \\ &\quad + \left[\sum_{n=1}^{N} Y^{x,y}_{t_n}\big(\tfrac{t-t_{n-1}}{t_n-t_{n-1}}\big)\,\mathbb{1}_{(t_{n-1},t_n]}(t)\right]. \end{aligned} \tag{350}$$

Combining this with (345) implies that for all $t \in [0, T]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} Y^{x,y}_t &= Y^{x,y}_{t_0}\,\mathbb{1}_{\{t_0\}}(t) + Y^{x,y}_{t_0}\big(\tfrac{t-t_1}{t_0-t_1}\big)\,\mathbb{1}_{(t_0,t_1]}(t) + Y^{x,y}_{t_N}\big(\tfrac{t-t_{N-1}}{t_N-t_{N-1}}\big)\,\mathbb{1}_{(t_{N-1},t_N]}(t) \\ &\quad + \sum_{n=1}^{N-1} Y^{x,y}_{t_n}\Big[\big(\tfrac{t-t_{n+1}}{t_n-t_{n+1}}\big)\,\mathbb{1}_{(t_n,t_{n+1}]}(t) + \big(\tfrac{t-t_{n-1}}{t_n-t_{n-1}}\big)\,\mathbb{1}_{(t_{n-1},t_n]}(t)\Big] \\ &= Y^{x,y}_{t_0}\big(\tfrac{t_1-t}{t_1-t_0}\big)\,\mathbb{1}_{[t_0,t_1]}(t) + Y^{x,y}_{t_N}\,f_N(t) + \sum_{n=1}^{N-1} f_n(t)\,Y^{x,y}_{t_n} \\ &= \sum_{n=0}^{N} f_n(t)\,Y^{x,y}_{t_n}. \end{aligned} \tag{351}$$

Next we claim that for all $n \in \{0, 1, \ldots, N\}$ it holds that

$$\big(\mathbb{R}^d \times (\mathbb{R}^d)^N \ni (x, y) \mapsto Y^{x,y}_{t_n} \in \mathbb{R}^d\big) \in C(\mathbb{R}^d \times (\mathbb{R}^d)^N, \mathbb{R}^d). \tag{352}$$

We now prove (352) by induction on $n \in \{0, 1, \ldots, N\}$. Note that the fact that for all $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that $Y^{x,y}_{t_0} = Y^{x,y}_0 = x$ proves (352) in the base case $n = 0$. For the induction step assume there exists $n \in \{0, 1, \ldots, N-1\}$ which satisfies that

$$\big(\mathbb{R}^d \times (\mathbb{R}^d)^N \ni (x, y) \mapsto Y^{x,y}_{t_n} \in \mathbb{R}^d\big) \in C(\mathbb{R}^d \times (\mathbb{R}^d)^N, \mathbb{R}^d). \tag{353}$$

Observe that (346) ensures that for all $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$Y^{x,y}_{t_{n+1}} = Y^{x,y}_{t_n} + (t_{n+1} - t_n)\,\mu\big(Y^{x,y}_{t_n}\big) + y_{n+1}. \tag{354}$$

Combining this with (353) and the hypothesis that $\mu \in C(\mathbb{R}^d, \mathbb{R}^d)$ demonstrates that

$$\big(\mathbb{R}^d \times (\mathbb{R}^d)^N \ni (x, y) \mapsto Y^{x,y}_{t_{n+1}} \in \mathbb{R}^d\big) \in C(\mathbb{R}^d \times (\mathbb{R}^d)^N, \mathbb{R}^d). \tag{355}$$

Induction thus proves (352). Next observe that (351), (352), and the fact that for all $n \in \{0, 1, \dots, N\}$ it holds that $f_n \in C(\mathbb{R}, \mathbb{R})$ show that

$$\begin{aligned}\big([0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \ni (t, x, y) \mapsto Y^{x,y}_t \in \mathbb{R}^d\big)\\ \in C([0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N, \mathbb{R}^d).\end{aligned} \tag{356}$$

Combining this with (351) establishes items (i)–(ii). The proof of Lemma 3.8 is thus completed. □

3.3.2 ANN representations for hat functions

Lemma 3.9 *Let $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$, let $\alpha, \beta, \gamma, h \in \mathbb{R}$ satisfy that $\alpha < \beta < \gamma$, let $W_1 \in \mathbb{R}^{4\times 1}$, $B_1 \in \mathbb{R}^4$, $W_2 \in \mathbb{R}^{1\times 4}$, $B_2 \in \mathbb{R}$ satisfy that*

$$W_1 = \begin{pmatrix} \frac{1}{(\beta-\alpha)} \\ \frac{1}{(\beta-\alpha)} \\ \frac{1}{(\gamma-\beta)} \\ \frac{1}{(\gamma-\beta)} \end{pmatrix}, \qquad B_1 = \begin{pmatrix} -\frac{\alpha}{(\beta-\alpha)} \\ -\frac{\beta}{(\beta-\alpha)} \\ -\frac{\beta}{(\gamma-\beta)} \\ -\frac{\gamma}{(\gamma-\beta)} \end{pmatrix}, \tag{357}$$

$$W_2 = \begin{pmatrix} h & -h & -h & h \end{pmatrix}, \qquad B_2 = 0, \tag{358}$$

and let $\Phi \in (\mathbb{R}^{4\times 1} \times \mathbb{R}^4) \times (\mathbb{R}^{1\times 4} \times \mathbb{R}) \subseteq \mathbf{N}$ satisfy that $\Phi = ((W_1, B_1), (W_2, B_2))$ (cf. Definition 2.1). Then

(i) *it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}, \mathbb{R})$ and*
(ii) *it holds for all $t \in \mathbb{R}$ that*

$$\begin{aligned}(\mathcal{R}_a(\Phi))(t) &= \left[\tfrac{(t-\alpha)h}{(\beta-\alpha)}\right] \mathbb{1}_{(\alpha,\beta]}(t) + \left[\tfrac{(\gamma-t)h}{(\gamma-\beta)}\right] \mathbb{1}_{(\beta,\gamma)}(t) \\ &= \begin{cases} 0 & : t \in (-\infty, \alpha] \cup [\gamma, \infty) \\ \frac{(t-\alpha)h}{(\beta-\alpha)} & : t \in (\alpha, \beta] \\ \frac{(\gamma-t)h}{(\gamma-\beta)} & : t \in (\beta, \gamma) \end{cases}\end{aligned} \tag{359}$$

(cf. Definition 2.3).

Proof of Lemma 3.9 Observe that for all $t \in \mathbb{R}$ it holds that $\mathcal{R}_a(\Phi) \in C(\mathbb{R}, \mathbb{R})$ and

$$\begin{aligned}
&(\mathcal{R}_a(\Phi))(t) = W_2\big(\mathfrak{M}_{a,4}(W_1 t + B_1)\big) + B_2 \\
&= h \max\Big\{\tfrac{(t-\alpha)}{(\beta-\alpha)}, 0\Big\} - h \max\Big\{\tfrac{(t-\beta)}{(\beta-\alpha)}, 0\Big\} - h \max\Big\{\tfrac{(t-\beta)}{(\gamma-\beta)}, 0\Big\} + h \max\Big\{\tfrac{(t-\gamma)}{(\gamma-\beta)}, 0\Big\} \\
&= h[0 - 0 - 0 + 0]\mathbb{1}_{(-\infty,\alpha]}(t) + h\left[\tfrac{(t-\alpha)}{(\beta-\alpha)} - 0 - 0 + 0\right]\mathbb{1}_{(\alpha,\beta]}(t) \\
&\quad + h\left[\tfrac{(t-\alpha)}{(\beta-\alpha)} - \tfrac{(t-\beta)}{(\beta-\alpha)} - \tfrac{(t-\beta)}{(\gamma-\beta)} + 0\right]\mathbb{1}_{(\beta,\gamma)}(t) \\
&\quad + h\left[\tfrac{(t-\alpha)}{(\beta-\alpha)} - \tfrac{(t-\beta)}{(\beta-\alpha)} - \tfrac{(t-\beta)}{(\gamma-\beta)} + \tfrac{(t-\gamma)}{(\gamma-\beta)}\right]\mathbb{1}_{[\gamma,\infty)}(t) \\
&= h\left[\tfrac{(t-\alpha)}{(\beta-\alpha)}\right]\mathbb{1}_{(\alpha,\beta]}(t) + h\left[1 - \tfrac{(t-\beta)}{(\gamma-\beta)}\right]\mathbb{1}_{(\beta,\gamma)}(t) \\
&= \left[\tfrac{(t-\alpha)h}{(\beta-\alpha)}\right]\mathbb{1}_{(\alpha,\beta]}(t) + \left[\tfrac{(\gamma-t)h}{(\gamma-\beta)}\right]\mathbb{1}_{(\beta,\gamma)}(t)
\end{aligned} \tag{360}$$

(cf. Definition 2.2). The proof of Lemma 3.9 is thus completed. □

3.3.3 A posteriori error estimates for space-time ANN approximations

Proposition 3.10 *Let $N, d \in \mathbb{N}$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$, let $T \in (0, \infty)$, $(t_n)_{n \in \{0,1,\dots,N\}} \subseteq \mathbb{R}$ satisfy for all $n \in \{0, 1, \dots, N\}$ that $t_n = \frac{nT}{N}$, let $\mathfrak{D} \in [1, \infty)$, $\varepsilon \in (0, 1]$, $q \in (2, \infty)$ satisfy that*

$$\mathfrak{D} = \left[\tfrac{720q}{(q-2)}\right]\left[\log_2(\varepsilon^{-1}) + q + 1\right] - 504, \tag{361}$$

let $\Phi \in \mathbf{N}$ satisfy that $\mathcal{I}(\Phi) = \mathcal{O}(\Phi) = d$, and let $Y = (Y_t^{x,y})_{(t,x,y) \in [0,T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N} : [0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$ be the function which satisfies for all $n \in \{0, 1, \dots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \dots, y_N) \in (\mathbb{R}^d)^N$ that $Y_0^{x,y} = x$ and

$$Y_t^{x,y} = Y_{t_n}^{x,y} + \left(\tfrac{tN}{T} - n\right)\left[\tfrac{T}{N}(\mathcal{R}_a(\Phi))\big(Y_{t_n}^{x,y}\big) + y_{n+1}\right] \tag{362}$$

(cf. Definition 2.1 and Definition 2.3). Then there exist $\Psi_y \in \mathbf{N}$, $y \in (\mathbb{R}^d)^N$, such that

(i) *it holds for all $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Psi_y) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,*

(ii) *it holds for all $n \in \{0, 1, \dots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that*

$$\|Y_t^{x,y} - (\mathcal{R}_a(\Psi_y))(t, x)\| \leqslant \varepsilon\big(2\sqrt{d} + \|Y_{t_n}^{x,y}\|^q + \|Y_{t_{n+1}}^{x,y}\|^q\big), \tag{363}$$

(iii) *it holds for all $n \in \{0, 1, \dots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that*

$$\|(\mathcal{R}_a(\Psi_y))(t, x)\| \leqslant 6\sqrt{d} + 2\big(\|Y_{t_n}^{x,y}\|^2 + \|Y_{t_{n+1}}^{x,y}\|^2\big), \tag{364}$$

(iv) *it holds for all $y \in (\mathbb{R}^d)^N$ that*

$$\begin{aligned}
\mathcal{P}(\Psi_y) \leqslant \tfrac{1}{2}\Big[&6d^2N^2\mathcal{H}(\Phi) \\
&+ 3N\Big[d^2\mathfrak{D} + \big(23 + 6N\mathcal{H}(\Phi) + 7d^2 + N\big[4d^2 + \mathcal{P}(\Phi)\big]^2\big)^2\Big]\Big]^2,
\end{aligned} \tag{365}$$

(v) *it holds for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that*

$$\big[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Psi_y))(t, x) \in \mathbb{R}^d\big] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big), \tag{366}$$

and

(vi) *it holds for all* $n \in \{0, 1, \ldots, N\}$, $t \in [0, t_n]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ *with* $\forall\, k \in \mathbb{N} \cap [0, n]\colon y_k = z_k$ *that*

$$(\mathcal{R}_a(\Psi_y))(t, x) = (\mathcal{R}_a(\Psi_z))(t, x) \tag{367}$$

(cf. Definition 3.6).

Proof of Proposition 3.10 Throughout this proof let $t_n \in \mathbb{R}$, $n \in \{-1, N+1\}$, satisfy for all $n \in \{-1, N+1\}$ that $t_n = \frac{nT}{N}$, let $(\mathbb{I}_{\mathfrak{d}})_{\mathfrak{d}\in\mathbb{N}} \subseteq \mathbf{N}$ satisfy for all $\mathfrak{d} \in \mathbb{N}$, $x \in \mathbb{R}^{\mathfrak{d}}$ that $\mathcal{R}_a(\mathbb{I}_{\mathfrak{d}}) \in C(\mathbb{R}^{\mathfrak{d}}, \mathbb{R}^{\mathfrak{d}})$, $\mathcal{D}(\mathbb{I}_{\mathfrak{d}}) = (\mathfrak{d}, 2\mathfrak{d}, \mathfrak{d})$, and

$$(\mathcal{R}_a(\mathbb{I}_{\mathfrak{d}}))(x) = x \tag{368}$$

(cf., e.g., [27, Lemma 5.4]), let $(\Pi_n)_{n\in\{0,1,\ldots,N\}} \subseteq \mathbf{N}$ satisfy for all $n \in \{0, 1, \ldots, N\}$, $t \in \mathbb{R}$ that $\mathcal{I}(\Pi_n) = \mathcal{O}(\Pi_n) = 1$, $\mathcal{H}(\Pi_n) = 1$, $\mathcal{P}(\Pi_n) = 13$, and

$$(\mathcal{R}_a(\Pi_n))(t) = \left[\tfrac{(t-t_{n-1})}{(t_n-t_{n-1})}\right] \mathbb{1}_{(t_{n-1},t_n]}(t) + \left[\tfrac{(t_{n+1}-t)}{(t_{n+1}-t_n)}\right] \mathbb{1}_{(t_n,t_{n+1})}(t) \tag{369}$$

(cf. Lemma 3.9), let $(\Xi_{n,y})_{(n,y)\in\{0,1,\ldots,N\}\times(\mathbb{R}^d)^N} \subseteq \mathbf{N}$ satisfy that

(I) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Xi_{n,y}) \in C(\mathbb{R}^d, \mathbb{R}^d)$,

(II) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Xi_{n,y}))(x) = Y^{x,y}_{t_n}$,

(III) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ that $\mathcal{H}(\Xi_{n,y}) = 1 + n\mathcal{H}(\Phi)$,

(IV) it holds for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ that

$$\mathcal{P}(\Xi_{n,y}) \leqslant \mathcal{P}(\mathbb{I}_d) + n\big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}_d) + \mathcal{P}(\Phi)\big]^2, \tag{370}$$

(V) it holds for all $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$ that

$$\big[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Xi_{n,y}))(x) \in \mathbb{R}^d\big] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big), \tag{371}$$

and

(VI) it holds for all $n \in \{0, 1, \ldots, N\}$, $m \in \mathbb{N}_0 \cap [0, n]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ with $\forall\, k \in \mathbb{N} \cap [0, n]\colon y_k = z_k$ that

$$(\mathcal{R}_a(\Xi_{m,y}))(x) = (\mathcal{R}_a(\Xi_{m,z}))(x) \tag{372}$$

(cf. Proposition 2.32), let $\Gamma \in \mathbf{N}$ satisfy that

(a) it holds that $\mathcal{R}_a(\Gamma) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,

(b) it holds for all $t \in \mathbb{R}$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_a(\Gamma))(t, 0) = (\mathcal{R}_a(\Gamma))(0, x) = 0$,

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

$$\|tx - (\mathcal{R}_a(\Gamma))(t, x)\| \leqslant \varepsilon\big(\sqrt{d}\,\big[\max\{1, |t|^q\}\big] + \|x\|^q\big), \tag{373}$$

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

$$\|(\mathcal{R}_a(\Gamma))(t, x)\| \leqslant \sqrt{d}\big(1 + 2t^2\big) + 2\|x\|^2, \tag{374}$$

(e) it holds that $\mathcal{P}(\Gamma) \leqslant d^2\big[\frac{360q}{(q-2)}\big]\big[\log_2(\varepsilon^{-1}) + q + 1\big] - 252d^2$, and

(f) it holds that $\mathcal{L}(\Gamma) \leqslant \frac{q}{(q-2)}[\log_2(\varepsilon^{-1}) + q]$

(cf. Proposition 3.7), let $(\Psi_{n,y})_{(n,y)\in\{0,1,\ldots,N\}\times(\mathbb{R}^d)^N} \subseteq \mathbf{N}$ satisfy for all $n \in \{0,1,\ldots,N\}$, $y \in (\mathbb{R}^d)^N$ that $\mathcal{I}(\Psi_{n,y}) = d+1$, $\mathcal{O}(\Psi_{n,y}) = d$, and

$$\Psi_{n,y} = \Gamma \odot_{\mathbb{I}_{d+1}} \left[\mathrm{P}_{2,(\mathbb{I}_1,\mathbb{I}_d)}(\Pi_n, \Xi_{n,y})\right] \tag{375}$$

(cf. Definition 2.15, Definition 2.22, Proposition 2.16, and Corollary 2.23), let $L_y \in \mathbb{N}$, $y \in (\mathbb{R}^d)^N$, satisfy for all $y \in (\mathbb{R}^d)^N$ that $L_y = \max_{n\in\{0,1,\ldots,N\}} \mathcal{L}(\Psi_{n,y})$, and let $(\Phi_y)_{y\in(\mathbb{R}^d)^N} \subseteq \mathbf{N}$ satisfy that

(A) it holds for all $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Phi_y) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,
(B) it holds for all $y \in (\mathbb{R}^d)^N$, $z \in \mathbb{R}^{d+1}$ that

$$(\mathcal{R}_a(\Phi_y))(z) = \sum_{n=0}^{N} (\mathcal{R}_a(\Psi_{n,y}))(z), \tag{376}$$

and
(C) it holds for all $y \in (\mathbb{R}^d)^N$ that

$$\begin{aligned}\mathcal{P}(\Phi_y) \leqslant \tfrac{1}{2}\Bigg[&\Big[\textstyle\sum_{n=0}^{N} 2\,\mathcal{P}(\Psi_{n,y})\mathbb{1}_{(\mathcal{L}(\Psi_{n,y}),\infty)}(L_y)\Big] \\ &+ \Big[\textstyle\sum_{n=0}^{N} \big((L_y - \mathcal{L}(\Psi_{n,y}) - 1)\,2d(2d+1) + d(2d+1)\big)\mathbb{1}_{(\mathcal{L}(\Psi_{n,y}),\infty)}(L_y)\Big] \\ &+ \Big[\textstyle\sum_{n=0}^{N} \mathcal{P}(\Psi_{n,y})\mathbb{1}_{\{\mathcal{L}(\Psi_{n,y})\}}(L_y)\Big]\Bigg]^2\end{aligned} \tag{377}$$

(cf. Proposition 2.26). Note that (III) and the fact that for all $n \in \{0,1,\ldots,N\}$ it holds that $\mathcal{H}(\Pi_n) = 1$ ensure that for all $n \in \{0,1,\ldots,N\}$, $y \in (\mathbb{R}^d)^N$ it holds that $\mathcal{L}(\Xi_{n,y}) = 2 + n\mathcal{H}(\Phi) \geqslant 2$, $\mathcal{L}(\Pi_n) = 2$, and

$$\max\{\mathcal{L}(\Pi_n), \mathcal{L}(\Xi_{n,y})\} = \max\{2, 2+n\mathcal{H}(\Phi)\} = 2 + n\mathcal{H}(\phi) = \mathcal{L}(\Xi_{n,y}). \tag{378}$$

Corollary 2.24 (with $a = a$, $n = 2$, $L = \max\{\mathcal{L}(\Pi_n), \mathcal{L}(\Xi_{n,y})\}$, $\mathfrak{i}_1 = 2$, $\mathfrak{i}_2 = 2d$, $\Psi = (\mathbb{I}_1, \mathbb{I}_d)$, $\Phi = (\Pi_n, \Xi_{n,y})$ for $n \in \{0,1,\ldots,N\}$, $y \in (\mathbb{R}^d)^N$ in the notation of Corollary 2.24), (IV), and the fact that for all $n \in \{0,1,\ldots,N\}$ it holds that $\mathcal{P}(\Pi_n) = 13$ hence prove that for all $n \in \{0,1,\ldots,N\}$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}&\mathcal{P}\big(\mathrm{P}_{2,(\mathbb{I}_1,\mathbb{I}_d)}(\Pi_n, \Xi_{n,y})\big) \leqslant \tfrac{1}{2}\big(2\mathcal{P}(\Pi_n) + 6(\mathcal{L}(\Xi_{n,y}) - 3) + 3 + \mathcal{P}(\Xi_{n,y})\big)^2 \\ &= \tfrac{1}{2}\big(11 + 6\mathcal{L}(\Xi_{n,y}) + \mathcal{P}(\Xi_{n,y})\big)^2 \\ &\leqslant \tfrac{1}{2}\big(11 + 6\big(2 + n\mathcal{H}(\Phi)\big) + \mathcal{P}(\mathbb{I}_d) + n\big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}_d) + \mathcal{P}(\Phi)\big]^2\big)^2 \\ &= \tfrac{1}{2}\big(23 + 6n\mathcal{H}(\Phi) + \mathcal{P}(\mathbb{I}_d) + n\big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}_d) + \mathcal{P}(\Phi)\big]^2\big)^2.\end{aligned} \tag{379}$$

Moreover, observe that (361) and (e) imply that $2\mathcal{P}(\Gamma) \leqslant d^2\mathfrak{D}$. Combining this with Proposition 2.16, (379), and the fact that $\mathcal{P}(\mathbb{I}_d) = 4d^2 + 3d \leqslant 4(d^2+d)$ ensures that for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}\mathcal{P}(\Psi_{n,y}) &= \mathcal{P}\big(\Gamma \odot_{\mathbb{I}_{d+1}} [\mathrm{P}_{2,(\mathbb{I}_1,\mathbb{I}_d)}(\Pi_n, \Xi_{n,y})]\big) \\ &\leqslant \max\Big\{1, \tfrac{2(d+1)}{(d+1)}\Big\} \big(\mathcal{P}(\Gamma) + \mathcal{P}(\mathrm{P}_{2,(\mathbb{I}_1,\mathbb{I}_d)}(\Pi_n, \Xi_{n,y}))\big) \\ &\leqslant d^2\mathfrak{D} + \big(23 + 6n\mathcal{H}(\Phi) + \mathcal{P}(\mathbb{I}_d) + n\big[\tfrac{1}{2}\mathcal{P}(\mathbb{I}_d) + \mathcal{P}(\Phi)\big]^2\big)^2 \\ &\leqslant d^2\mathfrak{D} + \big(23 + 6n\mathcal{H}(\Phi) + 4d^2 + 3d + n\big[2(d^2 + d) + \mathcal{P}(\Phi)\big]^2\big)^2.\end{aligned} \tag{380}$$

Next note that (III), (378), (84), (133), (116), item (ii) in Proposition 2.16, and item (i) in Lemma 2.13 demonstrate that for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}\mathcal{L}(\Psi_{n,y}) &= \mathcal{L}(\Gamma) + \mathcal{L}\big(\mathrm{P}_{2,(\mathbb{I}_1,\mathbb{I}_d)}(\Pi_n, \Xi_{n,y})\big) \\ &= \mathcal{L}(\Gamma) + \mathcal{L}\big(\mathbf{P}_2(\mathcal{E}_{\max\{\mathcal{L}(\Pi_n),\mathcal{L}(\Xi_{n,y})\},\mathbb{I}_1}(\Pi_n), \mathcal{E}_{\max\{\mathcal{L}(\Pi_n),\mathcal{L}(\Xi_{n,y})\},\mathbb{I}_d}(\Xi_{n,y}))\big) \\ &= \mathcal{L}(\Gamma) + \mathcal{L}\big(\mathbf{P}_2(\mathcal{E}_{\mathcal{L}(\Xi_{n,y}),\mathbb{I}_1}(\Pi_n), \mathcal{E}_{\mathcal{L}(\Xi_{n,y}),\mathbb{I}_d}(\Xi_{n,y}))\big) \\ &= \mathcal{L}(\Gamma) + \mathcal{L}\big(\mathcal{E}_{\mathcal{L}(\Xi_{n,y}),\mathbb{I}_d}(\Xi_{n,y})\big) \\ &= \mathcal{L}(\Gamma) + \mathcal{L}\big(((\mathbb{I}_d)^{\bullet 0}) \bullet \Xi_{n,y}\big) \\ &= \mathcal{L}(\Gamma) + \mathcal{L}\big((\mathbb{I}_d)^{\bullet 0}\big) + \mathcal{L}\big(\Xi_{n,y}\big) - 1 \\ &= \mathcal{L}(\Gamma) + \mathcal{L}(\Xi_{n,y}) = \mathcal{L}(\Gamma) + \mathcal{H}(\Xi_{n,y}) + 1 \\ &= \mathcal{L}(\Gamma) + 2 + n\mathcal{H}(\Phi).\end{aligned} \tag{381}$$

Therefore, we obtain that for all $n \in \{0, 1, \ldots, N\}$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}\mathcal{L}(\Psi_{N,y}) - \mathcal{L}(\Psi_{n,y}) - 1 &= (\mathcal{L}(\Gamma) + 2 + N\mathcal{H}(\Phi)) - (\mathcal{L}(\Gamma) + 2 + n\mathcal{H}(\Phi)) - 1 \\ &= (N - n)\mathcal{H}(\Phi) - 1.\end{aligned} \tag{382}$$

In addition, note that (381) proves that for all $y \in (\mathbb{R}^d)^N$ it holds that $L_y = \mathcal{L}(\Psi_{N,y}) = \mathcal{L}(\Psi_{N,0}) = L_0$. The fact that $\sum_{n=0}^{N-1}(N-n) = \sum_{m=1}^{N} m = \frac{1}{2}N(N+1)$, (377), and (382) hence assure that for all $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}&\mathcal{P}(\Phi_y) \\ &\leqslant \tfrac{1}{2}\Bigg[\Big[\textstyle\sum_{n=0}^{N-1} \big(2\mathcal{P}(\Psi_{n,y}) \\ &+ \max\big\{\big(\mathcal{L}(\Psi_{N,y}) - \mathcal{L}(\Psi_{n,y}) - 1\big)2d(2d+1) + d(2d+1), 0\big\}\big)\Big] + \mathcal{P}(\Psi_{N,y})\Bigg]^2 \\ &= \tfrac{1}{2}\Bigg[\Big[\textstyle\sum_{n=0}^{N-1} \big(2\mathcal{P}(\Psi_{n,y}) + \max\big\{(N-n)\mathcal{H}(\Phi)2d(2d+1) - d(2d+1), 0\big\}\big)\Big] \\ &+ \mathcal{P}(\Psi_{N,y})\Bigg]^2 \\ &\leqslant \tfrac{1}{2}\Bigg[(2N+1)\mathcal{P}(\Psi_{N,y}) \\ &+ \max\Big\{\mathcal{H}(\Phi)2d(2d+1)\Big[\textstyle\sum_{n=0}^{N-1}(N-n)\Big] - Nd(2d+1), 0\Big\}\Bigg]^2 \\ &= \tfrac{1}{2}\Big[(2N+1)\mathcal{P}(\Psi_{N,y}) + \max\big\{\mathcal{H}(\Phi)d(2d+1)N(N+1) - Nd(2d+1), 0\big\}\Big]^2.\end{aligned} \tag{383}$$

This and (380) imply that for all $y \in (\mathbb{R}^d)^N$ it holds that

$$\mathcal{P}(\Phi_y) \leqslant \tfrac{1}{2}\Big[(2N+1)\Big[d^2\mathfrak{D}+\big(23+6N\mathcal{H}(\Phi)+4d^2+3d+N\big[2(d^2+d)+\mathcal{P}(\Phi)\big]^2\big)^2\Big] + \max\big\{\mathcal{H}(\Phi)d(2d+1)N(N+1)-Nd(2d+1),0\big\}\Big]^2. \quad (384)$$

Therefore, we obtain that for all $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}
&\mathcal{P}(\Phi_y)\\
&\leqslant \tfrac{1}{2}\Big[(2N+1)\Big[d^2\mathfrak{D}+\big(23+6N\mathcal{H}(\Phi)+7d^2+N\big[4d^2+\mathcal{P}(\Phi)\big]^2\big)^2\Big]\\
&\quad + \max\big\{\mathcal{H}(\Phi)d(2d+1)N(N+1)-Nd(2d+1),0\big\}\Big]^2\\
&\leqslant \tfrac{1}{2}\Big[3N\Big[d^2\mathfrak{D}+\big(23+6N\mathcal{H}(\Phi)+7d^2+N\big[4d^2+\mathcal{P}(\Phi)\big]^2\big)^2\Big]+6d^2N^2\mathcal{H}(\Phi)\Big]^2.
\end{aligned} \quad (385)$$

In addition, note that (369), (II), and Lemma 3.8 (with $N = N$, $d = d$, $\mu = \mathcal{R}_a(\Phi)$, $T = T$, $(\{-1, 0, 1, \dots, N+1\} \ni n \mapsto t_n \in \mathbb{R}) = (\{-1, 0, 1, \dots, N+1\} \ni n \mapsto t_n \in \mathbb{R})$, $(\{0, 1, \dots, N\} \ni n \mapsto f_n \in C(\mathbb{R}, \mathbb{R})) = (\{0, 1, \dots, N\} \ni n \mapsto \mathcal{R}_a(\Pi_n) \in C(\mathbb{R}, \mathbb{R}))$, $Y = Y$ in the notation of Lemma 3.8) ensure that for all $t \in [0, T]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$Y_t^{x,y} = \sum_{n=0}^{N} [(\mathcal{R}_a(\Pi_n))(t)]\, Y_{t_n}^{x,y} = \sum_{n=0}^{N} [(\mathcal{R}_a(\Pi_n))(t)]\, [(\mathcal{R}_a(\Xi_{n,y}))(x)]. \quad (386)$$

Moreover, observe that (375), (376), item (iv) in Proposition 2.16 (with $\Psi = \mathbb{I}_{d+1}$, $\Phi_1 = \Gamma$, $\Phi_2 = \mathrm{P}_{2,(\mathbb{I}_1,\mathbb{I}_d)}(\Pi_n, \Xi_{n,y})$, $\mathfrak{i} = 2(d+1)$ for $n \in \{0, 1, \dots, N\}$, $y \in (\mathbb{R}^d)^N$ in the notation of Proposition 2.16), and Corollary 2.23 (with $a = a$, $n = 2$, $\mathbb{I} = (\mathbb{I}_1, \mathbb{I}_d)$, $\Phi = (\Pi_n, \Xi_{n,y})$ for $n \in \{0, 1, \dots, N\}$, $y \in (\mathbb{R}^d)^N$ in the notation of Corollary 2.23) demonstrate that for all $t \in [0, T]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$(\mathcal{R}_a(\Phi_y))(t, x) = \sum_{n=0}^{N} (\mathcal{R}_a(\Gamma))\big((\mathcal{R}_a(\Pi_n))(t), (\mathcal{R}_a(\Xi_{n,y}))(x)\big). \quad (387)$$

Next note that (369) shows that for all $k \in \{0, 1, \dots, N\}$, $t \in \mathbb{R}\backslash(t_{k-1}, t_{k+1})$ it holds that

$$(\mathcal{R}_a(\Pi_k))(t) = 0. \quad (388)$$

Combining this, (386), and (387) with (b) proves that for all $k \in \{0, 1, \dots, N-1\}$, $t \in [t_k, t_{k+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}
Y_t^{x,y} &= \sum_{n=0}^{N} [(\mathcal{R}_a(\Pi_n))(t)]\, [(\mathcal{R}_a(\Xi_{n,y}))(x)]\\
&= [(\mathcal{R}_a(\Pi_k))(t)]\, [(\mathcal{R}_a(\Xi_{k,y}))(x)] + [(\mathcal{R}_a(\Pi_{k+1}))(t)]\, [(\mathcal{R}_a(\Xi_{k+1,y}))(x)]
\end{aligned} \quad (389)$$

and

$$\begin{aligned}(\mathcal{R}_a(\Phi_y))(t,x) &= \sum_{n=0}^{N} (\mathcal{R}_a(\Gamma))\big((\mathcal{R}_a(\Pi_n))(t), (\mathcal{R}_a(\Xi_{n,y}))(x)\big) \\ &= (\mathcal{R}_a(\Gamma))\big((\mathcal{R}_a(\Pi_k))(t), (\mathcal{R}_a(\Xi_{k,y}))(x)\big) \\ &\quad + (\mathcal{R}_a(\Gamma))\big((\mathcal{R}_a(\Pi_{k+1}))(t), (\mathcal{R}_a(\Xi_{k+1,y}))(x)\big).\end{aligned} \tag{390}$$

The triangle inequality, (c), and (d) hence establish that for all $k \in \{0, 1, \ldots, N-1\}$, $t \in [t_k, t_{k+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}&\|Y_t^{x,y} - (\mathcal{R}_a(\Phi_y))(t,x)\| \\ &\leqslant \sum_{n=k}^{k+1} \big\|[(\mathcal{R}_a(\Pi_n))(t)]\,[(\mathcal{R}_a(\Xi_{n,y}))(x)] - (\mathcal{R}_a(\Gamma))\big((\mathcal{R}_a(\Pi_n))(t), (\mathcal{R}_a(\Xi_{n,y}))(x)\big)\big\| \\ &\leqslant \sum_{n=k}^{k+1} \varepsilon\big(\sqrt{d}\,\big[\max\{1, |(\mathcal{R}_a(\Pi_n))(t)|^q\}\big] + \|(\mathcal{R}_a(\Xi_{n,y}))(x)\|^q\big)\end{aligned} \tag{391}$$

and

$$\begin{aligned}\|(\mathcal{R}_a(\Phi_y))(t,x)\| &\leqslant \sum_{n=k}^{k+1} \big\|(\mathcal{R}_a(\Gamma))\big((\mathcal{R}_a(\Pi_n))(t), (\mathcal{R}_a(\Xi_{n,y}))(x)\big)\big\| \\ &\leqslant \sum_{n=k}^{k+1} \big(\sqrt{d}\big(1 + 2|(\mathcal{R}_a(\Pi_n))(t)|^2\big) + 2\|(\mathcal{R}_a(\Xi_{n,y}))(x)\|^2\big).\end{aligned} \tag{392}$$

Next note that (369) ensures that for all $n \in \{0, 1, \ldots, N\}$, $t \in \mathbb{R}$ it holds that $0 \leqslant (\mathcal{R}_a(\Pi_n))(t) \leqslant 1$. Combining this with (391), (392), and (II) demonstrates that for all $k \in \{0, 1, \ldots, N-1\}$, $t \in [t_k, t_{k+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}\|Y_t^{x,y} - (\mathcal{R}_a(\Phi_y))(t,x)\| &\leqslant \sum_{n=k}^{k+1} \varepsilon\big(\sqrt{d} + \|(\mathcal{R}_a(\Xi_{n,y}))(x)\|^q\big) \\ &= \varepsilon\big(\sqrt{d} + \|Y_{t_k}^{x,y}\|^q\big) + \varepsilon\big(\sqrt{d} + \|Y_{t_{k+1}}^{x,y}\|^q\big) \\ &= \varepsilon\big(2\sqrt{d} + \|Y_{t_k}^{x,y}\|^q + \|Y_{t_{k+1}}^{x,y}\|^q\big)\end{aligned} \tag{393}$$

and

$$\begin{aligned}\|(\mathcal{R}_a(\Phi_y))(t,x)\| &\leqslant \sum_{n=k}^{k+1} \big(3\sqrt{d} + 2\|(\mathcal{R}_a(\Xi_{n,y}))(x)\|^2\big) \\ &= 3\sqrt{d} + 2\|Y_{t_k}^{x,y}\|^2 + 3\sqrt{d} + 2\|Y_{t_{k+1}}^{x,y}\|^2 \\ &= 6\sqrt{d} + 2\big(\|Y_{t_k}^{x,y}\|^2 + \|Y_{t_{k+1}}^{x,y}\|^2\big).\end{aligned} \tag{394}$$

Furthermore, observe that (387), (V), and (a) ensure that for all $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\big[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Phi_y))(t,x) \in \mathbb{R}^d\big] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big). \tag{395}$$

In addition, observe that (b), (387) and (388) demonstrate that for all $n \in \{0, 1, \ldots, N\}$, $t \in [0, t_n]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$(\mathcal{R}_a(\Phi_y))(t,x) = \sum_{k=0}^{n} (\mathcal{R}_a(\Gamma))\big((\mathcal{R}_a(\Pi_k))(t), (\mathcal{R}_a(\Xi_{k,y}))(x)\big). \tag{396}$$

This and (VI) show that for all $n \in \{0, 1, \ldots, N\}$, $t \in [0, t_n]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ with $\forall\, k \in \mathbb{N} \cap [0, n]\colon y_k = z_k$ it holds that

$$\begin{aligned}(\mathcal{R}_a(\Phi_y))(t,x) &= \sum_{m=0}^{n} (\mathcal{R}_a(\Gamma))((\mathcal{R}_a(\Pi_m))(t), (\mathcal{R}_a(\Xi_{m,y}))(x)) \\ &= \sum_{m=0}^{n} (\mathcal{R}_a(\Gamma))((\mathcal{R}_a(\Pi_m))(t), (\mathcal{R}_a(\Xi_{m,z}))(x)) \\ &= (\mathcal{R}_a(\Phi_z))(t,x).\end{aligned} \tag{397}$$

Combining this with (A), (385), (393), (394), and (395) establishes items (i)–(vi). The proof of Proposition 3.10 is thus completed. □

3.3.4 A priori estimates for Euler approximations

Lemma 3.11 *Let $N, d \in \mathbb{N}$, $c, C \in [0, \infty)$, $A_1, A_2, \ldots, A_N \in \mathbb{R}^{d\times d}$, let $\|\cdot\| : \mathbb{R}^d \to [0, \infty)$ be a norm on $\mathbb{R}^d$, let $|||\cdot|||\colon \mathbb{R}^{d\times d} \to [0, \infty)$ be the function which satisfies for all $A \in \mathbb{R}^{d\times d}$ that $|||A||| = \sup_{\{x\in\mathbb{R}^d\colon \|x\|\leqslant 1\}} \|Ax\|$, let $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ be a function which satisfies for all $x \in \mathbb{R}^d$ that*

$$\|\mu(x)\| \leqslant C + c\|x\|, \tag{398}$$

and let $Y_n = (Y_n^{x,y})_{(x,y)\in\mathbb{R}^d\times(\mathbb{R}^d)^N} : \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$, $n \in \{0, 1, \ldots, N\}$, be the functions which satisfy for all $n \in \{0, 1, \ldots, N-1\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that $Y_0^{x,y} = x$ and

$$Y_{n+1}^{x,y} = Y_n^{x,y} + A_{n+1}\,\mu\big(Y_n^{x,y}\big) + y_{n+1}. \tag{399}$$

Then

(i) *it holds for all $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, , y_N) \in (\mathbb{R}^d)^N$ that*

$$Y_n^{x,y} = x + \sum_{k=0}^{n-1} \big[A_{k+1}\,\mu\big(Y_k^{x,y}\big) + y_{k+1}\big] \tag{400}$$

and

(ii) *it holds for all $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that*

$$\begin{aligned}&\|Y_n^{x,y}\| \\ &\leqslant \left(\|x\| + C\left[\sum_{k=1}^{n} |||A_k|||\right] + \max\nolimits_{m\in\{0,1,\ldots,n\}} \left\|\sum_{k=1}^{m} y_k\right\|\right) \exp\!\left(c\left[\sum_{k=1}^{n} |||A_k|||\right]\right).\end{aligned} \tag{401}$$

Proof of Lemma 3.11 We claim that for all $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$Y_n^{x,y} = x + \sum_{k=0}^{n-1} \big[A_{k+1}\,\mu\big(Y_k^{x,y}\big) + y_{k+1}\big]. \tag{402}$$

We now prove (402) by induction on $n \in \{0, 1, \ldots, N\}$. Observe that the hypothesis that for all $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that $Y_0^{x,y} = x$ proves (402) in the base case

$n = 0$. For the induction step note that (399) implies that for all $n \in \{0, 1, \ldots, N-1\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ with

$$Y_n^{x,y} = x + \sum_{k=0}^{n-1} \left[A_{k+1}\, \mu\big(Y_k^{x,y}\big) + y_{k+1} \right] \tag{403}$$

it holds that

$$\begin{aligned} Y_{n+1}^{x,y} &= Y_n^{x,y} + A_{n+1}\, \mu\big(Y_n^{x,y}\big) + y_{n+1} \\ &= x + \left[\sum_{k=0}^{n-1} \big(A_{k+1}\, \mu\big(Y_k^{x,y}\big) + y_{k+1} \big) \right] + \big(A_{n+1}\, \mu\big(Y_n^{x,y}\big) + y_{n+1} \big) \\ &= x + \left[\sum_{k=0}^{n} \big(A_{k+1}\, \mu\big(Y_k^{x,y}\big) + y_{k+1} \big) \right]. \end{aligned} \tag{404}$$

Induction thus proves (402). Observe that (402) establishes item (i). In addition, note that (402), the triangle inequality, and the fact that for all $A \in \mathbb{R}^{d\times d}$, $x \in \mathbb{R}^d$ it holds that $\|Ax\| \leqslant |||A|||\, \|x\|$ demonstrate that for all $m \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\|Y_m^{x,y}\| \leqslant \|x\| + \left[\sum_{k=0}^{m-1} |||A_{k+1}|||\, \big\|\mu\big(Y_k^{x,y}\big)\big\| \right] + \left\| \sum_{k=0}^{m-1} y_{k+1} \right\|. \tag{405}$$

Combining this with (398) ensures that for all $n \in \{0, 1, \ldots, N\}$, $m \in \{0, 1, \ldots, n\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} &\|Y_m^{x,y}\| \leqslant \|x\| + \left[\sum_{k=0}^{m-1} |||A_{k+1}|||\, \big(C + c\|Y_k^{x,y}\| \big) \right] + \left\| \sum_{k=1}^{m} y_k \right\| \\ &= \|x\| + C\left[\sum_{k=1}^{m} |||A_k||| \right] + \left\| \sum_{k=1}^{m} y_k \right\| + c\left[\sum_{k=0}^{m-1} |||A_{k+1}|||\, \|Y_k^{x,y}\| \right] \\ &\leqslant \|x\| + C\left[\sum_{k=1}^{n} |||A_k||| \right] + \left[\max\nolimits_{m\in\{0,1,\ldots,n\}} \left\| \sum_{k=1}^{m} y_k \right\| \right] + c\left[\sum_{k=0}^{m-1} |||A_{k+1}|||\, \|Y_k^{x,y}\| \right]. \end{aligned} \tag{406}$$

The time-discrete Gronwall inequality (cf., e.g., Hutzenthaler et al. [25, Lemma 2.1] (with $N = n$, $\alpha = \big(\|x\| + C\big[\sum_{k=1}^{n} |||A_k|||\big] + \max_{m\in\{0,1,\ldots,n\}} \|\sum_{k=1}^{m} y_k\|\big)$, $\beta_0 = c|||A_1|||$, $\beta_1 = c|||A_2|||, \ldots, \beta_{n-1} = c|||A_n|||$, $\epsilon_0 = \|Y_0^{x,y}\|$, $\epsilon_1 = \|Y_1^{x,y}\|, \ldots, \epsilon_n = \|Y_n^{x,y}\|$ for $n \in \{1, 2, \ldots, N\}$ in the notation of Hutzenthaler et al. [25, Lemma 2.1])) hence implies that for all $n \in \{1, 2, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\|Y_n^{x,y}\| \leqslant \left(\|x\| + C\left[\sum_{k=1}^{n} |||A_k||| \right] + \max\nolimits_{m\in\{0,1,\ldots,n\}} \left\| \sum_{k=1}^{m} y_k \right\| \right) \exp\!\left(c\left[\sum_{k=0}^{n-1} |||A_{k+1}||| \right] \right). \tag{407}$$

The hypothesis that for all $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that $Y_0^{x,y} = x$ therefore assures that for all $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\|Y_n^{x,y}\| \leqslant \left(\|x\| + C\left[\sum_{k=1}^{n} |||A_k|||\right] + \max_{m\in\{0,1,\ldots,n\}} \left\|\sum_{k=1}^{m} y_k\right\|\right) \exp\left(c\left[\sum_{k=1}^{n} |||A_k|||\right]\right). \tag{408}$$

This establishes item (ii). The proof of Lemma 3.11 is thus completed. □

3.3.5 A priori error estimates for space-time ANN approximations

Theorem 3.12 *Let $N, d \in \mathbb{N}$, $\mathfrak{C} \in [0,\infty)$, $a \in C(\mathbb{R},\mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$, let $T \in (0,\infty)$, $(t_n)_{n\in\{0,1,\ldots,N\}} \subseteq \mathbb{R}$ satisfy for all $n \in \{0,1,\ldots,N\}$ that $t_n = \frac{nT}{N}$, let $\mathfrak{D} \in [1,\infty)$, $\varepsilon \in (0,1]$, $q \in (2,\infty)$ satisfy that*

$$\mathfrak{D} = \left[\tfrac{720q}{(q-2)}\right]\left[\log_2(\varepsilon^{-1}) + q + 1\right] - 504, \tag{409}$$

let $\Phi \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}^d$ that $\mathcal{I}(\Phi) = \mathcal{O}(\Phi) = d$ and $\|(\mathcal{R}_a(\Phi))(x)\| \leqslant \mathfrak{C}(1 + \|x\|)$, let $Y = (Y_t^{x,y})_{(t,x,y)\in[0,T]\times\mathbb{R}^d\times(\mathbb{R}^d)^N} \colon [0,T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$ be the function which satisfies for all $n \in \{0,1,\ldots,N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that $Y_0^{x,y} = x$ and

$$Y_t^{x,y} = Y_{t_n}^{x,y} + \left(\tfrac{tN}{T} - n\right)\left[\tfrac{T}{N}(\mathcal{R}_a(\Phi))\big(Y_{t_n}^{x,y}\big) + y_{n+1}\right], \tag{410}$$

and let $g_n \colon \mathbb{R}^d \times (\mathbb{R}^d)^N \to [0,\infty)$, $n \in \{0,1,\ldots,N\}$, be the functions which satisfy for all $n \in \{0,1,\ldots,N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that

$$g_n(x,y) = \left(\|x\| + \mathfrak{C}t_n + \max_{m\in\{0,1,\ldots,n\}} \left\|\sum_{k=1}^{m} y_k\right\|\right)\exp(\mathfrak{C}t_n) \tag{411}$$

(cf. Definition 2.1, Definition 2.3, and Definition 3.6). Then there exist $\Psi_y \in \mathbf{N}$, $y \in (\mathbb{R}^d)^N$, such that

(i) *it holds for all $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Psi_y) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,*
(ii) *it holds for all $n \in \{0,1,\ldots,N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that*

$$\|Y_t^{x,y} - (\mathcal{R}_a(\Psi_y))(t,x)\| \leqslant \varepsilon\big(2\sqrt{d} + (g_n(x,y))^q + (g_{n+1}(x,y))^q\big), \tag{412}$$

(iii) *it holds for all $n \in \{0,1,\ldots,N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that*

$$\|(\mathcal{R}_a(\Psi_y))(t,x)\| \leqslant 6\sqrt{d} + 2\big((g_n(x,y))^2 + (g_{n+1}(x,y))^2\big), \tag{413}$$

(iv) *it holds for all $y \in (\mathbb{R}^d)^N$ that*

$$\mathcal{P}(\Psi_y) \leqslant \tfrac{9}{2} N^6 d^{16}\Big[2\mathcal{H}(\Phi) + \mathfrak{D} + \big(30 + 6\mathcal{H}(\Phi) + [4 + \mathcal{P}(\Phi)]^2\big)^2\Big]^2, \tag{414}$$

(v) *it holds for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that*

$$\left[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Psi_y))(t,x) \in \mathbb{R}^d\right] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big), \tag{415}$$

and

(vi) *it holds for all* $n \in \{0, 1, \ldots, N\}$, $t \in [0, t_n]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ *with* $\forall\, k \in \mathbb{N} \cap [0, n]\colon y_k = z_k$ *that*

$$(\mathcal{R}_a(\Psi_y))(t, x) = (\mathcal{R}_a(\Psi_z))(t, x). \tag{416}$$

Proof of Theorem 3.12 Throughout this proof let $\Psi_y \in \mathbf{N}$, $y \in (\mathbb{R}^d)^N$, satisfy that

(I) it holds for all $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Psi_y) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,
(II) it holds for all $n \in \{0, 1, \ldots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that

$$\|Y_t^{x,y} - (\mathcal{R}_a(\Psi_y))(t, x)\| \leqslant \varepsilon\big(2\sqrt{d} + \|Y_{t_n}^{x,y}\|^q + \|Y_{t_{n+1}}^{x,y}\|^q\big), \tag{417}$$

(III) it holds for all $n \in \{0, 1, \ldots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that

$$\|(\mathcal{R}_a(\Psi_y))(t, x)\| \leqslant 6\sqrt{d} + 2\big(\|Y_{t_n}^{x,y}\|^2 + \|Y_{t_{n+1}}^{x,y}\|^2\big), \tag{418}$$

(IV) it holds for all $y \in (\mathbb{R}^d)^N$ that

$$\begin{aligned} \mathcal{P}(\Psi_y) \leqslant \tfrac{1}{2}\Big[&6d^2N^2\mathcal{H}(\Phi) \\ &+ 3N\Big[d^2\mathfrak{D} + \big(23 + 6N\mathcal{H}(\Phi) + 7d^2 + N\big[4d^2 + \mathcal{P}(\Phi)\big]^2\big)^2\Big]\Big]^2, \end{aligned} \tag{419}$$

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

$$\big[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Psi_y))(t, x) \in \mathbb{R}^d\big] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big), \tag{420}$$

and
(VI) it holds for all $n \in \{0, 1, \ldots, N\}$, $t \in [0, t_n]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ with $\forall\, k \in \mathbb{N} \cap [0, n]\colon y_k = z_k$ that

$$(\mathcal{R}_a(\Psi_y))(t, x) = (\mathcal{R}_a(\Psi_z))(t, x) \tag{421}$$

(cf. Proposition 3.10). Note that (IV) ensures for all $y \in (\mathbb{R}^d)^N$ that

$$\begin{aligned} &\mathcal{P}(\Psi_y) \\ &\leqslant \tfrac{1}{2}\Big[6d^2N^2\mathcal{H}(\Phi) + 3N\Big[d^2\mathfrak{D} + \big(23 + 6N\mathcal{H}(\Phi) + 7d^2 + Nd^4\big[4 + \mathcal{P}(\Phi)\big]^2\big)^2\Big]\Big]^2 \\ &\leqslant \tfrac{1}{2}\Big[6d^2N^2\mathcal{H}(\Phi) + 3N\Big[d^2\mathfrak{D} + N^2d^8\big(30 + 6\mathcal{H}(\Phi) + \big[4 + \mathcal{P}(\Phi)\big]^2\big)^2\Big]\Big]^2. \end{aligned} \tag{422}$$

Hence, we obtain that for all $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} \mathcal{P}(\Psi_y) &\leqslant \tfrac{1}{2}\Big[6d^2N^2\mathcal{H}(\Phi) + 3N^3d^8\Big[\mathfrak{D} + \big(30 + 6\mathcal{H}(\Phi) + \big[4 + \mathcal{P}(\Phi)\big]^2\big)^2\Big]\Big]^2 \\ &\leqslant \tfrac{9}{2}\,N^6d^{16}\Big[2\mathcal{H}(\Phi) + \mathfrak{D} + \big(30 + 6\mathcal{H}(\Phi) + \big[4 + \mathcal{P}(\Phi)\big]^2\big)^2\Big]^2. \end{aligned} \tag{423}$$

In addition, observe that Lemma 3.11 and the hypothesis that for all $n \in \{0, 1, \ldots, N\}$ it holds that $t_n = \frac{nT}{N}$ demonstrate that for all $n \in \{0, 1, \ldots, N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} \|Y^{x,y}_{t_n}\| &\leqslant \left[\|x\| + \tfrac{\mathfrak{C}nT}{N} + \max\nolimits_{m\in\{0,1,\ldots,n\}} \left\|\sum_{k=1}^{m} y_k\right\|\right] \exp\!\left(\tfrac{\mathfrak{C}nT}{N}\right) \\ &= \left[\|x\| + \mathfrak{C}t_n + \max\nolimits_{m\in\{0,1,\ldots,n\}} \left\|\sum_{k=1}^{m} y_k\right\|\right] \exp(\mathfrak{C}t_n) = g_n(x, y). \end{aligned} \tag{424}$$

Combining this with (II) and (III) ensures that for all $n \in \{0, 1, \ldots, N-1\}$, $t \in [t_n, t_{n+1}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} \|Y^{x,y}_t - (\mathcal{R}_a(\Psi_y))(t, x)\| &\leqslant \varepsilon\big(2\sqrt{d} + \|Y^{x,y}_{t_n}\|^q + \|Y^{x,y}_{t_{n+1}}\|^q\big) \\ &\leqslant \varepsilon\big(2\sqrt{d} + (g_n(x, y))^q + (g_{n+1}(x, y))^q\big) \end{aligned} \tag{425}$$

and

$$\begin{aligned} \|(\mathcal{R}_a(\Psi_y))(t, x)\| &\leqslant 6\sqrt{d} + 2\big(\|Y^{x,y}_{t_n}\|^2 + \|Y^{x,y}_{t_{n+1}}\|^2\big) \\ &\leqslant 6\sqrt{d} + 2\big((g_n(x, y))^2 + (g_{n+1}(x, y))^2\big). \end{aligned} \tag{426}$$

This, (I), (V), (VI), and (423) establish items (i)-(vi). The proof of Theorem 3.12 is thus completed. □

Corollary 3.13 *Let $\mathfrak{C}, T, \mathfrak{d} \in (0, \infty)$, $a \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $a(x) = \max\{x, 0\}$, let $\Phi_d \in \mathbf{N}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that $\mathcal{I}(\Phi_d) = \mathcal{O}(\Phi_d) = d$, $\|(\mathcal{R}_a(\Phi_d))(x)\| \leqslant \mathfrak{C}(1 + \|x\|)$, and $\mathcal{P}(\Phi_d) \leqslant \mathfrak{C}d^{\mathfrak{d}}$, let $Y^{d,N} = (Y^{d,N}_{t,x,y})_{(t,x,y)\in[0,T]\times\mathbb{R}^d\times(\mathbb{R}^d)^N} : [0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^N \to \mathbb{R}^d$, $N, d \in \mathbb{N}$, be the functions which satisfy for all $d, N \in \mathbb{N}$, $n \in \{0, 1, \ldots, N-1\}$, $t \in \big[\frac{nT}{N}, \frac{(n+1)T}{N}\big]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that $Y^{d,N}_{0,x,y} = x$ and*

$$Y^{d,N}_{t,x,y} = Y^{d,N}_{\frac{nT}{N},x,y} + \big(\tfrac{tN}{T} - n\big)\left[\tfrac{T}{N}(\mathcal{R}_a(\Phi_d))(Y^{d,N}_{\frac{nT}{N},x,y}) + y_{n+1}\right]. \tag{427}$$

(cf. Definition 2.1, Definition 2.3, and Definition 3.6). Then there exist $C \in \mathbb{R}$ and $\Psi_{\varepsilon,d,N,y} \in \mathbf{N}$, $y \in (\mathbb{R}^d)^N$, $N, d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, such that

(i) *it holds for all $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,*

(ii) *it holds for all $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that*

$$\|Y^{d,N}_{t,x,y} - (\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t, x)\| \leqslant C d^{1/2} N^{3/2} \varepsilon(1 + \|x\|^3 + \|y\|^3), \tag{428}$$

(iii) *it holds for all $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that*

$$\|(\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t, x)\| \leqslant C d^{1/2} N(1 + \|x\|^2 + \|y\|^2), \tag{429}$$

(iv) *it holds for all $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $y \in (\mathbb{R}^d)^N$ that*

$$\mathcal{P}(\Psi_{\varepsilon,d,N,y}) \leqslant C d^{16+8\mathfrak{d}} N^6 \big[1 + |\ln(\varepsilon)|^2\big], \tag{430}$$

(v) *it holds for all $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that*

$$\big[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t, x) \in \mathbb{R}^d\big] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big), \tag{431}$$

and

(vi) *it holds for all* $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, $n \in \{0,1,\ldots,N\}$, $t \in [0, \frac{nT}{N}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ *with* $\forall k \in \mathbb{N} \cap [0,n]\colon y_k = z_k$ *that*

$$(\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t,x) = (\mathcal{R}_a(\Psi_{\varepsilon,d,N,z}))(t,x). \tag{432}$$

Proof of Corollary 3.13 Throughout this proof let $\mathfrak{D}_{\varepsilon,q} \in [1,\infty)$, $q \in (2,\infty)$, $\varepsilon \in (0,1]$, satisfy for all $\varepsilon \in (0,1]$, $q \in (2,\infty)$ that

$$\mathfrak{D}_{\varepsilon,q} = \left[\frac{720q}{(q-2)}\right]\left[\log_2(\varepsilon^{-1}) + q + 1\right] - 504, \tag{433}$$

let $c = \max\{\exp(\mathfrak{C}T), \mathfrak{D}_{1,3}, 62 + 6\mathfrak{C}(\mathfrak{C}+1)\}$, and let $g_n^{d,N}\colon \mathbb{R}^d \times (\mathbb{R}^d)^N \to [0,\infty)$, $n \in \{0,1,\ldots,N\}$, $N, d \in \mathbb{N}$, be the functions which satisfy for all $d, N \in \mathbb{N}$, $n \in \{0,1,\ldots,N\}$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ that

$$g_n^{d,N}(x,y) = \left(\|x\| + \frac{\mathfrak{C}nT}{N} + \max_{m\in\{0,1,\ldots,n\}} \left\|\sum_{k=1}^{m} y_k\right\|\right) \exp\!\left(\frac{\mathfrak{C}nT}{N}\right). \tag{434}$$

Note that Theorem 3.12 (with $N = N$, $d = d$, $\mathfrak{C} = \mathfrak{C}$, $a = a$, $T = T$, $t_n = \frac{nT}{N}$, $\mathfrak{D} = \mathfrak{D}_{\varepsilon,3}$, $\varepsilon = \varepsilon$, $q = 3$, $\Phi = \Phi_d$, $Y = Y^{d,N}$, $g_n = g_n^{d,N}$ for $N, d \in \mathbb{N}$, $n \in \{0,1,\ldots,N\}$, $\varepsilon \in (0,1]$ in the notation of Theorem 3.12) implies that there exist $\Psi_{\varepsilon,d,N,y} \in \mathbf{N}$, $y \in (\mathbb{R}^d)^N$, $N, d \in \mathbb{N}$, $\varepsilon \in (0,1]$, which satisfy that

(I) it holds for all $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, $y \in (\mathbb{R}^d)^N$ that $\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}) \in C(\mathbb{R}^{d+1}, \mathbb{R}^d)$,

(II) it holds for all $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, $n \in \{0,1,\ldots,N-1\}$, $t \in [\frac{nT}{N}, \frac{(n+1)T}{N}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that

$$\|Y_{t,x,y}^{d,N} - (\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t,x)\| \leqslant \varepsilon\big(2\sqrt{d} + (g_n^{d,N}(x,y))^3 + (g_{n+1}^{d,N}(x,y))^3\big), \tag{435}$$

(III) it holds for all $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, $n \in \{0,1,\ldots,N-1\}$, $t \in [\frac{nT}{N}, \frac{(n+1)T}{N}]$, $x \in \mathbb{R}^d$, $y \in (\mathbb{R}^d)^N$ that

$$\|(\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t,x)\| \leqslant 6\sqrt{d} + 2\big((g_n^{d,N}(x,y))^2 + (g_{n+1}^{d,N}(x,y))^2\big), \tag{436}$$

(IV) it holds for all $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, $y \in (\mathbb{R}^d)^N$ that

$$\begin{aligned}&\mathcal{P}(\Psi_{\varepsilon,d,N,y})\\ &\leqslant \tfrac{9}{2} N^6 d^{16}\Big[2\mathcal{H}(\Phi_d) + \mathfrak{D}_{\varepsilon,3} + \big(30 + 6\mathcal{H}(\Phi_d) + [4 + \mathcal{P}(\Phi_d)]^2\big)^2\Big]^2,\end{aligned} \tag{437}$$

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

$$\big[(\mathbb{R}^d)^N \ni y \mapsto (\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t,x) \in \mathbb{R}^d\big] \in C\big((\mathbb{R}^d)^N, \mathbb{R}^d\big), \tag{438}$$

and

(VI) it holds for all $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, $n \in \{0,1,\ldots,N\}$, $t \in [0, \frac{nT}{N}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N)$, $z = (z_1, z_2, \ldots, z_N) \in (\mathbb{R}^d)^N$ with $\forall k \in \mathbb{N} \cap$

$[0, n]\colon y_k = z_k$ it holds that

$$(\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t, x) = (\mathcal{R}_a(\Psi_{\varepsilon,d,N,z}))(t, x). \tag{439}$$

Observe that Jensen's inequality implies that for all $n \in \mathbb{N}$, $p \in [1, \infty)$, $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ it holds that

$$|x_1 + \cdots + x_n|^p \leqslant n^{p-1}(|x_1|^p + \cdots + |x_n|^p). \tag{440}$$

Moreover, note that Hölder's inequality shows that for all $N \in \mathbb{N}$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\sum_{k=1}^{N} \|y_k\| = \sum_{k=1}^{N} (1\|y_k\|) \leqslant N^{1/2}\Big(\sum_{k=1}^{N} \|y_k\|^2\Big)^{1/2} = N^{1/2}\|y\|. \tag{441}$$

Combining (440), (II), and (434) therefore ensures that for all $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $n \in \{0, 1, \ldots, N-1\}$, $t \in [\frac{nT}{N}, \frac{(n+1)T}{N}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}
&\|Y^{d,N}_{t,x,y} - (\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t, x)\| \leqslant 2d^{1/2}\varepsilon(1 + (g^{d,N}_N(x, y))^3) \\
&= 2d^{1/2}\varepsilon\left(1 + \left(\|x\| + \mathfrak{C}T + \max_{m\in\{0,1,\ldots,N\}}\left\|\sum_{k=1}^{m} y_k\right\|\right)^3 \exp(3\mathfrak{C}T)\right) \\
&\leqslant 2d^{1/2}\varepsilon\left(1 + 9\left(\|x\|^3 + c^3 + \Big(\sum_{k=1}^{N}\|y_k\|\Big)^3\right)c^3\right) \\
&\leqslant 2d^{1/2}\varepsilon\big(1 + 9\big(\|x\|^3 + c^3 + N^{3/2}\|y\|^3\big)c^3\big) \\
&\leqslant 2c^6 d^{1/2} N^{3/2}\varepsilon\big(1 + 9\big(\|x\|^3 + 1 + \|y\|^3\big)\big) \\
&= 2c^6 d^{1/2} N^{3/2}\varepsilon\big(10 + 9\|x\|^3 + 9\|y\|^3\big) \\
&\leqslant 20c^6 d^{1/2} N^{3/2}\varepsilon(1 + \|x\|^3 + \|y\|^3).
\end{aligned} \tag{442}$$

Next note that (III), (434), (440), and (441) imply that for all $\varepsilon \in (0, 1]$, $d, N \in \mathbb{N}$, $n \in \{0, 1, \ldots, N-1\}$, $t \in [\frac{nT}{N}, \frac{(n+1)T}{N}]$, $x \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned}
&\|(\mathcal{R}_a(\Psi_{\varepsilon,d,N,y}))(t, x)\| \leqslant 6\sqrt{d} + 4(g^{d,N}_N(x, y))^2 \\
&= 6\sqrt{d} + 4\left(\|x\| + \mathfrak{C}T + \max_{m\in\{0,1,\ldots,N\}}\left\|\sum_{k=1}^{m} y_k\right\|\right)^2 \exp(2\mathfrak{C}T) \\
&\leqslant 6\sqrt{d} + 12\left(\|x\|^2 + c^2 + \Big(\sum_{k=1}^{N}\|y_k\|\Big)^2\right)c^2 \\
&\leqslant 6\sqrt{d} + 12\big(\|x\|^2 + c^2 + N\|y\|^2\big)c^2 \\
&\leqslant 18c^4\sqrt{d}N(1 + \|x\|^2 + \|y\|^2).
\end{aligned} \tag{443}$$

Furthermore, observe that (433) shows that for all $\varepsilon \in (0,1]$ it holds that

$$(\mathfrak{D}_{\varepsilon,3})^2 = \big(\tfrac{2160}{\ln(2)}\ln(\varepsilon^{-1}) + \mathfrak{D}_{1,3}\big)^2 \leqslant (\mathfrak{D}_{1,3})^2(\ln(\varepsilon^{-1})+1)^2 \\ \leqslant c^2|1-\ln(\varepsilon)|^2 \leqslant 2c^2(1+|\ln(\varepsilon)|^2). \tag{444}$$

This, (IV), the hypothesis that for all $d \in \mathbb{N}$ it holds that $\mathcal{P}(\Phi_d) \leqslant \mathfrak{C}d^{\mathfrak{d}}$, and (440) assure that for all $\varepsilon \in (0,1]$, $d, N \in \mathbb{N}$, $y \in (\mathbb{R}^d)^N$ it holds that

$$\begin{aligned} &\mathcal{P}(\Psi_{\varepsilon,d,N,y}) \leqslant \tfrac{9}{2}N^6 d^{16}\Big[2\mathfrak{C}d^{\mathfrak{d}} + \mathfrak{D}_{\varepsilon,3} + \big(30 + 6\mathfrak{C}d^{\mathfrak{d}} + [4+\mathfrak{C}d^{\mathfrak{d}}]^2\big)^2\Big]^2 \\ &\leqslant \tfrac{27}{2}N^6 d^{16}\Big[4\mathfrak{C}^2 d^{2\mathfrak{d}} + (\mathfrak{D}_{\varepsilon,3})^2 + \big(30 + 6\mathfrak{C}d^{\mathfrak{d}} + 2\big[16 + \mathfrak{C}^2 d^{2\mathfrak{d}}\big]\big)^4\Big] \\ &\leqslant \tfrac{27}{2}N^6 d^{16}\Big[4\mathfrak{C}^2 d^{2\mathfrak{d}} + 2c^2\big(1+|\ln(\varepsilon)|^2\big) + \big(62 + 6\mathfrak{C}d^{\mathfrak{d}} + 2\mathfrak{C}^2 d^{2\mathfrak{d}}\big)^4\Big] \\ &\leqslant \tfrac{27}{2}N^6 d^{16}\Big[4\mathfrak{C}^2 d^{2\mathfrak{d}} + 2c^2\big(1+|\ln(\varepsilon)|^2\big) + \big(62 + 6\mathfrak{C}(\mathfrak{C}+1)d^{2\mathfrak{d}}\big)^4\Big] \\ &\leqslant \tfrac{27}{2}N^6 d^{16}\Big[cd^{2\mathfrak{d}} + 2c^2\big(1+|\ln(\varepsilon)|^2\big) + \big(cd^{2\mathfrak{d}}\big)^4\Big] \\ &\leqslant 27N^6 d^{16}\Big[c^2\big(1+|\ln(\varepsilon)|^2\big) + \big(cd^{2\mathfrak{d}}\big)^4\Big] \\ &\leqslant 54c^4 N^6 d^{16+8\mathfrak{d}}\Big[1+|\ln(\varepsilon)|^2\Big]. \end{aligned} \tag{445}$$

Combining (I), (442), (443), (445), (V), and (VI) establishes items (i)-(vi). The proof of Corollary 3.13 is thus completed. □

Funding The third author acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure. The third author gratefully acknowledges the startup fund project of Shenzhen Research Institute of Big Data under grant No. T00120220001.

Declarations

Conflict of Interests Hereby we declare there are no conflict of interests.

References

1. Beck, C., Becker, S., Cheridito, P., Jentzen, A., Neufeld, A.: Deep Splitting Method for Parabolic PDEs. SIAM J. Sci Comput. **43**(5), A3135–A3154 (2021)
2. Beck, C., Becker, S., Grohs, P., Jaafari, N., Jentzen, A.: Solving the Kolmogorov PDE by Means of Deep Learning. J. Sci. Comput. **88**(3), Paper No. 73 (2021)

3. Beck, C., E.W., Jentzen, A.: Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. J. Nonlinear Sci. **29**(4), 1563–1619 (2019)
4. Becker, S., Cheridito, P., Jentzen, A.: Deep optimal stopping. J. Mach. Learn. Res. **20**, Paper No. 74, 25 (2019)
5. Becker, S., Cheridito, P., Jentzen, A., Welti, T.: Solving high-dimensional optimal stopping problems using deep learning. European J. Appl. Math. **32**(3), 470–514 (2021)
6. Berg, J., Nyström, K.: A unified deep artificial neural network approach to partial differential equations in complex geometries. Neurocomputing **317**, 28–41 (2018). https://doi.org/10.1016/j.neucom.2018.06.056
7. Berner, J., Grohs, P., Jentzen, A.: Analysis of the generalization error: empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of black-Scholes partial differential equations. SIAM J. Math. Data Sci. **2**(3), 631–657 (2020)
8. Chan-Wai-Nam, Q., Mikael, J., Warin, X.: Machine Learning for Semi Linear PDEs. J. Sci. Comput. **79**(3), 1667–1712 (2019). https://doi.org/10.1007/s10915-019-00908-3
9. E, W., Han, J., Jentzen, A.: Deep Learning-Based Numerical Methods for High-Dimensional Parabolic Partial Differential Equations and Backward Stochastic Differential Equations. Commun. Math. Stat. **5**(4), 349–380 (2017). https://doi.org/10.1007/s40304-017-0117-6
10. E, W., Yu, B.: The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. Commun. Math. Stat. **6,**(1), 1–12 (2018). https://doi.org/10.1007/s40304-018-0127-z
11. Elbrächter, D., Grohs, P., Jentzen, A., Schwab, C.: DNN Expression Rate Analysis Of High-dimensional PDEs: Application to Option Pricing. Constr Approx. https://doi.org/10.1007/s00365-021-09541-6 (2021)
12. Farahmand, A.-m., Nabi, S., Nikovski, D.N.: Deep reinforcement learning for partial differential equation control. In: American Control Conference (ACC), pp. 3120–3127 (2017). https://doi.org/10.23919/ACC.2017.7963427
13. Fujii, M., Takahashi, A., Takahashi, M.: Asymptotic Expansion as Prior Knowledge in Deep Learning Method for High dimensional BSDEs. Asia-Pacific Financial Markets. https://doi.org/10.1007/s10690-019-09271-7 (2019)
14. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press. http://www.deeplearningbook.org (2016)
15. Goudenege, L., Molent, A., Zanette, A.: Machine Learning for Pricing American Options in High Dimension. arXiv:1903.11275, 11 pages (2019)
16. Grohs, P., Hornung, F., Jentzen, A., von Wurstemberger, P.: A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. To appear in Mem. Am. Math. Soc., arXiv:1809.02362, 124 pages (2018)
17. Grohs, P., Perekrestenko, D., Elbrächter, D., Bölcskei, H.: Deep Neural Network Approximation Theory. arXiv:1901.02220, 60 pages (2019)
18. Hairer, E., NØRsett, S.P., Wanner, G.: Solving Ordinary Differential Equations. I, Second Ed., Vol. 8 of Springer Series in Computational Mathematics. Springer, Berlin (1993). Nonstiff problems
19. Han, J., Jentzen, A., E, W.: Solving high-dimensional partial differential equations using deep learning. Proc. Natl. Acad. Sci. **115**(34), 8505–8510 (2018). https://doi.org/10.1073/pnas.1718942115
20. Han, J., Long, J.: Convergence of the deep BSDE method for coupled FBSDEs. arXiv:1811.01165, 26 pages (2018)
21. Henry-Labordere, P.: Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. Preprint, SSRN–id3071506, 16 pages (2017)
22. Hornung, F., Jentzen, A., Salimova, D.: Space-time deep neural network approximations for high-dimensional partial differential equations. arXiv:2006.02199, 52 pages (2020)
23. Huré, C., Pham, H., Warin, X.: Some machine learning schemes for high-dimensional nonlinear PDEs. arXiv:1902.01599, 33 pages (2019)
24. Hutzenthaler, M., Jentzen, A., Kruse, T., Nguyen, T.A.: A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. Partial Differ. Equ. Appl. 1 **2**, Paper No. 10 (2020)
25. Hutzenthaler, M., Jentzen, A., Wurstemberger, v.on.: P. Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. Electron. J. Probab. **25**, Paper No. 101, 73 (2020)
26. Jacquier, A., Oumgari, M.: Deep PPDEs for rough local stochastic volatility. arXiv:1906.02551, 21 pages (2019)

27. Jentzen, A., Salimova, D., Welti, T.: A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. Commun. Math. Sci. **19**(5), 1167–1205 (2021)
28. Kutyniok, G., Petersen, P., Raslan, M., Schneider, R.: A theoretical analysis of deep neural networks and parametric PDEs. *Constr. Approx* (2021)
29. Li, B., Shanshan, T., Yu, H.: Better Approximations of High Dimensional Smooth Functions by Deep Neural Networks with Rectified Power Units. arXiv:1903.05858v3, 21 pages (2019)
30. Long, Z., Lu, Y., Ma, X., Dong, B.: PDE-Net: Learning PDEs from Data. arXiv:1710.09668, 15 pages (2017)
31. Lye, K.O., Mishra, S., Ray, D.: Deep learning observables in computational fluid dynamics. arXiv:1903.03040, 57 pages (2019)
32. Magill, M., Qureshi, F., de Haan, H.: Neural networks trained to solve differential equations learn general representations. Adv. Neural Inf. Process. Syst., 4075–4085 (2018)
33. Petersen, P., Raslan, M., Voigtlaender, F.: Topological properties of the set of functions generated by neural networks of fixed size. Found Comput. Math. **21**(2), 375–444 (2021)
34. Petersen, P., Voigtlaender, F.: Optimal approximation of piecewise smooth functions using deep reLU neural networks. Neural Netw. **108**, 296–330 (2018)
35. Pham, H., Warin, X.: Neural networks-based backward scheme for fully nonlinear PDEs. arXiv:1908.00412v1, 15 pages (2019)
36. Raissi, M.: Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations. arXiv:1801.06637, 26 pages (2018)
37. Reisinger, C., Zhang, Y.: Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. arXiv:1903.06652, 34 pages (2019)
38. Schmidt-Hieber, J.: Nonparametric regression using deep neural networks with reLU activation function. Ann. Stat. **48**, 4 (2020)
39. Sirignano, J., Spiliopoulos, K.: DGM: A deep learning algorithm for solving partial differential equations. J. Comput. Phys. **375**, 1339–1364 (2018). https://doi.org/10.1016/j.jcp.2018.08.029
40. Yarotsky, D.: Error bounds for approximations with deep reLU networks. Neural Netw. **94**, 103–114 (2017). https://doi.org/10.1016/j.neunet.2017.07.002

Affiliations

Philipp Grohs[1] · **Fabian Hornung**[2,3] · **Arnulf Jentzen**[2,4,5] · **Philipp Zimmermann**[2]

Philipp Grohs
philipp.grohs@univie.ac.at

Fabian Hornung
fabianhornung89@gmail.com

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

1 Faculty of Mathematics and Research Platform Data Science, University of Vienna, Vienna, Austria

2 Department of Mathematics, ETH Zurich, Zürich, Switzerland

3 Faculty of Mathematics, Karlsruhe Institute of Technology, Karlsruhe, Germany

4 School of Data Science and Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China

5 Applied Mathematics: Institute for Analysis and Numerics, University of Münster, Münster, Germany