Analysis and Construction of Artificial for the Heat Equations, and Their Associated
Parameters, Depths, and Accuracies.

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Mathematics

by

Shakil Ahmed Rafi
Troy University
Bachelor of Science in Mathematics, 2015
University of Arkansas
Master of Science in Mathematics, 2019

May 2024
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Joshua Lee Padgett, Ph.D.
Dissertation Director, *ex-officio*

Ukash Nakarmi, Ph.D.
Committee Member

Jiahui Chen, Ph.D.
Committee Member

Tulin Kaman, Ph.D.
Committee Member

Abstract

This dissertation seeks to explore a certain calculus for artificial neural networks. Specifically we will be looking at versions of the heat equation, and exploring strategies on how to approximate them.

Our strategy towards the beginning will be to take a technique called Multi-Level Picard (MLP), and present a simplified version of it showing that it converges to a solution of the equation $\left(\frac{\partial}{\partial t} u_d\right)(t,x) = \left(\nabla_x^2 u_d\right)(t,x)$.

We will then take a small detour exploring the viscosity super-solution properties of solutions to such equations. It is here that we will first encounter Feynman-Kac, and see that solutions to these equations can be expressed the expected value of a certain stochastic integral.

The final and last part of the dissertation will be dedicated to expanding a certain neural network framework. We will build on this framework by introducing new operations, namely raising to a power, and use this to build out neural network polynomials. This opens the gateway for approximating transcendental functions such as $\exp(x), \sin(x)$, and $\cos(x)$. This, coupled with a trapezoidal rule mechanism for integration allows us to approximate expressions of the form $\exp\left(\int_a^b \square dt\right)$.

We will, in the last chapter, look at how the technology of neural networks developed in the previous two chapters work towards approximating the expression that Feynman-Kac asserts must be the solution to these modified heat equations. We will then end by giving approximate bounds for the error in the Monte Carlo method. All the while we will maintain that the parameter estimates and depth estimates remain polynomial on $\frac{1}{\varepsilon}$.

As an added bonus we will also look at the simplified MLP technque from the previous chapters of this dissertation and show that yes, they can indeed be approximated with artificial neural networks, and that yes, they can be done so with neural networks whose parameters and depth counts grow only polynomially on $\frac{1}{\varepsilon}$.

Our appendix will contain code listings of these neural network operations, some of the architectures, and some small scale simulation results.

Acknowledgements

I would like to acknowledge my advisor Dr. Joshua Padgett who has been instrumental in me Ph.D. journey. I am incredibly thankful for him taking the time out of his busy schedule to meet with me over the weekends and helping me finish my dissertation. Without his help, guidance, and patience I would never have been where I am today. You not only taught me mathematics, but also how to be a mathematician. Thank you.

I would also like to thank my department, and everyone there, including, but not limited to Dr. Andrew Raich, for his incredible patience and helpful guidance throughout the years. I would also like to thank Dr. Ukash Nakarmi for the excellent collaboartions I've had. I would also like to thank Egan Meaux for all the little things he does to keep the department going.

I would like to acknowledge Marufa Mumu for believing in me when I didn't. You really made the last few months of writing this dissertation, less painful.

I would like to acknowledge my cat, a beautiful Turkish Angora, Tommy. He was pretty useless, but stroking his fur made me stress a little less.

I would like to acknowledge my office-mate Eric Walker, without whom I would never have realized that rage and spite are equally as valid motivators as encouragement and praise.

Finally, I would like to thank Valetta Ventures, Inc. and their product Texifier. It is marvel of software engineering and made the process of creating this dissertation much less painful than it already was.

Dedication

To my grandparents,

M.A. Hye, M.A., & Nilufar Hye

who would've love to see this but can't;

to my parents,

Kamal Uddin Ahmed, M.A. & Shahnaz Parveen, M.A.,

who kept faith in me, always;

and finally to my brothers,

Wakil Ahmed Shabi, BBA & Nabbil Ahmed Sami, B.Eng.,

for whom I have been somewhat imperfect a role model.

Epigraph

Read, in the name of your Lord

—Surah Al-Alaq: 1

The conquest of nature must be achieved with number and measure.

—René Descartes

Contents

II   A Structural Description of Artificial Neural Networks   69

4 Introduction and Basic Notions About Neural Networks   70

5 ANN Product Approximations and Their Consequences   116

List of Figures

List of Tables

List of Published Papers

Parts of Chapter 5 have been made into a paper as *An Algebraic Framework for Understanding Fully Connected Feedforward Artificial Neural Networks, and Their Associated Parameter, Depth, and Accuracy Properties* by Rafi S., Padgett, J.L., and Nakarmi, U. and is currently undergoing review for publication for ICML 2024 at Vienna, Austria.

Parts of the simulation codebase have been submitted for review as *nnR: Neural Networks Made Algebraic* at *The R Journal.* They have further been published as a package `nnR` currently available on `CRAN`.

Part I

On Convergence of Brownian Motion Monte Carlo

Chapter 1

Introduction

1.1 Motivation

Artificial neural networks represent a sea change in computing. They have successfully been used in a wide range of applications, from protein-folding in Tsaban et al. (2022), knot theory in Davies et al. (2022), and extracting data from gravitational waves in Zhao et al. (2023).

As neural networks become more ubiquitous, we see that the number of parameters required to train them increases, which poses two problems: accessibility on low-power devices and the amount of energy needed to train these models, see for instance Wu et al. (2022) and Strubell et al. (2019). Parameter estimates become increasingly crucial in an increasingly climate-challenged world. That we know strict and precise upper bounds on parameter estimates tells us when training becomes wasteful, in some sense, and when, perhaps, different approaches may be needed.

Our goal in this dissertation is threefold:

(i) Firstly, we will take something called Multi-Level Picard first introduced in E et al. (2019) and E et al. (2021), and in particular, the version of Multi-Level Picard that

appears in Hutzenthaler et al. (2021). We show that dropping the drift term and substantially simplifying the process still results in convergence of the method and polynomial bounds for the number of computations required and rather nice properties for the approximations, such as integrability and measurability.

(ii) We will then go on to realize that the solution to a modified version of the heat equation has a solution represented as a stochastic differential equation by Feynman-Kac and further that a version of this can be realized by the modified multi-level Picard technique mentioned in Item (i), with certain simplifying assumptions since we dropped the drift term. A substantial amount of this is inspired by Beck et al. (2021c) and much earlier work in Karatzas and Shreve (1991) and Da Prato and Zabczyk (2002).

(iii) By far, the most significant part of this dissertation is dedicated to expanding and building upon a framework of neural networks as appears in Grohs et al. (2023). We modify this definition highly and introduce several new neural network architectures to this framework ($\mathsf{Pwr}_n^{q,\varepsilon}$, $\mathsf{Pnm}_C^{q,\varepsilon}$, Tun_n^d, $\mathsf{Etr}^{N,h}$, $\mathsf{Xpn}_n^{q,\varepsilon}$, $\mathsf{Csn}_n^{q,\varepsilon}$, $\mathsf{Sne}_n^{q,\varepsilon}$, $\mathsf{E}_n^{N,h,q,\varepsilon}$, $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}$, $\mathsf{UEX}_{n,\mathsf{G}_d,\omega}^{N,h,q,\varepsilon}$, and $\mathsf{UES}_{n,\mathsf{G}_d,\Omega}^{N,h,q,\varepsilon}$, among others) and show, for all these neural networks, that the parameter count grows only polynomially as the accuracy of our model increases, thus beating the curse of dimensionality. This finally paves the way for giving neural network approximations to the techniques realized in Item (ii). We show that it is not too wasteful (defined on the polynomiality of parameter counts) to use neural networks to approximate MLP to approximate a stochastic differential equation equivalent to certain parabolic PDEs as Feynman-Kac necessitates.

We end this dissertation by proposing two avenues of further research: analytical and algebraic. This framework of understanding neural networks as ordered tuples of ordered pairs may be extended to give neural network approximation of classical PDE approximation techniques such as Runge-Kutta, Adams-Moulton, and Bashforth. We

also propose three conjectures about neural networks, as defined in Grohs et al. (2023). They form a bimodule, and that instantiation is a functor.

This dissertation is broken down into three parts. At the end of each part, we will encounter tent-pole theorems, which will eventually lead to the final neural network approximation outcome. These tentpole theorems are Theorem 2.3.4, Theorem 3.3.1, and Theorem 7.4.4. Finally, the culmination of these three theorems is Corollary 7.4.4.1, the end product of the dissertation. We hope, you, the reader will enjoy this.

1.2   Notation, Definitions & Basic notions.

We introduce here basic notations that we will be using throughout this dissertation. Large parts are taken from standard literature inspired by *Matrix Computations* by Golub & van Loan, Golub and Van Loan (2013), *Probability: Theory & Examples* by Rick Durrett, Durrett (2019), and *Concrete Mathematics* by Knuth, Graham & Patashnik, Graham et al. (1994).

1.2.1   Norms and Inner Products

Definition 1.2.1 (Euclidean Norm)**.** *Let $\|\cdot\|_E : \mathbb{R}^d \to [0, \infty)$ denote the Euclidean norm defined for every $d \in \mathbb{N}_0$ and for all $x = \{x_1, x_2, \cdots, x_d\} \in \mathbb{R}^d$ as:*

$$\|x\|_E = \left(\sum_{i=1}^{d} x_i^2 \right)^{\frac{1}{2}} \tag{1.2.1}$$

For the particular case that $d = 1$ and where it is clear from context, we will denote $\|\cdot\|_E$ as $|\cdot|$.

Definition 1.2.2 (Max Norm)**.** *Let $\|\cdot\|_\infty : \mathbb{R}^d \to [0, \infty)$ denote the max norm defined for every $d \in \mathbb{N}$ and for all $x = \{x_1, x_2, \cdots, x_d\} \in \mathbb{R}^d$ as:*

$$\|x\|_\infty = \max_{i \in \{1, 2, \cdots, d\}} \{|x_i|\} \tag{1.2.2}$$

We will denote the max norm $\|\cdot\|_{\max} : \mathbb{R}^{m \times n} \to [0, \infty)$ *defined for every* $m, n \in \mathbb{N}$ *and for all* $A \in \mathbb{R}^{m \times n}$ *as:*

$$\|A\|_{\max} := \max_{\substack{i \in \{1,2,...,m\} \\ j \in \{1,2,...,n\}}} \left| [A]_{i,j} \right| \tag{1.2.3}$$

Definition 1.2.3 (Frobenius Norm). *Let* $\|\cdot\|_F : \mathbb{R}^{m \times n} \to [0, \infty)$ *denote the Frobenius norm defined for every* $m, n \in \mathbb{N}$ *and for all* $A \in \mathbb{R}^{m \times n}$ *as:*

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} [A]_{i,j}^2 \right)^{\frac{1}{2}} \tag{1.2.4}$$

Definition 1.2.4 (Euclidean Inner Product). *Let* $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *denote the Euclidean inner product defined for every* $d \in \mathbb{N}$, *for all* $\mathbb{R}^d \ni x = \{x_1, x_2, ..., x_d\}$, *and for all* $\mathbb{R}^d \ni y = \{y_1, y_2, ..., y_d\}$ *as:*

$$\langle x, y \rangle = \sum_{i=1}^{d} |x_i y_i| \tag{1.2.5}$$

1.2.2 Probability Space and Brownian Motion

Definition 1.2.5 (Probability Space). *A probability space is a triple* $(\Omega, \mathcal{F}, \mathbb{P})$ *where:*

(i) Ω *is a set of outcomes called the **sample space**.*

(ii) $\mathcal{F}$ *is a set of events called the **event space**, where each event is a set of outcomes from the sample space. More specifically, it is a* σ*-algebra on the set* Ω.

(iii) *A measurable function* $\mathbb{P} : \mathcal{F} \to [0,1]$ *assigning each event in the **event space** a probability. More specifically,* $\mathbb{P}$ *is a measure on* Ω *with the caveat that the measure of the entire space is* 1, *i.e.,* $\mathbb{P}(\Omega) = 1$.

Definition 1.2.6 (Random Variable). *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $d \in \mathbb{N}_0$. *For some* $d \in \mathbb{N}_0$ *a random variable is a measurable function* $\mathcal{X} : \Omega \to \mathbb{R}^d$.

Definition 1.2.7 (Expectation). *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the expected value of a random variable X, denoted $\mathbb{E}[X]$ is the Lebesgue integral given by:*

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} \tag{1.2.6}$$

Definition 1.2.8 (Variance). *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the variance of variable X, assuming $\mathbb{E}[X] < \infty$, denoted $\mathbb{V}[X]$, is the identity given by:*

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{1.2.7}$$

Definition 1.2.9 (Stochastic Process). *A stochastic process is a family of random variables over a fixed probability space $(\Omega, \mathcal{F}, \mathbb{R})$, indexed over a set, usually $[0, T]$ for $T \in (0, \infty)$.*

Definition 1.2.10 (Stochastic Basis). *A stochastic basis is a tuple $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ where:*

(i) *$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space equipped with a filtration $\mathbb{F}$ where,*

(ii) *$\mathbb{F} = (\mathcal{F}_i)_{i \in I}$, is a collection of non-decreasing sets under inclusion where for every $i \in I$, I being equipped in total order, it is the case that $\mathcal{F}_i$ is a sub σ-algebra of $\mathcal{F}$.*

Definition 1.2.11 (Brownian Motion Over a Stochastic Basis). *Given a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ a standard $(\mathbb{F}_t)_{t \in [0,T]}$-Brownian motion $\mathcal{W}_t$ is a mapping $\mathcal{W}_t : [0, T] \times \Omega \to \mathbb{R}^d$ satisfying:*

(i) *$\mathcal{W}_t$ is $\mathcal{F}_t$ measurable for all $t \in [0, \infty)$*

(ii) *$\mathcal{W}_0 = 0$ with $\mathbb{P}$-a.s.*

(iii) *$\mathcal{W}_t - \mathcal{W}_s \sim \mathrm{Norm}(0, t - s)$ when $s \in (0, t)$.*

(iv) *$\mathcal{W}_t - \mathcal{W}_s$ is independent of $\mathcal{W}_s$ whenever $s < t$.*

(v) *The paths that $\mathcal{W}_t$ take are $\mathbb{P}$-a.s. continuous.*

Definition 1.2.12 $((\mathbb{F}_t)_{t\in[0,T]}$-adapted Stochastic Process)**.** *Let $T \in (0,\infty)$. Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a filtered probability space with the filtration indexed over $[0,T]$. Let (S, Σ) be a measurable space. Let $\mathcal{X} : [0,T] \times \Omega \to S$ be a stochastic process. We say that $\mathcal{X}$ is an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process if it is the case that $\mathcal{X}_t : \Omega \to S$ is $(\mathcal{F}_t, \Sigma)$ measurable for each $t \in [0,T]$.*

Definition 1.2.13 $((\mathbb{F}_t)_{t\in[0,T]}$-adapted stopping time)**.** *Let $T \in (0,\infty)$, $\tau \in [0,T]$. Assume a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$. It is then the case that $\tau \in \mathbb{R}$ is a stopping time if the stochastic process $\mathcal{X} = (\mathcal{X}_t)_{t\in[0,T]}$ define as:*

$$
\mathcal{X}_t := \begin{cases} 1 : t < \tau \\ \\ 0 : t \geqslant \tau \end{cases} \tag{1.2.8}
$$

is adapted to the filtration $\mathbb{F} := (\mathcal{F}_i)_{i\in[0,T]}$

Definition 1.2.14 (Strong Solution of Stochastic Differential Equation)**.** *Let $d, m \in \mathbb{N}$. Let $\mu : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be Borel-measurable. Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochastic basis, and let $\mathcal{W} : [0,T] \times \Omega \to \mathbb{R}^d$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion. For all $t \in [0,T]$, $x \in \mathbb{R}^d$, let $\mathcal{X}^{t,x} = (\mathcal{X}_s^{t,x})_{s\in[t,T]} \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $t \in [0,T]$ we have $\mathbb{P}$-a.s. that:*

$$
\mathcal{X}^{t,x} = \mathcal{X}_0 + \int_0^t \mu(r, \mathcal{X}_r^{t,x})dr + \int_0^t \sigma(r, \mathcal{X}_r^{t,x})d\mathcal{W}_r \tag{1.2.9}
$$

A strong solution to the stochastic differential equation (1.2.9) on probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$, w.r.t Brownian motion $\mathcal{W}$, w.r.t to initial condition $\mathcal{X}_0 = 0$ is a stochastic process $(\mathcal{X}_t)_{t\in[0,\infty)}$ satisfying that:

(i) *$\mathcal{X}_t$ is adapted to the filtration $(\mathbb{F}_t)_{t\in[0,T]}$.*

(ii) *$\mathbb{P}(\mathcal{X}_0 = 0) = 1$.*

(iii) for all $t \in [0,T]$ it is the case that $\mathbb{P}\left(\int_0^t \|\mu(r, \mathcal{X}_r^{t,x})\|_E + \|\sigma(r, \mathcal{X}_r^{t,x})\|_F d\mathcal{W}_r < \infty \right) = 1$

(iv) it holds with $\mathbb{P}$-a.s. that $\mathcal{X}$ satisfies the equation:

$$\mathcal{X}^{t,x} = \mathcal{X}_0 + \int_0^t \mu(r, \mathcal{X}_r^{t,x})dr + \int_0^t \sigma(r, \mathcal{X}_r^{t,x})d\mathcal{W}_r \qquad (1.2.10)$$

Definition 1.2.15 (Strong Uniqueness Property for Solutions to Stochastic Differential Equations). *Let it be the case that whenever we have two strong solutions $\mathcal{X}$ and $\widetilde{\mathcal{X}}$, w.r.t. process $\mathcal{W}$ and initial condition $\mathcal{X}_0 = 0$, as defined in Definition 1.2.14, it is also the case that $\mathbb{P}(\mathcal{X}_t = \widetilde{\mathcal{X}}_t) = 1$ for all $t \in [0,T]$. We then say that the pair (μ, σ) exhibits a strong uniqueness property.*

1.2.3 Lipschitz and Related Notions

Definition 1.2.16. *Given a function $f : \mathbb{R} \to \mathbb{R}$. We will say that this function is continuous everywhere if the Lebesgue measure of the subsets of the domain where it is not continuous is 0. We will say that $f \in C_{ae}(\mathbb{R}, \mathbb{R})$.*

Definition 1.2.17 (Globally Lipschitz Function). *Let $d \in \mathbb{N}_0$. For every $d \in \mathbb{N}_0$, we say a function $f : \mathbb{R}^d \to \mathbb{R}^d$ is (globally) Lipschitz if there exists an $L \in (0, \infty)$ such that for all $x, y \in \mathbb{R}^d$ it is the case that :*

$$\|f(x) - f(y)\|_E \leqslant L \cdot \|x - y\|_E \qquad (1.2.11)$$

The set of globally Lipschitz functions over set X will be denoted $\mathrm{Lip}_G(X)$

Corollary 1.2.17.1. *Let $d \in \mathbb{N}_0$. For every $d \in \mathbb{N}_0$, a continuous function $f \in C(\mathbb{R}^d, \mathbb{R}^d)$ over a compact set $\mathcal{K} \subsetneq \mathbb{R}^d$ is Lipschitz over that set.*

Proof. By Hiene-Cantor, f is uniformly continuous over set $\mathcal{K}$. Fix an arbitrary ϵ and let δ be from the definition of uniform continuity. By compactness we have a finite cover of $\mathcal{K}$ by

balls of radius δ, centered around $x_i \in \mathcal{K}$:

$$\mathcal{K} \subseteq \bigcup_{i=1}^{N} B_\delta(x_i) \tag{1.2.12}$$

Note that within a given ball, no point x_j is such that $|x_i - x_j| > \delta$. Thus, by uniform continuity, we have the following:

$$|f(x_i) - f(x_j)| < \epsilon \quad \forall i, j \in \{1, 2, ..., N\} \tag{1.2.13}$$

and thus let $\mathfrak{L}$ be defined as:

$$\mathfrak{L} = \max_{\substack{i,j \in \{1,2,...,N\} \\ i \neq j}} \left| \frac{f(x_i) - f(x_j)}{x_i - x_j} \right| \tag{1.2.14}$$

$\mathfrak{L}$ satisfies the Lipschitz property. To see this, let x_1, x_2 be two arbitrary points within $\mathcal{K}$. Let $B_\delta(x_i)$ and $B_\delta(x_j)$ be two points such that $x_1 \in B_\delta(x_i)$ and $x_2 \in B_\delta(x_j)$. The triangle inequality then yields that:

$$|f(x_1) - f(x_2)| \leqslant |f(x_1) - f(x_i)| + |f(x_i) - f(x_j)| + |f(x_j) - f(x_2)|$$

$$\leqslant |f(x_i) - f(x_j)| + 2\epsilon$$

$$\leqslant \mathfrak{L} \cdot |x_i - x_j| + 2\epsilon$$

$$\leqslant \mathfrak{L} \cdot |x_1 - x_2| + 2\epsilon$$

for all $\epsilon \in (0, \infty)$. $\qquad\square$

Definition 1.2.18 (Locally Lipschitz Function)**.** *Let $d \in \mathbb{N}_0$. For every $d \in \mathbb{N}_0$ a function $f : \mathbb{R}^d \to \mathbb{R}^d$ is locally Lipschitz if for all $x_0 \in \mathbb{R}^d$ there exists a compact set $\mathcal{K} \subseteq \text{Domain}(f)$*

containing x_0, and a constant $L \in (0, \infty)$ for that compact set such that

$$\sup_{\substack{x,y \in \mathcal{K} \\ x \neq y}} \left\| \frac{f(x) - f(y)}{x - y} \right\|_E \leqslant L \qquad (1.2.15)$$

The set of locally Lipschitz functions over set X will be denoted $\mathrm{Lip}_L(X)$.

Corollary 1.2.18.1. *A function $f : \mathbb{R}^d \to \mathbb{R}^d$ that is globally Lipschitz is also locally Lipschitz. More concisely $\mathrm{Lip}_G(X) \subsetneq \mathrm{Lip}_L(X)$.*

Proof. Assume not, that is to say, there exists a point $x \in \mathrm{Domain}(f)$, a compact set $\mathcal{K} \subseteq \mathrm{Domain}(f)$, and points $x_1, x_2 \in \mathcal{K}$ such that:

$$\frac{|f(x_1) - f(x_2)|}{x_1 - x_2} \geqslant \mathfrak{L} \qquad (1.2.16)$$

This directly contradicts Definition 1.2.17. $\qquad \square$

1.2.4 Kolmogorov Equations

Definition 1.2.19 (Kolmogorov Equation)**.** *We take our definition from (Da Prato and Zabczyk, 2002, (7.0.1)) with, $u \curvearrowleft u$, $G \curvearrowleft \sigma$, $F \curvearrowleft \mu$, and $\varphi \curvearrowleft g$, and for our purposes we set $A : \mathbb{R}^d \to 0$. Given a separable Hilbert space H (in our case $\mathbb{R}^d$), and letting $\mu : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$, and $g : \mathbb{R}^d \to \mathbb{R}$ be at least Lipschitz, a Kolmogorov Equation is an equation of the form:*

$$\begin{cases} \left(\frac{\partial}{\partial t} u\right)(t, x) = \frac{1}{2} \mathrm{Trace}\left(\sigma(t, x) \left[\sigma(t, x)\right]^* (\mathrm{Hess}_x u)(t, x)\right) + \langle \mu(t, x), (\nabla_x u)(t, x) \rangle \\ u(0, x) = g(x) \end{cases}$$

$$(1.2.17)$$

Definition 1.2.20 (Strict Solution to Kolmogorov Equation)**.** *Let $d \in \mathbb{N}_0$. For every $d \in \mathbb{N}_0$ a function $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ is a strict solution to (1.2.17) if and only if:*

(i) $u \in C^{1,1}\left([0,T] \times \mathbb{R}^d\right)$ *and* $u(0,\cdot) = g$

(ii) $u(t,\cdot) \in UC^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$

(iii) For all $x \in \mathrm{Domain}(A)$*,* $u(\cdot, x)$ *is continuously differentiable on* $[0,\infty)$ *and satisfies (1.2.17).*

Definition 1.2.21 (Generalized Solution to Kolmogorov Equation). *A generalized solution to (1.2.17) is defined as:*

$$u(t,x) = \mathbb{E}\left[g\left(\mathcal{X}^{t,x}\right)\right] \tag{1.2.18}$$

Where the stochastic process $\mathcal{X}^{t,x}$ *is the solution to the stochastic differential equation, for* $x \in \mathbb{R}^d$*,* $t \in [0,T]$*:*

$$\mathcal{X}^{t,x} = \int_0^t \mu\left(\mathcal{X}_r^{t,x}\right) dr + \int_0^t \sigma\left(\mathcal{X}_r^{t,x}\right) dW_r \tag{1.2.19}$$

Definition 1.2.22 (Laplace Operator w.r.t. x). *Let* $d \in \mathbb{N}_0$*, and* $f \in C^2\left(\mathbb{R}^d, \mathbb{R}\right)$*. For every* $d \in \mathbb{N}_0$*, the Laplace operator* $\nabla_x^2 : C^2(\mathbb{R}^d, \mathbb{R}) \to \mathbb{R}$ *is defined as:*

$$\Delta_x f = \nabla_x^2 f := \nabla \cdot \nabla f = \sum_{i=1}^d \frac{\partial f}{\partial x_i} \tag{1.2.20}$$

1.2.5 Linear Algebra Notation and Definitions

Definition 1.2.23 (Identity, Zero Matrix, and the 1-matrix). *Let* $d \in \mathbb{N}$*. We will define the identity matrix for every* $d \in \mathbb{N}$ *as the matrix* $\mathbb{I}_d \in \mathbb{R}^{d \times d}$ *given by:*

$$\mathbb{I}_d = [\mathbb{I}_d]_{i,j} = \begin{cases} 1 & i = j \\ 0 & else \end{cases} \tag{1.2.21}$$

Note that $\mathbb{I}_1 = 1$*.*

Let $m, n, i, j \in \mathbb{N}$. For every $m, n \in \mathbb{N}$, $i \in \{1, 2, \ldots, m\}$, and $j \in \{1, 2, \ldots, n\}$ we define the zero matrix $\mathbb{0}_{m,n} \in \mathbb{R}^{m \times n}$ as:

$$\mathbb{0}_{m,n} = [\mathbb{0}_{m,n}]_{i,j} = 0 \tag{1.2.22}$$

Where we only have a column of zeros, it is convenient to denote $\mathbb{0}_d$ where d is the height of the column.

Let $m, n, i, j \in \mathbb{N}$. For every $m, n \in \mathbb{N}$, $i \in \{1, 2, \ldots, m\}$, and $j \in \{1, 2, \ldots, n\}$ we define matrix of ones $\mathbb{e}_{m,n} \in \mathbb{R}^{m \times n}$ as:

$$\mathbb{e}_{m,n} = [\mathbb{e}]_{i,j} = 1 \tag{1.2.23}$$

Where we only have a column of ones, it is convenient to denote $\mathbb{e}_d$ where d is the height of the column.

Definition 1.2.24 (Single-entry matrix). *Let $m, n, k, l \in \mathbb{N}$ and let $c \in \mathbb{R}$. For $k \in \mathbb{N} \cap [1, m]$ and $l \in \mathbb{N} \cap [1, n]$, we will denote by $\mathbb{k}_{k,l,c}^{m,n} \in \mathbb{R}^{m \times n}$ as the matrix defined by:*

$$\mathbb{k}_{k,l,c}^{m,n} = \left[\mathbb{k}_{k,l}^{m,n}\right]_{i,j} = \begin{cases} c & : k = i \wedge l = j \\ 0 & : else \end{cases} \tag{1.2.24}$$

Definition 1.2.25 (Complex conjugate and transpose). *Let $m, n, i, j \in \mathbb{N}$, and $A \in \mathbb{C}^{m \times n}$. For every $m, n \in \mathbb{N}$, $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$, we denote by $A^* \in \mathbb{C}^{n \times m}$ the matrix:*

$$A^* := [A^*]_{i,j} = \overline{[A]_{j,i}} \tag{1.2.25}$$

Where it is clear that we are dealing with real matrices, i.e., $A \in \mathbb{R}^{m \times n}$, we will denote this as A^T.

Definition 1.2.26 (Column and Row Notation)**.** *Let $m, n, i, j \in \mathbb{N}$ and let $A \in \mathbb{R}^{m \times n}$. For*

every $m, n \in N$ and $i \in \{1, 2, \ldots, m\}$ we denote i-th row as:

$$[A]_{i,*} = \begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,n} \end{bmatrix} \tag{1.2.26}$$

Similarly for every $m, n \in \mathbb{N}$ and $j \in \{1, 2, \ldots, n\}$, we done the j-th row as:

$$[A]_{*,j} = \begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix} \tag{1.2.27}$$

Definition 1.2.27 (Component-wise notation)**.** *Let $m, n, i, j \in \mathbb{N}$, and let $A \in \mathbb{R}^{m \times n}$. Let*

$f : \mathbb{R} \to \mathbb{R}$. *For all $m, n \in \mathbb{N}, i \in \{1, 2, \ldots, m\}$, and $j \in \{1, 2, \ldots, n\}$ we will define*

$f\left([A]_{*,*}\right) \in \mathbb{R}^{m \times n}$ *as:*

$$f\left([A]_{*,*}\right) := \left[f\left([A]_{i,j}\right) \right]_{i,j} \tag{1.2.28}$$

Thus under this notation the component-wise square of A is $\left([A]_{*,*}\right)^2$, the component-wise $\sin$

is $\sin\left([A]_{*,*}\right)$ and the Hadamard product of $A, B \in \mathbb{R}^{m \times n}$ then becomes $A \odot B = [A]_{*,*} \times [B]_{*,*}$.

Remark 1.2.28. *Where we are dealing with a column vector $x \in \mathbb{R}^{d \times 1}$ and it is evident*

from the context we may choose to write $f\left([x]_*\right)$.

Definition 1.2.29 (The Diagonalization Operator)**.** *Let $m_1, m_2, n_1, n_2 \in \mathbb{N}$. Given $A \in$*

$\mathbb{R}^{m_1 \times n_1}$ *and $B \in \mathbb{R}^{m_2 \times n_2}$, we will denote by $\mathrm{diag}\,(A, B)$ the matrix:*

$$\mathrm{diag}\,(A, B) = \begin{bmatrix} A & \mathbb{0}_{m_1, n_2} \\ \mathbb{0}_{m_2, n_1} & B \end{bmatrix} \tag{1.2.29}$$

Remark 1.2.30. $\mathrm{diag}\,(A_1, A_2, \ldots, A_n)$ *is defined analogously for a finite set of matrices*

$A_1, A_2, \ldots, A_n$.

Definition 1.2.31 (Number of rows and columns notation)**.** *Let* $m, n \in \mathbb{N}$. *Let* $A \in \mathbb{R}^{m \times n}$. *Let* rows $: \mathbb{R}^{m \times n} \to \mathbb{N}$ *and* columns $: \mathbb{R}^{m \times n} \to \mathbb{N}$, *be the functions respectively* rows $(A) = m$ *and* columns $(A) = n$.

1.2.6 O-type Notation and Function Growth

Definition 1.2.32 (O-type notation)**.** *Let* $g \in C(\mathbb{R}, \mathbb{R})$. *We say that* $f \in C(\mathbb{R}, \mathbb{R})$ *is in* $O(g(x))$, *denoted* $f \in O(g(x))$, *if there exists* $c \in (0, \infty)$ *and* $x_0 \in (0, \infty)$ *such that for all* $x \in [x_0, \infty)$ *it is the case that:*

$$0 \leqslant f(x) \leqslant c \cdot g(x) \tag{1.2.30}$$

We say that $f \in \Omega(g(x))$ *if there exists* $c \in (0, \infty)$ *and* $x_0 \in (0, \infty)$ *such that for all* $x \in [x_0, \infty)$ *it is the case that:*

$$0 \leqslant cg(x) \leqslant f(x) \tag{1.2.31}$$

We say that $f \in \Theta(g(x))$ *if there exists* $c_1, c_2, x_0 \in (0, \infty)$ *such that for all* $x \in [x_0, \infty)$ *it is the case that:*

$$0 \leqslant c_1 g(x) \leqslant f \leqslant c_2 g(x) \tag{1.2.32}$$

Corollary 1.2.32.1 (Bounded functions and O-type notation)**.** *Let* $f(x) \in C(\mathbb{R}, \mathbb{R})$, *then:*

(i) *if* f *is bounded above for all* $x \in \mathbb{R}$, *it is in* $O(1)$ *for some constant* $c \in \mathbb{R}$.

(ii) *if* f *is bounded below for all* $x \in \mathbb{R}$, *it is in* $\Omega(1)$ *for some constant* $c \in \mathbb{R}$.

(iii) *if* f *is bounded above and below for all* $x \in \mathbb{R}$, *it is in* $\Theta(1)$ *for some constant* $c \in \mathbb{R}$.

Proof. Assume $f \in C(\mathbb{R}, \mathbb{R})$, then:

(i) Assume for all $x \in \mathbb{R}$ it is the case that $f(x) \leqslant M$ for some $M \in \mathbb{R}$, then there exists an $x_0 \in (0, \infty)$ such that for all $x \in (x_0, \infty)$ it is also the case that $0 \leqslant f(x) \leqslant M$, whence $f(x) \in O(1)$.

(ii) Assume for all $x \in \mathbb{R}$ it is the case that $f(x) \geqslant M$ for some $M \in \mathbb{R}$, then there exists an $x_0 \in (0, \infty)$ such that for all $x \in [x_0, \infty)$ it is also the case that $f(x) \geqslant M \geqslant 0$, whence $f(x) \in \Omega(1)$.

(iii) This is a consequence of items (i) and (ii).

$\square$

Corollary 1.2.32.2. *Let $n \in \mathbb{N}_0$. For some $n \in \mathbb{N}_0$, let $f \in O(x^n)$. It is then also the case that $f \in O\left(x^{n+1}\right)$.*

Proof. Let $f \in O(x^n)$. Then there exists $c_0, x_0 \in (0, \infty)$, such that for all $x \in [x_0, \infty)$ it is the case that:

$$f(x) \leqslant c_0 \cdot x^n \tag{1.2.33}$$

Note however that for all $n \in \mathbb{N}_0$, there also exists $c_1, x_1 \in (0, \infty)$ such that for all $x \in (x_1, \infty)$ it is the case that:

$$x^n \leqslant c_1 \cdot x^{n+1} \tag{1.2.34}$$

Thus taken together this implies that for all $x \in (\max\{x_0, x_1\}, \infty)$ it is the case that:

$$f(x) \leqslant c_0 \cdot x^n \leqslant c_0 \cdot c_1 \cdot x^{n+1} \tag{1.2.35}$$

$\square$

Definition 1.2.33 (The floor and ceiling functions)**.** *We denote by* $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$ *and* $\lceil \cdot \rceil :$ $\mathbb{R} \to \mathbb{Z}$ *the functions satisfying for all* $x \in \mathbb{R}$ *that* $\lfloor x \rfloor = \max\left(\mathbb{Z} \cap (-\infty, x]\right)$ *and* $\lceil x \rceil =$ $\min\left(\mathbb{Z} \cap (-\infty, x]\right)$.

1.2.7 The Concatenation of Vectors & Functions

Definition 1.2.34 (Vertical Vector Concatenation)**.** *Let* $m, n \in \mathbb{N}$. *Let* $x = [x_1\ x_2\ \dots\ x_m]^\mathsf{T} \in$ $\mathbb{R}^m$ *and* $y = [y_1, y_2, \dots, y_n]^\mathsf{T} \in \mathbb{R}^n$. *For every* $m, n \in \mathbb{N}$, *we will denote by* $x \frown y \in \mathbb{R}^m \times \mathbb{R}^n$ *the vector given as:*

$$
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_m \\
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}
\tag{1.2.36}
$$

Remark 1.2.35. *We will stipulate that when concatenating vectors as* $x_1 \frown x_2$, x_1 *is on top, and* x_2 *is at the bottom.*

Corollary 1.2.35.1. *Let* $m_1, m_2, n_1, n_2 \in \mathbb{N}$, *such that* $m_1 = n_1$, $m_2 = n_2$, *and let* $x \in \mathbb{R}^{m_1}$, $y \in \mathbb{R}^{n_1}$, $\mathfrak{x} \in \mathbb{R}^{m_2}$, *and* $\mathfrak{y} \in \mathbb{R}^{n_2}$. *It is then the case that* $[x \frown \mathfrak{x}] + [y \frown \mathfrak{y}] = [x + y] \frown [\mathfrak{x} + \mathfrak{y}]$.

Proof. This follows straightforwardly from the fact that:

$$[x \frown \mathfrak{x}] + [y \frown \mathfrak{y}] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m_1} \\ \mathfrak{x}_1 \\ \mathfrak{x}_2 \\ \vdots \\ \mathfrak{x}_{m_2} \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_1} \\ \mathfrak{y}_1 \\ \mathfrak{y}_2 \\ \vdots \\ \mathfrak{y}_{n_2} \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_{m_1} + y_{n_1} \\ \mathfrak{x}_1 + \mathfrak{y}_1 \\ \mathfrak{x}_2 + \mathfrak{y}_2 \\ \vdots \\ \mathfrak{x}_{m_2} + \mathfrak{y}_{n_2} \end{bmatrix} = [x + y] \frown [\mathfrak{x} + \mathfrak{y}] \qquad (1.2.37)$$

$\square$

Definition 1.2.36 (Function Concatenation). *Let $m_1, n_1, m_2, n_2 \in \mathbb{N}$. Let $f : \mathbb{R}^{m_1} \to \mathbb{R}^{n_1}$ and $g : \mathbb{R}^{m_2} \to \mathbb{R}^{n_2}$. We will denote by $f \frown g : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ as the function given for all $x = \{x_1, x_2, \ldots, x_{m_1}\} \in \mathbb{R}^{m_1}$, $\overline{x} \in \{\overline{x_1}, \overline{x_2}, \ldots, \overline{x}_{m_2}\} \in \mathbb{R}^{m_2}$, and $x \frown \overline{x} = \{x_1, x_2, \ldots, x_{m_1}, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m_2}\} \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ by:*

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m_1} \\ \overline{x_1} \\ \overline{x_2} \\ \vdots \\ \overline{x_{m_2}} \end{bmatrix} \longrightarrow \begin{bmatrix} f(x) \\ g(\overline{x}) \end{bmatrix} \qquad (1.2.38)$$

Corollary 1.2.36.1. *Let $m, n \in \mathbb{N}$. Let $x_1 \in \mathbb{R}^m$, $x_2 \in \mathbb{R}^n$, and $f \in C(\mathbb{R}, \mathbb{R})$. It is then the case that $f(x_1 \frown x_2) = f(x_1) \frown f(x_2)$.*

Proof. This follows straightforwardly from the definition of function concatenation. $\square$

Lemma 1.2.37. *Let $m_1, m_2, n_1, n_2 \in \mathbb{N}$. Let $f \in C(\mathbb{R}^{m_1}, \mathbb{R}^{n_1})$ and $g \in C(\mathbb{R}^{m_2}, \mathbb{R}^{n_2})$. It is then also the case that $f \frown g \in C(\mathbb{R}^{m_1} \times \mathbb{R}^{n_1}, \mathbb{R}^{m_2} \times \mathbb{R}^{n_2})$.*

Proof. Let $\mathbb{R}^{m_2} \times \mathbb{R}^{n_2}$ be equipped with the usual product topology, i.e., the topology generated by all products $X \times Y$ of open subsets $X \in \mathbb{R}^{m_2}$ and $Y \in \mathbb{R}^{n_2}$. In such a case let $V \subsetneq \mathbb{R}^{m_2} \times \mathbb{R}^{n_2}$ be an open subset. Then let it be that V_f and V_g are the canonical projections to the first and second factors respectively. Since projection under the usual topology is continuous, it is the case that $V_f \subsetneq \mathbb{R}^{m_2}$ and $V_g \subsetneq \mathbb{R}^{n_2}$ are open sets, respectively. As such it is then also the case that $f^{-1}(V_f) \subsetneq \mathbb{R}^{m_1}$ and $g^{-1}(V_g) \subsetneq \mathbb{R}^{n_1}$ are open sets as well by continuity of f and g. Thus, their product is open as well, proving the lemma. $\square$

Chapter 2

Brownian Motion Monte Carlo

2.1 Brownian Motion Preliminaries

We will present here some standard invariants of Brownian motions. The proofs are standard and can be found in for instance Durrett (2019) and Karatzas and Shreve (1991).

Lemma 2.1.1 (Markov property of Brownian motions)**.** *Let $T \in \mathbb{R}$, $t \in [0, T]$, and $d \in \mathbb{N}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{W}_t : [0, T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion. Fix $s \in [0, \infty)$. Let $\mathfrak{W}_t = \mathcal{W}_{s+t} - \mathcal{W}_s$. Then $\mathfrak{W} = \{\mathfrak{W}_t : t \in [0, \infty)\}$ is also a standard Brownian motion independent of $\mathcal{W}$.*

Proof. We check against the Brownian motion axioms. First note that $\mathfrak{W}_0 = \mathcal{W}_{s+0} - \mathcal{W}_s = 0$ with $\mathbb{P}$-a.s.

Note that $t \mapsto \mathcal{W}_{s+t} - \mathcal{W}_s$ is $\mathbb{P}$-a.s. continuous as it is the difference of two functions that are also $\mathbb{P}$-a.s. continuous.

Note next that for $h \in (0, \infty)$ it is the case that:

$$
\begin{aligned}
\mathbb{E}\left[\mathfrak{W}_{t+h} - \mathfrak{W}_t\right] &= \mathbb{E}\left[\mathcal{W}_{s+t+h} - \mathcal{W}_{s+h} - \mathcal{W}_{s+t} + \mathcal{W}_s\right] \\
&= \mathbb{E}\left[\mathcal{W}_{s+t+h} - \mathcal{W}_{s+t}\right] - \mathbb{E}\left[\mathcal{W}_{s+h} - \mathcal{W}_s\right] \\
&= 0 - 0 = 0
\end{aligned}
\tag{2.1.1}
$$

We note finally that:

$$\mathbb{V}\left[\mathfrak{W}_{t+h} - \mathfrak{M}_t\right] = \mathbb{V}\left[\mathcal{W}_{s+t+h} - \mathcal{W}_s - \mathcal{W}_{s+t} + \mathcal{W}_s\right]$$

$$= \mathbb{V}\left[\mathcal{W}_{s+t+h} - \mathcal{W}_{s+t}\right] - \mathbb{V}\left[\mathcal{W}_s - \mathcal{W}_s\right] + \mathrm{Cov}\left(\mathcal{W}_{s+t+h} - \mathcal{W}_{s+t}, \mathcal{W}_{s+h} - \mathcal{W}_s\right)$$

$$= h - 0 = h$$

Finally note that two stochastic processes $\mathcal{W}$, $\mathcal{X}$ are independent whenever given a set of sample points $t_1, t_2, \ldots, t_n \in [0,T]$ it is the case that the vectors $[\mathcal{W}_{t_1}, \mathcal{W}_{t_2}, \ldots, \mathcal{W}_{t_n}]^\intercal$ and $[\mathcal{X}_{t_1}, \mathcal{X}_{t_2}, \ldots, \mathcal{X}_{t_n}]^\intercal$ are independent vectors.

That being the case note that the independent increments property of Brownian motions yields that, $\mathcal{W}_{s+t_1} - \mathcal{W}_s, \mathcal{W}_{s+t_2} - \mathcal{W}_s, \ldots, \mathcal{W}_{s+t_n} - \mathcal{W}_s$ is independent of $\mathcal{W}_{t_1}, \mathcal{W}_{t_2}, \ldots, \mathcal{W}_{t_n}$, i.e. $\mathfrak{W}$ and $\mathcal{W}$ are independent. $\qquad\square$

Lemma 2.1.2 (Independence of Brownian Motion)**.** *Let $T \in (0, \infty)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{X}, \mathcal{Y} : [0,T] \times \Omega \to \mathbb{R}^d$ be standard Brownian motions. It is then the case that they are independent of each other.*

Proof. We say that two Brownian motions are independent of each of each other if given a sampling vector of times $(t_1, t_2, \ldots, t_n)$, the vectors $(\mathcal{X}_{t_1}, \mathcal{X}_{t_2}, \ldots \mathcal{X}_{t_n})$ and $(\mathcal{Y}_{t_1}, \mathcal{Y}_{t_2}, \ldots, \mathcal{Y}_{t_n})$ are independent. As such let $n \in \mathbb{N}$ and let $(t_1, t_2, \ldots t_n)$ be a vector or times with samples as given above. Consider now a new Brownian motion $\mathcal{X} - \mathcal{Y}$, wherein our samples are now $(\mathcal{X}_{t_1} - \mathcal{Y}_{t_1}, \mathcal{X}_{t_2} - \mathcal{Y}_{t_2}, \ldots, \mathcal{X}_{t_n} - \mathcal{Y}_{t_n})$. By the independence property of Brownian motions, these increments must be independent of each other. Whence it is the case that the vectors $(\mathcal{X}_{t_1}, \mathcal{X}_{t_2}, \ldots, \mathcal{X}_{t_n})$ and $(\mathcal{Y}_{t_1}, \mathcal{Y}_{t_2}, \ldots, \mathcal{Y}_{t_n})$ are independent. $\qquad\square$

Lemma 2.1.3 (Scaling Invariance)**.** *Let $T \in \mathbb{R}$, $t \in [0,T]$, and $d \in \mathbb{N}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{W}_t : [0,T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion. Let $a \in \mathbb{R} \setminus \{0\}$. It is then the case that $\mathcal{X}_t := \frac{1}{a}\mathcal{W}_{a^2 \cdot t}$ is also a standard Brownian motion.*

Proof. We check against the Brownian motion axioms. Note for instance that the function

$t \mapsto \mathcal{X}_t$ is a product of a constant with a function that is $\mathbb{P}$-a.s. continuous yielding a function that is also $\mathbb{P}$-a.s. continuous.

Note also for instance that $\mathcal{X}_0 = \frac{1}{a} \cdot \mathcal{W}_{a^2 \cdot 0} = 0$ with $\mathbb{P}$-a.s.

Note that for all $h \in (0, \infty)$, and $t \in [0, T]$ it is the case that:

$$\mathbb{E}\left[\mathcal{X}_{t+h} - \mathcal{X}_t\right] = \mathbb{E}\left[\frac{1}{a}\mathcal{W}_{a^2 \cdot (t+h)} - \frac{1}{a}\mathcal{W}_{a^2 \cdot t}\right]$$

$$= \frac{1}{a}\mathbb{E}\left[\mathcal{W}_{a^2 \cdot (t+h)} - \mathcal{W}_{a^2 \cdot t}\right]$$

$$= 0$$

Note that for all $h \in (0, \infty)$, and $t \in [0, T]$ it is the case that:

$$\mathbb{V}\left[\mathcal{X}_{t+h} - \mathcal{X}_t\right] = \mathbb{V}\left[\frac{1}{a}\mathcal{W}_{a^2 \cdot (t+h)} - \frac{1}{a}\mathcal{W}_{a^2 \cdot t}\right]$$

$$= \frac{1}{a^2}\mathbb{V}\left[\mathcal{W}_{a^2 \cdot (t+h)} - \mathcal{W}_{a^2 \cdot t}\right]$$

$$= \frac{1}{\cancel{a^2}}\cancel{a^2}\left(\cancel{t} + h - \cancel{t}\right)$$

$$= h \tag{2.1.2}$$

Finally note that for $t \in [0, T]$ and $s \in [0, t)$ it is the case that $\mathcal{W}_{a^2 \cdot t} - \mathcal{W}_{a^2 \cdot s}$ is independent of $\mathcal{W}_{a^2 \cdot s}$. Whence it is also the case that $\mathcal{X}_t - \mathcal{X}_s$ is independent of $\mathcal{X}_s$. $\qquad \square$

Lemma 2.1.4 (Summation of Brownian Motions). *Let $T \in \mathbb{R}$, $t \in [0, T]$ and $d \in \mathbb{N}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{W}_t, \mathcal{X}_t : [0, T] \times \Omega \to \mathbb{R}^d$ be a standard independent Brownian motions. It is then the case that the process $\mathcal{Y}_t$ defined as $\mathcal{Y}_t = \frac{1}{\sqrt{2}}(\mathcal{W}_t + \mathcal{X}_t)$ is also a standard Brownian motion.*

Proof. Note that $t \mapsto \frac{1}{\sqrt{2}}(\mathcal{W}_t + \mathcal{X}_t)$ is $\mathbb{P}$-a.s. continuous as it is the linear combination of two functions that are also $\mathbb{R}$-a.s. continuous.

Note also that $\mathcal{Y}_0 = \frac{1}{\sqrt{2}}(\mathcal{W}_0 + \mathcal{X}_0) = 0 + 0 = 0$ with $\mathbb{P}$-a.s.

21

Note that for all $h \in (0, \infty)$ and $t \in [t, T]$ it is the case that:

$$\mathbb{E}\left[\frac{1}{\sqrt{2}}\left(\mathcal{Y}_{t+h} - \mathcal{Y}_t\right)\right] = \mathbb{E}\left[\frac{1}{\sqrt{2}}\left(\mathcal{W}_{t+h} + \mathcal{X}_{t+h} - \mathcal{W}_t - \mathcal{X}_t\right)\right]$$

$$= \frac{1}{\sqrt{2}}\mathbb{E}\left[\mathcal{W}_{t+h} - \mathcal{W}_t\right] + \frac{1}{\sqrt{2}}\mathbb{E}\left[\mathcal{X}_{t+h} - \mathcal{X}_t\right]$$

$$= 0$$

Note that for all $h \in (0, \infty)$, and $t \in [0, T]$ it is the case that:

$$\mathbb{V}\left[\frac{1}{\sqrt{2}}\left(\mathcal{Y}_{t+h} - \mathcal{Y}_t\right)\right] = \mathbb{V}\left[\frac{1}{\sqrt{2}}\left(\mathcal{W}_{t+h} + \mathcal{X}_{t+h} - \mathcal{W}_t - \mathcal{X}_t\right)\right]$$

$$= \mathbb{V}\left[\frac{1}{\sqrt{2}}\left(\mathcal{W}_{t+h} - \mathcal{W}_t\right) + \frac{1}{\sqrt{2}}\left(\mathcal{X}_{t+h} - \mathcal{X}_t\right)\right]$$

$$= \frac{1}{2}\mathbb{V}\left[\mathcal{W}_{t+h} - \mathcal{W}_t\right] + \frac{1}{2}\mathbb{V}\left[\mathcal{X}_{t+h} - \mathcal{X}_t\right] + \underline{\mathrm{Cov}\,(\mathcal{W}, \mathcal{X})}$$

$$= h$$

$$\square$$

Definition 2.1.5 (Of $\mathfrak{k}$, the modified Kahane–Kintchine constant). *Let $p \in [2, \infty)$. We denote by $\mathfrak{k}_p \in \mathbb{R}$ the real number given by $\mathfrak{k} := \inf\{c \in \mathbb{R}\}$ where it holds that for every probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and every random variable $\mathcal{X} : \Omega \to \mathbb{R}$ with $\mathbb{E}[|\mathcal{X}|] < \infty$ that $(\mathbb{E}\left[(\mathcal{X} - \mathbb{E}\left[\mathcal{X}\right])^p\right])^{\frac{1}{p}} \leqslant c\left(\mathbb{E}\left[|\mathcal{X}|^p\right]\right)^{\frac{1}{p}}.$*

Definition 2.1.6 (Primary Setting For This Chapter). *Let $d, m \in \mathbb{N}$, $T, \mathfrak{L}, p \in [0, \infty)$, $\mathfrak{p} \in [2, \infty)$ $\mathfrak{m} = \mathfrak{k}_\mathfrak{p}\sqrt{\mathfrak{p} - 1}$, $\Theta = \mathbb{Z}$, $g \in C(\mathbb{R}^d, \mathbb{R})$, assume for all $t \in [0, T], x \in \mathbb{R}^d$ that:*

$$\max\{|g(x)|\} \leqslant \mathfrak{L}\left(1 + \|x\|_E^p\right) \tag{2.1.3}$$

and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{W}^\theta : [0, T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$ be independent standard Brownian motions, let $u \in C([0, T] \times \mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0, T]$, $x \in \mathbb{R}^d$, that

$\mathbb{E}[|g(x + \mathcal{W}^0_{T-t})|] < \infty$ and:

$$u(t, x) = \mathbb{E}\left[g\left(x + \mathcal{W}^0_{T-t}\right)\right] \tag{2.1.4}$$

and let let $U^\theta : [0, T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $\theta \in \Theta$ satisfy, $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$, that:

$$U^\theta_m(t, x) = \frac{1}{m}\left[\sum_{k=1}^{m} g\left(x + \mathcal{W}^\theta_{T-t}\right)\right] \tag{2.1.5}$$

Lemma 2.1.7. *Assume Setting 2.1.6 then:*

(i) *it holds for all $n \in \mathbb{N}_0$, $\theta \in \Theta$ that $U^\theta : [0, T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$ is a continuous random field.*

(ii) *it holds that for all $\theta \in \Theta$ that $\sigma\left(U^\theta\right) \subseteq \sigma\left(\left(\mathcal{W}^{(\theta, \mathcal{V})}\right)_{\mathcal{V} \in \Theta}\right).$*

(iii) *it holds that $\left(U^\theta\right)_{\theta \in \Theta}, \left(\mathcal{W}^\theta\right)_{\theta \in \Theta}$, are independent.*

(iv) *it holds for all $n, m \in$, $i, k, \mathfrak{i}, \mathfrak{k} \in \mathbb{Z}$, with $(i, k) \neq (\mathfrak{i}, \mathfrak{k})$ that $(U^{(\theta, i, k)})_{\theta \in \Theta}$ and $\left(U^{(\theta, \mathfrak{i}, \mathfrak{k})}\right)_{\theta \in \Theta}$ are independent and,*

(v) *it holds that $\left(U^\theta\right)_{\theta \in \Theta}$ are identically distributed random variables.*

Proof. For (i) Consider that $\mathcal{W}^\theta_{T-t}$ are continuous random fields and that $g \in C(\mathbb{R}^d, \mathbb{R})$, we have that $U^\theta(t, x)$ is the composition of continuous functions with $m > 0$ by hypothesis, ensuring no singularities. Thus $U^\theta : [0, T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$ is a continuous random field.

For (ii) observe that for all $\theta \in \Theta$ it holds that $\mathcal{W}^\theta$ is $\mathcal{B}\left([0, T] \otimes \sigma\left(W^\theta\right)\right) / \mathcal{B}\left(\mathbb{R}^d\right)$-measurable, this, and induction on prove item (ii).

Moreover observe that item (ii) and the fact that for all $\theta \in \Theta$ it holds that $\left(\mathcal{W}^{(\theta, \vartheta)}_{\vartheta \in \Theta}\right), \mathcal{W}^\theta$ are independent establish item (iii).

Furthermore, note that (ii) and the fact that for all $i, k, \mathfrak{i}, \mathfrak{k} \in \mathbb{Z}$, $\theta \in \Theta$, with $(i, k) \neq (\mathfrak{i}, \mathfrak{k})$ it holds that $\left(\mathcal{W}^{(\theta, i, k, \vartheta)}\right)_{\vartheta \in \Theta}$ and $\left(\mathcal{W}^{(\theta, \mathfrak{i}, \mathfrak{k}, \vartheta)}\right)_{\vartheta \in \Theta}$ are independent, establish item (iv).

Hutzenhaler (Hutzenthaler et al., 2020a, Corollary 2.5) establish item (v). This completes the proof of Lemma 1.1. $\qquad\square$

Lemma 2.1.8. *Assume Setting 2.1.6. Then it holds for $\theta \in \Theta$, $s \in [0, T]$, $t \in [s, T]$, $x \in \mathbb{R}^d$ that:*

$$\mathbb{E}\left[\left|U^\theta\left(t, x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \mathbb{E}\left[\left|g\left(x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \int_s^T \mathbb{E}\left[\left|U^\theta\left(r, x + \mathcal{W}^\theta_{r-s}\right)\right|\right] dr < \infty \quad (2.1.6)$$

Proof. Note that (2.1.3), the fact that for all $r, a, b \in [0, \infty)$ it holds that $(a + b)^r \leqslant 2^{\max\{r-1,0\}}(a^r + b^r)$, and the fact that for all $\theta \in \Theta$ it holds that $\mathbb{E}\left[\|\mathcal{W}^\theta_T\|\right] < \infty$, assure that for all $s \in [0, T]$, $t \in [s, T]$, $\theta \in \Theta$ it holds that:

$$\mathbb{E}\left[\left|g(x + \mathcal{W}^\theta_{t-s})\right|\right] \leqslant \mathbb{E}\left[\mathfrak{L}\left(1 + \|x + \mathcal{W}^\theta_{t-s}\|^p_E\right)\right]$$

$$\leqslant \mathfrak{L}\left[1 + 2^{\max\{p-1,0\}}\left(\|x\|^p_E + \mathbb{E}\left[\|\mathcal{W}^\theta_T\|^p_E\right]\right)\right] < \infty \quad (2.1.7)$$

We next claim that for all $s \in [0, T]$, $t \in [s, T]$, $\theta \in \Theta$ it holds that:

$$\mathbb{E}\left[\left|U^\theta\left(t, x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \int_s^T \mathbb{E}\left[\left|U^\theta\left(r, x + \mathcal{W}^\theta_{r-s}\right)\right|\right] dr < \infty \quad (2.1.8)$$

To prove this claim observe the triangle inequality and (2.1.5), demonstrate that for all $s \in [0, T]$, $t \in [s, T]$, $\theta \in \Theta$, it holds that:

$$\mathbb{E}\left[\left|U^\theta\left(t, x + \mathcal{W}^\theta_{t-s}\right)\right|\right] \leqslant \frac{1}{m}\left[\sum_{i=1}^m \mathbb{E}\left[\left|g\left(x + \mathcal{W}^\theta_{t-s} + \mathcal{W}^\theta_{T-t}\right)\right|\right]\right] \quad (2.1.9)$$

Now observe that (2.1.7) and the fact that $(W^\theta)_{\theta \in \Theta}$ are independent imply that for all $s \in [0, T]$, $t \in [s, T]$, $\theta \in \Theta$, $i \in \mathbb{Z}$ it holds that:

$$\mathbb{E}\left[\left|g\left(x + \mathcal{W}^\theta_{t-s} + \mathcal{W}^{(\theta,0,i)}_{T-t}\right)\right|\right] = \mathbb{E}\left[\left|g\left(x + \mathcal{W}^\theta_{(t-s)+(T-t)}\right)\right|\right] = \mathbb{E}\left[\left|g\left(x + \mathcal{W}^\theta_{T-s}\right)\right|\right] < \infty$$

$$(2.1.10)$$

Combining (2.1.9) and (2.1.10) demonstrate that for all $s \in [0, T]$, $t \in [s, T]$, $\theta \in \Theta$ it holds that:

$$\mathbb{E}\left[\left|U^\theta(t, x + \mathcal{W}^\theta_{t-s})\right|\right] < \infty \tag{2.1.11}$$

Finally observe that for all $s \in [0, T]$ $\theta \in \Theta$ it holds that:

$$\int_s^T \mathbb{E}\left[\left|U^\theta\left(r, x + \mathcal{W}^\theta_{r-s}\right)\right|\right] \leqslant (T - s) \sup_{r \in [s, T]} \mathbb{E}\left[\left|U^\theta\left(r, x + \mathcal{W}^\theta_{r-s}\right)\right|\right] < \infty \tag{2.1.12}$$

Combining (??), (2.1.11), and (2.1.12) completes the proof of Lemma 2.1.8.

$\square$

Corollary 2.1.8.1. *Assume Setting 2.1.6, then we have:*

(i) it holds that $t \in [0, T]$, $x \in \mathbb{R}^d$ that:

$$\mathbb{E}\left[\left|U^0(t, x)\right|\right] + \mathbb{E}\left[\left|g\left(x + \mathcal{W}^0_{T-t}\right)\right|\right] < \infty \tag{2.1.13}$$

(ii) it holds that $t \in [0, T]$, $x \in \mathbb{R}^d$ that:

$$\mathbb{E}\left[U^0(t, x)\right] = \mathbb{E}\left[g\left(x + \mathcal{W}^0_{T-t}\right)\right] \tag{2.1.14}$$

Proof. (i) is a restatement of Lemma 2.1.8 in that for all $t \in [0, T]$:

$$\mathbb{E}\left[\left|U^0(t, x)\right|\right] + \mathbb{E}\left[\left|g\left(x + \mathcal{W}^0_{T-t}\right)\right|\right]$$

$$< \mathbb{E}\left[\left|U^\theta\left(t, x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \mathbb{E}\left[\left|g\left(x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \int_s^T \mathbb{E}\left[\left|U^\theta\left(r, x + \mathcal{W}^\theta_{r-s}\right)\right|\right] dr$$

$$< \infty \tag{2.1.15}$$

Furthermore (ii) is a restatement of Lemma 2.1.8 with $\theta = 0$, $m = 1$, and $k = 1$. This

completes the proof of Corollary 2.1.8.1. $\qquad\square$

2.2 Monte Carlo Approximations

Lemma 2.2.1. *Let $p \in (2, \infty)$, $n \in \mathbb{N}$, let $(\Omega, \mathcal{F}, \mathbb{P})$, be a probability space and let $\mathcal{X}_i : \Omega \to \mathbb{R}$, $i \in \{1, 2, ..., n\}$ be i.i.d. random variables with $\mathbb{E}[|\mathcal{X}_1|] < \infty$. Then it holds that:*

$$\left(\mathbb{E}\left[\left| \mathbb{E}\left[\mathcal{X}_1\right] - \frac{1}{n}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \right|^p \right] \right)^{\frac{1}{p}} \leqslant \left[\frac{p-1}{n}\right]^{\frac{1}{2}} \left(\mathbb{E}\left[|\mathcal{X}_1 - \mathbb{E}\left[\mathcal{X}_1\right]|^p\right]\right)^{\frac{1}{p}} \qquad (2.2.1)$$

Proof. The hypothesis that for all $i \in \{1, 2, ..., n\}$ it holds that $\mathcal{X}_i : \Omega \to \mathbb{R}$ are i.i.d. random variables ensures that:

$$\mathbb{E}\left[\left| \mathbb{E}\left[\mathcal{X}_1\right] - \frac{1}{n}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \right|^p \right] = \mathbb{E}\left[\left| \frac{1}{n}\left(\sum_{i=1}^{n} \left(\mathbb{E}\left[\mathcal{X}_1\right] - \mathcal{X}_i\right)\right) \right|^p \right] = \frac{1}{n^p}\mathbb{E}\left[\left| \sum_{i=1}^{n} \left(\mathbb{E}\left[\mathcal{X}_i\right] - \mathcal{X}_i\right) \right|^p \right]$$

$$(2.2.2)$$

This combined with the fact that for all $i \in \{1, 2, ..., n\}$ it is the case that $\mathcal{X}_i : \Omega \to \mathbb{R}$ are i.i.d. random variables and e.g. (Rio, 2009, Theorem 2.1) (with $p \curvearrowleft p$, $(S_i)_{i \in \{0,1,...,n\}} \curvearrowleft$ $(\sum_{k=1}^{i}(\mathbb{E}[X_k] - X_k))$, $(X_i)_{i \in \{1,2,...,n\}} \curvearrowleft (\mathbb{E}[X_i] - X_i)_{i \in \{1,2,...,n\}}$ in the notation of (Rio, 2009, Theorem 2.1) ensures that:

$$\begin{aligned} \left(\mathbb{E}\left[\left| \mathbb{E}\left[\mathcal{X}_1\right] - \frac{1}{n}\left(\sum_{i=1}^{n} \mathcal{X}_i\right) \right|^p \right] \right)^{\frac{2}{p}} &= \frac{1}{n^2}\left(\mathbb{E}\left[\left| \sum_{i=1}^{n} \left(\mathbb{E}\left[\mathcal{X}_i\right] - \mathcal{X}_i\right) \right|^p \right] \right)^{\frac{2}{p}} \\ &\leqslant \frac{p-1}{n^2}\left[\sum_{i=1}^{n} \left(\mathbb{E}\left[|\mathbb{E}\left[\mathcal{X}_i\right] - \mathcal{X}_i|^p\right]\right)^{\frac{2}{p}} \right] \\ &= \frac{p-1}{n^2}\left[n\left(\mathbb{E}\left[|\mathbb{E}\left[\mathcal{X}_1\right] - \mathcal{X}_1|^p\right]\right)^{\frac{2}{p}} \right] \qquad (2.2.3) \\ &= \frac{p-1}{n}\left(\mathbb{E}\left[|\mathbb{E}\left[\mathcal{X}_1\right] - \mathcal{X}_1|^p\right]\right)^{\frac{2}{p}} \qquad (2.2.4) \end{aligned}$$

This completes the proof of the lemma. $\qquad\square$

Corollary 2.2.1.1. *Let $p \in [2, \infty)$, $n \in \mathbb{N}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{X}_i : \Omega \to \mathbb{R}$, $i \in \{1, 2, ..., n\}$ be i.i.d random variables with $\mathbb{E}[|\mathcal{X}_1|] < \infty$. Then it holds that:*

$$\left(\mathbb{E}\left[\left| \mathbb{E}[\mathcal{X}_1] - \frac{1}{n}\left(\sum_{i=1}^{n} \mathcal{X}_i \right) \right|^p \right] \right)^{\frac{1}{p}} \leqslant \left[\frac{p-1}{n} \right]^{\frac{1}{2}} \left(\mathbb{E}[|\mathcal{X}_1 - \mathbb{E}[\mathcal{X}_1]|^p] \right)^{\frac{1}{p}} \qquad (2.2.5)$$

Proof. Observe that e.g. (Grohs et al., 2018, Proposition 2.3) and Lemma 2.3.1 establish (2.2.5). $\qquad \square$

Corollary 2.2.1.2. *Let $p \in [2, \infty)$, $n \in \mathbb{N}$, let $(\Omega, \mathcal{F}, \mathbb{P})$, be a probability space, and let $\mathcal{X}_i : \Omega \to \mathbb{R}$, $i \in \{1, 2, ..., n\}$, be i.i.d. random variables with $\mathbb{E}[|\mathcal{X}_1|] < \infty$, then:*

$$\left(\mathbb{E}\left[\left| \mathbb{E}[\mathcal{X}_1] - \frac{1}{n}\left(\sum_{i=1}^{n} \mathcal{X}_i \right) \right|^p \right] \right)^{\frac{1}{p}} \leqslant \frac{\mathfrak{k}_p \sqrt{p-1}}{n^{\frac{1}{2}}} \left(\mathbb{E}[|\mathcal{X}_1|^p] \right)^{\frac{1}{p}} \qquad (2.2.6)$$

Proof. This a direct consequence of Definition 2.1.5 and Corollary 2.2.1.1. $\qquad \square$

2.3 Bounds and Covnvergence

Lemma 2.3.1. *Assume Setting 2.1.6. Then it holds for all $t \in [0, T]$, $x \in \mathbb{R}^d$*

$$\left(\mathbb{E}\left[\left| U^0(t, x + \mathcal{W}_t^0) - \mathbb{E}\left[U^0\left(t, x + \mathcal{W}_t^0 \right) \right] \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}}$$
$$\leqslant \frac{\mathfrak{m}}{m^{\frac{1}{2}}} \left[\left(\mathbb{E}\left[\left| g\left(x + \mathcal{W}_T^0 \right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \right] \qquad (2.3.1)$$

Proof. For notational simplicity, let $G_k : [0, T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $k \in \mathbb{Z}$, satisfy for all $k \in \mathbb{Z}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that:

$$G_k(t, x) = g\left(x + \mathcal{W}_{T-t}^{(0,0,-k)} \right) \qquad (2.3.2)$$

Observe that the hypothesis that $(\mathcal{W}^\theta)_{\theta \in \Theta}$ are independent Brownian motions and the hypothesis that $g \in C(\mathbb{R}^d, \mathbb{R})$ assure that for all $t \in [0, T], x \in \mathbb{R}^d$ it holds that $(G_k(t, x))_{k \in \mathbb{Z}}$

are i.i.d. random variables. This and Corollary 2.2.1.2 (applied for every $t \in [0,T]$, $x \in \mathbb{R}^d$ with $p \curvearrowleft \mathfrak{p}$, $n \curvearrowleft m$, $(X_k)_{k \in \{1,2,\dots,m\}} \curvearrowleft (G_k(t,x))_{k \in \{1,2,\dots,m\}}$), with the notation of Corollary 2.2.1.2 ensure that for all $t \in [0,T]$, $x \in \mathbb{R}^d$, it holds that:

$$\left(\mathbb{E}\left[\left| \frac{1}{m}\left[\sum_{k=1}^{m} G_k(t,x) \right] - \mathbb{E}\left[G_1(t,x) \right] \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \frac{\mathfrak{m}}{m^{\frac{1}{2}}} \left(\mathbb{E}\left[|G_1(t,x)|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.3}$$

Combining this, with (2.1.4), (2.1.5), and Item (ii) of Corollary 2.1.8.1 yields that:

$$\left(\mathbb{E}\left[\left| U^0(t,x) - \mathbb{E}\left[U^0(t,x) \right] \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}}$$

$$= \left(\mathbb{E}\left[\left| \frac{1}{m}\left[\sum_{k=1}^{m} G_k(t,x) \right] - \mathbb{E}\left[G_1(t,x) \right] \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.4}$$

$$\leqslant \frac{\mathfrak{m}}{m^{\frac{1}{2}}} \left(\mathbb{E}\left[|G_1(t,x)|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.5}$$

$$= \frac{\mathfrak{m}}{m^{\frac{1}{2}}} \left[\left(\mathbb{E}\left[\left| g\left(x + \mathcal{W}^1_{T-t} \right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \right] \tag{2.3.6}$$

This and the fact that $\mathcal{W}^0$ has independent increments ensure that for all $m \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that:

$$\left(\mathbb{E}\left[\left| U^0\left(t, x + \mathcal{W}^0_t \right) - \mathbb{E}\left[U^0\left(t, x + \mathcal{W}^0_t \right) \right] \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \frac{\mathfrak{m}}{m^{\frac{1}{2}}} \left[\left(\mathbb{E}\left[\left| g\left(x + \mathcal{W}^0_T \right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \right] \tag{2.3.7}$$

This completes the proof of Lemma 2.3.1. $\qquad \square$

Lemma 2.3.2. *Assume Setting 2.1.6. Then it holds for all, $t \in [0,T]$, $x \in \mathbb{R}^d$ that:*

$$\left(\mathbb{E}\left[\left| U^0\left(t, x + \mathcal{W}^0_t \right) - u\left(t, x + \mathcal{W}^0_t \right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\mathbb{E}\left[\left| g\left(x + \mathcal{W}^0_T \right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.8}$$

Proof. Observe that from Corollary 2.1.8.1 item (ii) we have:

$$\mathbb{E}\left[U^0(t,x) \right] = \mathbb{E}\left[g\left(x + \mathcal{W}^{(0,0,-1)}_{T-t} \right) \right] \tag{2.3.9}$$

28

This and (2.1.4) ensure that:

$$u(t, x) - \mathbb{E}\left[U^0(t, x)\right] = 0$$

$$\mathbb{E}\left[U^0(t, x)\right] - u(t, x) = 0 \tag{2.3.10}$$

This, and the fact that $\mathcal{W}^0$ has independent increments, assure that for all, $t \in [0, T]$, $x \in \mathbb{R}^d$, it holds that:

$$\left(\mathbb{E}\left[\left|\mathbb{E}\left[U^0\left(t, x + \mathcal{W}_t^0\right)\right] - u\left(t, x + \mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} = 0 \leqslant \left(\mathbb{E}\left[\left|u\left(t, x + \mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right) \tag{2.3.11}$$

This along with (2.1.4) ensure that:

$$\left(\mathbb{E}\left[\left|\mathbb{E}\left[U^0\left(t, x + \mathcal{W}_t^0\right)\right] - u\left(t, x + \mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} = 0 \leqslant \left(\mathbb{E}\left[\left|g\left(x + \mathcal{W}_T^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.12}$$

Notice that the triangle inequality gives us:

$$\left(\mathbb{E}\left[\left|U^0\left(t, x + \mathcal{W}_t^0\right) - u\left(t, x + \mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \leqslant \left(\mathbb{E}\left[\left|U^0(t, x + W_t^0) - \mathbb{E}\left[U^0(t, x + \mathcal{W}_t^0)\right]\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}$$

$$+ \left(\mathbb{E}\left[\left|\mathbb{E}\left[U^0\left(t, x + \mathcal{W}_t^0\right)\right] - u\left(t, x + \mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.13}$$

This, combined with (1.26), (1.21), the independence of Brownian motions, gives us:

$$\left(\mathbb{E}\left[\left|U^0\left(t, x + \mathcal{W}_t^0\right) - u\left(t, x + \mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \leqslant \left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\mathbb{E}\left[\left|g\left(x + \mathcal{W}_T^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}$$

$$= \left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\mathbb{E}\left[\left|g\left(x + \mathcal{W}_T^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.14}$$

This completes the proof of Lemma 2.3.2. $\square$

Lemma 2.3.3. *Assume Setting 2.1.6. Then it holds for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that:*

$$\left(\mathbb{E}\left[\left| U^0\left(t, x + \mathcal{W}_t^0\right) - u\left(t, x + \mathcal{W}_t^0\right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\sup_{s \in [0,T]} \mathbb{E}\left[\left(1 + \|x + \mathcal{W}_s^0\|_E^p\right)^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}}$$

$$(2.3.15)$$

Proof. Observe that Lemma 2.3.2 ensures that:

$$\left(\mathbb{E}\left[\left| U^0\left(t, x + \mathcal{W}_t^0\right) - u\left(t, x + \mathcal{W}_t^0\right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\mathbb{E}\left[\left| g\left(x + \mathcal{W}_T^0\right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \qquad (2.3.16)$$

Observe next that (2.1.4) ensures that:

$$\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\mathbb{E}\left[\left| g\left(x + \mathcal{W}_T^0\right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\mathbb{E}\left[\left(1 + \|x + \mathcal{W}_T^0\|_E^p\right)^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \qquad (2.3.17)$$

Which in turn yields that:

$$\mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\mathbb{E}\left[\left(1 + \|x + \mathcal{W}_T^0\|_E^p\right)^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\sup_{s \in [0,T]} \mathbb{E}\left[\left(1 + \|x + \mathcal{W}_s^0\|_E^p\right)^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}}$$

$$(2.3.18)$$

Combining (2.3.16), (2.3.17), and (2.3.18) yields that:

$$\left(\mathbb{E}\left[\left| U^0\left(t, x + \mathcal{W}_t^0\right) - u\left(t, x + \mathcal{W}_t^0\right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \leqslant \left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\mathbb{E}\left[\left| g\left(x + \mathcal{W}_T^0\right) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}}$$

$$\leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}} \right) \left(\sup_{s \in [0,T]} \mathbb{E}\left[\left(1 + \|x + \mathcal{W}_s^0\|_E^p\right)^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}}$$

$$(2.3.19)$$

This completes the proof of Lemma 2.3.3. $\qquad \square$

Corollary 2.3.3.1. *Assume Setting 2.1.6. Then it holds for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that:*

$$\left(\mathbb{E}\left[\left|U^0\left(t, x\right) - u(t, x)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s \in [0,T]} \mathbb{E}\left[(1 + \| x + \mathcal{W}_s^0 \|_E^p)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.20}$$

Proof. Observe that for all $t \in [0, T - \mathfrak{t}]$ and $\mathfrak{t} \in [0, T]$, and the fact that W^0 has independent increments it is the case that:

$$u(t + \mathfrak{t}, x) = \mathbb{E}\left[g\left(x + \mathcal{W}_{T-(t+\mathfrak{t})}^0\right)\right] = \mathbb{E}\left[g\left(x + \mathcal{W}_{(T-\mathfrak{t})-t}^0\right)\right] \tag{2.3.21}$$

It is also the case that:

$$U^\theta(t + \mathfrak{t}, x) = \frac{1}{m}\left[\sum_{k=1}^m g\left(x + \mathcal{W}_{T-(t+\mathfrak{t})}^{(\theta,0,-k)}\right)\right] = \frac{1}{m}\left[\sum_{k=1}^m g\left(x + \mathcal{W}_{(T-\mathfrak{t})-t}^{(\theta,0,-k)}\right)\right]$$

Then, applying Lemma 2.3.3, applied for all $\mathfrak{t} \in [0, T]$, with $\mathfrak{L} \curvearrowright \mathfrak{L}$, $p \curvearrowright p$, $\mathfrak{p} \curvearrowright \mathfrak{p}$, $T \curvearrowright (T - \mathfrak{t})$ is such that for all $\mathfrak{t} \in [0, T]$, $t \in [0, T - \mathfrak{t}]$, $x \in \mathbb{R}^d$ we have:

$$\left(\mathbb{E}\left[\left|U^0\left(t + \mathfrak{t}, x + \mathcal{W}_t^0\right) - u\left(t + \mathfrak{t}, x + \mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}$$

$$\leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s \in [0,T-\mathfrak{t}]} \mathbb{E}\left[\left(1 + \| x + \mathcal{W}_s^0 \|_E^p\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}$$

$$\leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s \in [0,T]} \mathbb{E}\left[\left(1 + \| x + \mathcal{W}_s^0 \|_E^p\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.22}$$

Thus we get for all $\mathfrak{t} \in [0, T]$, $x \in \mathbb{R}^d$, $n \in$:

$$\left(\mathbb{E}\left[\left|U^0\left(\mathfrak{t}, x\right) - u\left(\mathfrak{t}, x\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} = \left(\mathbb{E}\left[\left|U^0\left(\mathfrak{t}, x + \mathcal{W}_0^0\right) - u\left(\mathfrak{t}, x + \mathcal{W}_0^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}$$

$$\leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s \in [0,T]} \mathbb{E}\left[\left(1 + \| x + \mathcal{W}_s^0 \|_E^p\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.23}$$

This completes the proof of Corollary 2.3.3.1. $\qquad\square$

Theorem 2.3.4. *Let* $T, L, p, q, \mathfrak{d} \in [0, \infty), m \in \mathbb{N}, \Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, *let* $g_d \in C(\mathbb{R}^d, \mathbb{R})$, *and assume that* $d \in \mathbb{N}, t \in [0, T], x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d, v, w \in \mathbb{R}$ *and that* $\max\{|g_d(x)|\} \leqslant Ld^p \left(1 + \Sigma_{k=1}^d |x_k|^q\right)$, *let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $\mathcal{W}^{d,\theta} : [0, T] \times \Omega \to \mathbb{R}^d, d \in \mathbb{N}, \theta \in \Theta$, *be independent standard Brownian motions, assume for every* $d \in \mathbb{N}$ *that* $\left(\mathcal{W}^{d,\theta}\right)_{\theta \in \Theta}$ *are independent, let* $u_d \in C([0, T] \times \mathbb{R}^d, \mathbb{R}), d \in \mathbb{N}$, *satisfy for all* $d \in \mathbb{N}, t \in [0, T], x \in \mathbb{R}^d$ *that* $\mathbb{E}\left[g_x \left(x + \mathcal{W}^{d,0}_{T-t}\right)\right] < \infty$ *and:*

$$u_d(t, x) = \mathbb{E}\left[g_d \left(x + \mathcal{W}^d_{T-t}\right)\right] \tag{2.3.24}$$

Let $U^{d,\theta}_m : [0, T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}, d \in \mathbb{N}, m \in \mathbb{Z}, \theta \in \Theta$, *satisfy for all,* $d \in \mathbb{N}, m \in \mathbb{Z}, \theta \in \Theta$, $t \in [0, T], x \in \mathbb{R}^d$ *that:*

$$U^d_m(t, x) = \frac{1}{m}\left[\sum_{k=1}^m g_d \left(x + \mathcal{W}^d_{T-t}\right)\right] \tag{2.3.25}$$

and for every $d, n, m \in \mathbb{N}$ *let* $\mathfrak{C}_{d,n,m} \in \mathbb{Z}$ *be the number of function evaluations of* $u_d(0, \cdot)$ *and the number of realizations of scalar random variables which are used to compute one realization of* $U^{d,0}_m(T, 0) : \Omega \to \mathbb{R}$.

There then exists $c \in \mathbb{R}$, *and* $\mathfrak{N} : \mathbb{N} \times (0, 1] \to \mathbb{N}$ *such that for all* $d \in \mathbb{N}, \varepsilon \in (0, 1]$ *it holds that:*

$$\sup_{t \in [0,T]} \sup_{x \in [-L,L]^d} \left(\mathbb{E}\left[\left|u_d(t, x) - U^{d,0}_{\mathfrak{N}(d,\epsilon)}\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \leqslant \varepsilon \tag{2.3.26}$$

and:

$$\mathfrak{C}_{d,\mathfrak{N}(d,\varepsilon),\mathfrak{N}(d,\varepsilon)} \leqslant cd^c \varepsilon^{-(2+\delta)} \tag{2.3.27}$$

Proof. Throughout the proof let $\mathfrak{m}_{\mathfrak{p}} = \mathfrak{k}_p \sqrt{\mathfrak{p} - 1}, \mathfrak{p} \in [2, \infty)$, let $\mathbb{F}^d_t \subseteq \mathcal{F}, d \in \mathbb{N}, t \in [0, T]$

satisfy for all $d \in \mathbb{N}$, $t \in [0, T]$ that:

$$\mathbb{F}_t^d = \begin{cases} \bigcap_{s \in [t,T]} \sigma\left(\sigma\left(W_r^{d,0} : r \in [0,s]\right) \cup \{A \in \mathcal{F} : \mathbb{P}(A) = 0\}\right) & : t < T \\ \sigma\left(\sigma\left(W_s^{d,0} : s \in [0,T]\right) \cup \{A \in \mathcal{F} : \mathbb{P}(A) = 0\}\right) & : t = T \end{cases} \tag{2.3.28}$$

Observe that (2.3.28) guarantees that $\mathbb{F}_t^d \subseteq \mathcal{F}$, $d \in \mathbb{N}$, $t \in [0, T]$ satisfies that:

(I) it holds for all $d \in \mathbb{N}$ that $\{A \in \mathcal{F} : \mathbb{P}(A) = 0\} \subseteq \mathbb{F}_0^d$

(II) it holds for all $d \in \mathbb{N}$, $t \in [0, T)$, that $\mathbb{F}_t^d = \bigcap_{s \in (t,T]} \mathbb{F}_s^d$.

Combining item (I), item (II), (2.3.28) and (Hutzenthaler et al., 2020b, Lemma 2.17) assures us that for all $d \in \mathbb{N}$ it holds that $W^{d,0} : [0, T] \times \Omega \to \mathbb{R}^d$ is a standard $\left(\Omega, \mathcal{F}, \mathbb{P}, \left(\mathbb{F}_t^d\right)_{t \in [0,T]}\right)$-Brownian motion. In addition (58) ensures that it is the case that for all $d \in N$, $x \in \mathbb{R}^d$ it holds that $[0, T] \times \Omega \ni (t, \omega) \mapsto x + W_t^{d,0}(\omega) \in \mathbb{R}^d$ is an $\left(\mathbb{F}_t^d\right)_{t \in [0,T]} / \mathcal{B}\left(\mathbb{R}^d\right)$-adapted stochastic process with continuous sample paths.

This and the fact that for all $d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that $a_d(t, x) = 0$, and the fact that for all $d \in \mathbb{N}$, $t \in [0, T]$, $x, v \in \mathbb{R}^d$ it holds that $b_d(t, x)v = v$ yield that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that $[0, T] \times \Omega \ni (t, \omega) \mapsto x + W_t^{d,0}(\omega) \in \mathbb{R}^d$ satisfies for all $t \in [0, T]$ it holds $\mathbb{P}$-a.s. that:

$$x + W_t^{d,0} = x + \int_0^t 0 \, ds + \int_0^t dW_s^{d,0} = x + \int_0^t a_d(s, x + W_s^{d,0}) ds + \int_0^t b_d(s, x + W_s^{d,0}) dW_s^{d,0} \tag{2.3.29}$$

This and (Hutzenthaler et al., 2020b, Lemma 2.6) (applied for every $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ with $d \curvearrowright d$, $m \curvearrowright d$, $T \curvearrowright T$, $C_1 \curvearrowright d$, $C_2 \curvearrowright 0$, $\mathbb{F} \curvearrowright \mathbb{F}^d$, $\xi \curvearrowright x$, $\mu \curvearrowright a_d$, $\sigma \curvearrowright b_d$, $W \curvearrowright W^{d,0}$, $X \curvearrowright \left([0, T] \times \Omega \ni (t, \omega) \mapsto x + W_t^{d,0}(\omega) \in \mathbb{R}^d\right)$ in the notation of (Hutzenthaler et al.,

2020b, Lemma 2.6) ensures that for all $r \in [0, \infty)$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $t \in [0, T]$ it holds that

$$\mathbb{E}\left[\left\|x + W_t^{d,0}\right\|^r\right] \leqslant \max\{T, 1\} \left(\left(1 + \|x\|^2\right)^{\frac{r}{2}} + (r+1)d^{\frac{r}{2}}\right) \exp\left(\frac{r(r+3)T}{2}\right) < \infty \quad (2.3.30)$$

This, the triangle inequality, and the fact that for all $v, w \in [0, \infty)$, $r \in (0, 1]$, it holds that $(v + w)^r \leqslant v^r + w^r$ assure that for all $\mathfrak{p} \in [2, \infty)$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ it holds that:

$$\sup_{s \in [0,T]} \left(\mathbb{E}\left[\left(1 + \left\|x + W_s^{d,0}\right\|_E^q\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \leqslant 1 + \sup_{s \in [0,T]} \left(\mathbb{E}\left[\left\|x + W_s^{d,0}\right\|_E^{q\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}$$

$$\leqslant 1 + \sup_{s \in [0,T]} \left(\max\{T, 1\}\left(\left(1 + \|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p} + 1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{q\mathfrak{p}(q\mathfrak{p} + 3)T}{2}\right)\right)^{\frac{1}{\mathfrak{p}}}$$

$$\leqslant 1 + \max\{T^{\frac{1}{\mathfrak{p}}}, 1\}\left(\left(1 + \|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p} + 1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{q(q\mathfrak{p} + 3)T}{2}\right)$$

$$\leqslant 2\left(\left(1 + \|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p} + 1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{q(q\mathfrak{p} + 3)T}{2} + \frac{T}{\mathfrak{p}}\right)$$

$$\leqslant 2\left(\left(1 + \|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p} + 1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{[q(q\mathfrak{p} + 3) + 1]T}{2}\right) \quad (2.3.31)$$

Given that for all $d \in \mathbb{N}$, $x \in [-L, L]^d$ it holds that $\|x\|_E \leqslant Ld^{\frac{1}{2}}$, this demonstrates for all $\mathfrak{p} \in [2, \infty)$, $d \in \mathbb{N}$, it holds that:

$$L\left(\frac{\mathfrak{m}_\mathfrak{p}}{m^{\frac{1}{2}}}\right)\left(\sup_{x \in [-L,L]^d} \sup_{s \in [0,T]} \left(\mathbb{E}\left[\left(1 + \left\|x + W_s^{d,0}\right\|_E^q\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}\right)$$

$$\leqslant L\left(\frac{\mathfrak{m}_\mathfrak{p}}{m^{\frac{1}{2}}}\right)\left(\sup_{x \in [-L,L]^d}\left[\left(\left(1 + \|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p} + 1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{[q(q\mathfrak{p} + 3) + 1]T}{2}\right)\right]\right)$$

$$\leqslant L\left(\frac{\mathfrak{m}_\mathfrak{p}}{m^{\frac{1}{2}}}\right)\left(\left(1 + L^2d\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p} + 1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{[q(q\mathfrak{p} + 3) + 1]T}{2}\right) \quad (2.3.32)$$

Combining this with Corollary 2.3.3.1 tells us that:

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left| u_d(t,x) - U_m^{d,0}(t,x) \right|^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \\
&\leqslant L \left(\frac{\mathfrak{m}_{\mathfrak{p}}}{m^{\frac{1}{2}}} \right) \left(\sup_{x \in [-L,L]^d} \sup_{s \in [0,T]} \left(\mathbb{E}\left[\left(1 + \left\| x + W_s^{d,0} \right\|_E^q \right)^{\mathfrak{p}} \right] \right)^{\frac{1}{\mathfrak{p}}} \right) \\
&\leqslant L \left(\frac{\mathfrak{m}_{\mathfrak{p}}}{m^{\frac{1}{2}}} \right) \left(\left(1 + L^2 d \right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}} \right) \exp\left(\frac{[q(q\mathfrak{p}+3)+1]\,T}{2} \right) \quad (2.3.33)
\end{aligned}
$$

This and the fact that for all $d \in \mathbb{N}$ and $\varepsilon \in (0,\infty)$ and the fact that $\mathfrak{m}_{\mathfrak{p}} \leqslant 2$, it holds that for fixed $L, q, \mathfrak{p}, d, T$ there exists an $\mathfrak{M}_{L,q,\mathfrak{p},d,T} \in \mathbb{R}$ such that $\mathfrak{N}_{d,\epsilon} \geqslant \mathfrak{M}_{L,q,\mathfrak{p},d,T}$ forces:

$$
L \left[\frac{\mathfrak{m}_{\mathfrak{p}}}{\mathfrak{N}_{d,\epsilon}^{\frac{1}{2}}} \right] \left(\left(1 + L^2 d \right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}} \right) \exp\left(\frac{[q(q\mathfrak{p}+3)+1]\,T}{2} \right) \leqslant \varepsilon \quad (2.3.34)
$$

Thus (2.3.33) and (2.3.34) together proves (2.3.26).

Note that $\mathfrak{C}_{d,\mathfrak{N}_{d,\epsilon},\mathfrak{N}_{d,\epsilon}}$ is the number of function evaluations of $u_d(0,\cdot)$ and the number of realizations of scalar random variables which are used to compute one realization of $U_{\mathfrak{N}_{d,\epsilon}}^{d,0}(T,0) : \Omega \to \mathbb{R}$. Let $\widetilde{\mathfrak{N}_{d,\varepsilon}}$ be the value of $\mathfrak{N}_{d,\varepsilon}$ that causes equality in (2.3.34). In such a situation the number of evaluations of $u_d(0,\cdot)$ do not exceed $\widetilde{\mathfrak{N}_{d,\varepsilon}}$. Each evaluation of $u_d(0,\cdot)$ requires at most one realization of scalar random variables. Thus we do not exceed $2\widetilde{\mathfrak{N}_{d,\epsilon}}$. Thus note that:

$$
\mathfrak{C}_{d,\mathfrak{N}_{d,\varepsilon},\mathfrak{N}_{d,\varepsilon}} \leqslant \left\lceil 2 \left[L\mathfrak{m}_{\mathfrak{p}} \left(\left(1 + L^2 d \right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}} \right) \exp\left(\frac{[q(q\mathfrak{p}+3)+1]\,T}{2} \right) \right] \varepsilon^{-1} \right\rceil \quad (2.3.35)
$$

Note that other than d and ε everything on the right-hand side is constant or fixed. Hence (2.3.35) can be rendered as:

$$
\mathfrak{C}_{d,\mathfrak{N}_{d,\varepsilon},\mathfrak{N}_{d,\varepsilon}} \leq cd^k \varepsilon^{-1} \quad (2.3.36)
$$

Where both c and k are dependent on $L, \mathfrak{p}, \mathfrak{m}, L$ $\qquad\square$

Chapter 3

That u is a Viscosity Solution

Our goal this chapter is to use Feynman-Kac to see that the solutions to certain versions of the heat equations can be expressed as the expectation of a stocahstic integral. Parts of this work is heavily inspired from Crandall et al. (1992) and esp. Beck et al. (2021a).

3.1   Some Preliminaries

We take work previously pioneered by Itô (1942a) and Itô (1942b), and then seek to re-apply concepts applied in Beck et al. (2021a) and Beck et al. (2021b).

Lemma 3.1.1. *Let $d, m \in \mathbb{N}$, $T \in (0, \infty)$. Let $\mu \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R}^{d \times m})$ satisfying that they have non-empty compact supports and let $\mathfrak{S} = \mathrm{supp}(\mu) \cup \mathrm{supp}(\sigma) \subseteq [0,T] \times \mathbb{R}^d$. Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0,T]})$ be a filtered probability space satisfying usual conditions. Let $W : [0,T] \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t \in [0,T]}$ -Brownian motion, and let $\mathcal{X} : [0,T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_t)_{t \in [0,T]}$-adapted stochastic process with continuous sample paths satisfying for all $t \in [0,T]$ with $\mathbb{P}$-a.s. that:*

$$\mathcal{X}_t = \mathcal{X}_0 + \int_0^t \mu(s, \mathcal{X}_s)ds + \int_0^t \sigma(s, \mathcal{X}_s)dW_s \tag{3.1.1}$$

It then holds that:

(i) $[(\mathbb{P}\left(\mathcal{X}_0 \notin \mathfrak{S}\right) = 1) \implies (\mathbb{P}\left(\forall t \in [0,T] : \mathcal{X}_t = \mathcal{X}_0\right) = 1)]$

(ii) $[(\mathbb{P}\left(\mathcal{X}_0 \in \mathfrak{S}\right) = 1) \implies (\mathbb{P}\left(\forall t \in [0,T] : \mathcal{X}_t \in \mathfrak{S}\right) = 1)]$

Proof. Assume that $\mathbb{P}(\mathcal{X}_0 \notin \mathfrak{S}) = 1$, meaning that the particle almost surely starts outside $\mathfrak{S}$. It is then the case that $\mathbb{P}(\forall t \in [0,T] : \|\mu(t, \mathcal{X}_0)\|_E + \|\sigma(t, \mathcal{X}_0)\|_F = 0) = 1$ as the μ and σ are outside their supports, and we integrate over zero over time.

It is then the case that:

$$\mathcal{Y} := \left([0,T] \times \Omega \ni (t, \omega) \mapsto \mathcal{X}_0(\omega) \in \mathbb{R}^d\right) \tag{3.1.2}$$

is an $(\mathbb{F}_t)_{t \in [0,T]}$ adapted stochastic process with continuous sample paths satisfying that for all $t \in [0,T]$ with $\mathbb{P}$-almost surety that:

$$\mathcal{Y}_t = \mathcal{X}_0 + \int_0^t 0 ds + \int_0^t 0 dW_s = \mathcal{X}_0 + \int_0^t \mu(s, \mathcal{X}_0) ds + \int_0^t \sigma(s, \mathcal{X}_0) dW_s$$
$$= \mathcal{X}_0 + \int_0^t \mu(s, \mathcal{Y}_s) ds + \int_0^t \sigma(s, \mathcal{Y}_s) dW_s \tag{3.1.3}$$

Note that since $\mu \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R}^{d \times m})$, and since continuous functions are locally Lipschitz, and since this is especially true in the space variable for μ and σ, the fact that $\mathfrak{S}$ is compact and continuous functions over compact sets are Lipschitz and bounded, and (Karatzas and Shreve, 1991, Theorem 5.2.5) allows us to conclude that strong uniqueness holds, that is to say:

$$\mathbb{P}\left(\forall t \in [0,T] : \mathcal{X}_t = \mathcal{X}_0\right) = \mathbb{P}\left(\forall t \in [0,T] : \mathcal{X}_t = \mathcal{Y}_t\right) = 1 \tag{3.1.4}$$

establishing the case (i).

Assume now that $\mathbb{P}(\mathcal{X}_0 \in \mathfrak{S}) = 1$ that is to say that the particle almost surely starts inside $\mathfrak{S}$. We define $\tau : \Omega \to [0,T]$ as $\tau = \inf\{t \in [0,T] : \mathcal{X}_t \notin \overline{\mathfrak{S}}\}$. τ is an $(\mathbb{F}_t)_{t \in [0,T]}$-adapted stopping time. On top of τ we can define $\mathcal{Y} : [0,T] \times \Omega \to \mathbb{R}^d$, for all $t \in [0,T]$, $\omega \in \Omega$

37

as $\mathcal{Y}_t(\omega) = \mathcal{X}_{\min\{t,\tau\}}(\omega)$. $\mathcal{Y}$ is thus an $(\mathbb{F}_t)_{t\in[0,T]}$-adapted stochastic process with continuous sample paths. Note however that for $t > \tau$ it is the case $\|\mu(t,\mathcal{Y}_t) + \sigma(t,\mathcal{Y}_t)\|_E = 0$ as we are outside their supports. For $t < \tau$ it is also the case that $\mathcal{Y}_t = \mathcal{X}_t$. This yields with $\mathbb{P}$-a.s. that:

$$
\begin{aligned}
\mathcal{Y}_t = \mathcal{X}_{\min\{t,\tau\}} &= \mathcal{X}_0 + \int_0^{\min\{t,\tau\}} \mu(s,\mathcal{X}_s)ds + \int_0^{\min\{t,\tau\}} \sigma(s,\mathcal{X}_s)dW_s \\
&= \mathcal{X}_0 + \int_0^t \mathbb{1}_{\{0<s\leqslant\tau\}}\mu(s,\mathcal{X}_s)ds + \int_0^t \mathbb{1}_{\{0<s\leqslant\tau\}}\sigma(s,\mathcal{X}_s)dW_s \\
&= \mathcal{X}_0 + \int_0^t \mu(s,\mathcal{Y}_s)ds + \int_0^t \sigma(s,\mathcal{Y}_s)dW_s
\end{aligned}
\tag{3.1.5}
$$

Thus another application of (Karatzas and Shreve, 1991, Theorem 5.2.5) and the fact that within our compact support $\mathfrak{S}$, the continuous functions μ and σ are Lipschitz and hence locally Lipschitz, and also bounded gives us:

$$
\mathbb{P}(\forall t \in [0,T] : \mathcal{X}_t = \mathcal{Y}_t) = 1
\tag{3.1.6}
$$

Proving case (ii). $\qquad\square$

Lemma 3.1.2. *Let $d, m \in \mathbb{N}$, $T \in (0,\infty)$. Let $g \in C^2(\mathbb{R}^d,\mathbb{R})$. Let $\mu \in C^{1,3}([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^{1,3}([0,T] \times \mathbb{R}^d, \mathbb{R}^{d\times m})$ have non-empty compact supports and let $\mathfrak{S} = \mathrm{supp}(\mu) \cup \mathrm{supp}(\sigma)$. Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$ be a stochaastic basis and let $W : [0,T] \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion. For every $t \in [0,T]$, $x \in \mathbb{R}^d$, let $\mathcal{X}^{t,x} = (\mathcal{X}_s^{t,x})_{s\in[t,T]} : [t,T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying for all $s \in [t,T]$ with $\mathbb{P}$-almost surety that:*

$$
\mathcal{X}_s^{t,x} = x + \int_t^s \mu(r,\mathcal{X}_r^{t,x})dr + \int_t^s \sigma(r,\mathcal{X}_s^{t,x})dW_r
\tag{3.1.7}
$$

also let $u : \mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that:

$$u(t,x) = \mathbb{E}\left[g(\mathcal{X}_T^{t,x})\right] \tag{3.1.8}$$

then it is the case that we have:

(i) $u \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ and

(ii) for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that $u(T,x) = g(x)$ and:

$$\left(\frac{\partial}{\partial t} u\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma\left(t,x\right)\left[\sigma\left(t,x\right)\right]^*\left(\operatorname{Hess}_x u\right)(t,x)\right) + \left\langle \mu\left(t,x\right), \left(\nabla_x u\right)(t,x)\right\rangle = 0 \tag{3.1.9}$$

Proof. We break the proof down into two cases, inside the support $\mathfrak{S} = \operatorname{supp}(\mu) \cup \operatorname{supp}(\sigma)$ and outside the support: $[0,T] \times (\mathbb{R}^d \setminus \mathfrak{S})$.

For the case inside $\mathfrak{S}$. Note that we may deduce from Item (i) of Lemma 3.1.1 that for all $t \in [0,T]$, $x \in \mathbb{R}^d \setminus \mathfrak{S}$ it is the case that $\mathbb{P}(\forall s \in [t,T] : \mathcal{X}_s^{t,x} = x) = 1$. Thus for all $t \in [0,T]$, $x \in \mathbb{R}^d \setminus \mathfrak{S}$ we have, deriving from (3.1.8):

$$u(t,x) = \mathbb{E}\left[g\left(\mathcal{X}_T^{t,x}\right)\right] = g(x) \tag{3.1.10}$$

Note that $g(x)$ only has a space parameter and so derivatives w.r.t. t is 0. Inhereting from the regularity properties of g and (3.1.10), we may assume for all $t \in [0,T]$, $x \in \mathbb{R}^d \setminus \mathfrak{S}$, that $u|_{[0,T]\times(\mathbb{R}^d\setminus\mathfrak{S})} \in C^{1,2}([0,T] \times (\mathbb{R}^d \setminus \mathfrak{S}))$. Note that the hypotheses that $\mu \in C^{1,3}([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^{1,3}([0,T] \times \mathbb{R}^d, \mathbb{R}^{d\times m})$ allow us to apply Theorem 7.4.3, Theorem 7.4.5 and Theorem 7.5.1 from Da Prato and Zabczyk (2002) for $t \in [0,T]$, $x \in \mathbb{R}^d \setminus \mathfrak{S}$, to give us:

(i) $u \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$.

(ii)

$$0 = \left(\frac{\partial}{\partial t} u\right)(t, x)$$
$$= \left(\frac{\partial}{\partial t} u\right)(t, x) + \frac{1}{2} \operatorname{Trace}\left(\sigma(t, x) \left[\sigma(t, x)\right]^* (\operatorname{Hess}_x u)(t, x)\right) + \langle \mu(t, x), (\nabla_x u)(t, x)\rangle$$

$$(3.1.11)$$

Now consider the case within support $\mathfrak{S}$. Note that by hypothesis μ and σ must at least be locally Lipschitz. Thus (Karatzas and Shreve, 1991, Theorem 5.2.5) allows us to conclude that within $\mathfrak{S}$ the pair (μ, σ) for our our stochastic process $\mathcal{X}_s^{t,x}$ defined in (3.1.7) must exhibit a strong uniqueness property.

Further note that Item (ii) from Lemma 3.1.1 tells us that:

$$\mathbb{P}(\forall t \in [0, T] : \mathcal{X}_s^{t,x} \in \mathfrak{S}) = 1. \tag{3.1.12}$$

Note that again the hypotheses that $\mu \in C^{1,3}([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^{1,3}([0, T] \times \mathbb{R}^d, \mathbb{R}^{d \times m})$, and $g \in C^2(\mathbb{R}^d)$ allow us to apply Theorem 7.4.3, Theorem 7.4.5 and Theorem 7.5.1 from Da Prato and Zabczyk (2002) for $t \in [0, T]$, $x \in \mathfrak{S}$, to give us:

(i) $u \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$.

(ii)

$$\left(\frac{\partial}{\partial t} u\right)(t, x) + \frac{1}{2} \operatorname{Trace}\left(\sigma(t, x) \left[\sigma(t, x)\right]^* (\operatorname{Hess}_x u)(t, x)\right) + \langle \mu(t, x), (\nabla_x u)(t, x)\rangle = 0$$

$$(3.1.13)$$

Note that (3.1.7) and (3.1.8) together prove that $u(T, x) = g(x)$. This completes the proof.

$\square$

40

3.2 Viscosity Solutions

Definition 3.2.1 (Symmetric Matrices). *Let $d \in \mathbb{N}$. The set of symmetric matrices is denoted $\mathbb{S}_d$ given by $\mathbb{S}_d = \{A \in \mathbb{S}_d : A^* = A\}$.*

Definition 3.2.2 (Upper semi-continuity). *A function $f : U \to \mathbb{R}$ is upper semi-continuous at x_0 if for every $\varepsilon > 0$, there exists $\delta > 0$ such that:*

$$f(x) < f(x_0) + \varepsilon \text{ for all } x \in B\left(x_0, \delta\right) \cap U \tag{3.2.1}$$

Definition 3.2.3 (Lower semi-continuity). *A function $f : U \to \mathbb{R}$ is lower semi-continuous at x_0 if for every $\varepsilon > 0$, there exists $\delta > 0$ such that:*

$$f(x) > f(x_0) - \varepsilon \text{ for all } x \in B\left(x_0, \delta\right) \cap U \tag{3.2.2}$$

Corollary 3.2.3.1. *Given two upper semi-continuous functions $f, g : \mathbb{R}^d \to \mathbb{R}$, their sum $(f + g) : \mathbb{R}^d \to \mathbb{R}$ is also upper semi-continuous.*

Proof. From definitions, at any given $x_0 \in \mathbb{R}^d$, for any $\varepsilon \in (0, \infty)$ there exist neighborhoods U and V around x_0 such that:

$$(\forall x \in U)\left(f(x) \leqslant f(x_0) + \varepsilon\right) \tag{3.2.3}$$

$$(\forall x \in V)\left(g(x) \leqslant g(x_0) + \varepsilon\right) \tag{3.2.4}$$

and hence:

$$(\forall x \in U \cap V)\left(f(x) + g(x) \leqslant f(x_0) + g(x_0) + 2\varepsilon\right) \tag{3.2.5}$$

$\square$

Corollary 3.2.3.2. *Given an upper semi-continuous function $f : \mathbb{R}^d \to \mathbb{R}$, it is the case*

that $(-f) : \mathbb{R}^d \to \mathbb{R}$ *is lower semi-continuous.*

Proof. Let $f : \mathbb{R}^d \to \mathbb{R}$ be upper semi-continuous. At any given $x_0 \in \mathbb{R}^d$, for any $\varepsilon \in (0, \infty)$ there exists a neighborhood U around x_0 such that:

$$(\forall x \in U) \, (f(x) \leqslant f(x_0) + \varepsilon) \tag{3.2.6}$$

This also means that:

$$(\forall x \in U) \, (-f(x) \geqslant -f(x_0) - \varepsilon)$$

$$\tag{3.2.7}$$

This completes the proof. $\qquad\square$

Definition 3.2.4 (Degenerate Elliptic Functions). *Let $d \in \mathbb{N}$, $T \in (0, \infty)$, let $\mathcal{O} \subseteq \mathbb{R}^d$ be a non-empty open set, and let $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the standard Euclidean inner product on $\mathbb{R}^d$. G is degenerate elliptic on $(0, T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$ if and only if:*

(i) $G : (0, T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ is a function, and

(ii) for all $t \in (0, T)$, $x \in \mathcal{O}$, $r \in \mathbb{R}$, $p \in \mathbb{R}^d$, $A, B \in \mathbb{S}_d$, with $\forall y \in \mathbb{R}^d$: $\langle Ay, y \rangle \leqslant \langle By, y \rangle$ that $G(t, x, r, p, A) \leqslant G(t, x, r, p, B)$.

Remark 3.2.5. *Let $t \in (0, T)$, $x \in \mathbb{R}^d$, $r \in \mathbb{R}$, $p \in \mathbb{R}^d$, $A \in \mathbb{S}_d$. Let $u \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, and let $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ and $\mu : \mathbb{R}^d \to \mathbb{R}^d$ be infinitely often differentiable. The function $G : (0, T) \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ given by:*

$$G(t, x, r, p, A) = \frac{1}{2} \operatorname{Trace} \left(\sigma(x) \, [\sigma(x)]^* \, (\operatorname{Hess}_x u) \, (t, x) \right) + \langle \mu(t, x), \nabla_x u \, (t, x) \rangle \tag{3.2.8}$$

where $(t, x, u(t, x), \mu(x), \sigma(x) \, [\sigma(x)]^*) \in (0, T) \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$, is degenerate elliptic.

Lemma 3.2.6. *Given a function $G : (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ that is degerate elliptic on $(0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$ it is also the case that $H : (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ given by $H(t,x,r,p,A) = -G(t,x,-r,-p,-A)$ is degenerate elliptic on $(0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$.*

Proof. Note that H is a function. Assume for $y \in \mathbb{R}^d$ it is the case that $\langle Ay, y \rangle \leqslant \langle By, y \rangle$ then it is also the case by **(??)** that $\langle -Ay, y \rangle \geqslant \langle -By, y \rangle$ for $y \in \mathbb{R}^d$. However since G is monotoically increasing over the subset of $(0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$ where $\langle Ay, y \rangle \leqslant \langle By, y \rangle$ then it is also the case that $H(t,x,r,p,A) = -G(t,x,-r,-p,-A) \geqslant -G(t,x,-r,-p,-B) = H(t,x,r,p,B)$. $\qquad\square$

Definition 3.2.7 (Viscosity subsolutions)**.** *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $\mathcal{O} \subseteq \mathbb{R}^d$ be a non-empty open set, and let $G : (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ be degenrate elliptic. Then we say that u is a viscosity solution of $\left(\frac{\partial}{\partial t}u\right)(t,x) + G\left(t,x,u(t,x),(\nabla_x u)(t,x),(\mathrm{Hess}_x u)(t,x)\right) \geqslant 0$ for $(t,x,) \in (0,T) \times \mathcal{O}$ if and only if there exists a set A such that:*

(i) *we have that $(0,T) \times \mathcal{O} \subseteq A$.*

(ii) *we have that $u : A \to \mathbb{R}$ is an upper semi-continuous function from A to $\mathbb{R}$, and*

(iii) *we have that for all $t \in (0,T)$, $x \in \mathcal{O}$, $\phi \in C^{1,2}\left((0,T) \times \mathcal{O}, \mathbb{R}\right)$ with $\phi(t,x) = u(t,x)$ and $\phi \geqslant u$ that:*

$$\left(\frac{\partial}{\partial t}u_d\right)(t,x) + G\left(t,x,\phi(t,x),(\nabla_x \phi)(t,x),(\mathrm{Hess}_x \phi)(t,x)\right) \geqslant 0 \qquad (3.2.9)$$

Definition 3.2.8 (Viscosity supersolutions)**.** *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, let $\mathcal{O} \subseteq \mathbb{R}^d$ be a non-empty open set, and let $G : (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ be degenrate elliptic. Then we say that u is a viscosity solution of $\left(\frac{\partial}{\partial t}u\right)(t,x) + G\left(t,x,u(t,x),(\nabla_x u)(t,x),(\mathrm{Hess}_x u)(t,x)\right) \leqslant 0$ for $(t,x,) \in (0,T) \times \mathcal{O}$ if and only if there exists a set A such that:*

(i) *we have that $(0,T) \times \mathcal{O} \subseteq A$.*

(ii) we have that $u : A \to \mathbb{R}$ is an upper semi-continuous function from A to $\mathbb{R}$, and

(iii) we have that for all $t \in (0, T)$, $x \in \mathcal{O}$, $\phi \in C^{1,2}((0,T) \times \mathcal{O}, \mathbb{R})$ with $\phi(t,x) = u(t,x)$ and $\phi \leqslant u$ that:

$$\left(\frac{\partial}{\partial t} u_d \right)(t,x) + G\left(t, x, \phi(t,x), (\nabla_x \phi)(t,x), (\mathrm{Hess}_x\, \phi)(t,x)\right) \leqslant 0 \qquad (3.2.10)$$

Definition 3.2.9 (Viscosity solution). *Let $d \in \mathbb{N}$, $T \in (0, \infty)$, $\mathcal{O} \subseteq \mathbb{R}^d$ be a non-empty open set and let $G : (0,T) \times \mathcal{O} \times R \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ be degenerate elliptic. Then we say that u_d is a viscosity solution to $\left(\frac{\partial}{\partial t} u_d \right)(t,x) + G(t, x, u(t,x), \nabla_x(x,t), (\mathrm{Hess}_x\, u_d)(t,x))$ if and only if:*

(i) *u is a viscosity subsolution of $\left(\frac{\partial}{\partial t} u_d \right)(t,x) + G(t, x, u(t,x), \nabla_x(x,t), (\mathrm{Hess}_x\, u_d)(t,x)) = 0$ for $(t,x) \in (0,T) \times \mathcal{O}$*

(ii) *u is a viscosity supersolution of $\left(\frac{\partial}{\partial t} u_d \right)(t,x) + G(t, x, u(t,x), \nabla_x(x,t), (\mathrm{Hess}_x\, u_d)(t,x)) = 0$ for $(t,x) \in (0,T) \times \mathcal{O}$*

Lemma 3.2.10. *Let $d \in \mathbb{N}$, $T \in (0, \infty)$, $\mathfrak{t} \in (0,T)$, let $\mathcal{O} \subseteq \mathbb{R}^d$ be an open set, let $\mathfrak{r} \in \mathcal{O}$, $\phi \in C^{1,2}((0,T) \times \mathcal{O}, \mathbb{R})$, let $G : (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ be degenerate elliptic and let $u_d : (0,T) \times \mathcal{O} \to \mathbb{R}$ be a viscosity solution of $\left(\frac{\partial}{\partial t} u_d \right)(t,x) + G\left(t, x, u(t,x), (\nabla_x u_D)(t,x), (\mathrm{Hess}_x\, u_d)(t,x)\right) \geqslant 0$ for $(t,x) \in (0,T) \times \mathcal{O}$, and assume that $u - \phi$ has a local maximum at $(\mathfrak{t}, \mathfrak{r}) \in (0,T) \times \mathcal{O}$, then:*

$$\left(\frac{\partial}{\partial t} \phi \right)(\mathfrak{t}, \mathfrak{r}) + G\left(\mathfrak{t}, \mathfrak{r}, u(\mathfrak{t}, \mathfrak{r}), (\nabla_x \phi)(\mathfrak{t}, \mathfrak{r}), (\mathrm{Hess}_x\, \phi)(\mathfrak{t}, \mathfrak{r})\right) \geqslant 0 \qquad (3.2.11)$$

Proof. That u is upper semi-continuous ensures that there exists as a neighborhood U around $(\mathfrak{t}, \mathfrak{r})$ and $\psi \in C^{1,2}((0,T) \times \mathcal{O}, \mathbb{R})$ where:

(i) for all $(t,x) \in (0,T) \times \mathcal{O}$ that $u(\mathfrak{t}, \mathfrak{r}) - \psi(\mathfrak{t}, \mathfrak{r}) \geqslant u(t,x) - \psi(t,x)$

(ii) for all $(t,x) \in U$ that $\phi(t,x) = \phi(t,x)$.

We therefore obtain that:

$$\left(\frac{\partial}{\partial t}\phi\right)(\mathfrak{t},\mathfrak{r}) + G\left(\mathfrak{t},\mathfrak{r},u(\mathfrak{t},\mathfrak{r}),(\nabla_x)(\mathfrak{t},\mathfrak{r}),(\mathrm{Hess}_x\,\phi)(\mathfrak{t},\mathfrak{r})\right)$$
$$= \left(\frac{\partial}{\partial t}\psi\right)(\mathfrak{t},\mathfrak{r}) + G\left(\mathfrak{t},\mathfrak{r},u(\mathfrak{t},\mathfrak{r}),(\nabla_x)(\mathfrak{t},\mathfrak{r}),(\mathrm{Hess}_x\,\psi)(\mathfrak{t},\mathfrak{r})\right) \geq 0 \qquad (3.2.12)$$

$\square$

Lemma 3.2.11. *Let* $d \in \mathbb{N}$, $T \in (0,\infty)$, *let* $\mathcal{O} \subseteq \mathbb{R}^d$ *be a non-empty open set, let* $u_n :$ $(0,T) \times \mathcal{O} \to \mathbb{R}$, $n \in \mathbb{N}_0$ *be functions, let* $G_n : (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$, $n \in \mathbb{N}$ *be degenerate elliptic, assume that* G_∞ *is upper semi-continuous for all non-empty compact* $\mathcal{K} \subseteq (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$ *that:*

$$\limsup_{n\to\infty}\left[\sup_{(t,x,r,p,A)\in\mathcal{K}}\left(|u_n(t,x) - u_0(t,x)| + |G_n(t,x,r,p,A) - G_0(t,x,r,p,A)|\right)\right] = 0 \quad (3.2.13)$$

and assume for all $n \in \mathbb{N}$ *that* u_n *is a viscosity solution of:*

$$\left(\frac{\partial}{\partial t}u_n\right)(t,x) + G_n\left(t,x,u_n(t,x),(\nabla_x u_n)(t,x),(\mathrm{Hess}_x\,u_n)(t,x)\right) \geq 0 \qquad (3.2.14)$$

then u_0 *is a viscosity solution of:*

$$\left(\frac{\partial}{\partial t}u_0\right)(t,x) + G_n\left(t,x,u_0(t,x),(\nabla_x u_0)(t,x),(\mathrm{Hess}_x\,u_0)(t,x)\right) \geq 0 \qquad (3.2.15)$$

Proof. Let $(t_0,x_0) \in (0,T) \times \mathcal{O}$. Let $\phi_\epsilon \in C^{1,2}((0,T) \times \mathcal{O},\mathbb{R})$ satisfy for all $\epsilon \in (0,\infty)$, $s \in (0,T)$, $y \in \mathcal{O}$ that $\phi_0(t_0,x_0) = u_0(t_0,x_0)$, $\phi_0(t_0,x_0) \geq u_0(t_0,x_0)$, and:

$$\phi_\varepsilon(s,y) = \phi_o(s,y) + \varepsilon\left(|s - t_0| + \|y - x_0\|_E\right) \qquad (3.2.16)$$

Let $\delta \in (0,\infty)$ be such that $\{(s,y) \in \mathbb{R}^d \times \mathbb{R} : \max\left(|s - t_0|^2, \|y - x_0\|_E^2\right) \leq \delta\}$. Note that this and (3.2.27) then imply for all $\varepsilon \in (0,\infty)$ there exists an $\nu_\varepsilon \in \mathbb{N}$ such that for all $n \geq \nu_\varepsilon$,

45

and $\max\left(|s - t_0|, \|y - x_0\|_E\right) \leqslant \delta$, it is the case that:

$$\sup\left(|u_n(s, y) - u_0(s, y)|\right) \leqslant \frac{\varepsilon\delta}{2} \tag{3.2.17}$$

Note that this combined with (3.2.16) tells us that for all $\varepsilon \in (0, \infty)$, $n \in \mathbb{N} \cap [\nu_\epsilon, \infty)$, $s \in (0, T)$, $y \in \mathcal{O}$, with $|s - t_0| < \delta$, $\|y - x_0\|_E \leqslant \delta$, $|s - t_0| + \|y - x_0\|_E > \delta$ that:

$$
\begin{aligned}
u_n(t_0, x_0) - \phi_\varepsilon(t_0, x_0) &= u_n(t_0, x_0) - \phi_0(t_0, x_0) \tag{3.2.18} \\
&= u_n(t_0, x_0) - u_0(t_0, x_0) \\
&\geqslant \frac{-\varepsilon\delta}{2} \\
&\geqslant u_n(s, y) - u_0(s, y) - \varepsilon\left(|s - t_0| + \|y - x_0\|_E\right) \\
&\geqslant u_n(s, y) - \phi_0(s, y) - \varepsilon\left(|s - t_0| + \|y - x_0\|_E\right) \\
&= u_n(s, y) - \phi_\varepsilon(s, y) \tag{3.2.19}
\end{aligned}
$$

Note that Corollary 3.2.3.1 implies that for all $\epsilon \in (0, \infty)$ and $n \in \mathbb{N}$ that $u_n - \phi_\varepsilon$ is upper semi-continuous. There therefore exists for all $\epsilon \in (0, \infty)$ and $n \in \mathbb{N}$, a $\tau_n^\varepsilon \in (t_0 - \delta, t_0 + \delta)$ and a ρ_n^ε, where $\|\rho_n^\varepsilon - x_0\| \leqslant \delta$ such that:

$$u_n(\tau_n^\varepsilon, \rho_n^\varepsilon) - \phi_\epsilon(\tau_n^\varepsilon, \rho_n^\varepsilon) \geqslant u_n(s, y) - \phi_\varepsilon(s, y) \tag{3.2.20}$$

By Lemma 3.2.10, it must be the case that for all $\varepsilon \in (0, \infty)$ and $n \in \mathbb{N} \cap [\nu_\varepsilon, \infty)$:

$$\left(\frac{\partial}{\partial t}\phi_\varepsilon\right)(\tau_n^\varepsilon, \rho_n^\varepsilon) + G_n\left(\tau_n^\varepsilon, \rho_n^\varepsilon, u_n\left(\tau_n^\varepsilon, \rho_n^\varepsilon\right), \left(\nabla_x\phi_\varepsilon\right)\left(\tau_n^\varepsilon, \rho_n^\varepsilon\right), \left(\text{Hess}_x\,\phi_\varepsilon\right)\left(\tau_n^\varepsilon, \rho_n^\varepsilon\right)\right) \geqslant 0 \quad (3.2.21)$$

Note however that (3.2.20) along with (3.2.16) and (3.2.27) yields that for all $\varepsilon \in (0, \infty)$

that:

$$\limsup_{n\to\infty} \left[u_n(\tau_n^\varepsilon, \rho_n^\varepsilon) - \phi_\epsilon(\tau_n^\varepsilon, \rho_n^\varepsilon) \right]$$

$$\geqslant \limsup_{n\to\infty} \left[u_n(\tau_n^\varepsilon, \rho_n^\varepsilon) - \left(\phi_0(\tau_n^\varepsilon, \rho_n^\varepsilon) + \varepsilon \left(|\tau_n^\varepsilon - t_0| + \|\rho_n^\varepsilon - x_0\|_E \right) \right) \right]$$

$$\geqslant \limsup_{n\to\infty} \left[u_n(\tau_n^\varepsilon, \rho_n^\varepsilon) - u_0(\tau_n^\varepsilon, \rho_n^\varepsilon) - \varepsilon \left(|\tau_n^\varepsilon - t_0| + \|\rho_n^\varepsilon - x_0\|_E \right) \right]$$

$$= \limsup_{n\to\infty} \left[-\varepsilon \left(|\tau_n^\varepsilon - t_0| + \|\rho_n^\varepsilon - x_0\|_E \right) \right] \leqslant 0 \qquad (3.2.22)$$

However note also that since G_0 is upper semi-continuous, further the fact that, $\phi_0 \in ((0,T) \times \mathcal{O}, \mathbb{R})$, and then (3.2.27), and (3.2.16), imply for all $\varepsilon \in (0,\infty)$ we have that: $\limsup_{n\to\infty} \left| \left(\frac{\partial}{\partial t} \phi_\varepsilon \right) (\tau_n^\varepsilon, \rho_n^\varepsilon) - \left(\frac{\partial}{\partial t} \phi_0 \right) (t_0, x_0) \right| = 0$ and:

$$G_0\left(t_0, x_0, \phi_0(t_0, x_0), (\nabla_x \phi_0)(t_0, x_0), (\mathrm{Hess}_x \phi_0)(t_0, x_0) + \mathrm{Id}_{\mathbb{R}^d}\right)$$

$$= G_0\left(t_0, x_0, u_0(t_0, x_0), (\nabla_x \phi_\varepsilon)(t_0, x_0), (\mathrm{Hess}_x \phi_\varepsilon)(t_0, x_0)\right)$$

$$\geqslant \limsup_{n\to\infty} \left[G_0\left(\tau_n^\varepsilon, \rho_n^\varepsilon, u_n(\tau_n^\varepsilon, \rho_n^\varepsilon), (\nabla_x \phi_\varepsilon)(\tau_n^\varepsilon, \rho_n^\varepsilon), (\mathrm{Hess}_x \phi_\varepsilon)(\tau_n^\varepsilon, \rho_n^\varepsilon)\right) \right] \qquad (3.2.23)$$

$$\geqslant \limsup_{n\to\infty} \left[G_n\left(\tau_n^\varepsilon, \rho_n^\varepsilon, u_n(\tau_n^\varepsilon, \rho_n^\varepsilon), (\nabla_x \phi_\varepsilon)(\tau_n^\varepsilon, \rho_n^\varepsilon), (\mathrm{Hess}_x \phi_\varepsilon)(\tau_n^\varepsilon, \rho_n^\varepsilon)\right) \right] \qquad (3.2.24)$$

This with (3.2.20) assures for all $\epsilon \in (0,\infty)$ that:

$$\left(\frac{\partial}{\partial t} \phi_0 \right)(t_0, x_0) + G_0\left(t_0, x_0, \phi_0(t_0, x_0), (\nabla_x \phi_0)(t_0, x_0), (\mathrm{Hess}_x \phi_0)(t_0, x_0) + \varepsilon \mathrm{Id}_{\mathbb{R}^d}\right) \geqslant 0$$

$$(3.2.25)$$

That G_0 is upper semi-continuous then yields that:

$$\left(\frac{\partial}{\partial t} \phi_0 \right)(t_0, x_0) + G_0\left(t_0, x_0, \phi_0(t_0, x_0), (\nabla_x \phi_0)(t_0, x_0), (\mathrm{Hess}_x \phi_0)(t_0, x_0) + \varepsilon \mathrm{Id}_{\mathbb{R}^d}\right) \geqslant 0$$

$$(3.2.26)$$

This establishes (3.2.29) which establishes the lemma.

$\square$

Corollary 3.2.11.1. *Let $d \in \mathbb{N}$, $T \in (0, \infty)$, let $\mathcal{O} \subseteq \mathbb{R}^d$ be a non-empty open set, let $u_n : (0, T) \times \mathcal{O} \to \mathbb{R}$, $n \in \mathbb{N}_0$ be functions, let $G_n : (0, T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$, $n \in \mathbb{N}_0$ be degenerate elliptic, assume that G_0 is lower semi-continuous for all non-empty compact $\mathcal{K} \subseteq (0, T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$ that:*

$$\limsup_{n \to \infty} \left[\sup_{(t,x,r,p,A) \in \mathcal{K}} \left(|u_n(t, x) - u_0(t, x)| + |G_n(t, x, r, p, A) - G_0(t, x, r, p, A)| \right) \right] = 0 \quad (3.2.27)$$

and assume for all $n \in \mathbb{N}$ that u_n is a viscosity solution of:

$$\left(\frac{\partial}{\partial t} u_n \right)(t, x) + G_n\left(t, x, u_n(t, x), (\nabla_x u_n)(t, x), (\operatorname{Hess}_x u_n)(t, x)\right) \leqslant 0 \quad (3.2.28)$$

then u_0 is a viscosity solution of:

$$\left(\frac{\partial}{\partial t} u_0 \right)(t, x) + G_n\left(t, x, u_0(t, x), (\nabla_x u_0)(t, x), (\operatorname{Hess}_x u_0)(t, x)\right) \leqslant 0 \quad (3.2.29)$$

Proof. Let $v_n : (0, T) \times \mathcal{O} \to \mathbb{R}$, $n \in \mathbb{N}_0$ and $H_n : (0, T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ satisfy for all $n \in \mathbb{N}_0$, $t \in (0, T)$, $x \in \mathcal{O}$, $r \in \mathbb{R}$, $p \in \mathbb{R}^d$, $A \in \mathbb{S}_d$ that $v_n(t, x) = -u_n(t, x)$ and that $H_n(t, x) = -G_n(t, x, -r, -p, -A)$.

Note that Corollary 3.2.3.2 gives us that H_0 is upper semi-continuous. Note also that since it is the case that for all $n \in \mathbb{N}_0$, G_n is degenerate elliptic then it is also the case by Lemma 3.2.6 that H_n is degenerate elliptic for all $n \in \mathbb{N}_0$. These together with (3.2.28) ensure that for all $n \in \mathbb{N}$, v_n is a viscosity solution of:

$$\left(\frac{\partial}{\partial t} v_n \right)(t, x) + H_n\left(t, x, v_n(t, x), (\nabla_x v_n)(t, x), (\operatorname{Hess}_x v_n)(t, x)\right) \geqslant 0 \quad (3.2.30)$$

This together with (3.2.27) establish that:

$$\limsup_{n\to\infty}\left[\sup_{(t,x,r,p,A)\in\mathcal{K}}(|u_n(t,x)-u_0(t,x)|+|H_n(t,x,r,p,A)-H_0(t,x,r,p,A)|)\right]=0 \quad (3.2.31)$$

This (3.2.30) and the fact that H_0 is upper semi-continuous then establish that:

$$\left(\frac{\partial}{\partial t}v_0\right)(t,x)+H_0\left(t,x,v_0(t,x),(\nabla_x v_0)(t,x),(\text{Hess}_x v_0)(t,x)\right)\geqslant 0 \quad (3.2.32)$$

for $(t,x)\in(0,T)\times\mathcal{O}$. Hence v_0 is a viscosity solution of:

$$\left(\frac{\partial}{\partial t}u_0\right)(t,x)+H_0\left(t,x,u_0(t,x),(\nabla_x u_0)(t,x),(\text{Hess}_x u_0)(t,x)\right)\leqslant 0 \quad (3.2.33)$$

This completes the proof. $\square$

Corollary 3.2.11.2. *Let $d\in\mathbb{N}$, $T\in(0,\infty)$, let $\mathcal{O}\subseteq\mathbb{R}^d$ be a non-empty set, let $u_n:$ $(0,T)\times\mathcal{O}\to\mathbb{R}$, $n\in\mathbb{N}_0$, be functions, let $G_n:(0,T)\times\mathcal{O}\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}_d\to\mathbb{R}$, $n\in\mathbb{N}_0$ be degenerate elliptic, assume also that $G_0:(0,T)\times\mathcal{O}\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}_d\to\mathbb{R}$ be consinuous and assume for all non-empty compact $\mathcal{K}\subseteq(0,T)\times\mathcal{O}\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{S}_d$ it is the case that:*

$$\limsup_{n\to\infty}\left[\sup_{(t,x,r,p,A)\in\mathcal{K}}(|G_n(t,x,r,p,A)-G_0(t,x,r,p,A)|+|u_n(t,x)-u_0(t,x)|)\right]=0$$

$$(3.2.34)$$

and further assume for all $n\in\mathbb{N}$, that u_n is a viscosity solution of:

$$\left(\frac{\partial}{\partial t}u_n\right)(t,x)+G_n\left(t,x,u_n(t,x),(\nabla_x u_n)(t,x),(\text{Hess}_x u_n)(t,x)\right)=0 \quad (3.2.35)$$

for $(t,x)\in(0,T)\times\mathcal{O}$, then we have that u_0 is a viscosity solution of:

$$\left(\frac{\partial}{\partial t}u_0\right)(t,x)+G_0\left(t,x,u_0(t,x),(\nabla_x u_0)(t,x),(\text{Hess}_x u_0)(t,x)\right)=0 \quad (3.2.36)$$

Proof. Note that Lemma 3.2.11 gives us that u_0 is a viscosity solution of:

$$\left(\frac{\partial}{\partial t} u_0\right)(t, x) + G_n\left(t, x, u_0(t, x), (\nabla_x u_0)(t, x), (\text{Hess}_x u_0)(t, x)\right) \geqslant 0 \qquad (3.2.37)$$

for $(t, x) \in (0, T) \times \mathcal{O}$. Also note that Corollary 3.2.11.1 ensures that u_0 is a viscosity solution of:

$$\left(\frac{\partial}{\partial t} u_0\right)(t, x) + G_n\left(t, x, u_0(t, x), (\nabla_x u_0)(t, x), (\text{Hess}_x u_0)(t, x)\right) \leqslant 0 \qquad (3.2.38)$$

Taken together these prove the corollary. □

Lemma 3.2.12. *For all $a, b \in \mathbb{R}$ it is the case that $(a + b)^2 \leqslant 2a^2 + 2b^2$.*

Proof. Since for all $a, b \in \mathbb{R}$ it is the case that $(a - b)^2 \geqslant 0$ we then have that:

$$\begin{aligned}
(a + b)^2 &\leqslant (a + b)^2 + (a - b)^2 \\
&\leqslant a^2 + 2ab + b^2 + a^2 - 2ab + b^2 \\
&= 2a^2 + 2b^2
\end{aligned}$$

This completes the proof. □

Lemma 3.2.13. *Let $d, m \in \mathbb{N}$, $T \in (0, \infty)$. Let $\mathcal{O} \subseteq \mathbb{R}^d$ be a non-empty compact set, and for all $n \in \mathbb{N}_0$, $\mu_n \in C([0, T] \times \mathcal{O}, \mathbb{R})$, $\sigma_n \in C([0, T] \times \mathcal{O}, \mathbb{R}^{d \times m})$ assume also:*

$$\limsup_{n \to \infty} \left[\sup_{t \in [0, T]} \sup_{x \in \mathcal{O}} \left(\| \mu_n(t, x) - \mu_0(t, x) \|_E + \| \sigma_n(t, x) - \sigma_0(t, x) \|_F \right) \right] = 0 \qquad (3.2.39)$$

Let $(\Omega, \mathcal{F}, \mathbb{R})$ be a stochastic basis and let $W : [0, T] \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t \in [0, T]}$-Brownian motion for every $t \in [0, T]$, $x \in \mathcal{O}$, let $\mathcal{X}^{t, x} = (\mathcal{X}^{t, x}_s)_{s \in [t, T]} : [t, T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s \in [t, T]}$ adapted stochastic process with continuous sample paths, satisfying for

all $s \in [t, T]$ *we have* $\mathbb{P}$-*a.s.*

$$\mathcal{X}_s^{n,t,x} = x + \int_t^s \mu_n(r, \mathcal{X}_s^{n,t,x})dr + \int_t^s \sigma_n(r, \mathcal{X}_r^{n,t,x})dW_r \tag{3.2.40}$$

then it is the case that:

$$\limsup_{n \to \infty} \left[\sup_{t \in [0,T]} \sup_{s \in [t,T]} \sup_{x \in \mathcal{O}} \left(\mathbb{E} \left[\left\| \mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x} \right\|_E^2 \right] \right) \right] = 0 \tag{3.2.41}$$

for $(t, x) \in (0, T) \times \mathbb{R}^d$.

Proof. Since $\mathcal{O}$ is compact, let $L \in \mathbb{R}$ be such that for all $t \in [0, T]$, $x, y \in \mathcal{O}$ it is the case that:

$$\|\mu_0(t, x) - \mu_0(t, y)\|_E - \|\sigma_0(t, x) + \sigma_0(t, y)\|_F \leqslant L\|x - y\|_E \tag{3.2.42}$$

Furthermore (Karatzas and Shreve, 1991, Theorem 5.2.9) tells us that:

$$\sup_{s \in [t,T]} \mathbb{E} \left[\|\mathcal{X}_s^{n,t,x}\|_E \right] < \infty \tag{3.2.43}$$

Note now that (3.2.40) tells us that:

$$\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x} = \int_t^s \mu_n(r, \mathcal{X}_s^{n,t,x}) - \mu_0(r, \mathcal{X}_s^{0,t,x})dr + \int_t^s \sigma_n(r, \mathcal{X}_r^{n,t,x}) - \sigma_0(r, \mathcal{X}_r^{0,t,x})dW_r \tag{3.2.44}$$

Minkowski's Inequality applied to (3.2.44) then tells us for all $n \in \mathbb{N}$, $t \in [0, T]$, $s \in [t, T]$, and $x \in \mathcal{O}$ that:

$$\left(\mathbb{E} \left[\left\| \mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x} \right\|_E \right] \right)^{\frac{1}{2}} \leqslant \int_t^s \left(\mathbb{E} \left[\left\| \mu_n(r, \mathcal{X}_r^{n,t,x}) - \mu_0(r, \mathcal{X}_r^{0,t,x}) \right\|_E^2 \right] \right)^{\frac{1}{2}} dr +$$
$$\left(\mathbb{E} \left[\left\| \int_t^s (\sigma_n(r, \mathcal{X}_r^{n,t,x}) - \sigma_0(r, \mathcal{X}_r^{0,t,x}))dW_r \right\|_E^2 \right] \right)^{\frac{1}{2}} \tag{3.2.45}$$

Itô's isometry applied to the second summand yields:

$$\left(\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E\right]\right)^{\frac{1}{2}} \leqslant \int_t^s \left(\mathbb{E}\left[\left\|\mu_n(r,\mathcal{X}_r^{n,t,x}) - \mu_0(r,\mathcal{X}_r^{0,t,x})\right\|_E^2\right]\right)^{\frac{1}{2}} dr +$$

$$\left(\int_t^s \mathbb{E}\left[\left\|\sigma_n(r,\mathcal{X}_r^{n,t,x}) - \sigma_0(r,\mathcal{X}_r^{0,t,x})\right\|_F^2\right] dr\right)^{\frac{1}{2}} \qquad (3.2.46)$$

Applying Lemma 3.2.12 followed by the Cauchy-Schwarz Inequality then gives us for all $n \in \mathbb{N}$, $t \in [0,T]$, $s \in [t,T]$, and $x \in \mathcal{O}$ that:

$$\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{n,t,x}\right\|_E^2\right] \leqslant 2\left[\int_t^s \left(\mathbb{E}\left[\left\|\mu_n(r,\mathcal{X}_r^{n,t,x}) - \mu_0(r,\mathcal{X}_r^{0,t,x})\right\|_E^2\right]\right)^{\frac{1}{2}} dr\right]^2$$

$$+ 2\int_t^s \mathbb{E}\left[\left\|\sigma_n(r,\mathcal{X}_r^{nt,x}) - \sigma_0(r,\mathcal{X}_r^{0,t,x})\right\|_F^2\right] dr$$

$$\leqslant 2T\int_t^s \mathbb{E}\left[\left\|\mu_n(r,\mathcal{X}_r^{n,t,x}) - \mu_0(r,\mathcal{X}_r^{0,t,x})\right\|_E^2\right] dr$$

$$+ 2\int_t^s \mathbb{E}\left[\left\|\sigma_n(r,\mathcal{X}_r^{n,t,x}) - \sigma_0(r,\mathcal{X}_r^{0,t,x})\right\|_F^2\right] dr \qquad (3.2.47)$$

Applying Lemma 3.2.12 again to each summand then yields for all $n \in \mathbb{N}$, $t \in [0,T]$ $s \in [t,T]$, and $x \in \mathcal{O}$ it is the case that:

$$\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|^2\right]$$

$$\leqslant 2T\int_t^s \left(2\mathbb{E}\left[\left\|\mu_n(r,\mathcal{X}_r^{n,t,x}) - \mu_0(r,\mathcal{X}_r^{n,t,x})\right\|_E^2\right] + 2\mathbb{E}\left[\left\|\mu_0(r,\mathcal{X}_r^{n,t,x}) - \mu_0(r,\mathcal{X}_r^{0,t,x})\right\|_E^2\right]\right) dr$$

$$+ 2\int_t^2 \left(2\mathbb{E}\left[\left\|\sigma_n(r,\mathcal{X}_r^{n,t,x}) - \sigma_0(r,\mathcal{X}_r^{n,t,x})\right\|_F^2\right] + 2\mathbb{E}\left[\left\|\sigma_0(r,\mathcal{X}_r^{n,t,x}) - \sigma_0(r,\mathcal{X}_r^{0,t,x})\right\|_F\right]\right) dr$$

$$(3.2.48)$$

However assumption (3.2.42) then gives us that for all $n \in \mathbb{N}$, $t \in [0,T]$, $s \in [t,T]$, and $x \in \mathcal{O}$

that:

$$\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E^2\right] \leqslant 4L^2(T+1)\int_t^s \mathbb{E}\left[\left\|\mathcal{X}_r^{n,t,x} - \mathcal{X}_r^{0,t,x}\right\|_E^2\right] dr$$

$$+ 4T(T+1)\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_0(r,y)\|_F^2\right)\right]$$

Finally Gronwall's Inequality with assumption (3.2.43) gives us for all $n \in \mathbb{N}$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathcal{O}$ that:

$$\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E^2\right]$$

$$\leqslant 4T(T+1)\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_0(r,y)\|_F^2\right)\right]e^{4L^2T(T+1)}$$

$$(3.2.49)$$

Applying $\limsup_{n\to\infty}$ to both sides and applying (3.2.39) gives us for all $n \in \mathbb{N}$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathcal{O}$ that:

$$\limsup_{n\to\infty}\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E^2\right] \leqslant$$

$$\limsup_{n\to\infty}\left[4T(T+1)\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_0(r,y)\|_F^2\right)\right]e^{4L^2T(T+1)}\right]$$

$$\leqslant$$

$$4T(T+1)\left[\limsup_{n\to\infty}\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_0(r,y)\|_F^2\right)\right]\right]e^{4L^2T(T+1)}$$

$$\leqslant 0$$

This completes the proof. $\qquad\qquad\square$

Lemma 3.2.14. *Let $d, m \in \mathbb{N}$, $T \in (0,\infty)$, let $\mathcal{O} \subseteq [0,T] \times \mathbb{R}^d$, let $\mu \in C([0,T] \times \mathcal{O}, \mathbb{R}^d)$ and $\sigma \in C([0,T] \times \mathcal{O}, \mathbb{R}^{d\times m})$ have compact supports such that $\mathrm{supp}(\mu) \cup \mathrm{supp}(\sigma) \subseteq [0,T] \times \mathcal{O}$ let $g \in C(\mathbb{R}^d, \mathbb{R})$. Let $\left(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]}\right)$ be a stochastic basis, let $W : [0,T] \times \Omega \to \mathbb{R}^m$ be*

a standard $(\mathbb{F}_t)_{t \in [0,T]}$ *Brownian motion, for every* $t \in [0,T]$, $x \in \mathbb{R}^d$, *let* $\mathcal{X}^{t,x} = (\mathcal{X}^{t,x}_s)_{s \in [t,T]}$: $[t,T] \times \Omega \to \mathbb{R}^d$ *be an* $(\mathbb{F}_s)_{s \in [t,T]}$ *adapted stochastic process with continuous sample paths satisfying for all* $s \in [t,T]$ *with* $\mathbb{F}$-*a.s. that:*

$$\mathcal{X}^{t,x}_s = x + \int_t^s \mu\left(r, \mathcal{X}^{t,x}_r\right) dr + \int_t^s \sigma\left(r, \mathcal{X}^{t,x}_r\right) dW_r \tag{3.2.50}$$

and further let $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ *satisfy for all* $t \in [0,T]$, $x \in \mathbb{R}^d$ *that:*

$$u(t,x) = \mathbb{E}\left[g\left(\mathcal{X}^{t,x}_T\right)\right] \tag{3.2.51}$$

Then u is a viscosity solution of:

$$\left(\frac{\partial}{\partial t}u\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*(\operatorname{Hess}_x u)(t,x)\right) + \langle\mu(t,x),(\nabla_x u)(t,x)\rangle = 0 \tag{3.2.52}$$

and where $u(T,x) = g(x)$ *for* $(t,x) \in (0,T) \times \mathcal{O}$.

Proof. Let $\mathcal{S} = \operatorname{supp}(\mu) \cup \operatorname{supp}(\sigma) \subseteq [0,T] \times \mathcal{O}$ be bounded in space by $\rho \in (0,\infty)$, as $\mathcal{S} \subseteq [0,T] \times (-\rho,\rho)^d$. This exists as the supports are compact and thus by Hiene-Börel is closed and bounded. Let $\mathfrak{s}_n, \mathfrak{m}_n \in C^\infty([0,T] \times \mathbb{R}^d, \mathbb{R}^{d \times n})$ where $\bigcup_{n \in \mathbb{N}} [\operatorname{supp}(\mathfrak{s}_n) \cup \operatorname{supp}(\mathfrak{m}_n)] \subseteq [0,T] \times (-\rho,\rho)^d$ satisfy for $n \in \mathbb{N}$ that:

$$\limsup_{n \to \infty}\left[\sup_{t \in [0,T]}\sup_{x \in \mathbb{R}}\left(\|\mathfrak{m}_n(t,x) - \mu(t,x)\|_E + \|\mathfrak{s}_n - \sigma(t,x)\|_F\right)\right] = 0 \tag{3.2.53}$$

We construct a set of degenerate elliptic functions, $G^n : (0,T) \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \to \mathbb{R}$ with

$n \in \mathbb{N}_0$ such that:

$$G^0(t, x, r, p, A) = \frac{1}{2} \text{Trace} \left(\sigma(t, x)[\sigma(t, x)]^* A \right) + \langle \mu(t, x), p \rangle \tag{3.2.54}$$

and

$$G^n(t, x, r, p, A) = \frac{1}{2} \text{Trace} \left(\mathfrak{s}_n(t, x)[\mathfrak{s}_n(t, x)]^* A \right) + \langle \mu(t, x), p \rangle \tag{3.2.55}$$

Also let $\mathfrak{g}_n \in C^\infty(\mathbb{R}^d, \mathbb{R})$ for $n \in \mathbb{N}$ satisfy for all $n \in \mathbb{N}$ that:

$$\limsup_{n \to \infty} \sup_{t \in [0,T]} \sup_{x \in \mathbb{R}^d} \left(\| \mathfrak{g}_n(x) - g(x) \|_E \right) = 0 \tag{3.2.56}$$

Further let $\mathfrak{X}^{n,t,x} = (\mathfrak{X}_s^{n,t,x})_{s \in [t,T]} : [t, T] \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s \in [t,T]}$-adapted stochastic process with continuous sample paths that satisfy:

$$\mathfrak{X}_s^{n,t,x} = x + \int_t^s \mathfrak{m}_n(r, \mathfrak{X}_r^{n,t,x}) dr + \int_t^s \mathfrak{s}_n(r, \mathfrak{X}_r^{n,t,x}) dW_r \tag{3.2.57}$$

Finally let $\mathfrak{u}^n : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ for $n \in \mathbb{N}$ be:

$$\mathfrak{u}^n = \mathbb{E} \left[\mathfrak{g}_n \left(\mathfrak{X}_T^{n,t,x} \right) \right] \tag{3.2.58}$$

and:

$$\mathfrak{u}^0 = \mathbb{E} \left[\mathfrak{g}_n \left(\mathfrak{X}_T^{t,x} \right) \right] \tag{3.2.59}$$

Note that (Beck et al., 2021b, Lemma 2.2) with $g \curvearrowleft \mathfrak{g}_k$, $\mu \curvearrowleft \mathfrak{m}_n$, $\sigma \curvearrowleft \mathfrak{s}_n$, $\mathfrak{X}^{t,x} \curvearrowleft \mathfrak{X}^{n,t,x}$ gives us $\mathfrak{u}^n \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, and $\mathfrak{u}^n(t, x) = \mathfrak{g}_k(x)$ where:

$$\left(\frac{\partial}{\partial t} \mathfrak{u}^n \right)(t, x) + \frac{1}{2} \text{Trace} \left(\mathfrak{s}_n(t, x) [\mathfrak{s}_n(t, x)]^* (\text{Hess}_x \mathfrak{u}^n)(t, x) \right) + \langle \mathfrak{m}_n(t, x), (\nabla_x \mathfrak{u}^n)(t, x) \rangle = 0 \tag{3.2.60}$$

And by Definitions 3.2.7, 3.2.8, and 3.2.9 we have that $\mathfrak{u}^n$ is a viscosity solution of

$$\left(\frac{\partial}{\partial t}\mathfrak{u}^n\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\mathfrak{s}_n(t,x)\left[\mathfrak{s}_n(t,x)\right]^*(\operatorname{Hess}_x\mathfrak{u}^n)(t,x)\right) + \langle\mathfrak{m}_n(t,x),(\nabla_x\mathfrak{u}^n)(t,x)\rangle = 0$$

$$(3.2.61)$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$.

Since for all $n \in \mathbb{N}$, it is the case that $\mathcal{S} = (\operatorname{supp}(\mathfrak{m}_n) \cup \operatorname{supp}(\mathfrak{s}_n) \cup \operatorname{supp}(\mu) \cup \operatorname{supp}(\sigma)) \subseteq [0,T] \times (-\rho,\rho)^d$ and because of (3.2.50) of (3.2.57) we have that (Beck et al., 2021a, Lemma 3.2, Item (ii)) which yields that for all $n \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d \setminus (-\rho,\rho)^d$ that $\mathbb{P}(\forall s \in [t,T] : \mathfrak{X}_s^{n,t,x} = x = \mathcal{X}_s^{t,x}) = 1$. This in turn shows that for all $n \in \mathbb{N}$, $x \in \mathbb{R}^d \setminus (-\rho,\rho)^d$ that $\mathfrak{u}^n(t,x) = \mathfrak{u}^0(t,x)$ which along with (3.2.58) and (3.2.59) yields that:

$$\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\left[\left|\mathfrak{u}^n(t,x)-\mathfrak{u}^0(t,x)\right|\right] = \sup_{t\in[0,T]}\sup_{x\in(-\rho,\rho)^d}\left[\left|\mathfrak{u}^n(t,x)-\mathfrak{u}^0(t,x)\right|\right]$$

$$\leqslant \sup_{t\in[0,T]}\sup_{x\in(-\rho,\rho)^d}\left(\mathbb{E}\left[\left|\mathfrak{g}_k\left(\mathfrak{X}_T^{n,t,x}\right)-\mathfrak{g}\left(\mathcal{X}_T^{t,x}\right)\right|\right]\right) \quad (3.2.62)$$

Note that Lemma 3.2.13 allows us to conclude that:

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in(-\rho,\rho)^d}\left(\mathbb{E}\left[\left\|\mathfrak{X}_T^{n,t,x}-\mathcal{X}_s^{t,x}\right\|\right]\right)\right] = 0 \quad (3.2.63)$$

But then we have that (3.2.62) which yields that:

$$\limsup_{n\to 0}\left[\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\left(\left|\mathfrak{u}^n(t,x)-\mathfrak{u}^0(t,x)\right|\right)\right] = 0 \quad (3.2.64)$$

However now note that (3.2.55) and (3.2.61) thus yield that for $n \in \mathbb{N}_0$, $\mathfrak{u}^n$ is a viscosity

solution to:

$$\left(\frac{\partial}{\partial t}\mathfrak{u}^n\right)(t,x) + G^n\left(t,x,\mathfrak{u}^n\left(t,x\right),\left(\nabla_x\mathfrak{u}^n\right)\left(t,x\right),\left(\mathrm{Hess}_x\mathfrak{u}^n\right)\left(t,x\right)\right) = 0 \tag{3.2.65}$$

But since we've established (3.2.53) we have that for a non-empty compact set $\mathcal{C} \subseteq (0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$ that:

$$\begin{aligned}
&\limsup_{n\to\infty}\left[\sup_{(t,x,r,p,A)\in\mathcal{C}}\left|G^n\left(t,x,r,p,A\right) - G^0\left(t,x,r,p,A\right)\right|\right]\\
&\leqslant \limsup_{n\to\infty}\left[\sup_{(t,x,r,p,A)\in\mathcal{C}}\|\mu(t,x) - \mathfrak{m}_n(t,x)\|_E\|p\|_E\right]\\
&+ \limsup_{n\to\infty}\left[\sup_{(t,x,r,p,A)\in\mathcal{C}}\|\sigma(t,x)\left[\sigma(t,x)\right]^* - \mathfrak{s}_n(t,x)\left[\mathfrak{s}_n(t,x)\right]^*\|_F\|A\|_F\right] = 0
\end{aligned} \tag{3.2.66}$$

This, together with (3.2.64), (3.2.65) and Corollary 3.2.11.2 yields that $\mathfrak{u}^0$ is also a viscosity solution to:

$$\left(\frac{\partial}{\partial t}\mathfrak{u}^0\right)(t,x) + G^0\left(t,x,\mathfrak{u}^0(t,x),\left(\nabla_x\mathfrak{u}^0\right)(t,x),\left(\mathrm{Hess}_x\right)(t,x)\right) = 0 \tag{3.2.67}$$

Finally note that (3.2.53), (3.2.57), (3.2.59), and (3.2.67) yield that u is a viscosity solution of::

$$\left(\frac{\partial}{\partial t}u\right)(t,x) + \frac{1}{2}\mathrm{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*\left(\mathrm{Hess}_x u\right)(t,x)\right) + \langle\mu(t,x),\left(\nabla_x\right)(t,x)\rangle = 0 \tag{3.2.68}$$

for $(t,x) \in [0,T] \times \mathbb{R}^d$. Finally (3.2.50) and (3.2.51) allows us to conclude that for all $x \in \mathbb{R}^d$ it is the case that $u(T,x) = g(x)$. This concludes the proof. $\qquad\square$

Lemma 3.2.15. *Let $d,m \in \mathbb{N}$, $T \in (0,\infty)$, further let $\mathcal{O} \subseteq \mathbb{R}^d$ be a non, empty compact set. Let every $r \in (0,\infty)$ satisfy the condition that $O_r \subseteq \mathcal{O}$, where $O_r = \{x \in \mathcal{O} : \left(\|x\|_E \leqslant r \text{ and } \{y \in \mathbb{R}^d : \|y - x\|_E < \frac{1}{r}\} \subseteq \mathcal{O}\right)\}$ let $g \in C(\mathcal{O},\mathbb{R})$, $\mu \in C([0,T] \times \mathcal{O},\mathbb{R})$, $V \in$*

$C^{1,2}([0,T] \times \mathcal{O}, (0,\infty))$, assume that for all $t \in [0,T]$, $x \in \mathcal{O}$ that:

$$\sup\left(\left\{\frac{\|\mu(t,x) - \mu(t,y)\|_E + \|\sigma(t,x) - \sigma(t,y)\|_F}{\|x-y\|_E} : t \in [0,T], x, y \in O_r, x \neq y\right\} \cup \{0\}\right) < \infty$$

(3.2.69)

$$\left(\frac{\partial}{\partial t}V\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*(\operatorname{Hess}_x V)(t,x)\right) + \langle\mu(t,x), (\nabla_x V)(t,x)\rangle \leqslant 0$$

(3.2.70)

assume that $\sup_{r \in (0,\infty)}\left[\inf_{x \in \mathcal{O}\setminus O_r} V(t,x)\right] = \infty$ and $\inf_{r \in (0,\infty)}\left[\sup_{t \in [0,T]}\sup_{x \in \mathcal{O}\setminus O_r}\left(\frac{g(x)}{V(T,x)}\right)\right] = 0$. Let $\left(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0,T]}\right)$ be a stochastic basis and let $W : [0,T] \times \Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t \in [0,T]}$-Brownian motion, for every $t \in [0,T]$, $x \in \mathcal{O}$ let $\mathcal{X}^{t,x} = (\mathcal{X}_s^{t,x})_{s \in [t,T]} : [t,T] \times \Omega \to \mathcal{O}$ be an $(\mathbb{F}_s)_{s \in [t,T]}$-adapted stochastic process with continuous sample paths satisfying that for all $s \in [t,T]$, we have $\mathbb{P}$-a.s. that:

$$\mathcal{X}_s^{t,x} = x + \int_t^s \mu(r, \mathcal{X}_r^{t,x})dr + \int_t^s \sigma(r, \mathcal{X}_n^{t,x})dW_r$$

(3.2.71)

also let $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that:

$$u(t,x) = \mathbb{E}\left[u(T, \mathcal{X}_T^{t,x})\right]$$

(3.2.72)

It is then the case that u is a viscosity solution to:

$$\left(\frac{\partial}{\partial t}u\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*(\operatorname{Hess}_x u)(t,x)\right) + \langle\mu(t,x), (\nabla_x)(t,x)\rangle = 0 \quad (3.2.73)$$

for $(t,x) \in (0,T) \times \mathcal{O}$ with $u(T,x) = g(x)$.

Proof. Let it be the case, that throughout the proof, for $n \in \mathbb{N}$, we have that $\mathfrak{g}_n \in C(\mathbb{R}^d, \mathbb{R})$,

compactly supported and that $\left[\bigcup_{n\in\mathbb{N}}\operatorname{supp}(\mathfrak{g}_m)\right]\subseteq[0,T]\times\mathcal{O}$ and further that:

$$\limsup_{n\to\infty}\left[\sup_{t\in[0,T]}\sup_{x\in\mathcal{O}}\left(\frac{|\mathfrak{g}_n(x)-g(x)|}{V(T,x)}\right)\right]=0 \tag{3.2.74}$$

Let is also be the case that for $n\in\mathbb{N}$, $\mathfrak{m}_n\in C([0,T]\times\mathbb{R}^d,\mathbb{R}^d)$ and $\mathfrak{s}_n\in C([0,T]\times\mathbb{R}^d,\mathbb{R}^{d\times m})$ satisfy:

(i) for all $n\in\mathbb{N}$:

$$\sup_{t\in[0,T]}\sup_{x,y\in\mathbb{R}^d,x\neq y}\left[\frac{\|\mathfrak{m}_n(t,y)-\mathfrak{m}_n(t,y)\|_E+\|\mathfrak{s}_n(t,x)-\mathfrak{s}_n(t,y)\|_E}{\|x-y\|_E}\right]=0 \tag{3.2.75}$$

(ii) for all all $n\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathcal{O}$:

$$\mathbb{1}_{\{V\leqslant n\}}(t,x)\left[\|\mathfrak{m}_n(t,x)-\mu(t,x)\|_E+\|\mathfrak{s}_n(t,x)-\sigma(t,x)\|_F\right]=0 \tag{3.2.76}$$

and

(iii) for all $n\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d\setminus\{V\leqslant n+1\}$ that:

$$\|\mathfrak{m}_n(t,x)\|_E+\|\mathfrak{s}_n(t,x)\|_F=0 \tag{3.2.77}$$

Next for every $n\in\mathbb{N}$, $t\in[0,T]$ and $x\in\mathbb{R}^d$ let it be the case that $\mathfrak{X}_s^{n,t,x}=(\mathfrak{X}_s^{n,t,x})_{s\in[t,T]}$: $[t,t]\times\Omega.\to\mathbb{R}^d$ be a stochastic process with continuous sample paths satisfying:

$$\mathfrak{X}_s^{n,t,x}=x+\int_t^s\mathfrak{m}_n(r,\mathfrak{X}_s^{n,t,x})dr+\int_t^s\mathfrak{s}_n(r,\mathfrak{X}_s^{n,t,x})dW_r \tag{3.2.78}$$

Let $\mathfrak{u}^n:[0,T]\times\mathbb{R}^d\to\mathbb{R}$, $k\in\mathbb{N}$, $n\in\mathbb{N}_0$, satisfy for all $n\in\mathbb{N}$, $t\in[0,T]$, $x\in\mathbb{R}^d$ that:

$$\mathfrak{u}^{n,k}(t,x)=\mathbb{E}\left[\mathfrak{g}_k(\mathfrak{X}_T^{n,t,x})\right] \tag{3.2.79}$$

and

$$\mathfrak{u}^{0,k}(t,x) = \mathbb{E}\left[\mathfrak{g}_k\left(\mathcal{X}_T^{t,x}\right)\right] \tag{3.2.80}$$

and finally let, for every $n \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathcal{O}$, there be $\mathfrak{t}_n^{t,x} : \Omega \to [t,T]$ which satisfy $\mathfrak{t}_n^{t,x} = \inf\left(\{s \in [t,T], \max\{V(s,\mathfrak{X}_s^{t,x}), V(s,\mathcal{X}_s^{t,x})\} \geqslant n\} \cup \{T\}\right)$. We may apply Lemma 3.2.14 with $\mu \curvearrowleft \mathfrak{m}_n$, $\sigma \curvearrowleft \mathfrak{s}_n$, $g \curvearrowleft \mathfrak{g}_k$ to show that for all $n,k \in \mathbb{N}$ we have that $\mathfrak{u}^{n,k}$ is a viscosity solution to:

$$\left(\frac{\partial}{\partial t}\mathfrak{u}^{n,k}\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\mathfrak{s}_n(t,x)\left[\mathfrak{s}_n(t,x)\right]^*\left(\operatorname{Hess}_x\mathfrak{u}^{n,k}\right)(t,x)\right) + \left\langle\mathfrak{m}_n(t,x),\left(\nabla_x(\mathfrak{u}^{n,k})\right)(t,x)\right\rangle$$

$$= 0$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$. But note that items (i)-(iii) and 3.2.78 give us that, in line with (Beck et al., 2021a, Lemma 3.5):

$$\mathbb{P}\left(\forall s \in [t,T] : \mathbb{1}_{\{s \leqslant \mathfrak{t}_n^{t,x}\}}\mathfrak{X}_s^{n,t,x} = \mathbb{1}_{\{s \leqslant \mathfrak{t}_n^{t,x}\}}\mathcal{X}_s^{t,x}\right) = 1 \tag{3.2.81}$$

Further this implies that for all $n,k \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathcal{O}$ that:

$$\mathbb{E}\left[\left|\mathfrak{g}_k\left(\mathfrak{X}_T^{n,t,x}\right) - \mathfrak{g}_k(\mathcal{X}_T^{t,x})\right|\right] = \mathbb{E}\left[\mathbb{1}_{\{\mathfrak{t}_n^{t,x}<T\}}\left|\mathfrak{g}_k(\mathfrak{X}_T^{n,t,x}) - \mathfrak{g}_k(\mathcal{X}_T^{t,x})\right|\right]$$

$$\leqslant 2\left[\sup_{y\in\mathcal{O}}|\mathfrak{g}_k(y)|\right]\mathbb{P}\left(\mathfrak{t}_n^{t,x} < T\right)$$

Note that this combined with (Beck et al., 2021a, Lemma 3.1) implies for all $t \in [0,T]$,

$x \in \mathcal{O}$, $n \in \mathbb{N}$ we have that $\mathbb{E}\left[V\left(\mathfrak{t}_n^{t,x}, \mathcal{X}_{\mathfrak{t}_n^{t,x}}^{t,x}\right)\right] \leqslant V(t,x)$, which then further proves that:

$$
\begin{aligned}
\left|\mathfrak{u}^{n,k}(t,x) - \mathfrak{u}^{0,k}(t,x)\right| &\leqslant 2\left[\sup_{y \in \mathcal{O}} |\mathfrak{g}_k(y)|\right] \mathbb{P}\left(\mathfrak{t}_n^{t,x} < T\right) \\
&\leqslant 2\left[\sup_{y \in \mathcal{O}} |\mathfrak{g}_k(y)|\right] \mathbb{P}\left(V\left(\mathfrak{t}_n^{t,x}, \mathcal{X}_{\mathfrak{t}_n^{t,x}}^{t,x}\right) \geqslant n\right) \\
&\leqslant \frac{2}{n}\left[\sup_{y \in \mathcal{O}} |\mathfrak{g}_k(y)|\right] \mathbb{E}\left[V\left(\mathfrak{t}_n^{t,x}, \mathcal{X}_{\mathfrak{t}_n^{t,x}}^{t,x}\right)\right] \\
&\leqslant \frac{2}{n}\left[\sup_{y \in \mathcal{O}} |\mathfrak{g}_k(y)|\right] V(t,x,)
\end{aligned}
$$

Together these imply that for all $k \in \mathbb{N}$ and compact $\mathcal{K} \subseteq [0,T] \times \mathcal{O}$:

$$
\limsup_{k \to \infty}\left[\sup_{(t,x) \in \mathcal{K}}\left(\left|\mathfrak{u}^{n,k}(t,x) - \mathfrak{u}^{0,k}(t,x)\right|\right)\right] = 0 \tag{3.2.82}
$$

But again note that since have that $\sup_{r \in (0,\infty)}\left[\inf_{t \in [0,T], x \in \mathbb{R}^d \setminus O_r} V(t,x)\right] = \infty$ and (3.2.76) tell us that for all compact $\mathcal{K} \subseteq [0,T] \times \mathcal{O}$ we have that:

$$
\limsup_{n \to \infty}\left[\sup_{(t,x) \in \mathcal{K}}\left(\|\mathfrak{m}_n(t,x) - \mu(t,x)\|_E + \|\mathfrak{s}_n(t,x) - \sigma(t,x)\|_F\right)\right] = 0 \tag{3.2.83}
$$

Note that (3.2.81), (3.2.82) and Corollary 3.2.11.2 tell us that for all $k \in \mathbb{N}$ we have that $\mathfrak{u}^{0,k}$ is a viscosity solution to:

$$
\left(\frac{\partial}{\partial t}\mathfrak{u}^{0,k}\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*\left(\operatorname{Hess}_x \mathfrak{u}^{0,k}\right)(t,x)\right) + \left\langle \mu(t,x), \left(\nabla_x \mathfrak{u}^{0,k}\right)(t,x)\right\rangle = 0 \tag{3.2.84}
$$

for $(t,x) \in (0,T) \times \mathcal{O}$. However note that (3.2.71),(3.2.74), (3.2.80) prove that for all compact $\mathcal{K} \subseteq [0,T] \times \mathcal{O}$ we have:

$$
\limsup_{k \to \infty}\left[\sup_{(t,x) \in \mathcal{K}}\left|\mathfrak{u}^{0,k}(t,x) - u(t,x)\right|\right] = 0 \tag{3.2.85}
$$

61

This together with (3.2.84), (3.2.74), Corollary 3.2.11.2 shows that u_0 is a viscosity solution to:

$$\left(\frac{\partial}{\partial t}u\right)(t,x) + \frac{1}{2}\text{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*(\text{Hess}_x u)(t,x)\right) + \langle\mu(t,x),(\nabla_x u)\rangle = 0 \quad (3.2.86)$$

for $(t,x) \in (0,T) \times \mathcal{O}$. By (3.2.73) we are ensured that for all $x \in \mathbb{R}^d$ we have that $u(T,x) = g(x)$ which together with proves the proposition.

$\square$

3.3 Solutions, Characterization, and Computational Bounds

Theorem 3.3.1 (Existence and characterization of u_d). *Let $T \in (0,\infty)$. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. Let $\sigma_d \in C\left(\mathbb{R}^d,\mathbb{R}^{d\times d}\right)$ and $\mu_d \in C\left(\mathbb{R}^d,\mathbb{R}^d\right)$ for $d \in \mathbb{N}$, let $u_d \in C^{1,2}\left([0,T]\times\mathbb{R}^d,\mathbb{R}\right)$ satisfy for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that:*

$$\left(\frac{\partial}{\partial t}u_d\right)(t,x) + \frac{1}{2}\text{Trace}\left(\sigma_d(x)\left[\sigma_d(x)\right]^*(\text{Hess}_x u_d)(t,x)\right) + \langle\mu_d(x),(\nabla_x u_d)(t,x)\rangle = 0$$

$$(3.3.1)$$

let $\mathcal{W}^d : [0,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$ be a standard Brownian motions and let $\mathcal{X}^{d,t,x} : [t,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, $t \in [0,T]$, be a stochastic process with continuous sample paths satisfying for all $d \in \mathbb{N}$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$, we have $\mathbb{P}$-a.s. that:

$$\mathcal{X}^{d,t,x} = x + \int_s^t \mu_d\left(\mathcal{X}_r^{d,t,x}\right)dr + \int_s^t \sigma\left(\mathcal{X}_r^{d,t,x}\right)d\mathcal{W}_r^d \quad (3.3.2)$$

Then for all $d \in N$, $t \in [0,T]$, $x \in \mathbb{R}$, it holds that:

$$u_d(t,x) = \mathbb{E}\left[u_d\left(T,\mathcal{X}_t^{d,t,x}\right)\right] \quad (3.3.3)$$

Furthermore, u_d is a viscosity solution to (3.3.1).

Proof. This is a consequence of Lemma 3.1.2 and 3.2.14. $\qquad\square$

Corollary 3.3.1.1. *Let $T \in (0, \infty)$,*

let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $u_d \in C^{1,2}\left([0,T] \times \mathbb{R}^d, \mathbb{R}\right)$, $d \in \mathbb{N}$ satisfy for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that:

$$\left(\frac{\partial}{\partial t} u_d\right)(t,x) + \frac{1}{2}\left(\nabla_x^2 u_d\right)(t,x) = 0 \tag{3.3.4}$$

Let $\mathcal{W}^d : [0,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$ be standard Brownian motions, and let $\mathcal{X}^{d,t,x} : [t,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$, be a stochastic process with continuous sample paths satisfying that for all $d \in N$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ we have $\mathbb{P}$-a.s. that:

$$\mathcal{X}_s^{d,t,x} = x + \int_t^s d\mathcal{W}_r^d = x + \mathcal{W}_{t-s}^d \tag{3.3.5}$$

Then for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that:

$$u_d(t,x) = \mathbb{E}\left[u_d\left(T, \mathcal{X}_t^{d,T,x}\right)\right] \tag{3.3.6}$$

Proof. This is a special case of Theorem 3.3.1. It is the case where $\sigma_d(x) = \mathbb{I}_d$, the uniform identity function where $\mathbb{I}_d$ is the identity matrix in dimension d for $d \in \mathbb{N}$, and $\mu_d(x) = \mathbb{0}_d$ where $\mathbb{0}_d$ is the zero vector in dimension d for $d \in \mathbb{N}$. $\qquad\square$

Lemma 3.3.2. *Let $T \in (0, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$, be a probability space, let $\alpha_d \in C_b^2\left(\mathbb{R}^d, \mathbb{R}\right)$, and $\alpha \in \mathcal{O}\left(x^2\right)$ for $d \in N$, be infinitey often differentiable function, let $u_d \in C^{1,2}\left([0,T] \times \mathbb{R}^d, \mathbb{R}\right)$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$, that:*

$$\left(\frac{\partial}{\partial t} u_d\right)(t,x) + \frac{1}{2}\left(\nabla_x^2 u_d\right)(t,x) + \alpha_d(x) u_d(t,x) = 0 \tag{3.3.7}$$

Let $\mathcal{W}^d : [0,T] \times \Omega \to \mathbb{R}^d$ be standard Brownian motions and let $\mathcal{X}^{d,t,x} : [t,T] \times \Omega \to \mathbb{R}^d$,

$d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ be a stochastic process with continuous sample paths satisfying that for all $d \in \mathbb{N}$, $t \in [0, T]$, $s \in (t, T]$, $x \in \mathbb{R}^d$, we have $\mathbb{P}$-a.s. that:

$$\mathcal{X}_s^{d,t,x} = x + \int_s^t \frac{1}{2} d\mathcal{W}_r^d = \frac{1}{2} \mathcal{W}_{t-r}^d \tag{3.3.8}$$

Then for all $d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that:

$$u_d(t, x) = \mathbb{E}\left[\exp\left(\int_t^T \alpha_d\left(\mathcal{X}_r^{d,t,x}\right) dr \right) u_d\left(T, \mathcal{X}_T^{d,t,x}\right) \right] \tag{3.3.9}$$

Proof. Let $v_d : \mathbb{R}^d \to \mathbb{R}$ be continuous. Throughout the proof let $u_d(t, x) = e^{-t\alpha_d(x)} v_d(t, x)$ for all $d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$. For notational simplicity, we will drop the d, t, x wherever it is obvious. Therefore the derivatives become:

$$u_t = -\alpha e^{-t\alpha} v + e^{-t\alpha} v_t \tag{3.3.10}$$

$$\frac{1}{2}\nabla_x^2 u = \frac{1}{2}\left[e^{-t\alpha}\nabla_x^2 v + 2\langle \nabla_x v, \nabla_x e^{-t\alpha} \rangle + v\nabla_x^2 e^{-t\alpha} \right] \tag{3.3.11}$$

This then renders (3.3.7) as:

$$\cancel{-\alpha e^{-t\alpha} v} + e^{-t\alpha} v_t + \frac{1}{2}\left[e^{-t\alpha}\nabla_x^2 v + 2\langle \nabla_x v, \nabla_x e^{-t\alpha} \rangle + v\nabla_x^2 e^{-t\alpha} \right] + \cancel{\alpha e^{-t\alpha} v} = 0$$

$$e^{-t\alpha} v_t + \frac{1}{2}\left[e^{-t\alpha}\nabla_x^2 v - 2t e^{-t\alpha}\langle \nabla_x v, \nabla_x \alpha \rangle + v\nabla_x^2 e^{-t\alpha} \right] = 0$$

$$e^{-t\alpha} v_t + \frac{1}{2}\left[e^{-t\alpha}\nabla_x^2 v - 2t e^{-t\alpha}\langle \nabla_x v, \nabla_x \alpha \rangle - t v e^{-t\alpha}\nabla_x^2 \alpha \right] = 0$$

$$v_t + \frac{1}{2}\left[\nabla_x^2 v - 2t\langle \nabla_x v, \nabla_x \alpha \rangle - t v\nabla_x^2 \alpha \right] = 0$$

$$v_t + \frac{1}{2}\left[\nabla_x^2 v - 2t\langle \nabla_x \alpha, \nabla_x v \rangle - t v\nabla_x^2 \alpha \right] = 0$$

$$v_t + \frac{1}{2}\nabla_x^2 v + \langle -t\nabla_x \alpha, \nabla_x v \rangle - \frac{1}{2} t v\nabla_x^2 \alpha = 0 \tag{3.3.12}$$

Let $\sigma(t, x) = \mathbb{I}_d$, i.e. the uniform identity function. Let $\mu(t, x) = -t\nabla_x \alpha$ for $t \in [0, T]$, $x \in \mathbb{R}^d$, and for fixed α. Let $f(t, x, v) = -\frac{1}{2} t v\nabla_x^2 \alpha$ for $t \in [0, T]$, $x \in \mathbb{R}^d$.

Claim 3.3.3. *It is the case that for for all $x \in \mathbb{R}^d$ and $t \in [0,T]$ that $\langle x, \mu(t,x) \rangle \leqslant L(1 + \|x\|_E)$ for some constant $L \in (0, \infty)$.*

Proof. Since α has bounded first and second derivatives let:

$$\mathfrak{B} = \max \left\{ \sup_{x \in \mathbb{R}^d} \|\nabla_x \alpha\|_E, \sup_{x \in \mathbb{R}^d} |\nabla_x^2 \alpha| \right\} \tag{3.3.13}$$

Note that we then have the Cauchy-Schwarz inequality:

$$\langle x, \mu(t,x) \rangle \leqslant \|\langle x, -t\nabla_x \alpha \rangle\|_E \leqslant \|x\|_E \|t\nabla_x \alpha\|_E$$

$$\leqslant T\left(\|x\|_E \mathfrak{B}\right)$$

$$\leqslant T\left(\mathfrak{B} + d\right) \|x\|_E$$

$$= L\|x\|_E \leqslant L\left(1 + \|x\|_E^2\right) \tag{3.3.14}$$

It also follows that $\|\sigma(t,x)\|_F = \sqrt{d} \leqslant L \leqslant L(1 + \|x\|_E)$. $\qquad \square$

Claim 3.3.4. *It is the case that for all $x, y \in \mathbb{R}^d$, and $t \in [0,T]$ that: $\|\mu(t,x) - \mu(t,y)\|_E + \|\sigma(t,x) - \sigma(t,y)\|_E \leqslant \mathfrak{C}\left(\|x\|_E + \|y\|_E\right)\left(\|x - y\|_E\right)$ for some constant $\mathfrak{C} \in (0, \infty)$.*

Proof. The fact that for all $x, y \in \mathbb{R}^d$ and $t \in [0,T]$ it is the case that $\|\sigma(t,x) - \sigma(t,y)\|_F = 0$, the fact that for all $x, y \in \mathbb{R}^d$ it is the case that $(\|x\|_E + \|y\|_E)(\|x - y\|_E) \geqslant 0$ and (3.3.13) tells us that:

$$\|\mu(t,x) - \mu(t,y)\|_E + \|\sigma(t,x) - \sigma(t,y)\|_F = \|\mu(t,x) - \mu(t,y)\|_E + 0$$

$$= \|t\nabla_x \alpha(x) - t\nabla_x \alpha(y)\|_E$$

$$\leqslant T\|\nabla_x \alpha(x) - \nabla_x \alpha(y)\|_E$$

$$\leqslant 2T\mathfrak{B} \tag{3.3.15}$$

Now consider a function $\mathfrak{f} \in C\left([0,T] \times \mathbb{R}^d, \mathbb{R}^d\right)$, where for all $x, y \in \mathbb{R}^d$ it is the case that

65

$\mathfrak{f}(x) - \mathfrak{f}(y) \leqslant \mathscr{C}\left(\|x\|_E + \|y\|_E\right)\left(\|x + y\|_e\right)$. Note then that setting $y = x + h$ gives us:

$$\left|\frac{\mathfrak{f}(x + h) - \mathfrak{f}(x)}{h}\right| \leqslant \mathscr{C}\left(\|x\|_E + \|x + h\|_E\right)$$

$$\lim_{h \to 0}\left|\frac{\mathfrak{f}(x + h) - \mathfrak{f}(x)}{h}\right| \leqslant \lim_{h \to 0}\mathscr{C}\left(\|x\|_E + \|x + h\|_E\right)$$

$$|\nabla_x\mathfrak{f}(x)| \leqslant 2\mathscr{C}\|x\|_E = \mathscr{K}\|x\|_E \tag{3.3.16}$$

This suggests that $\nabla_x\mathfrak{f} \in O(x)$ and in particular that $\mathfrak{f} \in O(x^2)$. However with $\mathfrak{f} \curvearrowright \mu$ we first notice that because $\mu \leqslant 2T\mathfrak{B}$ in (3.3.15) it must also be that case that $\mu \in O(1)$ by Corollary 1.2.32.1. However since $O(c) \subseteq O(x) \subseteq O(x^2)$ by Corollary 1.2.32.2 it is also the case that $\mu \in O(x^2)$, and hence there exists a $\mathfrak{C}$ satisfying the claim. This proves the claim. $\qquad\square$

Claim 3.3.5. *It is the case that* $|f(t, x, v) - f(t, x, w)| \leqslant L|v - w|$

Proof. Note that by the absolute homogeneity property of norms, we have:

$$\begin{aligned}
|f(t, x, v) - f(t, x, w)| &= \left|\frac{1}{2}tv\nabla_x^2\alpha - \frac{1}{2}tw\nabla_x^2\alpha\right| \\
&= \left|\frac{1}{2}t\nabla_x^2\alpha\right||v - w| \\
&\leqslant \frac{1}{2}T\left|\nabla_x^2\alpha\right||v - w| \\
&\leqslant \frac{1}{2}T\mathfrak{B}|v - w| \\
&\leqslant T(\mathfrak{B} + d)|v - w| \\
&= L|v - w| \tag{3.3.17}
\end{aligned}$$

$\qquad\square$

Note that we may rewrite (3.3.12) as:

$$\left(\frac{\partial}{\partial t}v\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma\left(t,x\right)\left[\sigma\left(t,x\right)\right]^*\left(\operatorname{Hess}_x v\right)\left(t,x\right)\right) + \left\langle\mu\left(t,x\right),\left(\nabla_x v\right)\left(t,x\right)\right\rangle$$

$$+ f\left(t,x,v\left(t,x\right)\right) = 0$$

We realize that (3.3.12) is a case of (Beck et al., 2021c, Corollary 3.9) where it is the case that: $u(t,x) \curvearrowleft v(t,x)$, where $\sigma_d(x) = \mathbb{I}_d$ for all $x \in \mathbb{R}^d$, $d \in \mathbb{N}$, where $\mu(t,x) = -t\nabla_x\alpha$ for fixed α and for all $t \in [0,T]$, $x \in \mathbb{R}^d$, and where $f\left(t,x,u\left(t,x\right)\right) = -\frac{1}{2}tu\nabla_x^2\alpha$ for fixed α and for all $t \in [0,T]$, $x \in \mathbb{R}^d$.

We thus have that there exists a unique, at most polynomially growing viscosity solution $v \in C\left([0,T]\times\mathbb{R}^d,\mathbb{R}\right)$ given as:

$$v(t,x) = \mathbb{E}\left[v\left(T,\mathcal{Y}_T^{t,x}\right) + \int_t^T f\left(s,\mathcal{Y}_s^{t,x},v\left(s,\mathcal{Y}_s^{t,x}\right)\right)ds\right] \tag{3.3.18}$$

Let $\mathcal{V} : [0,T]\times\Omega \to \mathbb{R}^m$ be a standard $(\mathbb{F}_t)_{t\in[0,T]}$-Brownian motion. Note that this also implies that the $\mathcal{Y}$ in (3.3.18) is characterized as:

$$\mathcal{Y}_s^{t,x} = x + \int_t^s \mu\left(r,\mathcal{Y}_r^{t,x}\right)dr + \int_t^s \sigma\left(s,\mathcal{X}_r^{t,x}\right)d\mathcal{V}_r \tag{3.3.19}$$

With substitution, this is then:

$$\mathcal{Y}_s^{t,x} = x + \int_t^s -r\nabla_x\alpha\left(\mathcal{Y}_r^{t,x}\right)dr + \int_t^s \mathbb{I}d\mathcal{V}_r$$

$$\mathcal{Y}_s^{t,x} = x - \int_t^s r\nabla_x\alpha\left(\mathcal{Y}_s^{t,x}\right)dr + \mathcal{V}_{s-t}$$

Note that our initial substitution tells us: $v(t, x) = e^{t\alpha(x)} u(t, x)$. And so we have that:

$$v(t, x) = \mathbb{E}\left[v\left(T, \mathcal{X}_T^{t,x}\right) + \int_t^T f\left(s, \mathcal{X}_s^{t,x}, v\left(s, \mathcal{X}_s^{t,x}\right)\right) ds\right] \quad (3.3.20)$$

$$v(t, x) = \mathbb{E}\left[v\left(T, \mathcal{X}_T^{t,x}\right) - \frac{1}{2}\int_t^T tv\left(s, \mathcal{X}_s^{t,x}\right) \nabla_x^2 \alpha\left(\mathcal{X}_s^{t,x}\right) ds\right]$$

$$e^{t\alpha(x)} u(t, x) = \mathbb{E}\left[\exp\left[T\alpha\left(\mathcal{X}_T^{t,x}\right)\right] u\left(T, \mathcal{X}_T^{t,x}\right) - \frac{1}{2}\int_t^T t \exp\left[t\alpha\left(\mathcal{X}_s^{t,x}\right)\right] u\left(t, \mathcal{X}_s^{t,x}\right) \nabla_x^2 \alpha\left(\mathcal{X}_s^{t,x}\right) ds\right]$$

$$u(t, x) = \mathbb{E}\left[\exp\left[T\alpha\left(\mathcal{X}_T^{t,x}\right) - t\alpha(x)\right] u\left(T, \mathcal{X}_T^{t,x}\right)\right]$$

$$- \mathbb{E}\left[\frac{1}{2e^{t\alpha(x)}}\int_t^T t \exp\left[t\alpha\left(\mathcal{X}_s^{t,x}\right)\right] u\left(t, \mathcal{X}_s^{t,x}\right) \nabla_x^2 \alpha\left(\mathcal{X}_s^{t,x}\right) ds\right]$$

$\square$

Part II

A Structural Description of Artificial Neural Networks

Chapter 4

Introduction and Basic Notions About Neural Networks

We seek here to introduce a unified framework for artificial neural networks. This framework borrows from the work presented in Grohs et al. (2018), which was in turn inspired by work done in Petersen and Voigtlaender (2018). The most recent exposition of this framework can be found in Jentzen et al. (2023), and it is this exposition that our work will be based on and extended upon.

With this framework in place, we wish to study ANNs from the perspective of trying to see the number of parameters required to define a neural network to solve certain PDEs. The *curse of dimensionality* here refers to the number of parameters and depths of neural networks necessary to approximate functions to a certain accuracy. Specifically a scheme is said to have beat the curse of dimensionality if the number of parameters and depths necessary to approximate an underlying function to an accuracy (specifically the upper bound on the the 1-norm difference between the approximant and the function over the entire domain), only grows polynomially or at-least sub-exponentially on $\frac{1}{\varepsilon}$.

4.1 The Basic Definition of ANNs and instantiations of ANNs

Definition 4.1.1 (Rectifier Function)**.** *Let $d \in \mathbb{N}$ and $x \in \mathbb{R}^d$. We denote by $\mathfrak{r} : \mathbb{R} \to \mathbb{R}$ the function given by:*

$$\mathfrak{r}(x) = \max\{0, x\} \tag{4.1.1}$$

Remark 4.1.2. *By analogy the multidimensional rectifier function, defined for $x = [x_1\ x_2\ \cdots\ x_n]^\intercal \in \mathbb{R}^n$ is:*

$$\mathfrak{r}([x]_*) = [\max\{0, x_1\}\ \max\{0, x_2\}\ \cdots \max\{0, x_n\}]^\intercal \tag{4.1.2}$$

Definition 4.1.3 (Artificial Neural Networks)**.** *Denote by* NN *the set given by:*

$$\mathsf{NN} = \bigcup_{L \in \mathbb{N}} \bigcup_{l_0, l_1, \ldots, l_L \in \mathbb{N}} \left(\underset{k=1}{\overset{L}{\times}} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right) \tag{4.1.3}$$

An artificial neural network is a tuple $(\nu, \mathsf{P}, \mathsf{D}, \mathsf{I}, \mathsf{O}, \mathsf{H}, \mathsf{L}, \mathsf{W})$ where $\nu \in \mathsf{NN}$ and is equipped with the following functions (referred to as auxiliary functions) satisfying for all $\nu \in \left(\underset{k=1}{\overset{L}{\times}} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right)$:

(i) $\mathsf{P} : \mathsf{NN} \to \mathbb{N}$ *denoting the number of parameters of ν, given by:*

$$\mathsf{P}(\nu) = \sum_{k=1}^{L} l_k \left(l_{k-1} + 1 \right) \tag{4.1.4}$$

(ii) $\mathsf{D} : \mathsf{NN} \to \mathbb{N}$ *denoting the number of layers of ν other than the input layer given by:*

$$\mathsf{D}(\nu) = L \tag{4.1.5}$$

(iii) $\mathsf{I} : \mathsf{NN} \to \mathbb{N}$ *denoting the width of the input layer, given by:*

$$\mathsf{I}(\nu) = l_0 \qquad (4.1.6)$$

(iv) $\mathsf{O} : \mathsf{NN} \to \mathbb{N}$ *denoting the width of the output layer, given by:*

$$\mathsf{O}(\nu) = l_L \qquad (4.1.7)$$

(v) $\mathsf{H} : \mathsf{NN} \to \mathbb{N}_0$ *denoting the number of hidden layers (i.e., layers other than the input and output), given by:*

$$\mathsf{H}(\nu) = L - 1 \qquad (4.1.8)$$

(vi) $\mathsf{L} : \mathsf{NN} \to \bigcup_{L \in \mathbb{N}} \mathbb{N}^L$ *denoting the width of layers as an $(L+1)$-tuple, given by:*

$$\mathsf{L}(\nu) = (l_0, l_1, l_2, ..., l_L) \qquad (4.1.9)$$

We sometimes refer to this as the layer configuration or layer architecture of ν.

(vii) $\mathsf{W}_i : \mathsf{NN} \to \mathbb{N}_0$ *denoting the width of layer i, given by:*

$$\mathsf{W}_i(\nu) = \begin{cases} l_i & i \leqslant L \\ 0 & i > L \end{cases} \qquad (4.1.10)$$

Note that this implies that $\nu = ((W_1, b_1), (W_2, b_2), ...(W_L, b_L)) \in \left(\bigtimes_{k=1}^{L} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right)$. Note that we denote by $\mathrm{Weight}_{(\cdot),\nu} : (\mathrm{Weight}_{n,\nu})_{n \in \{1,2,...,L\}} : \{1, 2, ..., L\} \to \left(\bigcup_{m,k \in \mathbb{N}} \mathbb{R}^{m \times k} \right)$ and also $\mathrm{bias}_{(\cdot),\nu} : (\mathrm{bias}_{n,\nu})_{\{1,2,...,L\}} : \{1, 2, ..., L\} \to \left(\bigcup_{m \in \mathbb{N}} \mathbb{R}^m \right)$ the functions that satisfy for all $n \in \{1, 2, ..., L\}$ that $\mathrm{Weight}_{i,\nu} = W_i$ i.e. the weights matrix for neural network ν at layer i and $\mathrm{bias}_{i,\nu} = b_i$, i.e. the bias vector for neural network ν at layer i.

We will call l_0 the *starting width* and l_L the *finishing width*. Together, they will be referred to as *end-widths*.

Remark 4.1.4. *Notice that our definition varies somewhat from the conventional ones found in Petersen and Voigtlaender (2018) and Grohs et al. (2023) in that whereas the former talk about auxiliary functions as existing within the set* NN *we will talk about these auxiliary functions as something elements of* NN *are endowed with. In other words, elements of* NN *may exist whose depths and parameter counts, for instance, are undefined and non-determinate. Note that we develop this definition to closely align to popular deep-learning frameworks such as* PyTorch, TensorFlow, *and* Flux, *where, in principle, it is always possible to know the parameter count, depth, number of layers, and other auxiliary information.*

We will often say let $\nu \in$ NN, and it is implied that the tuple ν with the auxiliary functions is what is being referred to. This is analogous to when we say that X is a topological but we mean the pair (X, τ), i.e. X endowed with topology τ, or when we say that Y is a measurable space when we mean the triple (X, Ω, μ), i.e. X, endowed with $\sigma-$algebra Ω, and measure μ.

Definition 4.1.5 (Instantiations of Artificial Neural Networks with Activation Functions)**.** *Let $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, we denote by $\mathfrak{I}_\mathfrak{a} :$ NN $\to \left(\bigcup_{k,l \in \mathbb{N}} C\left(\mathbb{R}^k, \mathbb{R}^l\right) \right)$ the function satisfying for all $L \in \mathbb{N}$, $l_0, l_1, ..., l_L \in \mathbb{N}$, $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L)) \in \left(\bigtimes_{k=1}^{L} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right)$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, ..., x_{L-1} \in \mathbb{R}^{l_{L-1}}$ and with $\forall k \in \mathbb{N} \cap (0, L) : x_k = \mathfrak{a}\left(\left[W_k x_k + b_k \right]_{*,*} \right)$ such that:*

$$\mathfrak{I}_\mathfrak{a}(\nu) \in C\left(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}\right) \ and \qquad (\mathfrak{I}_\mathfrak{a}(\nu))(x_0) = W_L x_{L-1} + b_L \qquad (4.1.11)$$

Remark 4.1.6. *For an R implementation see Listings 9.1, 9.2, 9.3, and 9.4.*

Lemma 4.1.7. *Let $\nu \in$ NN, it is then the case that:*

(i) $\mathsf{L}(\nu) \in \mathbb{N}^{\mathsf{D}(\nu)+1}$, *and*
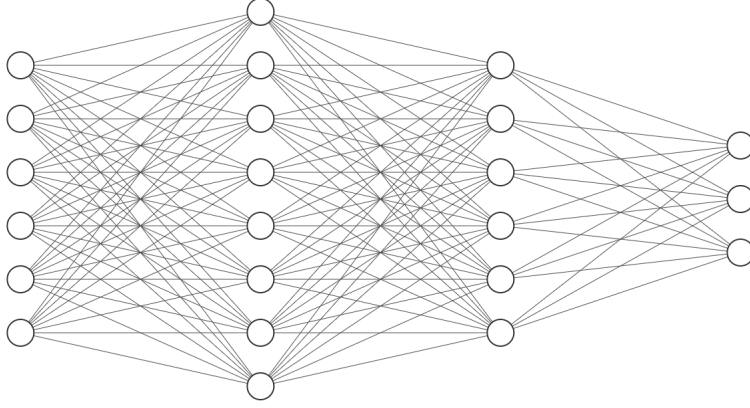
Figure 4.1: A neural network ν with $\mathsf{L}\left(\nu\right) = (6,8,6,3)$

(ii) for all $\mathfrak{a} \in C\left(\mathbb{R}, \mathbb{R}\right)$, $\mathfrak{I}_{\mathfrak{a}} \in C\left(\mathbb{R}^{\mathsf{I}(\nu)}, \mathbb{R}^{\mathsf{O}(\nu)}\right)$

Proof. By assumption:

$$\nu \in \mathsf{NN} = \bigcup_{L \in \mathbb{N}} \bigcup_{(l_0, l_1, ..., l_L) \in \mathbb{N}^{L+1}} \left(\underset{k=1}{\overset{L}{\times}} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right) \tag{4.1.12}$$

This ensures that there exist $l_0, l_1, ..., l_L, L \in \mathbb{N}$ such that:

$$\nu \in \left(\underset{j=1}{\overset{L}{\times}} \left[\mathbb{R}^{l_j \times l_{j-1}} \times \mathbb{R}^{B_j} \right] \right) \tag{4.1.13}$$

This also ensures that $\mathsf{L}(\nu) = (l_0, l_1, ..., l_L) \in \mathbb{N}^{L+1} = \mathbb{N}^{\mathsf{D}(\nu)+1}$ and further that $\mathsf{I}(\nu) = l_0$, $\mathsf{O}(\nu) = l_L$, and that $\mathsf{D}(\nu) = L$. Together with (4.1.11), this proves the lemma. $\square$

4.2 Compositions of ANNs

The first operation we want to be able to do is to compose neural networks. Note that the composition is not done in an obvious way; for instance, note that the last layer of the first component of the composition is superimposed with the first layer of the second component of the composition.

Definition 4.2.1 (Compositions of ANNs)**.** *We denote by* $(\cdot) \bullet (\cdot) : \{(\nu_1, \nu_2) \in \mathsf{NN} \times \mathsf{NN} :$ $\mathsf{I}(\nu_1) = \mathsf{O}(\nu_1)\} \to \mathsf{NN}$ *the function satisfying for all* $L, M \in \mathbb{N}, l_0, l_1, ..., l_L, m_0, m_1, ..., m_M \in$ $\mathbb{N}, \nu_1 = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L)) \in \left(\bigtimes_{k=1}^{L} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right),$ *and* $\nu_2 =$ $((W_1', b_1'), (W_2', b_2'), ... (W_M', b_M')) \in \left(\bigtimes_{k=1}^{M} [\mathbb{R}^{m_k \times m_{k-1}} \times \mathbb{R}^{m_k}] \right)$ *with* $l_0 = \mathsf{I}(\nu_1) = \mathsf{O}(\nu_2) = m_M$ *and :*

$$\nu_1 \bullet \nu_2 = \tag{4.2.1}$$

$$\begin{cases} ((W_1', b_1'), (W_2', b_2'), ...(W_{M-1}', b_{M-1}'), (W_1 W_M', W_1 b_M' + b_1), (W_2, b_2), \\ \quad ..., (W_L, b_L)) & : (L > 1) \wedge (M > 1) \\ \\ ((W_1 W_1', W_1 b_1' + b_1), (W_2, b_2), (W_3, b_3), ..., (W_L b_L)) & : (L > 1) \wedge (M = 1) \\ \\ ((W_1', b_1'), (W_2', b_2'), ..., (W_{M-1}', b_{M-1}')(W_1, b_M' + b_1)) & : (L = 1) \wedge (M > 1) \\ \\ ((W_1 W_1', W_1 b_1' + b_1)) & : (L = 1) \wedge (M = 1) \end{cases}$$

$$\tag{4.2.2}$$

Remark 4.2.2. *For an* R *implementation see Listing 9.7*

Lemma 4.2.3. *Let* $\nu, \mu \in \mathsf{NN}$ *be such that* $\mathsf{O}(\mu) = \mathsf{I}(\nu)$. *It is then the case that:*

(i) $\mathsf{D}(\nu \bullet \mu) = \mathsf{D}(\nu) + \mathsf{D}(\mu) - 1$

(ii) *For all* $i \in \{1, 2, ..., \mathsf{D}(\nu \bullet \mu)\}$ *that:*

$$\left(\mathrm{Weight}_{i,(\nu \bullet \mu)}, \mathrm{bias}_{i,(\nu \bullet \mu)} \right)$$

$$= \begin{cases} \left(\mathrm{Weight}_{i,\mu}, \mathrm{bias}_{i,\mu} \right) & : i < \mathsf{D}(\mu) \\ \\ \left(\mathrm{Weight}_{1,\nu} \, \mathrm{Weight}_{\mathsf{D}(\mu),\mu}, \mathrm{Weight}_{1,\nu} \, \mathrm{bias}_{\mathsf{D}(\mu),\mu} + \mathrm{bias}_{1,\nu} \right) & : i = \mathsf{D}(\mu) \\ \\ \left(\mathrm{Weight}_{i-\mathsf{D}(\mu)+1,\nu} \, \mathrm{bias}_{i-\mathsf{D}(\mu)+1,\nu} \right) & : i > \mathsf{D}(\mu) \end{cases}$$

Proof. This is a consequence of (4.2.1), which implies (i)—(ii). $\qquad\square$

Lemma 4.2.4. *Let $\nu_1, \nu_2, \nu_3 \in$ NN satisfy that $\mathsf{I}(\nu_1) = \mathsf{O}(\nu_2)$ and $\mathsf{I}(\nu_2) = \mathsf{O}(\nu_3)$, it is then the case that:*

$$(\nu_1 \bullet \nu_2) \bullet \nu_3 = \nu_1 \bullet (\nu_2 \bullet \nu_3) \tag{4.2.3}$$

Proof. This is a consequence of (Grohs et al., 2023, Lemma 2.8) with $\Phi_1 \curvearrowright \nu_1$, $\Phi_2 \curvearrowright \nu_2$, and $\Phi_3 \curvearrowright \nu_3$, and the functions $\mathcal{I} \curvearrowright \mathsf{I}$, $\mathcal{L} \curvearrowright \mathsf{D}$ and $\mathcal{O} \curvearrowright \mathsf{O}$. $\qquad\square$

The following Lemma will be important later on, referenced numerous times, and found in (Grohs et al., 2023, Proposition 2.6). For completion, we will include a simplified version of the proof here.

Lemma 4.2.5. *Let $\nu_1, \nu_2 \in$ NN. Let it also be that $\mathsf{O}(\nu_1) = \mathsf{I}(\nu_2)$. It is then the case that:*

(i) $\mathsf{D}(\nu_1 \bullet \nu_2) = \mathsf{D}(\nu_1) + \mathsf{D}(\nu_2) - 1$

(ii) $\mathsf{L}(\nu_1 \bullet \nu_2) = \left(\mathsf{W}_1(\nu_2), \mathsf{W}_2(\nu_2), \ldots, \mathsf{W}_{\mathsf{H}(\nu_2)}, \mathsf{W}_1(\nu_1), \mathsf{W}_2(\nu_1), \ldots, \mathsf{W}_{\mathsf{D}(\phi_1)}(\nu_1)\right)$

(iii) $\mathsf{H}(\nu_1 \bullet \nu_2) = \mathsf{H}(\nu_1) + \mathsf{H}(\nu_2)$

(iv) $\mathsf{P}(\nu_1 \bullet \nu_2) \leqslant \mathsf{P}(\nu_1) + \mathsf{P}(\nu_2) + \mathsf{W}_1(\nu_1) \cdot \mathsf{W}_{\mathsf{H}(\nu_2)}(\nu_2)$

(v) *for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ that $\mathfrak{I}_{\mathfrak{a}}(\nu_1 \bullet \nu_2)(x) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_2)}, \mathbb{R}^{\mathsf{O}(\nu_1)}\right)$ and further:*

$$\mathfrak{I}_{\mathfrak{a}}(\nu_1 \bullet \nu_2) = [\mathfrak{I}_{\mathfrak{a}}(\nu_1)] \circ [\mathfrak{I}_{\mathfrak{a}}(\nu_2)] \tag{4.2.4}$$

Proof. Note that Items (i)---(iii) are a simple consequence of Definition 4.2.1. Specifically, given neural networks $\nu_1, \nu_2 \in$ NN, and $\mathsf{D}(\nu_1) = n$ and $\mathsf{D}(\nu_2) = m$, note that for all four cases, we have that the depth of the composed neural network $\nu_1 \bullet \nu_2$ is given by $n - 1 + m - 1 = n + m - 1$ proving Item (i). Note that the outer neural network loses its

last layer, yielding Item (ii) in all four cases. Finally since, for all $\nu \in \mathsf{NN}$ it is the case that $\mathsf{H}(\nu) = \mathsf{D}(\nu) - 1$, Item (i) yields Item (iii).

Now, suppose it is the case that $\nu_3 = \nu_1 \bullet \nu_2$ and that:

$$\nu_1 = ((W_{1,1}, b_{1,1}), (W_{1,2}, b_{1,2}), \ldots, (W_{1,L_1}, b_{1,L_1}))$$
$$\nu_2 = ((W_{2,1}, b_{2,1}), (W_{2,2}, b_{2,2}), \ldots, (W_{2,L_2}, b_{2,L_2}))$$
$$\nu_3 = ((W_{3,1}, b_{3,1}), (W_{3,2}, b_{3,2}), \ldots, (W_{3,L_2}, b_{3,L_2}))$$

$$(4.2.5)$$

And that:

$$\mathsf{L}(\nu_1) = (l_{1,1}, l_{1,2}, \ldots, l_{1,L_1})$$
$$\mathsf{L}(\nu_2) = (l_{2,1}, l_{2,2}, \ldots, l_{2.L_2})$$
$$\mathsf{L}(\nu_1 \bullet \nu_2) = (l_{3,1}, l_{3,2}, \ldots, l_{3,L_3}) \qquad (4.2.6)$$

and further 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 the condition that:

$$\forall k \in \mathbb{N} \cap (0, L_2) : x_k = \mathfrak{a}\left([W_{2,k}x_{k-1} + b_{2,k}]_{*,*}\right) \qquad (4.2.7)$$

also let $y_0 \in \mathbb{R}^{l_{1,0}}, y_1 \in \mathbb{R}^{l_{1,1}}, \ldots, y_{L_1-1} \in \mathbb{R}^{l_{2,L_2-1}}$ satisfy:

$$\forall k \in \mathbb{N} \cap (0, L_1) : y_k = \mathfrak{a}\left([W_{1,k}y_{k-1} + b_{1,k}]_{*,*}\right) \qquad (4.2.8)$$

and finally 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:

$$\forall k \in \mathbb{N} \cap (0, L_3) : z_k = \mathfrak{a}\left([W_{3,k}z_{k-1} + b_{3,k}]_{*,*}\right) \qquad (4.2.9)$$

Note then that by Item (i) of Definition 4.1.3 we have that:

$$P\left(\nu_1 \bullet \nu_2\right) = \sum_{k=1}^{L_3} l_{3,k}\left(l_{3,k-1}+1\right)$$

$$= \left[\sum_{k=1}^{L_2-1} l_{3,k}\left(l_{3,k-1}+1\right)\right] + l_{3,L_2}\left(l_{3,L_2-1}+1\right) + \left[\sum_{k=L_2+1}^{L_3} l_{3,k}\left(l_{3,k-1}+1\right)\right]$$

$$= \left[\sum_{k=1}^{L_2-1} l_{2,j}\left(l_{2,j-1}+1\right)\right] + l_{1,1}\left(l_{2,L-1}+1\right) + \left[\sum_{k=L_2+1}^{L_3} l_{j-L_2+1}\left(l_{1,j-L_2}+1\right)\right]$$

$$= \left[\sum_{k=1}^{L_2-1} l_{2,j}\left(l_{2,k-1}+1\right)\right] + \left[\sum_{k=2}^{L_1} l_{1,j}\left(l_{1,k-1}+1\right)\right] + l_{1,1}\left(l_{2,L_2-1}+1\right)$$

$$= \left[\sum_{k=1}^{L_2} l_{2,k}\left(l_{2,k-1}+1\right)\right] + \left[\sum_{k=1}^{L_1} l_{1,k}\left(l_{1,k-1}+1\right)\right] + l_{1,1}\left(l_{2,L_2-1}+1\right)$$

$$- l_{2,L_2}\left(l_{2,L_2-1}+1\right) - l_{1,1}\left(l_{1,0}+1\right)$$

$$= P\left(\nu_1\right) + P\left(\nu_2\right) + l_{1,1}\cdot l_{2,L_2-1} \tag{4.2.10}$$

Thus establishing Item (iv). Note by Definition 4.2.1, and the fact that $\mathfrak{a} \in C\left(\mathbb{R},\mathbb{R}\right)$ it is the case that

$$\mathfrak{I}_{\mathfrak{a}}\left(\nu_1 \bullet \nu_2\right) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_1)}, \mathbb{R}^{\mathsf{O}(\nu_2)}\right) \tag{4.2.11}$$

Next note that by definition, it is the case that:

$$\mathsf{L}\left(\nu_1 \bullet \nu_2\right) = \left(l_{2,0}, l_{2,1}, \ldots, l_{2,L_2-1}, l_{1,1}, l_{1,2}, \ldots, l_{1,L_1}\right) \tag{4.2.12}$$

And further that:

$$\forall k \in \mathbb{N} \cap \left(0, L_2\right) : \left(W_{3,k}, b_{3,k}\right) = \left(W_{2,k}, b_{2,k}\right)$$

$$\left(W_{3,L_2}, b_{3,L_2}\right) = \left(W_{1,1}\cdot W_{2,L_2}, W_{1,1}b_{2,L_2} + b_{1,1}\right)$$

$$\text{and } \forall k \in \mathbb{N} \cap \left(L_2, L_1+L_2\right) : \left(W_{3,k}, b_{3,k}\right) = \left(W_{1,j+1-L_2}, b_{1,j+1-L_2}\right) \tag{4.2.13}$$

Since for all $k \in \mathbb{N} \cap [0, L_2)$ it is the case that $z_j = x_j$ and the fact that $y_0 = W_{2,l_2}x_{L_2-1} + b_{2,L_2}$ ensures us that:

$$W_{3,L_2}z_{L_2-1} + b_{3,L_2} = W_{3,L_2}x_{L_2-1} + b_{3,L_2}$$
$$= W_{1,1}W_{2,L_2}x_{L_2-1} + W_{1,1}b_{2,L_2} + b_{1,1}$$
$$= W_{1,1}\left(W_{2,L_2}x_{L_2-1} + b_{2,L_2}\right) + b_{1,1} = W_{1,1}y_0 + b_{1,1} \qquad (4.2.14)$$

We next claim that for all $k \in \mathbb{N} \cap [L_2, L_1 + L_2)$ it is the case that:

$$W_{3,k}z_{k-1} + b_{3,k} = W_{1,k+1-L_2}y_{k-L_2} + b_{1,k+1-L_2} \qquad (4.2.15)$$

This can be proved via induction on $k \in \mathbb{N} \cap [L_2, L_1 + L_2)$. Consider that our base case of $k = L_2$ in (4.2.15) is fulfilled by (4.2.14). Now note that for all $k \in \mathbb{N} \cap [L_2, \infty) \cap (0, L_1 + L_2 - 1)$ with:

$$W_{3,k}z_{k-1} + b_{3,k} = W_{1,k+1-L_2}y_{k-L_2} + b_{1,k+1-L_2} \qquad (4.2.16)$$

it holds that:

$$W_{3,k+1}z_k + b_{3,k+1} = W_{3,k+1}\left([W_{3,k}z_{k-1} + b_{3,k}]_{*,*}\right) + b_{3,k+1}$$
$$= W_{1,k+2-L_2}\left([W_{1,k+1-L_2}y_{k-L_2}] + b_{1,k+1-L_2}\right) + b_{1,k+2-L_2}$$
$$= W_{1,k+2-L_2}y_{k+1-L_2} + b_{1,k+2-L_2} \qquad (4.2.17)$$

Whence induction proves (4.2.15). This, along with the fact that $L_3 = L_1 + L_2 - 1$ then indicates that:

$$W_{3,L_3}z_{L_3-1} + b_{3,L_3} = W_{3,L_1+L_2-1}z_{L_1+L_2-2} + b_{3,L_1+L_2-1} = W_{1,L_1}y_{L_1-1} + b_{1,L_1} \qquad (4.2.18)$$

Finally, the fact that $\nu_3 = \nu_1 \bullet \nu_2$, in addition with (4.2.7),(4.2.8), and (4.2.9) then tells us that:

$$
\begin{aligned}
\left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_1 \bullet \nu_2\right)\right]\left(x_0\right) = \left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_3\right)\right]\left(x_0\right) &= \left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_3\right)\right]\left(z_0\right) = W_{3,L_3} z_{L_3-1} + b_{3,L_3} \\
&= W_{1,L_1} y_{L_1-1} + b_{1,L_1} = \left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_1\right)\right]\left(y_0\right) \\
&= \left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_1\right)\right]\left(\left[W_{2,L_2} x_{L_2-1} + b_{2,L_2}\right]_{*,*}\right) \\
&= \left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_1\right)\right]\left(\left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_2\right)\right]\left(x_0\right)\right) = \left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_1\right)\right] \circ \left[\mathfrak{I}_{\mathfrak{a}}\left(\nu_2\right)\right]\left(x_0\right) \quad (4.2.19)
\end{aligned}
$$

This and (4.2.11) then prove Item (v), hence proving the lemma. $\qquad\square$

4.3   Stacking of ANNs

We will introduce here the important concept of stacking of ANNs. Given an input vector $x \in \mathbb{R}^d$, it is sometimes very helpful to imagine two neural networks working on them simultaneously, whence we have stacking. Because vectors are ordered tuples, stacking ν_1 and ν_2 is not necessarily the same as stacking ν_2 and ν_1. We will thus forego the phrase "parallelization" used in e.g. Grohs et al. (2023) and Jentzen et al. (2023), and opt to use the term "stacking". This because parallelization implies commutativity, but it is clearly not the case that $\nu_1 \boxminus \nu_2$ is the ame as $\nu_2 \boxminus \nu_1$.

4.3.1   Stacking of ANNs of Equal Depth

Definition 4.3.1 (Stacking of ANNs of same depth)**.** *Let $L, n \in \mathbb{N}$, and let $\nu_1, \nu_2, \ldots, \nu_n \in$ NN, such that $\mathsf{D}\left(\nu_1\right) = \mathsf{D}\left(\nu_2\right) = \cdots = \mathsf{D}\left(\nu_n\right) = L$. As such, for all $i \in \{1, \ldots, n\}$, let it also be the case that $\mathsf{L}\left(\nu_i\right) = \left(\left(W_1^i, b_1^i\right), \left(W_2^i, b_2^i\right), \ldots, \left(W_L^i, b_L^i\right)\right)$. We then denote by $\boxminus_{i=1}^n \nu_i$, the*

neural network whose layer architecture is given by:

$$\mathsf{L}\left(\boxminus_{i=1}^{n}\nu_i\right) = \left(\left(\mathrm{diag}\left(W_1^1, W_1^2, \ldots, W_1^n\right), b_1^1 \frown b_1^2, \frown \cdots \frown b_1^n\right),\right.$$

$$\left(\mathrm{diag}\left(W_2^1, W_2^2, \ldots, W_2^n\right), b_2^1 \frown b_2^2, \frown \cdots \frown b_2^n\right),$$

$$\vdots$$

$$\left.\left(\mathrm{diag}\left(W_L^1, W_L^2, \ldots, W_L^n\right), b_L^1 \frown b_L^2, \frown \cdots \frown b_L^n\right)\right)$$

Remark 4.3.2. *For an* R *implementation see Listing 9.5*

Lemma 4.3.3. *Let* $\nu_1, \nu_2 \in \mathsf{NN}$, *with* $\mathsf{D}\left(\nu_1\right) = \mathsf{D}\left(\nu_2\right)$, $x_1 \in \mathbb{R}^{m_1}$, $x_2 \in \mathbb{R}^{m_2}$, *and* $\mathfrak{x} \in \mathbb{R}^{m_1+m_2}$. *Let* $\mathfrak{I}_{\mathfrak{r}}\left(\nu_1\right) : \mathbb{R}^{m_1} \to \mathbb{R}^{n_1}$, *and* $\mathfrak{I}_{\mathfrak{r}} : \mathbb{R}^{m_2} \to \mathbb{R}^{n_2}$. *It is then the case that* $\mathfrak{I}_{\mathfrak{r}}\left(\nu_1 \boxminus \nu_2\right)\left(\mathfrak{x}\right) = \mathfrak{I}_{\mathfrak{r}}\left(\nu_1\right)\left(x_1\right) \frown \mathfrak{I}_{\mathfrak{r}}\left(\nu_2\right)\left(x_2\right)$.

Proof. Let $\mathsf{L}\left(\nu_1\right) = \left(\left(W_1, b_1\right), \left(W_2, b_2\right), \ldots, \left(W_L, b_L\right)\right)$ and
$\mathsf{L}\left(\nu_2\right) = \left(\left(\mathfrak{W}_1, \mathfrak{b}_1\right), \left(\mathfrak{W}_2, \mathfrak{b}_2\right), \ldots, \left(\mathfrak{W}_L, \mathfrak{b}_L\right)\right)$, and as such it is the case according to Definition 4.3.1 that:

$$\mathsf{L}\left(\nu_1 \boxminus \nu_2\right) = \left(\left(\mathrm{diag}\left(W_1, \mathfrak{W}_1\right), b_1 \frown \mathfrak{b}_1\right),\right.$$

$$\left(\mathrm{diag}\left(W_2, \mathfrak{W}_2\right), b_2 \frown \mathfrak{b}_2\right),$$

$$\vdots$$

$$\left.\left(\mathrm{diag}\left(W_L, \mathfrak{W}_L\right), b_L^1 \frown \mathfrak{b}_L\right)\right)$$

Note that for all, $\mathfrak{a} \in \left(\mathbb{R}, \mathbb{R}\right)$, $j \in \{1, 2, \ldots, L-1\}$ and for all $x \in \mathbb{R}^{\mathrm{columns}(W_j)+\mathrm{columns}(\mathfrak{W}_j)}$, $x_1 \in \mathbb{R}^{\mathrm{columns}(W_j)}$, $x_2 \in \mathbb{R}^{\mathrm{columns}(\mathfrak{W}_j)}$, $y \in \mathbb{R}^{\mathrm{rows}(W_j)+\mathrm{rows}(\mathfrak{W}_j)}$, $y_1 \in \mathbb{R}^{\mathrm{rows}(W_j)}$, $y_2 \in \mathbb{R}^{\mathrm{rows}(\mathfrak{W}_j)}$, where
$y_1 = \mathfrak{a}\left(\left[W_j \cdot x_1 + b_1\right]_{*,*}\right)$, $y_2 = \mathfrak{a}\left(\left[\mathfrak{W}_j \cdot x_2 + \mathfrak{b}_j\right]_{*,*}\right)$, $y = \mathfrak{a}\left(\left[\mathrm{diag}\left(W_j, \mathfrak{W}_j\right) \cdot x + \left(b_j \frown \mathfrak{b}_j\right)\right]_{*,*}\right)$

it is the case that, Corollary 1.2.36.1 tells us that:

$$y = \mathfrak{a}\left([\mathrm{diag}\left(W_j, \mathfrak{W}_j\right) \cdot x + (b_j \frown \mathfrak{b}_j)]_{*,*}\right) = \mathfrak{a}\left([(W_j \cdot x_1 + b_j) \frown (\mathfrak{W}_j \cdot x_2 + \mathfrak{b}_j)]_{*,*}\right)$$

$$= \mathfrak{a}\left([W_j \cdot x_1 + b_j]_{*,*}\right) \frown \mathfrak{a}\left([\mathfrak{W}_j \cdot x_2 + \mathfrak{b}_j]_{*,*}\right)$$

$$= y_1 \frown y_2 \tag{4.3.1}$$

Note that this is repeated from one layer to the next, yielding that $\mathfrak{I}_{\mathfrak{r}}\left(\nu_1 \boxminus \nu_2\right)(\mathfrak{x}) = \mathfrak{I}_{\mathfrak{r}}\left(\nu_1\right)(x_1) \frown \mathfrak{I}_{\mathfrak{r}}\left(\nu_2\right)(x_2)$. $\qquad\qquad \square$

Remark 4.3.4. *Given $n, L \in \mathbb{N}$, $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$ such that $L = \mathsf{D}(\nu_1) = \mathsf{D}(\nu_2) = ... = \mathsf{D}(\nu_n)$ it is then the case, as seen from (??) that:*

$$\boxminus_{i=1}^n \nu_i \in \left(\underset{k=1}{\overset{L}{\times}} \left[\mathbb{R}^{\left(\sum_{j=1}^n \mathsf{W}_k(\nu_j)\right) \times \left(\sum_{j=1}^n \mathsf{W}_{k-1}(\nu_j)\right)} \times \mathbb{R}^{\left(\sum_{j=1}^n \mathsf{W}_k(\nu_j)\right)}\right]\right) \tag{4.3.2}$$

Lemma 4.3.5. *Let $n, L \in \mathbb{N}$, $\nu_1, \nu_2, \ldots, \nu_n \in \mathsf{NN}$ satisfty that $L = \mathsf{D}\left(\nu_1\right) = \mathsf{D}\left(\nu_2\right) = \cdots = \mathsf{D}\left(\nu_n\right)$. It is then the case that:*

$$\mathsf{P}\left([\boxminus_{i=1}^n \nu_i]\right) \leqslant \frac{1}{2}\left[\sum_{i=1}^n \mathsf{P}\left(\nu_i\right)\right]^2 \tag{4.3.3}$$

Proof. Note that by Remark 4.4.2 we have that:

$$
\begin{aligned}
\mathsf{P}\left(\left[\boxminus_{i=1}^n \nu_i\right]\right) &= \sum_{k=1}^L \left[\sum_{i=1}^n l_{i,k}\right]\left[\left(\sum_{i=1}^n l_{i,k-1}\right)+1\right] \\
&= \sum_{k=1}^L \left[\sum_{i=1}^n l_{i,k}\right]\left[\left(\sum_{j=1}^n l_{j,k-1}\right)+1\right] \\
&\leqslant \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^L l_{i,k}\left(l_{j,k-1}+1\right) \\
&\leqslant \sum_{i=1}^n\sum_{j=1}^n\sum_{\ell=1}^L l_{i,k}\left(l_{j,\ell-1}+1\right) \\
&= \sum_{i=1}^n\sum_{j=1}^n\left[\sum_{k=1}^L l_{i,k}\right]\left[\sum_{\ell=1}^L \left(l_{j,\ell-1}+1\right)\right] \\
&\leqslant \sum_{i=1}^n\sum_{j=1}^n\left[\sum_{k=1}^L \frac{1}{2}l_{i,k}\left(l_{i,k-1}+1\right)\right]\left[\sum_{\ell=1}^L l_{j,\ell}\left(l_{j,\ell-1}+1\right)\right] \\
&= \sum_{i=1}^n\sum_{j=1}^n\frac{1}{2}\mathsf{P}\left(\nu_i\right)\mathsf{P}\left(\nu_j\right) = \frac{1}{2}\left[\sum_{i=1}^n\mathsf{P}\left(\nu_i\right)\right]^2
\end{aligned}
\tag{4.3.4}
$$

This completes the proof of the lemma. $\qquad\square$

Corollary 4.3.5.1. *Let $n \in \mathbb{N}$. Let $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$ satisfy that $\mathsf{L}\left(\nu_1\right) = \mathsf{L}\left(\nu_2\right) = \cdots = \mathsf{L}\left(\nu_n\right)$. It is then the case that:*

$$
\mathsf{P}\left(\boxminus_{i=1}^n \nu_i\right) \leqslant n^2\,\mathsf{P}\left(\nu_1\right)
\tag{4.3.5}
$$

Proof. Since it is the case that for all $j \in \{1, 2, ..., n\}$ that: $\mathsf{L}\left(\nu_j\right) = \left(l_0, l_1, ..., l_L\right)$, where $l_0, l_1, ..., l_L, L \in \mathbb{N}$, we may say that:

$$
\begin{aligned}
\mathsf{P}\left(\boxminus_{j=1}^n \nu_j\right) &= \sum_{j=1}^L \left(nl_j\right)\left[\left(nl_{j-1}\right)+1\right] \leqslant \sum_{j=1}^L \left(nl_j\right)\left[\left(nl_{j-1}\right)+n\right] \\
&= n^2\left[\sum_{j=1}^L l_j\left(l_{j-1}+1\right)\right] = n^2\,\mathsf{P}\left(\nu_1\right)
\end{aligned}
\tag{4.3.6}
$$

$\square$

Lemma 4.3.6. *Let* $\nu_1, \nu_2 \in \mathsf{NN}$, *such that* $\mathsf{D}(\nu_1) = \mathsf{D}(\nu_2) = L$. *It is then the case that* $\mathsf{P}(\nu_1) + \mathsf{P}(\nu_2) \leqslant \mathsf{P}(\nu_1 \boxminus \nu_2)$.

Proof. Remark 4.3.4 tells us that:

$$\nu_1 \boxminus \nu_2 \in \left(\underset{k=1}{\overset{L}{\times}} \left[\mathbb{R}^{(\mathsf{W}_k(\nu_1)+\mathsf{W}_k(\nu_2)) \times (\mathsf{W}_{k-1}(\nu_1)+\mathsf{W}_{k-1}(\nu_2))} \times \mathbb{R}^{\mathsf{W}_k(\nu_1)+\mathsf{W}_k(\nu_2)} \right] \right) \qquad (4.3.7)$$

The definition of $\mathsf{P}()$ from Defition 4.1.3, and the fact that $\mathsf{W}_i \geqslant 1$ for all $i \in \{1, 2, \ldots, L\}$ tells us then that:

$$
\begin{aligned}
\mathsf{P}(\nu_1 \boxminus \nu_2) &= \sum_{k=1}^{L} \left[(\mathsf{W}_k(\nu_1) + \mathsf{W}_k(\nu_2)) \times (\mathsf{W}_{k-1}(\nu_1) + \mathsf{W}_{k-1}(\nu_2) + 1) \right] \\
&= \sum_{k=1}^{L} \left[\mathsf{W}_k(\nu_1)\mathsf{W}_{k-1}(\nu_1) + \mathsf{W}_k(\nu_1)\mathsf{W}_{k-1}(\nu_2) \right. \\
&\quad \left. + \mathsf{W}_k(\nu_1) + \mathsf{W}_k(\nu_2)\mathsf{W}_{k-1}(\nu_1) + \mathsf{W}_k(\nu_2)\mathsf{W}_{k-1}(\nu_2) + \mathsf{W}_k(\nu_2) \right] \\
&\geqslant \sum_{k=1}^{L} \left[\mathsf{W}_k(\nu_1)\mathsf{W}_{k-1}(\nu_1) + \mathsf{W}_k(\nu_1) + \mathsf{W}_k(\nu_2)\mathsf{W}_{k-1}(\nu_2) + \mathsf{W}_k(\nu_2) \right] \\
&= \sum_{k=1}^{L} \left[\mathsf{W}_k(\nu_1)(\mathsf{W}_{k-1}(\nu_1) + 1) \right] + \sum_{k=1}^{L} \left[\mathsf{W}_k(\nu_2)(\mathsf{W}_{k-1}(\nu_2) + 1) \right] \\
&= \mathsf{P}(\nu_1) + \mathsf{P}(\nu_2) \qquad (4.3.8)
\end{aligned}
$$

$\square$

Corollary 4.3.6.1. *Let* $\nu_1, \nu_2, \nu_3 \in \mathsf{NN}$. *Let* $\mathsf{P}(\nu_2) \leqslant \mathsf{P}(\nu_3)$. *It is then the case that* $\mathsf{P}(\nu_1 \boxminus \nu_2) \leqslant \mathsf{P}(\nu_1 \boxminus \nu_3)$.

Proof. Lemma 4.3.6 tells us that:

$$0 \leqslant \mathsf{P}(\nu_1) + \mathsf{P}(\nu_3) \leqslant \mathsf{P}(\nu_1 \boxminus \nu_3) \qquad (4.3.9)$$

$$0 \leqslant \mathsf{P}(\nu_1) + \mathsf{P}(\nu_2) \leqslant \mathsf{P}(\nu_1 \boxminus \nu_2) \qquad (4.3.10)$$

Subtracting (4.3.9) from (4.3.10) gives us that:

$$0 \leqslant \mathsf{P}\left(\nu_3\right) - \mathsf{P}\left(\nu_2\right) \leqslant \mathsf{P}\left(\nu_1 \boxminus \nu_3\right) - \mathsf{P}\left(\nu_1 \boxminus \nu_2\right)$$

$$\mathsf{P}\left(\nu_1 \boxminus \nu_2\right) \leqslant \mathsf{P}\left(\nu_1 \boxminus \nu_2\right)$$

This completes the proof of the Corollary. $\qquad\square$

Lemma 4.3.7. *Let* $m_1, m_2, n_1, n_2 \in \mathbb{N}$. *Let* $\nu_1, \nu_2 \in \mathsf{NN}$, *such that* $\mathfrak{I}_{\mathfrak{r}}\left(\nu_1\right) \in C\left(\mathbb{R}^{m_1}, \mathbb{R}^{n_1}\right)$ *and* $\mathfrak{I}_{\mathfrak{r}}\left(\nu_2\right) \in C\left(\mathbb{R}^{m_2}, \mathbb{R}^{n_2}\right)$. *It is then the case that* $\left(\mathfrak{I}_{\mathfrak{a}}(\nu_1 \boxminus \nu_2)\right)\left(\begin{bmatrix} x \\ x' \end{bmatrix}\right) = \left(\mathfrak{I}_{\mathfrak{a}}(\nu_2 \boxminus \nu_1)\right)\left(\begin{bmatrix} x' \\ x \end{bmatrix}\right)$ *for* $x \in \mathbb{R}^{m_1}, x' \in \mathbb{R}^{n_1}$, *upto transposition.*

Proof. Note that this is a consequence of the commutativity of summation in the exponents of (4.3.2), and the fact that switching ν_1 and ν_2 with a transposition results in a transposed output for transposed input. $\qquad\square$

Lemma 4.3.8. *Let* $\mathfrak{a} \in C\left(\mathbb{R}, \mathbb{R}\right)$, $n \in \mathbb{N}$, *and* $\nu = \boxminus_{i=1}^n \nu_i$ *satisfy the condition that* $\mathsf{D}(\nu_1) = \mathsf{D}(\nu_2) = ... = \mathsf{D}(\nu_n)$. *It is then the case that* $\mathfrak{I}_{\mathfrak{a}}\left(\nu\right) \in C\left(\mathbb{R}^{\sum_{i=1}^n \mathsf{I}(\nu_i)}, \mathbb{R}^{\sum_{i=1}^n \mathsf{O}(\nu_i)}\right)$

Proof. Let $L = \mathsf{D}(\nu_1)$, and let $l_{i,0}, l_{i,1}...l_{i,L} \in \mathbb{N}$ satisfy for all $i \in \{1, 2, ..., n\}$ that $\mathsf{L}(\nu_i) = (l_{i,0}, l_{i,1}, ..., l_{i,L})$. Furthermore let $\left(\left(W_{i,1}, b_{i,1}\right), \left(W_{i,2}, b_{i,2}\right), ..., \left(W_{i,L}, b_{i,L}\right)\right) \in \left(\bigtimes_{j=1}^L \left[\mathbb{R}^{l_{i,j} \times l_{i,j-1}} \times \mathbb{R}^{l_{i,j}}\right]\right)$ satisfy for all $i \in \{1, 2, ..., n\}$ that:

$$\nu_i = \left(\left(W_{i,1}, b_{i,1}\right), \left(W_{i,2}, b_{i,2}\right), ..., \left(W_{i,L}, b_{i,L}\right)\right) \qquad (4.3.11)$$

Let $\alpha_j \in \mathbb{N}$ with $j \in \{0, 1, ..., L\}$ satisfy that $\alpha_j = \sum_{i=1}^n l_{i,j}$ and let $\left(\left(A_1, b_1\right), \left(A_2, b_2\right), ..., \left(A_L, b_L\right)\right) \in \left(\bigtimes_{j=1}^L \left[\mathbb{R}^{\alpha_j \times \alpha_{j-1}} \times \mathbb{R}^{\alpha_j}\right]\right)$ satisfy that:

$$\boxminus_{i=1}^n \nu_i = \left(\left(A_1, b_1\right), \left(A_2, b_2\right), ..., \left(A_L, b_L\right)\right) \qquad (4.3.12)$$

See Remark 5.3.2. Let $x_{i,0}, x_{i,1}, ..., x_{i,L-1} \in \left(\mathbb{R}^{l_{i,0}} \times \mathbb{R}^{l_{i,1}} \times \cdots \times \mathbb{R}^{l_{i,L-1}}\right)$ satisfy for all $i \in$

$\{1, 2, ..., n\}$ $k \in \mathbb{N} \cap (0, L)$ that:

$$x_{i,j} = \text{Mult}_{\mathfrak{a}}^{l_{i,j}} \left(W_{i,j} x_{i,j-1} + b_{i,j} \right) \tag{4.3.13}$$

Note that (4.3.12) demonstrates that $\mathsf{I} \left(\boxminus_{i=1}^{n} \nu_i \right) = \alpha_0$ and $\mathsf{O} \left(\boxminus_{i=1}^{n} \nu_i \right) = \alpha_L$. This and Item(ii) of Lemma 4.1.7, and the fact that for all $i \in \{1, 2, ..., n\}$ it is the case that $\mathsf{I}(\nu_i) = l_{i,0}$ and $\mathsf{O}(\nu_i) = l_{i,L}$ ensures that:

$$\begin{aligned} \mathfrak{I}_{\mathfrak{a}} \left(\boxminus_{i=1}^{n} \right) \in C \left(\mathbb{R}^{\alpha_0}, \mathbb{R}^{\alpha_L} \right) &= C \left(\mathbb{R}^{\sum_{i=1}^{n} l_{i,0}}, \mathbb{R}^{\sum_{i=1}^{n} l_{i,L}} \right) \\ &= C \left(\mathbb{R}^{\sum_{i=1}^{n} \mathsf{I}(\nu_i)}, \mathbb{R}^{\sum_{i=1}^{n} \mathsf{O}(\nu_i)} \right) \end{aligned}$$

This proves the lemma. $\qquad\square$

4.3.2 Stacking of ANNs of Unequal Depth

We will often encounter neural networks that we want to stack but have unequal depth. Definition 4.3.1 only deals with neural networks of the same depth. We will facilitate this situation by introducing a form of padding for our shorter neural network. Hence, they come out to the same length before stacking them. This padding will be via the tunneling neural network, as shown below.

Definition 4.3.9 (Identity Neural Network)**.** *Let $d \in \mathbb{N}$. We will denote by $\mathsf{Id}_d \in \mathsf{NN}$ the neural network satisfying for all $d \in \mathbb{N}$ that:*

(i)

$$\mathsf{Id}_1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \in \left(\left(\mathbb{R}^{2 \times 1} \times \mathbb{R}^2 \right) \times \left(\mathbb{R}^{1 \times 2} \times \mathbb{R}^1 \right) \right) \tag{4.3.14}$$

86

(ii)

$$\mathsf{Id}_d = \boxminus_{i=1}^{d} \mathsf{Id}_1 \tag{4.3.15}$$

For $d \in \mathbb{N} \cap [2, \infty)$.

Remark 4.3.10. *We will discuss some properties of Id_d in Section 5.2.*

Definition 4.3.11 (The Tunneling Neural Network). *We define the tunneling neural network, denoted as Tun_n for $n \in \mathbb{N}$ and $d \in \mathbb{N}$ by:*

$$\mathsf{Tun}_n^d = \begin{cases} \mathsf{Aff}_{\mathbb{I}_d, 0} & : n = 1 \\ \mathsf{Id}_d & : n = 2 \\ \bullet^{n-2} \mathsf{Id}_d & n \in \mathbb{N} \cap [3, \infty) \end{cases} \tag{4.3.16}$$

We will drop the requirement for d and Tun_n by itself will be used to denote Tun_n^1.

Remark 4.3.12. *We will discuss some properties of the Tun_n^d network in Section 5.2. We will also discuss properties of wider tunneling neural network in Lemma 5.2.6.*

Definition 4.3.13. *Let $n \in \mathbb{N}$, and $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$. We will define the stacking of unequal length neural networks, denoted $\diamondsuit_{i=1}^{n} \nu_i$ as the neural network given by:*

$$\diamondsuit_{i=1}^{n} \nu_i = \boxminus_{i=1}^{n} \left[\mathsf{Tun}_{\max_i \{\mathsf{D}(\nu_i)\} + 1 - \mathsf{D}(\nu_i)}^{\mathsf{O}(\nu_i)} \bullet \nu_i \right] \tag{4.3.17}$$

Diagrammatically, this can be thought of as shown below.

Lemma 4.3.14. *Let $\nu_1, \nu_2 \in \mathsf{NN}$. It is then the case that:*

$$\mathsf{P}(\nu_1 \diamondsuit \nu_2) \leqslant 2 \cdot (\max \{\mathsf{P}(\nu_1), \mathsf{P}(\nu_2)\})^2 \tag{4.3.18}$$

Proof. This is a straightforward consequence of Lemma 4.3.5. $\qquad \square$
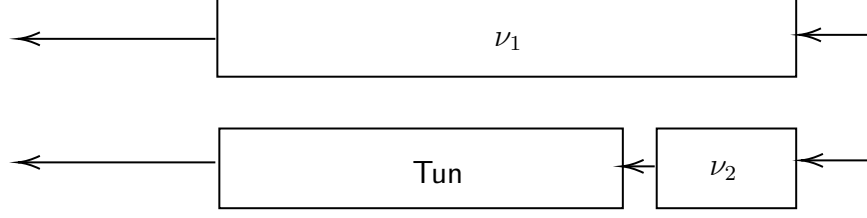
Figure 4.2: Diagrammmatic representation of the stacking of unequal depth neural networks

4.4 Affine Linear Transformations as ANNs and Their Properties.

Affine neural networks present an important class of neural networks. By virtue of them being only one layer deep, they may be instantiated with any activation function whatsoever and still retain their affine transformative properties, see Definition 4.1.5. In addition, when composing, they are subsumed into the function being somposed to, i.e. they do not change the depth of a neural network once composed into it, see Lemma 4.2.5.

Definition 4.4.1. *Let* $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. *We denote by* $\mathsf{Aff}_{W,b} \in (\mathbb{R}^{m \times n} \times \mathbb{R}^m) \subsetneq$ NN *the neural network given by* $\mathsf{Aff}_{W,b} = ((W, b))$.

Lemma 4.4.2. *Let* $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. *It is then the case that:*

(i) $\mathsf{L}(\mathsf{Aff}_{W,b}) = (n, m) \in \mathbb{N}^2$.

(ii) *for all* $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *it is the case that* $\mathfrak{I}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}) \in C(\mathbb{R}^n, \mathbb{R}^m)$

(iii) *for all* $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ *we have* $(\mathfrak{I}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}))(x) = Wx + b$

Proof. Note that (i) is a consequence of Definition 4.1.3 and 4.4.1. Note next that $\mathsf{Aff}_{W,b} = (W, b) \in (\mathbb{R}^{m \times n} \times \mathbb{R}^m) \subsetneq \mathsf{NN}$. Note that (4.1.11) then tells us that $\mathfrak{I}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}) = Wx + b$ which in turn proves (ii) and (iii) $\qquad\square$

Remark 4.4.3. *Given* $W \in \mathbb{R}^{m \times n}$, *and* $b \in \mathbb{R}^{m \times 1}$, *it is the case that according to Definition (4.1.4) we have:* $\mathsf{P}(\mathsf{Aff}_{W,b}) = m \times n + m$

Remark 4.4.4. *For an* R *implementation see Listing 9.6*

Lemma 4.4.5. *Let $\nu \in$ NN. It is then the case that:*

(i) *For all $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathsf{O}(\nu)}$*

$$\mathsf{L}(\mathsf{Aff}_{W,B} \bullet \nu) = \big(\mathsf{W}_0(\nu), \mathsf{W}_1(\nu), ..., \mathsf{W}_{\mathsf{D}(\nu)-1}(\nu), m\big) \in \mathbb{N}^{\mathsf{D}(\nu)+1} \qquad (4.4.1)$$

(ii) *For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathsf{O}(\nu)}$, $B \in \mathbb{R}^m$, we have that $\mathfrak{I}_\mathfrak{a}(\mathsf{Aff}_{W,B} \bullet \nu) \in$*
   $C\left(\mathbb{R}^{\mathsf{I}(\nu)}, \mathbb{R}^m\right)$.

(iii) *For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathsf{O}(\nu)}$, $B \in \mathbb{R}^m$, $x \in \mathbb{R}^{\mathsf{I}(\nu)}$ that:*

$$\left(\mathfrak{I}\left(\mathsf{Aff}_{W,b} \bullet \nu\right)\right)(x) = W\left(\mathfrak{I}_\mathfrak{a}(\nu)\right)(x) + b \qquad (4.4.2)$$

(iv) *For all $n \in \mathbb{N}$, $W \in \mathbb{R}^{\mathsf{I}(\nu) \times n}$, $b \in \mathbb{R}^{\mathsf{I}(\nu)}$ that:*

$$\mathsf{L}(\nu \bullet \mathsf{Aff}_{W,b}) = \big(n, \mathsf{W}_1(\nu), \mathsf{W}_2(\nu), ..., \mathsf{W}_{\mathsf{D}(\nu)}(\nu)\big) \in \mathbb{N}^{\mathsf{D}(\nu)+1} \qquad (4.4.3)$$

(v) *For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $W \in \mathbb{R}^{\mathsf{I}(\nu) \times n}$, $b \in \mathbb{R}^{\mathsf{I}(\nu)}$ that $\mathfrak{I}_\mathfrak{a}(\nu \bullet \mathsf{Aff}_{W,b}) \in C\left(\mathbb{R}^n, \mathbb{R}^{\mathsf{O}(\nu)}\right)$*
   and,

(vi) *For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $W \in \mathbb{R}^{\mathsf{I}(\nu) \times n}$, $b \in \mathbb{R}^{\mathsf{I}(\nu)}$, $x \in \mathbb{R}^n$ that:*

$$\left(\mathfrak{I}_\mathfrak{a}(\nu \bullet \mathsf{Aff}_{W,b})\right)(x) = \left(\mathfrak{I}_\mathfrak{a}(\nu)\right)(Wx + b) \qquad (4.4.4)$$

Proof. From Lemma 4.4.2 we see that $\mathfrak{I}_\mathfrak{a}(\mathsf{Aff}_{W,b}) \in C(\mathbb{R}^n, \mathbb{R}^m)$ given by $\mathfrak{I}_\mathfrak{a}(\mathsf{Aff}_{W,b}) = Wx + b$. This and Lemma 4.2.5 prove $(i) - (vi)$. $\qquad \square$

Corollary 4.4.5.1. *Let $m, n \in \mathbb{N}$, and $W \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Let $\nu \in$ NN. It is then the case that:*

(i) for all $\mathsf{Aff}_{W,b} \in \mathsf{NN}$ *with* $\mathsf{I}(\mathsf{Aff}_{W,b}) = \mathsf{O}(\nu)$ *that:*

$$\mathsf{P}(\mathsf{Aff}_{W,b} \bullet \nu) \leqslant \left[\max\left\{1, \frac{\mathsf{O}(\mathsf{Aff}_{W,b})}{l_L}\right\}\right] \mathsf{P}(\nu) \tag{4.4.5}$$

(ii) for all $\mathsf{Aff}_{W,b} \in \mathsf{NN}$ *with* $\mathsf{O}(\mathsf{Aff}_{W,b}) = \mathsf{I}(\nu)$ *that:*

$$\mathsf{P}(\nu \bullet \mathsf{Aff}_{W,b}) \leqslant \left[\max\left\{1, \frac{\mathsf{I}(\mathsf{Aff}_{W,b}) + 1}{\mathsf{I}(\nu) + 1}\right\}\right] \mathsf{P}(\nu) \tag{4.4.6}$$

Proof. Let it be the case that $\mathsf{L}(\nu) = (l_0, l_1, ..., l_L)$ for $l_0, l_1, ..., l_L, L \in \mathbb{N}$. Lemma 4.4.5, Item (i), and Lemma 4.2.5 then tells us that:

$$\mathsf{P}(\mathsf{Aff}_{W,b} \bullet \nu) \tag{4.4.7}$$

$$= \left[\sum_{m=1}^{L-1} l_m(l_{m-1} + 1)\right] + \mathsf{O}(\mathsf{Aff}_{W,b})(l_{L-1} + 1)$$

$$= \left[\sum_{m=1}^{L-1} l_m(l_{m-1} + 1)\right] + \left[\frac{\mathsf{O}(\mathsf{Aff}_{W,b})}{l_L}\right] l_L(l_{L-1} + 1)$$

$$\leqslant \left[\max\left\{1, \frac{\mathsf{O}(\mathsf{Aff}_{W,b})}{l_L}\right\}\right]\left[\sum_{m=1}^{L-1} l_m(l_{m-1} + 1)\right] + \left[\max\left\{1, \frac{\mathsf{O}(\mathsf{Aff}_{W,b})}{l_L}\right\}\right] l_L(l_{L-1} + 1)$$

$$= \left[\max\left\{1, \frac{\mathsf{O}(\mathsf{Aff}_{W,b})}{l_L}\right\}\right]\left[\sum_{m=1}^{L} l_m(l_{m-1} + 1)\right] = \left[\max\left\{1, \frac{\mathsf{O}(\mathsf{Aff}_{W,b})}{l_L}\right\}\right] \mathsf{P}(\nu)$$

and further that:

$$\mathsf{P}\left(\nu \bullet \mathsf{Aff}_{W,b}\right) \tag{4.4.8}$$

$$= \left[\sum_{m=2}^{L} l_m \left(l_{m-1} + 1\right)\right] + l_1 \left(\mathsf{I}\left(\mathsf{Aff}_{W,b}\right) + 1\right)$$

$$= \left[\sum_{m=2}^{L} l_m \left(l_{m-1} + 1\right)\right] + \left[\frac{\mathsf{I}\left(\mathsf{Aff}_{W,b}\right) + 1}{l_0 + 1}\right] l_1 \left(l_0 + 1\right)$$

$$\leqslant \left[\max\left\{1, \frac{\mathsf{I}(\mathsf{Aff}_{W,b}) + 1}{l_0 + 1}\right\}\right] \left[\sum_{m=2}^{L} l_m \left(l_{m-1} + 1\right)\right] + \left[\max\left\{1, \frac{\mathsf{I}\left(\mathsf{Aff}_{W,b}\right) + 1}{l_0 + 1}\right\}\right] l_1 \left(l_0 + 1\right)$$

$$= \left[\max\left\{1, \frac{\mathsf{I}\left(\mathsf{Aff}_{W,b}\right) + 1}{l_0 + 1}\right\}\right] \left[\sum_{m=1}^{L} l_m \left(l_{m-1} + 1\right)\right] = \left[\max\left\{1, \frac{\mathsf{I}\left(\mathsf{Aff}_{W,b}\right) + 1}{\mathsf{I}\left(\nu\right) + 1}\right\}\right] \mathsf{P}\left(\nu\right)$$

This completes the proof of the lemma. $\qquad\square$

Lemma 4.4.6. *Let $\mathfrak{a}_1, \mathfrak{a}_2$ be two affine neural networks as defined in Definition 4.4.1. It is then the case that $\mathfrak{a}_1 \boxminus \mathfrak{a}_2$ is also an affine neural network*

Proof. This follows straightforwardly from Definition 4.3.1, where, given that $\mathfrak{a}_1 = ((W_1, b_1))$, and $\mathfrak{a}_2 = ((W_2, b_2))$, their stackings is the neural network $((\mathrm{diag}\,(W_1, W_2), b_1 \frown b_2))$, which is clearly an affine neural network. $\qquad\square$

4.5   Sums of ANNs of Same End-widths

Definition 4.5.1 (The $\mathsf{Cpy}_{n,k}$ Network). *We define the neural network, $\mathsf{Cpy}_{n,k} \in \mathsf{NN}$ for $n, k \in \mathbb{N}$ as the neural network given by:*

$$\mathsf{Cpy}_{n,k} = \mathsf{Aff}_{\underbrace{\left[\mathbb{I}_k\ \mathbb{I}_k\ \cdots\ \mathbb{I}_k\right]^T}_{n-many}, \mathbb{0}_{nk}} \tag{4.5.1}$$

Where k represents the dimensions of the vectors being copied and n is the number of copies of the vector being made.

Remark 4.5.2. *See Listing 9.6.*

Lemma 4.5.3. *Let $n, k \in \mathbb{N}$ and let $\mathsf{Cpy}_{n,k} \in \mathsf{NN}$, it is then the case for all $n, k \in \mathbb{N}$ that:*

(i) $\mathsf{D}\left(\mathsf{Cpy}_{n,k}\right) = 1$

(ii) $\mathsf{P}\left(\mathsf{Cpy}_{n,k}\right) = nk^2 + nk$

Proof. Note that (i) is a consequence of Definition 4.4.1, and (ii) follows from the structure of $\mathsf{Cpy}_{n,k}$. $\qquad\qquad\square$

Definition 4.5.4 (The $\mathsf{Sum}_{n,k}$ Network). *We define the neural network $\mathsf{Sum}_{n,k}$ for $n, k \in \mathbb{N}$ as the neural network given by:*

$$\mathsf{Sum}_{n,k} = \mathsf{Aff}_{\underbrace{[\mathbb{I}_k \; \mathbb{I}_k \; \cdots \; \mathbb{I}_k]}_{n-many},\mathbb{0}_k} \tag{4.5.2}$$

Where k represents the dimensions of the vectors being added and n is the number of vectors being added.

Remark 4.5.5. *See again, Listing 9.6*

Lemma 4.5.6. *Let $n, k \in \mathbb{N}$ and $\mathsf{Sum}_{n,k} \in \mathsf{NN}$, it is then the case for all $n, k \in \mathbb{N}$ that:*

(i) $\mathsf{D}\left(\mathsf{Sum}_{n,k}\right) = 1$

(ii) $\mathsf{P}\left(\mathsf{Sum}_{n,k}\right) = nk^2 + k$

Proof. (i) is a consequence of Definition 4.4.1 and (ii) follows from the structure of $\mathsf{Sum}_{n,k}$. $\qquad\qquad\square$

Definition 4.5.7 (Sum of ANNs of the same depth and same end widths). *Let $u, v \in \mathbb{Z}$ with $u \leqslant v$. Let $\nu_u, \nu_{u+1}, ..., \nu_v \in \mathsf{NN}$ satisfy for all $i \in \mathbb{N} \cap [u, v]$ that $\mathsf{D}(\nu_i) = \mathsf{D}(\nu_u)$, $\mathsf{I}(\nu_i) = \mathsf{I}(\nu_u)$, and $\mathsf{O}(\nu_i) = \mathsf{O}(\nu_u)$. We then denote by $\oplus_{i=u}^n \nu_i$ or alternatively $\nu_u \oplus \nu_{u+1} \oplus \ldots \oplus \nu_v$ the neural network given by:*

$$\oplus_{i=u}^v \nu_i := \left(\mathsf{Sum}_{v-u+1,\mathsf{O}(\nu_2)} \bullet [\boxminus_{i=u}^v \nu_i] \bullet \mathsf{Cpy}_{(v-u+1),\mathsf{I}(\nu_1)}\right) \tag{4.5.3}$$

Remark 4.5.8. *For an* R *implementation, see Listing 9.9.*

Remark 4.5.9. *We may diagrammatically refer to this network as:*
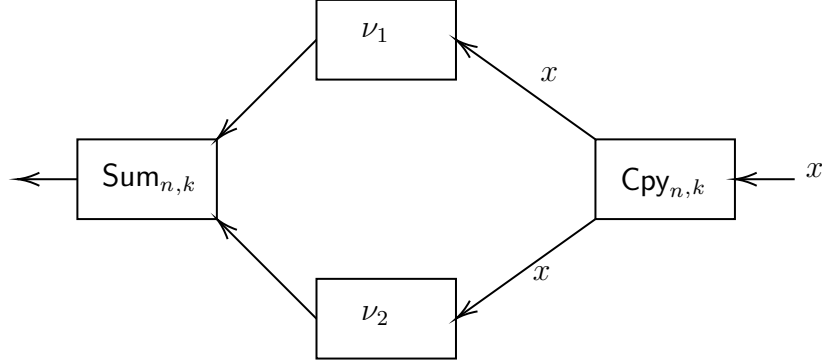
Figure 4.3: Neural Network diagram of a neural network sum.

4.5.1 Neural Network Sum Properties

Lemma 4.5.10. *Let* $\nu_1, \nu_2 \in \mathsf{NN}$ *satisfy that* $\mathsf{D}(\nu_1) = \mathsf{D}(\nu_2) = L$, $\mathsf{I}(\nu_1) = \mathsf{I}(\nu_2)$, *and* $\mathsf{O}(\nu_1) = \mathsf{O}(\nu_2)$, *and* $\mathsf{L}(\nu_1) = (l_{1,1}, l_{1,2}, ... l_{1,L})$ *and* $\mathsf{L}(\nu_2) = (l_{2,1}, l_{2,2}, ..., l_{2,L})$ *it is then the case that:*

$$\mathsf{P}\left(\nu_1 \oplus \nu_2\right) = \mathsf{P}\left(\mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{O}(\nu_2)} \, \mathbb{I}_{\mathsf{O}(\nu_2)}\right], \mathbb{0}_{\mathsf{O}(\nu_2)}} \bullet \left[\nu_1 \boxminus \nu_2\right] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{I}(\nu_1)} \, \mathbb{I}_{\mathsf{I}(\nu_1)}\right]^T, \mathbb{0}_{2 \cdot \mathsf{I}(\nu_1)}}\right) \qquad (4.5.4)$$

$$\leqslant \frac{1}{2}\left(\mathsf{P}\left(\nu_1\right) + \mathsf{P}\left(\nu_2\right)\right)^2$$

Proof. Note that by Lemma 4.3.5 we have that:

$$\mathsf{P}\left(\nu_1 \boxminus \nu_2\right) = \frac{1}{2}\left(\mathsf{P}\left(\nu_1\right) + \mathsf{P}\left(\nu_2\right)\right)^2 \qquad (4.5.5)$$

Note also that since Cpy and Sum are affine neural networks, from Corollary 4.4.5.1 we get

that:

$$P\left([\nu_1 \boxminus \nu_2] \bullet \mathsf{Cpy}_{2,\mathsf{I}(\nu_1)}\right) \leqslant \max\left\{1, \frac{\mathsf{I}(\nu_1)+1}{2\,\mathsf{I}(\nu_1)+1}\right\} \frac{1}{2}\left(\mathsf{P}(\nu_1) + \mathsf{P}(\nu_2)\right)^2$$

$$= \frac{1}{2}\left(\mathsf{P}(\nu_1) + \mathsf{P}(\nu_2)\right)^2 \tag{4.5.6}$$

and further that:

$$\mathsf{P}\left(\mathsf{Sum}_{2,\mathsf{O}(\nu_1\boxminus\nu_2)} \bullet [\nu_1 \boxminus \nu_2] \bullet \mathsf{Cpy}_{2,\mathsf{I}(\nu_1)}\right) \leqslant \left[\max\left\{1, \frac{\mathsf{O}(\mathsf{Aff}_{W,b})}{2\,\mathsf{O}(\nu_1)}\right\}\right] \frac{1}{2}\left(\mathsf{P}(\nu_1) + \mathsf{P}(\nu_2)\right)^2$$

$$= \frac{1}{2}\left(\mathsf{P}(\nu_1) + \mathsf{P}(\nu_2)\right)^2 \tag{4.5.7}$$

$\square$

Corollary 4.5.10.1. *Let* $n \in \mathbb{N}$. *Let* $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$ *satisfy that* $\mathsf{L}(\nu_1) = \mathsf{L}(\nu_2) = \cdots = \mathsf{L}(\nu_n)$. *It is then the case that:*

$$\mathsf{P}\left(\bigoplus_{i=1}^{n} \nu_i\right) \leqslant n^2 \cdot \mathsf{P}(\nu_1) \tag{4.5.8}$$

Proof. Let $\mathsf{L}(\nu_1) = (l_0, l_1, ..., l_L)$ where for all $i \in \{0, 1, ..., L\}$ it is the case that $l_i, L \in \mathbb{N}$. Corollary 4.3.5.1 then tells us that:

$$\mathsf{P}\left(\boxminus_{i=1}^{n}\nu_i\right) \leqslant n^2\,\mathsf{P}(\nu_i) \tag{4.5.9}$$

Then from Corollary 4.4.5.1, and (4.5.6) we get that:

$$\mathsf{P}\left([\boxminus_{i=1}^{n}\nu_i] \bullet \mathsf{Cpy}_{2,\mathsf{I}(\nu_1)}\right) \leqslant n^2\,\mathsf{P}(\nu_1) \tag{4.5.10}$$

And further that:

$$\mathsf{P}\left(\mathsf{Sum}_{2,\mathsf{O}\left(\boxminus_{i=1}^{n}\nu_i\right)} \bullet [\boxminus_{i=1}^{n}\nu_i] \bullet \mathsf{Cpy}_{2,\mathsf{I}(\nu_1)}\right) \leqslant n^2\,\mathsf{P}(\nu_1) \tag{4.5.11}$$

94

□

Lemma 4.5.11. *Let* $\nu_1, \nu_2 \in \mathsf{NN}$ *satisfy that* $\mathsf{D}(\nu_1) = \mathsf{D}(\nu_2) = L$, $\mathsf{I}(\nu_1) = \mathsf{I}(\nu_2)$, *and* $\mathsf{O}(\nu_1) = \mathsf{O}(\nu_2)$, *and* $\mathsf{L}(\nu_1) = (l_{1,1}, l_{1,2}, ...l_{1,L})$ *and* $\mathsf{L}(\nu_2) = (l_{2,1}, l_{2,2}, ..., l_{2,L})$ *it is then the case that:*

$$\mathsf{D}(\nu_1 \oplus \nu_2) = L \tag{4.5.12}$$

Proof. Note that $\mathsf{D}\left(\mathsf{Cpy}_{n,k}\right) = 1 = \mathsf{D}\left(\mathsf{Sum}_{n,k}\right)$ for all $n, k \in \mathbb{N}$. Note also that $\mathsf{D}\left(\nu_1 \boxminus \nu_2\right) = \mathsf{D}\left(\nu_1\right) = \mathsf{D}\left(\nu_2\right)$ and that for $\nu, \mu \in \mathsf{NN}$ it is the case that $\mathsf{D}\left(\nu \bullet \mu\right) = \mathsf{D}\left(\nu\right) + \mathsf{D}\left(\mu\right) - 1$. Thus:

$$\mathsf{D}(\nu_1 \oplus \nu_1) = \mathsf{D}(\nu_1 \oplus \nu_2) = \mathsf{D}\left(\mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{O}(\nu_2)} \; \mathbb{I}_{\mathsf{O}(\nu_2)}\right], \mathbb{0}_{\mathsf{O}(\nu_2)}} \bullet [\nu_1 \boxminus \nu_2] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{I}(\nu_1)} \; \mathbb{I}_{\mathsf{I}(\nu_1)}\right]^T, \mathbb{0}_{2 \cdot \mathsf{I}(\nu_1)}}\right)$$

$$= L$$

□

Lemma 4.5.12. *Let* $\nu_1, \nu_2 \in \mathsf{NN}$, *such that* $\mathsf{D}(\nu_1) = \mathsf{D}(\nu_2) = L$, $\mathsf{I}(\nu_1) = \mathsf{I}(\nu_2) = l_0$, *and* $\mathsf{O}(\nu_1) = \mathsf{O}(\nu_2) = l_L$. *It is then the case that* $\mathfrak{I}(\nu_1 \oplus \nu_2) = \mathfrak{I}(\nu_2 \oplus \nu_1)$, *i.e., the instantiated sum of ANNs of the same depth and same end widths is commutative.*

Proof. Let $\nu_1 = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$ and let $\nu_2 = ((W_1', b_1'), (W_2', b_2'), ..., (W_L', b_L'))$. Note that Definition 4.3.1 then tells us that:

$$\nu_1 \boxminus \nu_2 = \left(\left(\begin{bmatrix} W_1 & 0 \\ 0 & W_1' \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix}\right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix}\right), ...,\right.$$
$$\left.\left(\begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \begin{bmatrix} b_L \\ b_L' \end{bmatrix}\right)\right)$$

95

Note also that by Claims **??** and **??** and Definition 4.4.1 we know that:

$$\mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{I}(\nu_2)}\ \mathbb{I}_{\mathsf{I}(\nu_2)}\right]^T,\mathbb{0}_{2\,\mathsf{I}(\nu_2),1}} = \left(\begin{bmatrix} \mathbb{I}_{\mathsf{I}(\nu_2)} \\ \mathbb{I}_{\mathsf{I}(\nu_2)} \end{bmatrix}, \mathbb{0}_{2\,\mathsf{I}(\nu_2),1}\right) \tag{4.5.13}$$

and:

$$\mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{O}(\nu_1)}\ \mathbb{I}_{\mathsf{O}(\nu_1)}\right],\mathbb{0}_{2\,\mathsf{O}(\nu_1),1}} = \left(\begin{bmatrix} \mathbb{I}_{\mathsf{O}(\nu_1)}\ \mathbb{I}_{\mathsf{O}(\nu_1)} \end{bmatrix}, \mathbb{0}_{2\,\mathsf{O}(\nu_1),1}\right) \tag{4.5.14}$$

Applying Definition 4.2.1, specifically the second case, (4.5.3) and (**??**) yields that:

$$[\nu_1 \boxminus \nu_2] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{I}(\nu_2)}\ \mathbb{I}_{\mathsf{I}(\nu_2)}\right]^T,\mathbb{0}_{2\,\mathsf{I}(\nu_2),1}}$$

$$= \left(\left(\begin{bmatrix} W_1 & 0 \\ 0 & W_1' \end{bmatrix}\begin{bmatrix} \mathbb{I}_{\mathsf{I}(\nu_1)} \\ \mathbb{I}_{\mathsf{I}(\nu_1)} \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix}\right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix}\right), ..., \left(\begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \begin{bmatrix} b_L \\ b_L' \end{bmatrix}\right)\right)$$

$$= \left(\left(\begin{bmatrix} W_1 \\ W_1' \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix}\right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix}\right), ..., \left(\begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \begin{bmatrix} b_L \\ b_L' \end{bmatrix}\right)\right)$$

Applying Claim **??** and especially the third case of Definition 4.2.1 to to the above then gives us:

$$\mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{O}(\nu_1)}\ \mathbb{I}_{\mathsf{O}(\nu_1)}\right],0} \bullet [\nu_1 \boxminus \nu_2] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathsf{I}(\nu_2)}\ \mathbb{I}_{\mathsf{I}(\nu_2)}\right]^T,0} =$$

$$\left(\left(\begin{bmatrix} W_1 \\ W_1' \end{bmatrix}, \begin{bmatrix} B_1 \\ B_1' \end{bmatrix}\right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix}\right), ...,\right.$$

$$\left.\left(\begin{bmatrix} \mathbb{I}_{\mathsf{O}(\nu_2)}\ \mathbb{I}_{\mathsf{O}(\nu_2)} \end{bmatrix}\begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \begin{bmatrix} \mathbb{I}_{\mathsf{O}(\nu_2)}\ \mathbb{I}_{\mathsf{O}(\nu_2)} \end{bmatrix}\begin{bmatrix} b_L \\ b_L' \end{bmatrix}\right)\right)$$

$$= \left(\left(\begin{bmatrix} W_1 \\ W_1' \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix}\right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix}\right), ..., \left(\begin{bmatrix} W_L & W_L' \end{bmatrix}, b_L + b_L'\right)\right) \tag{4.5.15}$$

Now note that:

$$\nu_2 \boxminus \nu_1 = \left(\left(\begin{bmatrix} W_1' & 0 \\ 0 & W_1 \end{bmatrix}, \begin{bmatrix} b_1' \\ b_1 \end{bmatrix} \right), \left(\begin{bmatrix} W_2' & 0 \\ 0 & W_2 \end{bmatrix}, \begin{bmatrix} b_2' \\ b_2 \end{bmatrix} \right), ..., \right.$$
$$\left. \left(\begin{bmatrix} W_L' & 0 \\ 0 & W_L \end{bmatrix}, \begin{bmatrix} b_L' \\ b_L \end{bmatrix} \right) \right)$$

And thus:

$$\mathsf{Aff}_{[\mathbb{I}_{O(\nu_2)} \ \mathbb{I}_{O(\nu_2)}],0} \bullet [\nu_2 \boxminus \nu_1] \bullet \mathsf{Aff}_{[\mathbb{I}_{I(\nu_1)} \ \mathbb{I}_{I(\nu_1)}]^T,0}$$
$$= \left(\left(\begin{bmatrix} W_1' \\ W_1 \end{bmatrix}, \begin{bmatrix} b_1' \\ b_1 \end{bmatrix} \right), \left(\begin{bmatrix} W_2' & 0 \\ 0 & W_2 \end{bmatrix}, \begin{bmatrix} b_2' \\ b_2 \end{bmatrix} \right), ..., \left(\begin{bmatrix} W_L' & W_L \end{bmatrix}, \begin{bmatrix} b_L' + b_L \end{bmatrix} \right) \right) \quad (4.5.16)$$

Let $x \in \mathbb{R}^{I(\nu_1)}$, note then that:

$$\begin{bmatrix} W_1 \\ W_1' \end{bmatrix} x + \begin{bmatrix} b_1 \\ b_1' \end{bmatrix} = \begin{bmatrix} W_1 x + b_1 \\ W_1' x + b_1' \end{bmatrix}$$

The full instantiation of (4.5.15) with activation function $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ is then given by:

$$\begin{bmatrix} W_L & W_L' \end{bmatrix} \begin{bmatrix} \mathfrak{a} \left(W_{L-1}(... \mathfrak{a}(W_2 \left(\mathfrak{a} \left(W_1 x + b_1 \right) \right) + b_2) + ...) + b_{L-1} \right) \\ \mathfrak{a} \left(W_{L-1}'(... \mathfrak{a}(W_2' \left(\mathfrak{a} \left(W_1' x + b_1' \right) \right) + b_2') + ...) + b_{L-1}' \right) \end{bmatrix} + b_L + b_L' \quad (4.5.17)$$

The full instantiation of (4.5.16) is then given by:

$$\begin{bmatrix} W_L' & W_L \end{bmatrix} \begin{bmatrix} \mathfrak{a} \left(W_{L-1}'(... \mathfrak{a}(W_2' \left(\mathfrak{a} \left(W_1' x + b_1' \right) \right) + b_2') + ...) + b_{L-1}' \right) \\ \mathfrak{a} \left(W_{L-1}(... \mathfrak{a}(W_2 \left(\mathfrak{a} \left(W_1 x + b_1 \right) \right) + b_2) + ...) + b_{L-1} \right) \end{bmatrix} + b_L + b_L' \quad (4.5.18)$$

Since (4.5.27) and (4.5.18) are the same this proves that $\nu_1 \oplus \nu_2 = \nu_2 \oplus \nu_1$. $\qquad \square$

Remark 4.5.13. *This is a special case of (Grohs et al., 2022, Lemma 3.28).*

Lemma 4.5.14. *Let $l_0, l_1, ..., l_L \in \mathbb{N}$. Let $\nu \in \mathsf{NN}$ with $\mathsf{L}(\nu) = (l_0, l_1, ..., l_L)$. There then exists a neural network $\mathrm{Zr}_{l_0,l_1,...,l_L} \in \mathsf{NN}$ such that $\mathfrak{I}(\nu \oplus \mathrm{Zr}_{l_0,l_1,...,l_L}) = \mathfrak{I}(\mathrm{Zr}_{l_0,l_1,...,l_L} \oplus \nu) = \nu$.*

Proof. Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$, where $W_1 \in \mathbb{R}^{l_1 \times l_0}$, $b_1 \in \mathbb{R}^{l_1}$, $W_2 \in \mathbb{R}^{l_2 \times l_1}$, $b_2 \in \mathbb{R}^{l_2}, ..., W_L \in \mathbb{R}^{l_L \times l_{L-1}}$, $b_L \in \mathbb{R}^{l_L}$. Denote by $\mathrm{Zr}_{l_0,l_1,...,l_L}$ the neural network which for all $l_0, l_1, ..., l_L \in \mathbb{N}$ is given by:

$$\mathrm{Zr}_{l_0,l_1,...,l_L} = \left(\left(\mathbb{O}_{l_1,l_0}, \mathbb{O}_{l_1} \right), \left(\mathbb{O}_{l_2,l_1}, \mathbb{O}_{l_2} \right), ..., \left(\mathbb{O}_{l_L,l_{L-1}}, \mathbb{O}_{l_L} \right) \right) \tag{4.5.19}$$

Thus, by (4.5.27), we have that:

$$\mathfrak{I}(\mathrm{Zr}_{l_0,l_1,...,l_L} \oplus \nu) = \begin{bmatrix} 0 & W_L \end{bmatrix} \begin{bmatrix} 0 \\ W_{L-1}(...(W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1} \end{bmatrix} + b_L$$

$$= W_L(W_{L-1}(...W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1}) + b_L \tag{4.5.20}$$

$$\mathfrak{I}(\nu \oplus \mathrm{Zr}_{l_0,l_1,...,l_L}) = \begin{bmatrix} W_L & 0 \end{bmatrix} \begin{bmatrix} W_{L-1}(...(W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1} \\ 0 \end{bmatrix} + b_L$$

$$= W_L(W_{L-1}(...W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1}) + b_L \tag{4.5.21}$$

And finally:

$$\mathfrak{I}(\nu) = W_L(W_{L-1}(...W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1}) + b_L \tag{4.5.22}$$

This completes the proof. $\qquad \square$

Lemma 4.5.15. *Given neural networks $\nu_1, \nu_2, \nu_3 \in \mathsf{NN}$ with fixed depth L, fixed starting width of l_0 and fixed finishing width of l_L, it is then the case that $\mathfrak{I}((\nu_1 \oplus \nu_2) \oplus \nu_3) = \mathfrak{I}(\nu_1 \oplus (\nu_2 \oplus \nu_3))$, i.e. the instantiation with a continuous activation function of $\oplus$ is associative.*

Proof. Let $\nu_1 = ((W_1^1, b_1^1), (W_2^1, b_2^1), ..., (W_L^1, b_L^1))$, $\nu_2 = ((W_1^2, b_1^2), (W_2^2, b_2^2), ..., (W_L^2, b_L^2))$, and $\nu_3 = ((W_1^3, b_1^3), (W_2^3, b_2^3), ..., (W_L^3, b_L^3))$. Then (4.5.27) tells us that:

$$\mathfrak{I}(\nu_1 \oplus \nu_2) = \begin{bmatrix} W_L^1 & W_L^2 \end{bmatrix} \begin{bmatrix} W_{L-1}^1 (... (W_2^1 (W_1^1 x + b_1^1) + b_2^1) + ...) + b_{L-1}^1 \\ W_{L-1}^2 (... (W_2^2 (W_1^2 x + b_1^2) + b_2^2) + ...) + b_{L-1}^2 \end{bmatrix} + b_L^1 + b_L^2$$

And thus:

$$\mathfrak{I}((\nu_1 \oplus \nu_2) \oplus \nu_3)(x) =$$

$$\mathfrak{I}\left(\begin{bmatrix} \mathbb{I} & W_L^3 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} W_L^1 & W_L^2 \end{bmatrix} \begin{bmatrix} W_{L-1}^1 (... (W_2^1 (W_1^1 x + b_1^1) + b_2^1) + ...) + b_{L-1}^1 \\ W_{L-1}^2 (... (W_2^2 (W_1^2 x + b_1^2) + b_2^2) + ...) + b_{L-1}^2 \end{bmatrix} + b_L^1 + b_L^2 \\ W_{L-1}^3 (... (W_2^3 (W_1^3 x + b_1^3) + b_2^3) + ...) + b_{L-1}^3 \end{bmatrix} + b_L^3 \right)$$

(4.5.23)

Similarly, we have that:

$$\mathfrak{I}_{\mathfrak{a}}(\nu_1 \oplus (\nu_2 \oplus \nu_3))(x) =$$

$$\mathfrak{I}\left(\begin{bmatrix} W_L^1 & \mathbb{I} \end{bmatrix} \begin{bmatrix} W_{L-1}^1 (... (W_2^1 (W_1^1 x + b_1^1) + b_2^1) + ...) + b_{L-1}^1 \\ \begin{bmatrix} W_L^2 & W_L^3 \end{bmatrix} \begin{bmatrix} W_{L-1}^2 (... (W_2^2 (W_1^2 x + b_1^2) + b_2^2) + ...) + b_{L-1}^2 \\ W_{L-1}^3 (... (W_2^3 (W_1^3 x + b_1^3) + b_2^3) + ...) + b_{L-1}^3 \end{bmatrix} + b_L^2 + b_L^3 \end{bmatrix} + b_L^1 \right)$$

(4.5.24)

Note that the associativity of matrix-vector multiplication ensures that (4.5.23) and (4.5.24) are the same. $\square$

Definition 4.5.16 (Commutative Semi-group). *A set X equipped with a binary operation $*$ is called a monoid if:*

(i) for all $x, y, z \in X$ it is the case that $(x * y) * z = x * (y * z)$ and

(ii) for all $x, y \in X$ it is the case that $x * y = y * x$

Theorem 4.5.17. *For fixed depth and layer widths, the set of instantiated neural networks* $\nu \in \mathsf{NN}$ *form a commutative semi-group under the operation of* $\oplus$.

Proof. This is a consequence of Lemmas 4.5.12, 4.5.14, and 4.5.15. $\qquad\square$

Lemma 4.5.18. *Let* $\nu, \mu \in \mathsf{NN}$, *with the same length and end-widths. It is then the case that* $\mathfrak{I}_{\mathfrak{a}}(\nu \oplus \mu) = \mathfrak{I}_{\mathfrak{a}}(\nu) + \mathfrak{I}_{\mathfrak{a}}(\mu)$.

Proof. Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$ and $\mu = ((W_1', b_1'), (W_2', b_2'), ..., (W_L', b_L'))$. Note now that by (4.5.27) we have that:

$$\mathfrak{I}_{\mathfrak{a}}(\nu) = W_L \, \mathfrak{a}\,(W_{L-1}(... \mathfrak{a}(W_2 \, \mathfrak{a}\,(W_1 x + b_1) + b_2) + ...) + b_{L-1}) + b_L \tag{4.5.25}$$

And:

$$\mathfrak{I}_{\mathfrak{a}}(\mu) = W_L' \, \mathfrak{a}\,\left(W_{L-1}'(... \mathfrak{a}(W_2' \, \mathfrak{a}\,(W_1' x + b_1') + b_2') + ...) + b_{L-1}'\right) + b_L' \tag{4.5.26}$$

In addition, because of the block matrix structure of the weights of our summands:

$$\mathfrak{I}_{\mathfrak{a}}(\nu \oplus \mu)(x) = \begin{bmatrix} W_L & W_L' \end{bmatrix} \begin{bmatrix} \mathfrak{a}\,(W_{L-1}(... \mathfrak{a}(W_2 \, \mathfrak{a}\,(W_1 x + b_1) + b_2) + ...) + b_{L-1}) \\ \mathfrak{a}\,\left(W_{L-1}'(... \mathfrak{a}(W_2' \, \mathfrak{a}\,(W_1' x + b_1') + b_2') + ...) + b_{L-1}'\right) \end{bmatrix} + b_L + b_L'$$

$$= W_L \, \mathfrak{a}\,(W_{L-1}(... \mathfrak{a}(W_2 \, \mathfrak{a}\,(W_1 x + b_1) + b_2) + ...) + b_{L-1}) + b_L$$

$$+ W_L' \, \mathfrak{a}\,\left(W_{L-1}'(... \mathfrak{a}(W_2' \, \mathfrak{a}\,(W_1' x + b_1') + b_2') + ...) + b_{L-1}'\right) + b_L'$$

$$= \mathfrak{I}_{\mathfrak{a}}(\nu)(x) + \mathfrak{I}_{\mathfrak{a}}(\mu)(x) \tag{4.5.27}$$

This proves the lemma. $\qquad\square$

Lemma 4.5.19. *Let* $n \in \mathbb{N}$. *Let* $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$. *It is then the case that:*

$$\mathfrak{I}_{\mathfrak{a}}\left(\bigoplus_{i=1}^{n} \nu_i\right) = \sum_{i=1}^{n} \mathfrak{I}_{\mathfrak{r}}(\nu_i) \tag{4.5.28}$$

100

Proof. This is the consequence of a finite number of applications of Lemma 4.5.18. This proves the Lemma. □

4.5.2 Sum of ANNs of Unequal Depth But Same End-widths

Definition 4.5.20 (Sum of ANNs of different depths but same end widths). *Let $n \in \mathbb{N}$. Let $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$ such that they have the same end widths. We define the neural network $\bigoplus_{i=1}^{n} \nu_i \in \mathsf{NN}$, the neural network sum of neural networks of unequal depth as:*

$$\bigoplus_{i=1}^{n} \nu_i := \left(\mathsf{Sum}_{n, \mathsf{O}(\nu_2)} \bullet [\bigotimes_{i=u}^{v} \nu_i] \bullet \mathsf{Cpy}_{n, \mathsf{I}(\nu_1)} \right) \tag{4.5.29}$$

Lemma 4.5.21. *Let $n \in \mathbb{N}$. Let $\nu_1, \nu_2 \in \mathsf{NN}$ and assume also that they have the same end-widths. It is then the case that:*

$$\mathfrak{I}_\mathfrak{r} \left(\nu_1 \oplus \nu_2 \right)(x) = \mathfrak{I}_\mathfrak{r} \left(\nu_1 \right) + \mathfrak{I}_\mathfrak{r} \left(\nu_2 \right) \tag{4.5.30}$$

Proof. Note that Lemma 5.2.3 tellls us that for all $n \in \mathbb{N}$ it is the case that $\mathfrak{I}_\mathfrak{r} \left(\mathsf{Tun}_n \right)(x) = x$. This combined with Lemma 4.2.5 then tells us that for all $n \in \mathbb{N}$ it is the case for all $\nu \in \mathsf{NN}$ that:

$$\mathfrak{I}_\mathfrak{r} \left(\mathsf{Tun}_n \bullet \nu \right)(x) = \mathfrak{I}_\mathfrak{r} \left(\nu \right)(x) \tag{4.5.31}$$

Thus, this means that:

$$\mathfrak{I}_\mathfrak{r} \left(\nu_1 \oplus \nu_2 \right)(x) = \left(\mathsf{Sum}_{n, \mathsf{O}(\nu_2)} \bullet [\nu_1 \otimes \nu_2] \bullet \mathsf{Cpy}_{n, \mathsf{I}(\nu_1)} \right)$$
$$= \mathfrak{I}_\mathfrak{r} \left(\nu_1 \right)(x) + \mathfrak{I}_\mathfrak{r} \left(\nu_2 \right)(x) \tag{4.5.32}$$

This then proves the lemma. □

Lemma 4.5.22. *Let $n \in \mathbb{N}$. Let $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$. Let it also be the case that they have*

the same end-widths. It is then the case that:

$$\mathfrak{I}_{\mathfrak{r}}\left(\oplus_{i=1}^{n}\nu_{i}\right)(x) = \sum_{i=1}^{n}\mathfrak{I}_{\mathfrak{r}}\left(\nu_{i}\right)(x) \tag{4.5.33}$$

Proof. This is a consequence of a finite number of applications of Lemma 4.5.21. This proves the Lemma. $\square$

Remark 4.5.23. *We may represent this kind of sum as the neural network diagram shown below:*
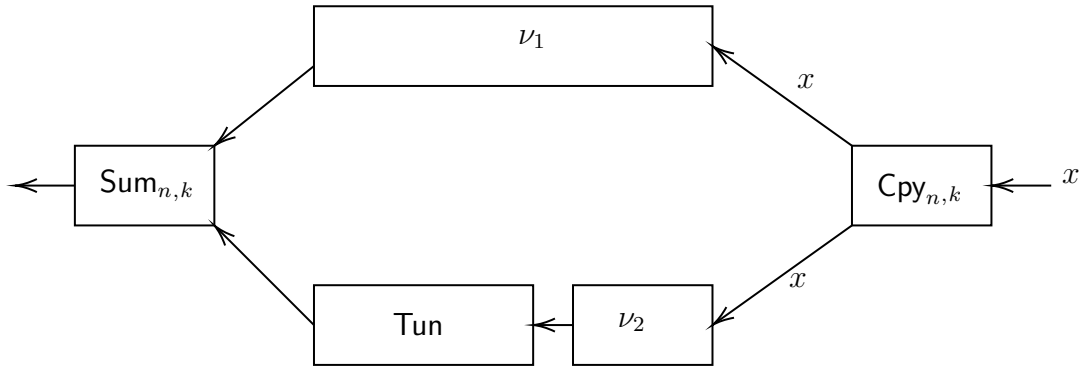
Figure 4.4: Neural network diagram of a neural network sum of unequal depth networks.

4.6 Linear Combinations of ANNs and Their Properties

Definition 4.6.1 (Scalar left-multiplication with an ANN)**.** *Let $\lambda \in \mathbb{R}$. We will denote by $(\cdot) \triangleright (\cdot) : \mathbb{R} \times \mathsf{NN} \rightarrow \mathsf{NN}$ the function that satisfy for all $\lambda \in \mathbb{R}$ and $\nu \in \mathsf{NN}$ that $\lambda \triangleright \nu = \mathsf{Aff}_{\lambda \mathbb{I}_{O(\nu)},0} \bullet \nu$.*

Definition 4.6.2 (Scalar right-multiplication with an ANN)**.** *Let $\lambda \in \mathbb{R}$. We will denote by $(\cdot) \triangleleft (\cdot) : \mathsf{NN} \times \mathbb{R} \rightarrow \mathsf{NN}$ the function satisfying for all $\nu \in \mathsf{NN}$ and $\lambda \in \mathbb{R}$ that $\nu \triangleleft \lambda = \nu \bullet \mathsf{Aff}_{\lambda \mathbb{I}_{I(\nu)},0}$.*

Remark 4.6.3. *Note that whereas $\lambda \in \mathbb{R}$, the actual neural network in question, properly speaking, must always be referred to as $\lambda \triangleright$ or $\triangleleft \lambda$, and we shall do so whenever this comes up in any neural network diagrams. This is by analogy with, for example, $\log_\lambda$ or $\sqrt[\lambda]{}$ for $\lambda \neq 0$, where the argument λ is generally always written except for $\lambda = 10$ for the logarithm or $\lambda = 2$ for the root.*

Remark 4.6.4. *For an* R *implementation, see Listing 9.8*

Lemma 4.6.5. *Let $\lambda \in \mathbb{R}$ and $\nu \in \mathsf{NN}$. it is then the case that:*

(i) $\mathsf{L}(\lambda \triangleright \nu) = \mathsf{L}(\nu)$

(ii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ that $\mathfrak{I}_\mathfrak{a}(\lambda \triangleright \nu) \in C\left(\mathbb{R}^{\mathsf{I}(\nu)}, \mathbb{R}^{\mathsf{O}(\nu)}\right)$

(iii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, and $x \in \mathbb{R}^{\mathsf{I}(\nu)}$ that:

$$\mathfrak{I}_\mathfrak{a}\left(\lambda \triangleright \nu\right) = \lambda\,\mathfrak{I}_\mathfrak{a}(\nu) \tag{4.6.1}$$

Proof. Let $\nu \in \mathsf{NN}$ such that $\mathsf{L}(\nu) = (l_1, l_2, ..., l_L)$ and $\mathsf{D}(\nu) = L$ where $l_1, l_2, ..., l_L, L \in \mathbb{N}$. Then Item (i) of Lemma 4.4.2 tells us that:

$$\mathsf{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{O}(\nu)},0}\right) = (\mathsf{O}(\nu), \mathsf{O}(\nu)) \tag{4.6.2}$$

This and Item (i) from Lemma 4.4.5 gives us that:

$$\mathsf{L}\left(\lambda \triangleright \nu\right) = \mathsf{L}\left(\mathsf{Aff}_{\lambda\mathbb{I}_{\mathsf{O}(\nu)},0} \bullet \nu\right) = (l_0, l_1, ..., l_{L-1}, \mathsf{O}(\nu)) = \mathsf{L}(\nu) \tag{4.6.3}$$

Which proves (i). Item (ii)—(iii) of Lemma 4.4.2 then prove that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$,

$x \in \mathbb{R}^{\mathsf{I}(\nu)}$, that $\mathfrak{I}_{\mathfrak{a}} \left(\lambda \triangleright \nu \right) \in C \left(\mathbb{R}^{\mathsf{I}(\nu),\mathsf{O}(\nu)} \right)$ given by:

$$
\begin{aligned}
\left(\mathfrak{I}_{\mathfrak{a}} \left(\lambda \triangleright \nu \right) \right) (x) &= \left(\mathfrak{I}_{\mathfrak{a}} \left(\mathsf{Aff}_{\lambda \mathbb{I}_{\mathsf{O}(\nu),0}} \bullet \nu \right) \right) (x) \\
&= \lambda \mathbb{I}_{\mathsf{O}(\nu)} \left(\left(\mathfrak{I}_{\mathfrak{a}} \left(\nu \right) \right) (x) \right) = \lambda \left(\left(\mathfrak{I}_{\mathfrak{a}} \left(\nu \right) \right) (x) \right)
\end{aligned} \tag{4.6.4}
$$

This establishes Items (ii)—(iii), completing the proof. □

Lemma 4.6.6. *Let $\lambda \in \mathbb{R}$ and $\nu \in \mathsf{NN}$. It is then the case that:*

(i) $\mathsf{L}(\nu \triangleleft \lambda) = \mathsf{L}(\nu)$

(ii) For all $\mathfrak{a} \in C \left(\mathbb{R}, \mathbb{R} \right)$ that $\mathfrak{I}_{\mathfrak{a}}(\nu \triangleleft \lambda) \in C \left(\mathbb{R}^{\mathsf{I}(\nu)}, \mathbb{R}^{\mathsf{O}(\nu)} \right)$

(iii) For all $\mathfrak{a} \in C \left(\mathbb{R}, \mathbb{R} \right)$, and $x \in \mathbb{R}^{\mathsf{I}(\nu)}$ that:

$$
\mathfrak{I}_{\mathfrak{a}} \left(\nu \triangleleft \lambda \right) = \mathfrak{I}_{\mathfrak{a}}(\nu) \left(\lambda x \right) \tag{4.6.5}
$$

Proof. Let $\nu \in \mathsf{NN}$ such that $\mathsf{L}(\nu) = (l_1, l_2, ..., l_L)$ and $\mathsf{D}(\nu) = L$ where $B_1, l_2, ..., l_L, L \in \mathbb{N}$. Then Item (i) of Lemma 4.4.2 tells us that:

$$
\mathsf{L} \left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu)},0} \right) = (\mathsf{I}(\nu), \mathsf{I}(\nu)) \tag{4.6.6}
$$

This and Item (iv) of Lemma 4.4.5 tells us that:

$$
\mathsf{L}(\nu \triangleleft \lambda) = \mathsf{L} \left(\nu \bullet \mathsf{Aff}_{\lambda \mathbb{I}_{\mathsf{I}(\nu)}} \right) = (\mathsf{I}(\nu), l_1, l_2, ..., l_L) = \mathsf{L}(\nu) \tag{4.6.7}
$$

Which proves (*i*). Item (v)—(vi) of Lemma 4.4.5 then prove that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$,

$x \in \mathbb{R}^{\mathsf{I}(\nu)}$ that $\mathfrak{I}_{\mathfrak{a}}\left(\nu \triangleleft \lambda\right) \in C\left(\mathbb{R}^{\mathsf{I}(\nu),\mathsf{O}(\nu)}\right)$ given by:

$$
\begin{aligned}
\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu \triangleleft \lambda\right)\right)(x) &= \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu \bullet \mathsf{Aff}_{\lambda \mathbb{I}_{\mathsf{I}(\nu)},0}\right)\right)(x) \\
&= \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu\right)\right)\left(\mathsf{Aff}_{\lambda \mathbb{I}_{\mathsf{I}(\nu)}}\right)(x) \\
&= \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu\right)\right)(\lambda x) \tag{4.6.8}
\end{aligned}
$$

This completes the proof. $\qquad\square$

Lemma 4.6.7. *Let $\nu, \mu \in \mathsf{NN}$ with the same length and end-widths, and $\lambda \in \mathbb{R}$. It is then the case, for all $\mathfrak{a} \in C\left(\mathbb{R}, \mathbb{R}\right)$ that:*

$$
\begin{aligned}
\mathfrak{I}_{\mathfrak{a}}\left(\left(\nu \oplus \mu\right) \triangleleft \lambda\right)(x) &= \mathfrak{I}_{\mathfrak{a}}\left(\left(\nu \triangleleft \lambda\right) \oplus \left(\mu \triangleleft \lambda\right)\right)(x) \\
&= \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu\right)\right)(\lambda x) + \left(\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\right)(\lambda x)
\end{aligned}
$$

Proof. Let $\nu = \left(\left(W_1, b_1\right), \left(W_2, b_2\right), ..., \left(W_L, b_L\right)\right)$ and $\mu = \left(\left(W_1', b_1'\right), \left(W_2', b_2'\right), ..., \left(W_L', b_L'\right)\right)$. Then from Lemma 4.6.6 and (4.5.27) we have that:

$$
\begin{aligned}
&\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu \oplus \mu\right) \triangleleft \lambda\right)(x) \\
&= \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu \oplus \mu\right)\right)(\lambda x) \\
&= \begin{bmatrix} W_L & W_L' \end{bmatrix} \begin{bmatrix} \mathfrak{a}\left(W_{L-1}(...(\mathfrak{a}\left(W_2\left(\mathfrak{a}\left(W_1 \lambda x + b_1\right)\right) + b_2\right)) + ...) + b_{L-1}\right) \\ \mathfrak{a}\left(W_{L-1}'(...(\mathfrak{a}\left(W_2'\left(\mathfrak{a}\left(W_1' \lambda x + b_1'\right)\right) + b_2'\right)) + ...) + b_{L-1}'\right) \end{bmatrix} + b_L + b_L'
\end{aligned}
$$

Note that:

$$
\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu\right)\right)(\lambda x) = W_L \cdot \mathfrak{a}\left(W_{L-1}(...(\mathfrak{a}\left(W_2\left(\mathfrak{a}\left(W_1 \lambda x + b_1\right)\right) + b_2\right)) + ...) + b_{L-1}\right) + b_L \tag{4.6.9}
$$

and that:

$$
\left(\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\right)(\lambda x) = W_L' \cdot \mathfrak{a}\left(W_{L-1}'(...(\mathfrak{a}\left(W_2'\left(\mathfrak{a}\left(W_1' \lambda x + b_1'\right)\right) + b_2'\right)) + ...) + b_{L-1}'\right) + b_L' \tag{4.6.10}
$$

This, together with Lemma 4.5.18, completes the proof. $\qquad\square$

Lemma 4.6.8. *Let* $\nu, \mu \in \mathsf{NN}$ *with the same length and end-widths, and* $\lambda \in \mathbb{R}$. *It is then the case, for all* $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *that:*

$$\mathfrak{I}_{\mathfrak{a}}\left(\lambda \triangleright (\nu \oplus \mu)\right)(x) = \mathfrak{I}_{\mathfrak{a}}\left((\lambda \triangleright \nu) \oplus (\lambda \triangleright \mu)\right)(x)$$
$$= \lambda \cdot \left(\mathfrak{I}_{\mathfrak{a}}(\nu)\right)(x) + \lambda \cdot \left(\mathfrak{I}_{\mathfrak{a}}(\mu)\right)(x)$$

Proof. Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$ and $\mu = ((W_1', b_1'), (W_2', b_2'), ..., (W_L', b_L'))$. Then from Lemma 4.6.6 and (4.5.27) we have that:

$$\mathfrak{I}_{\mathfrak{a}}\left(\lambda(\nu \oplus \mu)\right)(x)$$

$$= \mathfrak{I}_{\mathfrak{a}}\left(\lambda \triangleright (\nu \oplus \mu)\right)(\lambda x)$$

$$= \lambda \cdot \begin{bmatrix} W_L & W_L' \end{bmatrix} \begin{bmatrix} \mathfrak{I}_{\mathfrak{r}}\left(W_{L-1}(...(\mathfrak{I}_{\mathfrak{r}}\left(W_2\left(\mathfrak{I}_{\mathfrak{r}}\left(W_1 x + b_1\right)\right) + b_2\right)) + ...) + b_{L-1}\right) \\ \mathfrak{I}_{\mathfrak{r}}\left(W_{L-1}'(...(\mathfrak{I}_{\mathfrak{r}}\left(W_2'\left(\mathfrak{I}_{\mathfrak{r}}\left(W_1' x + b_1'\right)\right) + b_2'\right)) + ...) + b_{L-1}'\right) \end{bmatrix} + b_L + b_L'$$

Note that:

$$\lambda \cdot \left(\mathfrak{I}_{\mathfrak{a}}(\nu)\right)(x) = W_L \cdot \mathfrak{I}_{\mathfrak{r}}\left(W_{L-1}(...(\mathfrak{I}_{\mathfrak{r}}\left(W_2\left(\mathfrak{I}_{\mathfrak{r}}\left(W_1 x + b_1\right)\right) + b_2\right)) + ...) + b_{L-1}\right) + b_L \quad (4.6.11)$$

and that:

$$\lambda \cdot \left(\mathfrak{I}_{\mathfrak{a}}(\mu)\right)(x) = W_L' \cdot \mathfrak{I}_{\mathfrak{r}}\left(W_{L-1}'(...(\mathfrak{I}_{\mathfrak{r}}\left(W_2'\left(\mathfrak{I}_{\mathfrak{r}}\left(W_1' x + b_1'\right)\right) + b_2'\right)) + ...) + b_{L-1}'\right) + b_L'$$
$$(4.6.12)$$

This, together with Lemma 4.5.18, completes the proof. $\qquad\square$

Lemma 4.6.9. *Let* $u, v \in \mathbb{Z}$ *with* $u \leqslant v$ *and* $n = v - u + 1$. *Let* $\lambda_u, \lambda_{u+1}, ..., \lambda_v \in \mathbb{R}$. *Let* $\nu_u, \nu_{u+1}, ..., \nu_v, \mu \in \mathsf{NN}$, $B_u, B_{u+1}, ..., B_v \in \mathbb{R}^{\mathsf{I}(\mu)}$ *satisfy that* $\mathsf{L}(\nu_u) = \mathsf{L}(\nu_{u+1}) = ... = \mathsf{L}(\nu_v)$

and further that:

$$\mu = \left[\oplus_{i=u}^{v} \left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_1)}, B_i} \right) \right) \right] \tag{4.6.13}$$

It then holds:

(i) *That:*

$$\mathsf{L}(\mu) = \left(\mathsf{I}(\nu_u), \sum_{i=u}^{v} \mathsf{W}_1(\nu_u), \sum_{i=u}^{v} \mathsf{W}_2(\nu_u), ..., \sum_{i=u}^{v} \mathsf{W}_{\mathsf{D}(\nu_u)-1}(\nu_u), \mathsf{O}(\nu_u) \right)$$

$$= \left(\mathsf{I}(\nu_u), n\,\mathsf{W}_1(\nu_u), n\,\mathsf{W}_2(\nu_u), ..., n\,\mathsf{W}_{\mathsf{D}(\nu_u-1)}, \mathsf{O}(\nu_u) \right)$$

(ii) *that for all* $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, *that* $\mathfrak{I}_{\mathfrak{a}}(\mu) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)} \right)$, *and*

(iii) *for all* $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *and* $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$ *that:*

$$(\mathfrak{I}_{\mathfrak{a}}(\mu))(x) = \sum_{i=u}^{v} c_i \left(\mathfrak{I}_{\mathfrak{a}}(\nu_i) \right)(x + B_i) \tag{4.6.14}$$

Proof. Assume hypothesis that $\mathsf{L}(\nu_u) = \mathsf{L}(\nu_{u+1}) = ... = \mathsf{L}(\nu_v)$. Note that Item (i) of Lemma 4.4.2 gives us that for all $i \in \{u, u+1, ..., v\}$ that:

$$\mathsf{L}\left(\mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, B_i} \right) = \mathsf{L}\left(\mathsf{Aff}_{\mathbb{I}_{(\nu_u)}} \right) = \left(\mathsf{I}(\nu_u), \mathsf{I}(\nu_u) \right) \in \mathbb{N}^2 \tag{4.6.15}$$

This together with Lemma 4.2.5, Item (i), assures us that for all $i \in \{u, u+1, ..., v\}$ it is the case that:

$$\mathsf{L}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, B_i} \right) = \left(\mathsf{I}(\nu_u), \mathsf{W}_1(\nu_u), \mathsf{W}_2(\nu_u), ..., \mathsf{W}_{\mathsf{D}(\nu_u)}(\nu_u) \right) \tag{4.6.16}$$

This and (Grohs et al., 2022, Lemma 3.14, Item (i)) tells us that for all $i \in \{u, u+1, ..., v\}$

it is the case that:

$$\mathsf{L}\left(c_i \rhd \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, B_i}\right)\right) = \mathsf{L}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, B_i}\right) \tag{4.6.17}$$

This, (4.6.16), and (Grohs et al., 2022, Lemma 3.28, Item (ii)) then yield that:

$$\begin{aligned}
\mathsf{L}(\mu) &= \mathsf{L}\left(\oplus_{i=u}^{v}\left(c_i \rhd \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, B_i}\right)\right)\right)\\
&= \left(\mathsf{I}(\nu_u), \sum_{i=u}^{v}\mathsf{W}_1\left(\nu_u\right), \sum_{i=u}^{v}\mathsf{W}_2\left(\nu_u\right), ..., \sum_{i=u}^{v}\mathsf{W}_{\mathsf{D}(\nu_u)-1}\left(\nu_u\right), \mathsf{O}\left(\nu_u\right)\right)\\
&= \left(\mathsf{I}(\nu_u), n\,\mathsf{W}_1(\nu_u), n\,\mathsf{W}_2(\nu_u), ..., n\,\mathsf{W}_{\mathsf{D}(\nu_u)-1}(\nu_u), \mathsf{O}(\nu_u)\right)
\end{aligned} \tag{4.6.18}$$

This establishes item (i). Items (v) and (vi) from Lemma 4.4.5 tells us that for all $i \in \{u, u+1, ..., v\}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it is the case that $\mathfrak{I}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, B_i}\right) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right)$ and further that:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, b_i}\right)\right)(x) = \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)(x + b_i) \tag{4.6.19}$$

This along with (Grohs et al., 2022, Lemma 3.14) ensures that for all $i \in \{u, u+1, ..., v\}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it is the case that:

$$\mathfrak{I}_{\mathfrak{a}}\left(c_i \rhd \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, B_i}\right)\right) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right) \tag{4.6.20}$$

and:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(c_i \rhd \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{(\nu_i)}, b_i}\right)\right)\right)(x) = c_i\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)(x + b_i) \tag{4.6.21}$$

Now observe that (Grohs et al., 2022, Lemma 3.28) and (4.6.17) ensure that for all $\mathfrak{a} \in$

$C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it is the case that $\mathfrak{I}_\mathfrak{a}(\mu) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right)$ and that:

$$
\begin{aligned}
(\mathfrak{I}_\mathfrak{a}(\mu))(x) &= \left(\mathfrak{I}_\mathfrak{a}\left(\oplus_{i=u}^{v}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i}\right)\right)\right)\right)(x) \\
&= \sum_{i=u}^{v}\left(\mathfrak{I}_\mathfrak{a}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i}\right)\right)\right)(x) \\
&= \sum_{i=u}^{v} c_i\left(\mathfrak{I}_\mathfrak{a}(\nu_i)\right)(x + b_i)
\end{aligned}
$$

This establishes items (ii)--(iii); thus, the proof is complete. $\qquad\square$

Lemma 4.6.10. *Let* $u, v \in \mathbb{Z}$ *with* $u \leqslant v$. *Let* $\lambda_u, \lambda_{u+1}, ..., \lambda_v \in \mathbb{R}$. *Let* $\nu_u, \nu_{u+1}, ..., \nu_v, \mu \in$ *NN,* $B_u, B_{u+1}, ..., B_v \in \mathbb{R}^{\mathsf{I}(\mu)}$ *satisfy that* $\mathsf{L}(\nu_u) = \mathsf{L}(\nu_{u+1}) = ... = \mathsf{L}(\nu_v)$ *and further that:*

$$
\mu = \left[\oplus_{i=u}^{v}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_1)}, b_i} \bullet \nu\right) \triangleleft c_i\right)\right] \tag{4.6.22}
$$

It then holds:

(i) That:

$$
\begin{aligned}
\mathsf{L}(\mu) &= \left(\mathsf{I}(\nu_u), \sum_{i=u}^{v}\mathsf{W}_1(\nu_u), \sum_{i=u}^{v}\mathsf{W}_2(\nu_u), ..., \sum_{i=u}^{v}\mathsf{W}_{\mathsf{D}(\nu_u)-1}(\nu_u), \mathsf{O}(\nu_u)\right) \\
&= \left(\mathsf{I}(\nu_u), n\,\mathsf{W}_1(\nu_u), n\,\mathsf{W}_2(\nu_u), ..., n\,\mathsf{W}_{\mathsf{D}(\nu_u-1)}, \mathsf{O}(\nu_u)\right) \tag{4.6.23}
\end{aligned}
$$

(ii) that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$*, that* $\mathfrak{I}_\mathfrak{a}(\mu) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right)$*, and*

(iii) for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *and* $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$ *that:*

$$
(\mathfrak{I}_\mathfrak{a}(\mu))(x) = \sum_{i=u}^{v}(\mathfrak{I}_\mathfrak{a}(\nu_i))(c_i x + b_i) \tag{4.6.24}
$$

Proof. Assume hypothesis that $\mathsf{L}(\nu_u) = \mathsf{L}(\nu_{u+1}) = ... = \mathsf{L}(\nu_v)$. Note that Item (i) of Lemma

4.4.2 gives us that for all $i \in \{u, u+1, ..., v\}$ that:

$$\mathsf{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, B_i}\right) = \mathsf{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_u)}}\right) = (\mathsf{I}(\nu_u), \mathsf{I}(\nu_u)) \in \mathbb{N}^2 \tag{4.6.25}$$

Note then that Lemma 4.2.5, Item (ii), tells us that for all $i \in \{u, u+1, ..., v\}$ it is the case that:

$$\mathsf{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, B_i} \bullet \nu\right) = \left(\mathsf{I}(\nu_u), \mathsf{W}_1(\nu_u), \mathsf{W}_2(\nu_u), ..., \mathsf{W}_{\mathsf{D}(\nu_u)}(\nu_u)\right) \tag{4.6.26}$$

This and Item (i) of Lemma 4.6.6 tells us that for all $i \in \{u, u+1, ..., v\}$ it is the case that:

$$\mathsf{L}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i} \bullet \nu\right) \lhd c_i\right) = \mathsf{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i} \bullet \nu\right) \tag{4.6.27}$$

This, (4.6.26), and (Grohs et al., 2022, Lemma 3.28, Item (ii)) tell us that:

$$\begin{aligned}
\mathsf{L}(\mu) &= \mathsf{L}\left(\oplus_{i=u}^{v}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i} \bullet \nu_i\right) \lhd c_i\right)\right) \\
&= \left(\mathsf{I}(\nu_u), \sum_{i=u}^{v}\mathsf{W}_1(\nu_u), \sum_{i=u}^{v}\mathsf{W}_2(\nu_u), ..., \sum_{i=u}^{v}\mathsf{W}_{\mathsf{D}(\nu_u)-1}(\nu_u), \mathsf{O}(\nu_u)\right) \\
&= \left(\mathsf{I}(\nu_u), n\,\mathsf{W}_1(\nu_u), n\,\mathsf{W}_2(\nu_u), ..., n\,\mathsf{W}_{\mathsf{D}(\nu_u)-1}(\nu_u), \mathsf{O}(\nu_u)\right) \tag{4.6.28}
\end{aligned}$$

This establishes Item (i). Items (i) and (ii) from Lemma 4.4.5 tells us that for all $i \in \{u, u+1, ..., v\}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it is the case that $\mathfrak{I}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, B_i}\right) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right)$ and further that:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i} \bullet \nu_i\right)\right)(x) = \left(\mathfrak{I}_{\mathfrak{a}}(\nu_i)\right)(x) + b_i \tag{4.6.29}$$

This along with Lemma 4.6.6 ensures that for all $i \in \{u, u+1, ..., v\}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$,

it is the case that:

$$\mathfrak{I}_{\mathfrak{a}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\bullet\nu_i\right)\triangleleft c_i\right)\in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)},\mathbb{R}^{\mathsf{O}(\nu_u)}\right) \tag{4.6.30}$$

and:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\bullet\nu_i\right)\triangleleft c_i\right)\right)(x)=\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(c_i x+b_i\right) \tag{4.6.31}$$

Now observe that (Grohs et al., 2022, Lemma 3.28) and (**??**) ensure that for all $\mathfrak{a}\in C\left(\mathbb{R},\mathbb{R}\right)$, $x\in\mathbb{R}^{\mathsf{I}(\nu_u)}$, it is the case that $\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)},\mathbb{R}^{\mathsf{O}(\nu_u)}\right)$ and that:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\right)(x)=\left(\mathfrak{I}_{\mathfrak{a}}\left(\oplus_{i=u}^{v}\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\bullet\nu_i\right)\right)\triangleleft c_i\right)(x) \tag{4.6.32}$$

$$=\sum_{i=u}^{v}\left(\mathfrak{I}_{\mathfrak{a}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\bullet\nu_i\right)\triangleleft c_i\right)\right)(x) \tag{4.6.33}$$

$$=\sum_{i=u}^{v}\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(c_i x+b_i\right)$$

This establishes items (ii)—(iii); thus, the proof is complete. $\square$

Lemma 4.6.11. *Let $L\in\mathbb{N}$, $u,v\in\mathbb{Z}$ with $u\leqslant v$. Let $c_u,c_{u+1},...,c_v\in\mathbb{R}$. $\nu_u,\nu_{u+1},...,\nu_v,\mu\in$* **NN**, *$B_u,B_{u+1},...,B_v\in\mathbb{R}^{\mathsf{I}(\nu_u)}$, $\mathfrak{a}\in C\left(\mathbb{R},\mathbb{R}\right)$, satisfy for all $j\in\mathbb{N}\cap[u,v]$ that $L=\max_{i\in\mathbb{N}\cap[u,v]}$ $\mathsf{D}(\nu_i)$, $\mathsf{I}(\nu_j)=\mathsf{I}(\nu_u)$, $\mathsf{O}(\nu_j)=\mathsf{I}(\mathfrak{I})=\mathsf{O}(\mathfrak{I})$, $\mathsf{H}(\mathfrak{I})=1$, $\mathfrak{I}_{\mathfrak{a}}(\mathfrak{I})=\mathbb{I}_{\mathbb{R}}$, and that:*

$$\mu=\Diamond_{i=u,\mathfrak{I}}^{v}\left(c_i\triangleright\left(\nu_i\bullet\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\right)\right) \tag{4.6.34}$$

We then have that:

(i) it holds that $\mathfrak{I}_{\mathfrak{a}}(\mu)\in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)},\mathbb{R}^{\mathsf{O}(\nu_u)}\right)$, and that,

(ii) it holds for all $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$ that:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\right)\left(x\right) = \sum_{i=u}^{v} c_i \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(x + b_i\right) \tag{4.6.35}$$

Proof. Note that Item(i) from Lemma 4.6.9 establish Item(i) and (4.5.25); in addition, items (v) —(vi) from Lemma 4.4.5 tell us that for all $i \in \mathbb{N} \cap [u,v]$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it holds that $\mathfrak{I}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},B_i} \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right)\right)$ and further that:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},B_i}\right)\right)\left(x\right) = \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(x + b_k\right) \tag{4.6.36}$$

This, Lemma 4.6.5 and (Grohs et al., 2023, Lemma 2.14, Item (ii)) show that for all $i \in \mathbb{N} \cap [u,v]$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it holds that:

$$\mathfrak{I}_{\mathfrak{a}}\left(\mathfrak{E}_{L,\mathfrak{I}}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\right)\right)\right) = \mathfrak{I}_{\mathfrak{a}}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\right)\right) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right) \tag{4.6.37}$$

and:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\mathfrak{E}_{L,\mathfrak{I}}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\right)\right)\right)\right)\left(x\right) = \left(\mathfrak{I}_{\mathfrak{a}}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\right)\right)\right)\left(x\right)$$
$$= c_i \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(x + b_i\right) \tag{4.6.38}$$

This combined with (Grohs et al., 2022, Lemma 3.28) and (4.6.17) demonstrate that for all $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$ it holds that $\mathfrak{I}_{\mathfrak{a}}\left(\mu\right) \in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)}\right)$ and that:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\right)\left(x\right) = \left(\mathfrak{I}_{\mathfrak{a}}\left(\boxplus_{i=u,\mathfrak{I}}^{v}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}}\right)\right)\right)\right)\left(x\right)$$
$$= \left(\mathfrak{I}_{\mathfrak{a}}\left(\oplus_{i=u}^{v} \mathfrak{E}_{L,\mathfrak{I}}\left(c_i \triangleright \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\right)\right)\right)\right)\left(x\right)$$
$$= \sum_{i=u}^{v} c_i \left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(x + b_i\right) \tag{4.6.39}$$

This establishes Items(ii)--(iii), thus proving the lemma. $\qquad\square$

Lemma 4.6.12. *Let* $L \in \mathbb{N}$, $u, v \in \mathbb{Z}$ *with* $u \leqslant v$. *Let* $c_u, c_{u+1}, ..., c_v \in \mathbb{R}$. $\nu_u, \nu_{u+1}, ..., \nu_v, \mu, \mathfrak{I} \in$ NN, $B_u, B_{u+1}, ..., B_v \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, *satisfy for all* $j \in \mathbb{N} \cap [u, v]$ *that* $L = \max_{i \in \mathbb{N} \cap [u, v]} \mathsf{D}(\nu_i)$, $\mathsf{I}(\nu_j) = \mathsf{I}(\nu_u)$, $\mathsf{O}(\nu_j) = \mathsf{I}(\mathfrak{I}) = \mathsf{O}(\mathfrak{I})$, $\mathsf{H}(\mathfrak{I}) = 1$, $\mathfrak{I}_\mathfrak{a}(\mathfrak{I}) = \mathbb{I}_\mathbb{R}$, *and that:*

$$\mu = \boxplus_{i=u,\mathfrak{I}}^v \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i} \bullet \nu_i \right) \lhd c_i \right) \tag{4.6.40}$$

We then have:

(i) *it holds that:*

$$\mathsf{L}(\mu) = \left(\mathsf{I}(\nu_u), \sum_{i=u}^v \mathsf{W}_1 \left(\mathfrak{E}_{L,\mathfrak{I}} (\nu_i) \right), \sum_{i=u}^v \mathsf{W}_2 \left(\mathfrak{E}_{L,\mathfrak{I}} (\nu_i) \right), ..., \sum_{i=u}^v \mathsf{W}_{L-1} \left(\mathfrak{E}_{L,\mathfrak{I}} (\nu_i), \mathsf{O}(\nu_u) \right) \right) \tag{4.6.41}$$

(ii) *it holds that* $\mathfrak{I}_\mathfrak{a}(\mu) \in C \left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)} \right)$, *and that,*

(iii) *it holds for all* $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$ *that:*

$$\left(\mathfrak{I}_\mathfrak{a} (\mu) \right) (x) = \sum_{i=u}^v \left(\mathfrak{I}_\mathfrak{a} (\nu_i) \right) (c_i x + b_i) \tag{4.6.42}$$

Proof. Note that Item(i) from Lemma 4.6.10 establish Item(i) and (4.5.25); in addition, items (ii) and (iii) from Lemma 4.4.5 tell us that for all $i \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it holds that $\mathfrak{I}_\mathfrak{a} \left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, B_i} \bullet \nu_i \in C \left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)} \right) \right)$ and further that:

$$\left(\mathfrak{I}_\mathfrak{a} \left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, B_i} \bullet \nu_i \right) \right) (x) = \left(\mathfrak{I}_\mathfrak{a} (\nu_i) \right) (x) + b_k \tag{4.6.43}$$

This, Lemma 4.6.6 and (Grohs et al., 2023, Lemma 2.14, Item (ii)) show that for all $i \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathsf{I}(\nu_u)}$, it holds that:

$$\mathfrak{I}_\mathfrak{a} \left(\mathfrak{E}_{L,\mathfrak{I}} \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i} \bullet \nu_i \right) \lhd c_i \right) \right) = \mathfrak{I}_\mathfrak{a} \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}, b_i} \bullet \nu_i \right) \lhd c_i \right) \in C \left(\mathbb{R}^{\mathsf{I}(\nu_u)}, \mathbb{R}^{\mathsf{O}(\nu_u)} \right) \tag{4.6.44}$$

and:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\mathfrak{E}_{L,\mathfrak{I}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\bullet\nu_i\right)\triangleleft c_i\right)\right)\right)(x)=\left(\mathfrak{I}_{\mathfrak{a}}\left(c_i\triangleright\left(\nu_i\bullet\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\right)\right)\right)(x)$$

$$=\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(c_i x+b_i\right) \qquad (4.6.45)$$

This and (Grohs et al., 2022, Lemma 3.28) and (4.6.27) demonstrate that for all $x\in\mathbb{R}^{\mathsf{I}(\nu_u)}$ it holds that $\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\in C\left(\mathbb{R}^{\mathsf{I}(\nu_u)},\mathbb{R}^{\mathsf{O}(\nu_u)}\right)$ and that:

$$\left(\mathfrak{I}_{\mathfrak{a}}\left(\mu\right)\right)(x)=\left(\mathfrak{I}_{\mathfrak{a}}\left(\boxplus_{i=u,\mathfrak{I}}^{v}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)}}\bullet\nu_i\right)\triangleleft c_i\right)\right)\right)(x)$$

$$=\left(\mathfrak{I}_{\mathfrak{a}}\left(\oplus_{i=u}^{v}\mathfrak{E}_{L,\mathfrak{I}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathsf{I}(\nu_i)},b_i}\bullet\nu_i\right)\triangleleft c_i\right)\right)\right)(x)$$

$$=\sum_{i=u}^{v}\left(\mathfrak{I}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(c_i x+b_i\right) \qquad (4.6.46)$$

This completes the proof. □

4.7 Neural Network Diagrams

Conceptually, it will be helpful to construct what are called ``neural network diagrams''. They take inspiration from diagrams typically seen in the literature, for instance, Vaswani et al. (2017), Arik and Pfister (2021), and Chollet (2017). They are constructed as follows. Lines with arrows indicate the flow of data:

$$\xrightarrow{\quad x \quad}$$

$$\xleftarrow{\quad x \quad}$$
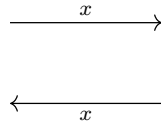
Named neural networks are always enclosed in boxes with serif fonts:

$$\boxed{\mathsf{Aff}_{a,b}}$$

Where possible, we seek to label the arrows going in and going out of a boxed neural network with the appropriate operations that take place:

114

$$\xleftarrow{ax+b} \boxed{\mathsf{Aff}_{a,b}} \xleftarrow{x}$$
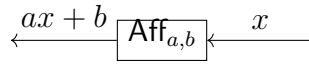
It is often more helpful to draw the arrows from right to left, as above.

Stacked neural networks are drawn in adjacent boxes.

$$\xleftarrow{ax+b} \boxed{\mathsf{Aff}_{a,b}} \xleftarrow{x}$$

$$\xleftarrow{\;} \boxed{\mathsf{Aff}_{c,d}} \xleftarrow{\;}$$
$$cx+d \qquad\qquad x$$
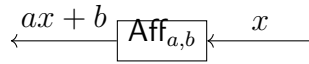
For neural networks that take in two inputs and give out one output, we use two arrows going in and one arrow going out:

$$\xleftarrow{x+y} \boxed{\mathsf{Sum}_{2,1}}$$

with x and y as the two incoming arrows.
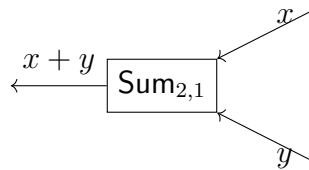
For neural networks that take in one input and give out two outputs, we use one arrow going in and two arrows going out:

with $\boxed{\mathsf{Cpy}_{1,2}}$ taking in x and outputting x and x.
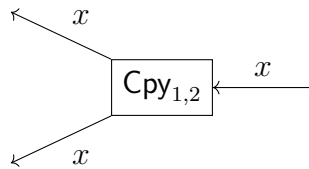
Thus taking this all together the sum of neural networks $\mathsf{Aff}_{a,b}, \mathsf{Aff}_{c,d} \in \mathsf{NN}$ is given by:

115
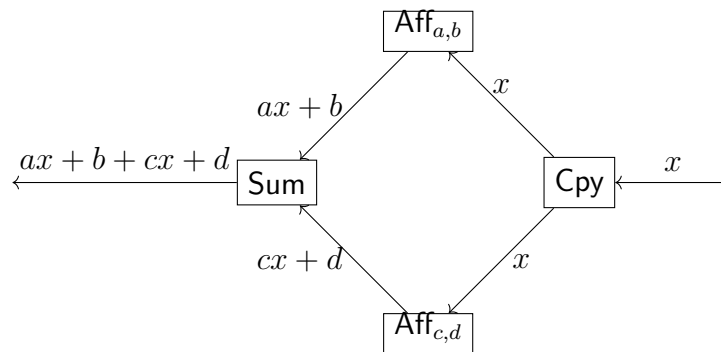
Chapter 5

ANN Product Approximations and Their Consequences

5.1 Approximation for Products of Two Real Numbers

We will build up the tools necessary to approximate e^x via neural networks in the framework described in the previous sections. While much of the foundation comes from, e.g., Grohs et al. (2023) way, we will, along the way, encounter neural networks not seen in the literature, such as the Tay, Pwr, Tun, and finally a neural network approximant for e^x. For each of these neural networks, we will be concerned with at least the following:

(i) whether their instantiations using the ReLU function (often just continuous functions) are continuous.

(ii) whether their depths are bounded, at most polynomially, on the type of accuracy we want, ε.

(iii) whether their parameter estimates are bounded at most polynomially on the type of accuracy we want, ε.

(iv) The accuracy of our neural networks.

The sections pertaining to squaring and taking the product of neural networks derive mostly from Yarotsky (2017) via Jentzen et al. (2023).

5.1.1   The squares of real numbers in $[0, 1]$

One of the most important operators we will approximate is the product operator $\times$ for two real numbers. The following sections takes a streamlined version of the proof given in (Grohs et al., 2023, Section 3.1). In particular we will assert the existence of the neural network Φ and ϕ_d and work our way towards its properties.

Definition 5.1.1 (The $\mathfrak{i}_d$ Network). *For all $d \in \mathbb{N}$ we will define the following set of neural networks as ``activation neural networks'' denoted $\mathfrak{i}_d$ as:*

$$\mathfrak{i}_d = ((\mathbb{I}_d, \mathbb{O}_d), (\mathbb{I}_d, \mathbb{O}_d)) \tag{5.1.1}$$

Lemma 5.1.2. *Let $d \in \mathbb{N}$. It is then the case that:*

(i) $\mathfrak{I}_{\mathfrak{r}}(\mathfrak{i}_4) \in C\left(\mathbb{R}^d, \mathbb{R}^d\right)$.

(ii) $\mathsf{L}(\mathfrak{i}_d) = (d, d, d)$

(iii) $\mathsf{P}(\mathfrak{i}_4) = 2d^2 + 2d$

Proof. Item (i) is straightforward from the fact that for all $d \in \mathbb{N}$ it is the case that $\mathfrak{I}_{\mathfrak{r}}(\mathfrak{i}_d) = \mathbb{I}_d\left(\mathfrak{I}_{\mathfrak{r}}\left([\mathbb{I}_d]_*\right) + \mathbb{O}_d\right) + \mathbb{O}_d$. Item (ii) is straightforward from the fact that $\mathbb{I}_d \in \mathbb{R}^{d \times d}$. We realize Item (iii) by observation. □

Lemma 5.1.3. *Let $(c_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$, $(A_k)_{k \in \mathbb{N}} \in \mathbb{R}^{4 \times 4}$, $\mathbb{B} \in \mathbb{R}^{4 \times 1}$, $(C_k)_{k \in \mathbb{N}}$ satisfy for all $k \in \mathbb{N}$*

that:

$$A_k = \begin{bmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{bmatrix} \quad C_k = \begin{bmatrix} -c_k & 2c_k & -c_k & 1 \end{bmatrix} \tag{5.1.2}$$

and that:

$$c_k = 2^{1-2k} \tag{5.1.3}$$

Let $\Phi_k \in \mathsf{NN}$, $k \in \mathbb{N}$ satisfy for all $k \in [2, \infty) \cap \mathbb{N}$ that $\Phi_1 = (\mathsf{Aff}_{C_1,0} \bullet \mathsf{i}_4) \bullet \mathsf{Aff}_{\mathfrak{e}_4,B}$, that for all $d \in \mathbb{N}$, $\mathsf{i}_d = ((\mathbb{I}_d, \mathbb{0}_d), (\mathbb{I}_d, \mathbb{0}_d))$ and that:

$$\Phi_k = (\mathsf{Aff}_{C_k,0} \bullet \mathsf{i}_4) \bullet \left(\mathsf{Aff}_{A_{k-1},B} \bullet \mathsf{i}_4\right) \bullet \cdots \bullet (\mathsf{Aff}_{A_1,B} \bullet \mathsf{i}_4) \bullet \mathsf{Aff}_{\mathfrak{e}_4,B}, \tag{5.1.4}$$

It is then the case that:

(i) for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ we have $\mathfrak{I}_\mathfrak{r}(\Phi_k)(x) \in C(\mathbb{R}, \mathbb{R})$

(ii) for all $k \in \mathbb{N}$ we have $\mathsf{L}(\Phi_k) = (1, 4, 4, ..., 4, 1) \in \mathbb{N}^{k+2}$

(iii) for all $k \in \mathbb{N}$, $x \in \mathbb{R} \setminus [0, 1]$ that $(\mathfrak{I}_\mathfrak{r}(\Phi_k))(x) = \mathfrak{r}(x)$

(iv) for all $k \in \mathbb{N}$, $x \in [0, 1]$, we have $|x^2 - (\mathfrak{I}_\mathfrak{r}(\Phi_k))(x)| \leqslant 2^{-2k-2}$, and

(v) for al $k \in \mathbb{N}$, we have that $\mathsf{P}(\Phi_k) = 20k - 7$

Proof. Firstly note that Lemma 4.4.2, Lemma 4.2.5, and Lemma 5.1.2 ensure that for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ it is the case that $\mathfrak{I}_\mathfrak{r}(\Phi_k)(x) \in C(\mathbb{R}, \mathbb{R})$. This proves Item (i).

Note next that Lemma 4.4.2, Lemma 5.1.2, and Lemma 4.2.5 tells us that:

$$\mathsf{L}(\Phi_1) = \mathsf{L}(\mathsf{Aff}_{\mathfrak{e}_4}, B) = (1, 4, 1) \tag{5.1.5}$$

118

and for all $k \in \mathbb{N}$ it is the case that:

$$\mathsf{L}\left(\mathsf{Aff}_{A_k,B} \bullet \mathsf{i}_4\right) = (4,4,4,4) \tag{5.1.6}$$

Whence it is straightforward to see that for Φ_k where $k \in \mathbb{N} \cap [2,\infty)$, Lemma 4.2.5 tells us then that:

$$\mathsf{L}\left(\Phi_k\right) = \mathsf{L}\left(\left(\mathsf{Aff}_{C_k,0} \bullet \mathsf{i}_4\right) \bullet \left(\mathsf{Aff}_{A_{k-1},B} \bullet \mathsf{i}_4\right) \bullet \cdots \bullet \left(\mathsf{Aff}_{A_1,B} \bullet \mathsf{i}_4\right) \bullet \mathsf{Aff}_{\mathbf{e}_4,B}\right)$$

$$= (1,4) \; \overbrace{(4}^{merged},4,4,\; \overbrace{4)(4}^{merged},4,4,\; \overbrace{4)}^{merged} \;\cdots\; \overbrace{(4}^{merged},4,4,\; \overbrace{4)}^{merged}(4,1) \tag{5.1.7}$$

$$\underbrace{}_{k-1 \text{ many}}$$

This thus finally yields that:

$$\mathsf{L}\left(\Phi_k\right) = (1,4,4,\ldots,4,1) \in \mathbb{N}^{k+2} \tag{5.1.8}$$

Let $g_k : \mathbb{R} \to [0,1]$, $k \in \mathbb{N}$ be the functions defined as such, satisfying for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ that:

$$g_1(x) = \begin{cases} 2x & : x \in \left[0, \frac{1}{2}\right) \\ 2 - 2x & : x \in \left[\frac{1}{2}, 1\right] \\ 0 & : x \in \mathbb{R} \setminus [0,1] \end{cases} \tag{5.1.9}$$

$$g_{k+1} = g_1(g_k)$$

and let $f_k : [0,1] \to [0,1]$, $k \in \mathbb{N}_0$ be the functions satisfying for all $k \in \mathbb{N}_0$, $n \in \{0,1,\ldots,2^k - 1\}$, $x \in \left[\frac{n}{2^k}, \frac{n+1}{2^k}\right)$ that $f_k(1) = 1$ and:

$$f_k(x) = \left[\frac{2n+1}{2^k}\right]x - \frac{n^2 + n}{2^{2k}} \tag{5.1.10}$$

119

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 which satisfy for all $x \in \mathbb{R}$, $k \in \mathbb{N}$ that:

$$r_1(x) = \begin{bmatrix} r_{1,1}(x) \\ r_{2,1}(x) \\ r_{3,1}(x) \\ r_{4,1}(x) \end{bmatrix} = \mathfrak{r} \left(\begin{bmatrix} x \\ x - \frac{1}{2} \\ x - 1 \\ x \end{bmatrix} \right) \tag{5.1.11}$$

$$r_{k+1} = \mathfrak{r}\left(A_{k+1} r_k(x) + B\right)$$

Note that since it is the case that for all $x \in \mathbb{R}$ that $\mathfrak{r}(x) = \max\{x, 0\}$, (5.1.9) and (5.1.11) shows that it holds for all $x \in \mathbb{R}$ that:

$$2r_{1,1}(x) - 4r_{2,1}(x) + 2r_{3,1}(x) = 2\,\mathfrak{r}(x) - 4\,\mathfrak{r}\left(x - \frac{1}{2}\right) + 2\,\mathfrak{r}(x-1)$$

$$= 2\max\{x,0\} - 4\max\left\{x - \frac{1}{2}, 0\right\} + 2\max\{x-1,0\}$$

$$= g_1(x) \tag{5.1.12}$$

Note also that combined with (5.1.10), the fact that for all $x \in [0,1]$ it holds that $f_0(x) = x = \max\{x,0\}$ tells us that for all $x \in \mathbb{R}$:

$$r_{4,1}(x) = \max\{x,0\} = \begin{cases} f_0(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R} \setminus [0,1] \end{cases} \tag{5.1.13}$$

We next claim that for all $k \in \mathbb{N}$, it is the case that:

$$(\forall x \in \mathbb{R} : 2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x) = g(x)) \tag{5.1.14}$$

and that:

$$\left(\forall x \in \mathbb{R} : r_{4,k}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R} \setminus [0,1] \end{cases} \right) \tag{5.1.15}$$

We prove (5.1.14) and (5.1.15) by induction. The base base of $k = 1$ is proved by (5.1.12) and (5.1.13) respectively. For the induction step $\mathbb{N} \ni k \to k+1$ assume there does exist a $k \in \mathbb{N}$ such that for all $x \in \mathbb{R}$ it is the case that:

$$2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x) = g_k(x) \tag{5.1.16}$$

and:

$$r_{4,k}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R} \setminus [0,1] \end{cases} \tag{5.1.17}$$

Note that then (5.1.9), (5.1.11), and (5.1.12) then tells us that for all $x \in \mathbb{R}$ it is the case that:

$$\begin{aligned}
g_{k+1}(x) &= g_1(g_k(x)) = g_1(2r_{1,k}(x) + 4r_{2,k}(x) + 2r_{3,k}(x)) \\
&= 2\,\mathfrak{r}\,(2r_{1,k}(x)) + 4r_{2,k} + 2r_{3,k}(x)) \\
&\quad - 4\,\mathfrak{r}\,\left(2r_{1,k}(x) - 4r_{2,k} + 2r_{3,k}(x) - \frac{1}{2}\right) \\
&\quad + 2\,\mathfrak{r}\,(2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x) - 1) \\
&= 2r_{1,k+1}(x) - 4r_{2,k+1}(x) + 2r_{3,k+1}(x)
\end{aligned} \tag{5.1.18}$$

In addition note that (5.1.10), (5.1.11), and (5.1.13) tells us that for all $x \in \mathbb{R}$:

$$r_{4,k+1}(x) = \mathfrak{r}\left((-2)^{3-2(k+1)} r_{1,k}(x) + 2^{4-2(k+1)} r_{2,k}(x) + (-2)^{3-2(k+1)} r_{3,k}(x) + r_{4,k}(x)\right)$$

$$= \mathfrak{r}\left((-2)^{1-2k} r_{1,k}(x) + 2^{2-2k} r_{k,2}(x) + (-2)^{1-2k} r_{3,k}(x) + r_{4,k}(x)\right)$$

$$= \mathfrak{r}\left(2^{-2k}\left[-2r_{1,k}(x) + 2^2 r_{2,k}(x) - 2r_{3,k}(x)\right] + r_{4,k}(x)\right)$$

$$= \mathfrak{r}\left(-\left[2^{-2k}\right]\left[2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x)\right] + r_{4,k}(x)\right)$$

$$= \mathfrak{r}\left(-\left[2^{-2k}\right] g_k(x) + r_{4,k}(x)\right) \tag{5.1.19}$$

This and the fact that for all $x \in \mathbb{R}$ it is the case that $\mathfrak{r}(x) = \max\{x, 0\}$, that for all $x \in [0, 1]$ it is the case that $f_k(x) \geqslant 0$, (5.1.17), shows that for all $x \in [0, 1]$ it holds that:

$$r_{4,k+1}(x) = \mathfrak{r}\left(-2\left[2^{-2k} g_k\right] + f_{k-1}(x)\right) = \mathfrak{r}\left(-2\left(2^{-2k} g_k(x)\right) + x - \left[\sum_{j=1}^{k-1}\left(2^{-2j} g_j(x)\right)\right]\right)$$

$$= \mathfrak{r}\left(x - \left[\sum_{j=1}^{k} 2^{-2j} g_j(x)\right]\right) = \mathfrak{r}(f_k(x)) = f_k(x) \tag{5.1.20}$$

Note next that (5.1.17) and (5.1.19) then tells us that for all $x \in \mathbb{R} \setminus [0, 1]$:

$$r_{4,k+1}(x) = \max\left\{-\left(2^{-2k} g_x(x)\right) + r_{4,k}(x)\right\} = \max\{\max\{x, 0\}, 0\} = \max\{x, 0\} \tag{5.1.21}$$

Combining (5.1.18) and (5.1.20) proves (5.1.14) and (5.1.15). Note that then (5.1.2) and

(5.1.14) assure that for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ it holds that $\mathfrak{I}_{\mathfrak{r}}(\Phi_k) \in C(\mathbb{R}, \mathbb{R})$ and that:

$$
\begin{aligned}
(\mathfrak{I}_{\mathfrak{r}}(\Phi_k))\,(x) \\
&= \left(\mathfrak{I}_{\mathfrak{r}}\left(\left(\mathsf{Aff}_{C_k,0} \bullet \mathfrak{i}_4\right) \bullet \left(\mathsf{Aff}_{A_{k-1},B} \bullet \mathfrak{i}_4\right) \bullet \cdots \bullet \left(\mathsf{Aff}_{A_1,B} \bullet \mathfrak{i}_4\right) \bullet \mathsf{Aff}_{\mathfrak{e}_4,B}\right)\right)(x) \\
&= (-2)^{1-2k}\, r_{1,k}\,(x) + 2^{2-2k} r_{2,k}\,(x) + (-2)^{1-2k}\, r_{3,k}\,(x) + r_{4,k}\,(x) \\
&= (-2)^{2-2k}\left(\left[\frac{r_{1,k}\,(x) + r_{3,k}\,(x)}{-2}\right] + r_{2,k}\,(x)\right) + r_{4,k}\,(x) \\
&= 2^{2-2k}\left(\left[\frac{r_{1,k}\,(x) + r_{3,k}\,(x)}{-2}\right] + r_{2,k}\,(x)\right) + r_{4,k}\,(x) \\
&= 2^{-2k}\left(4r_{2,k}\,(x) - 2r_{1,k}\,(x) - 2r_{3,k}\,(x)\right) + r_{4,k}\,(x) \\
&= -\left[2^{-2k}\right]\left[2r_{1,k}\,(x) - 4r_{2,k}\,(x) + 2r_{3,k}\,(x)\right] + r_{4,k}\,(x) = -\left[2^{-2k}\right] g_k\,(x) + r_{4,k}\,(x) \quad (5.1.22)
\end{aligned}
$$

This and (5.1.15) tell us that:

$$
\begin{aligned}
(\mathfrak{I}_{\mathfrak{r}}(\Phi_k))\,(x) = -\left(2^{-2k} g_k\,(x)\right) + f_{k-1}\,(x) &= -\left(2^{-2k} g_k\,(x)\right) + x - \left[\sum_{j=1}^{k-1} 2^{-2j} g_j\,(x)\right] \\
&= x - \left[\sum_{j=1}^{k} 2^{-2j} g_j\,(x)\right] = f_k\,(x)
\end{aligned}
$$

Which then implies for all $k \in \mathbb{N}$, $x \in [0,1]$ that it holds that:

$$
\left|x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi_k))\,(x)\right| \leqslant 2^{-2k-2} \tag{5.1.23}
$$

This, in turn, establishes Item (i).

Finally observe that (5.1.22) then tells us that for all $k \in \mathbb{N}$, $x \in \mathbb{R} \setminus [0,1]$ it holds that:

$$
(\mathfrak{I}_{\mathfrak{r}}(\Phi_k))\,(x) = -2^{-2k} g_k\,(x) + r_{4,k}\,(x) = r_{4,k}\,(x) = \max\{x, 0\} = \mathfrak{r}(x) \tag{5.1.24}
$$

This establishes Item(iv). Note next that Item(iii) ensures for all $k \in \mathbb{N}$ that $\mathsf{D}(\xi_k) = k+1$,

and:

$$\mathsf{P}\left(\Phi_k\right) = 4(1+1) + \left[\sum_{j=2}^{k} 4\left(4+1\right)\right] + (4+1) = 8 + 20\left(k-1\right) + 5 = 20k - 7 \qquad (5.1.25)$$

This, in turn, proves Item(vi). The proof of the lemma is thus complete. $\qquad\square$

Remark 5.1.4. *For an* R *implementation see Listing 9.13*
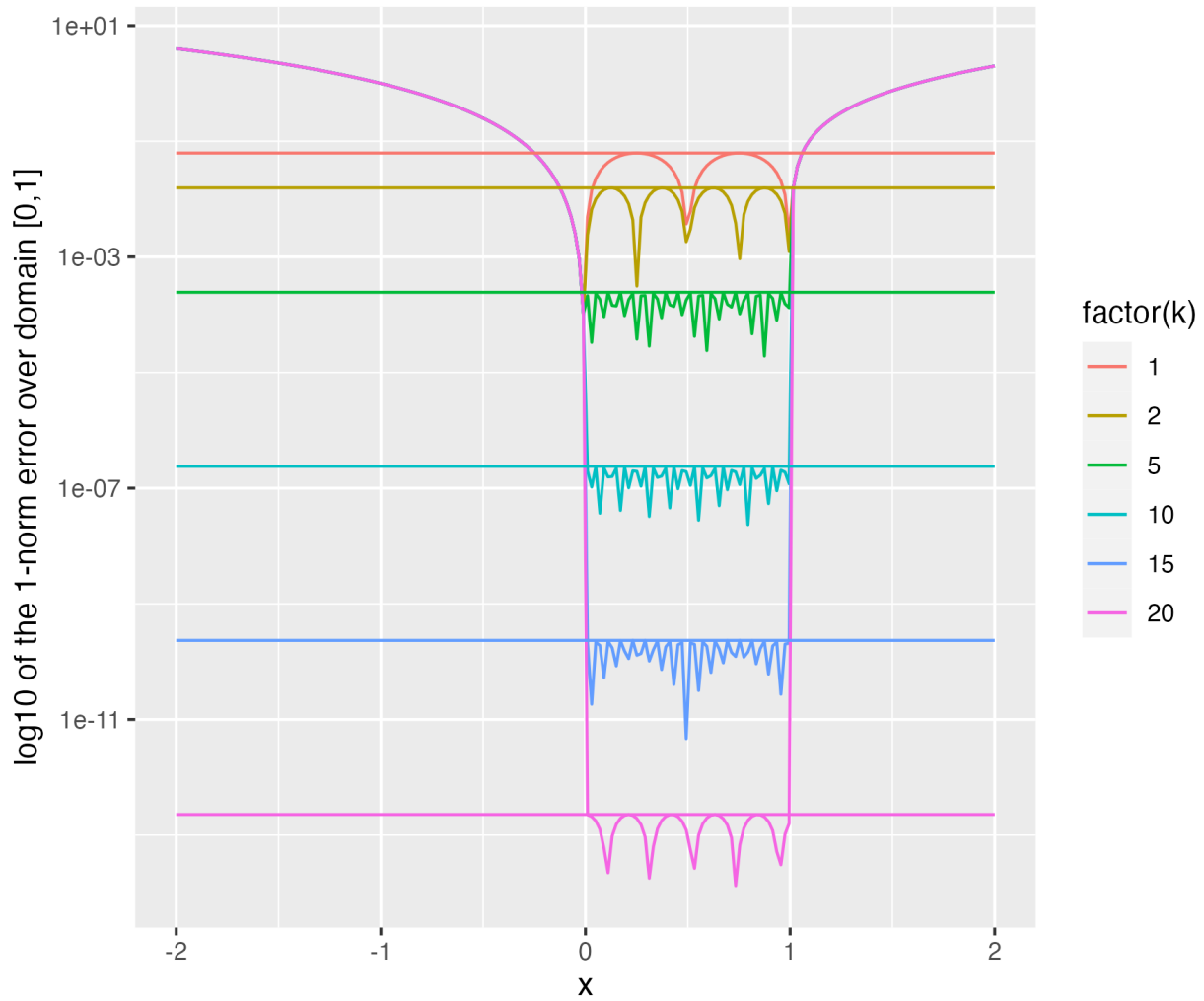
Figure 5.1: Plot of $\log_{10}$ of the L^1 difference between Φ_k and x^2 over $[0,1]$ for different values of k

Corollary 5.1.4.1. *Let* $\varepsilon \in (0,\infty)$, $M = \min\{\frac{1}{2}\log_2\left(\varepsilon^{-1}\right) - 1, \infty\} \cap \mathbb{N}$, $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$,

124

$(A_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}^{4\times 4}$, $B \in \mathbb{R}^{4\times 1}$, $(C_k)_{k\in\mathbb{N}}$ *satisfy for all* $k \in \mathbb{N}$ *that:*

$$
A_k = \begin{bmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{bmatrix} \quad C_k = \begin{bmatrix} -c_k & 2c)_k & -c_k & 1 \end{bmatrix} \tag{5.1.26}
$$

where:

$$
c_k = 2^{1-2k} \tag{5.1.27}
$$

and let $\Phi \in \mathsf{NN}$ *be defined as:*

$$
\Phi = \begin{cases} \left[\mathsf{Aff}_{C_1,0} \bullet \mathfrak{i}_4\right] \bullet \mathsf{Aff}_{\mathfrak{e}_4,B} & : M = 1 \\[2mm] \left[\mathsf{Aff}_{C_M,0} \bullet \mathfrak{i}_4\right] \bullet \left[\mathsf{Aff}_{A_{M-1},0} \bullet \mathfrak{i}_4\right] \bullet \cdots \bullet \left[\mathsf{Aff}_{A_1,B} \bullet \mathfrak{i}_4\right] \bullet \mathsf{Aff}_{\mathfrak{e}_4,B} & : M \in [2,\infty) \cap \mathbb{N} \end{cases} \tag{5.1.28}
$$

it is then the case that:

(i) $\mathfrak{I}_{\mathfrak{r}}(\Phi) \in C(\mathbb{R},\mathbb{R})$

(ii) $\mathsf{L}(\Phi) = (1,4,4,...,4,1) \in \mathbb{N}^{M+2}$

(iii) *it holds for all* $x \in \mathbb{R}\setminus[0,1]$ *that* $(\mathfrak{I}_{\mathfrak{r}}(\Phi))(x) = \mathfrak{r}(x)$

(iv) *it holds for all* $x \in [0,1]$ *that* $|x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi))(x)| \leqslant 2^{-2M-2} \leqslant \varepsilon$

(v) $\mathsf{D}(\Phi) \leqslant M + 1 \leqslant \max\{\frac{1}{2}\log_2(\varepsilon^{-1}) + 1, 2\}$, *and*

(vi) $\mathsf{P}(\Phi) = 20M - 7 \leqslant \max\{10\log_2(\varepsilon^{-1}) - 7, 13\}$

Proof. Items (i)--(iii) are direct consequences of Lemma 5.1.3, Items (i)--(iii). Note next the

fact that $M = \min\left\{\mathbb{N} \cap \left[\frac{1}{2}\log_2\left(\varepsilon^{-1}\right) - 1\right], \infty\right\}$ ensures that:

$$M = \min\left\{\mathbb{N} \cap \left[\frac{1}{2}\log_2\left(\varepsilon^{-1}\right) - 1\right], \infty\right\} \tag{5.1.29}$$

$$\geqslant \min\left\{\left[\max\left\{1, \frac{1}{2}\log_2\left(\varepsilon^{-1}\right) - 1\right\}, \infty\right]\right\} \tag{5.1.30}$$

$$\geqslant \frac{1}{2}\log_2\left(\varepsilon^{-1}\right) - 1 \tag{5.1.31}$$

This and Item (v) of Lemma 5.1.3 demonstrate that for all $x \in [0, 1]$ it then holds that:

$$\left|x^2 - \left(\mathfrak{I}_{\mathfrak{r}}\left(\Phi\right)\right)(x)\right| \leqslant 2^{-2M-2} = 2^{-2(M+1)} \leqslant 2^{-\log_2\left(\varepsilon^{-1}\right)} = \varepsilon \tag{5.1.32}$$

Thus establishing Item (iv). The fact that $M = \min\left\{\mathbb{N} \cap \left[\frac{1}{2}\log_2\left(\varepsilon^{-1}\right) - 1, \infty\right]\right\}$ and Item (ii) of Lemma 5.1.3 tell us that:

$$\mathsf{D}\left(\Phi\right) = M + 1 \leqslant \max\left\{\frac{1}{2}\log_2\left(\varepsilon^{-1}\right) + 1, 2\right\} \tag{5.1.33}$$

Which establishes Item(v). This and Item (v) of Lemma 5.1.3 then tell us that:

$$\mathsf{P}\left(\Phi_M\right) \leqslant 20M - 7 \leqslant 20\max\left\{\frac{1}{2}\log_2\left(\varepsilon^{-1}\right), 2\right\} - 7 = \max\left\{10\log_2\left(\varepsilon^{-1}\right) - 7, 13\right\} \tag{5.1.34}$$

This completes the proof of the corollary. $\qquad\square$

Remark 5.1.5. *For an implementation in* R*, see Listing 9.15*

Remark 5.1.6. *Note that (5.1.28) implies that* $\mathsf{D}\left(\Phi\right) \geqslant 4$*.*

Now that we have neural networks that perform the squaring operation inside $[-1, 1]$, we may extend to all of $\mathbb{R}$. Note that this neural network representation differs somewhat from the ones in Grohs et al. (2023).
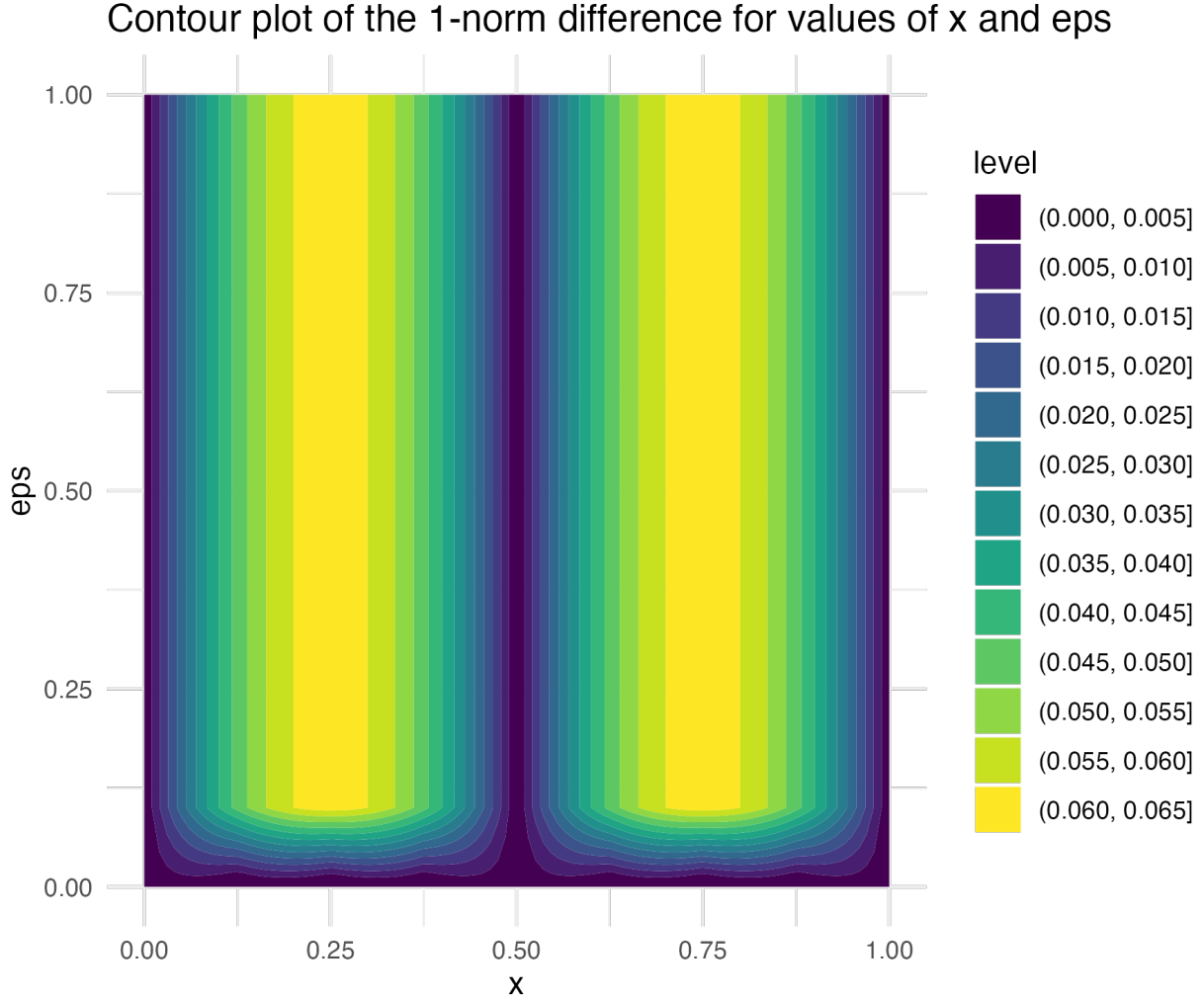
Figure 5.2: Contour plot of the L^1 difference between Φ and x^2 over $[0,1]$ for different values of ε.

5.1.2 The Sqr$^{q,\varepsilon}$ Neural Networks and Squares of Real Numbers

Lemma 5.1.7. *Let* $\delta, \epsilon \in (0,\infty)$, $\alpha \in (0,\infty)$, $q \in (2,\infty)$, $\Phi \in \mathsf{NN}$ *satisfy that* $\delta = 2^{\frac{-2}{q-2}} \varepsilon^{\frac{q}{q-2}}$, $\alpha = \left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}}$, $\mathfrak{I}_{\mathfrak{r}}(\Phi) \in C(\mathbb{R},\mathbb{R})$, $\mathsf{D}(\Phi) \leqslant \max\left\{\frac{1}{2}\log_2(\delta^{-1}) + 1, 2\right\}$, $\mathsf{P}(\Phi) \leqslant \max\left\{10\log_2\left(\delta^{-1}\right) - 7, 13\right\}$, $\sup_{x \in \mathbb{R}\backslash[0,1]}|(\mathfrak{I}_{\mathfrak{r}}(\Phi) - \mathfrak{r}(x)| = 0$, *and* $\sup_{x \in [0,1]}|x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi))(x)| \leqslant \delta$, *let* $\Psi \in \mathsf{NN}$ *be the neural network given by:*

$$\Psi = \left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}\right) \bigoplus \left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}\right) \tag{5.1.35}$$

(i) it holds that $\mathfrak{I}_{\mathfrak{r}}(\Psi) \in C(\mathbb{R}, \mathbb{R})$.

(ii) it holds that $(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0) = 0$

(iii) it holds for all $x \in \mathbb{R}$ that $0 \leqslant (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \leqslant \varepsilon + |x|^2$

(iv) it holds for all $x \in \mathbb{R}$ that $|x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x)| \leqslant \varepsilon \max\{1, |x|^q\}$

(v) it holds that $\mathsf{D}(\Psi) \leqslant \max\left\{1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \log_2(\varepsilon^{-1}), 2\right\}$, and

(vi) it holds that $\mathsf{P}(\Psi) \leqslant \max\left\{\left[\frac{40q}{q-2}\right] \log_2(\varepsilon^{-1}) + \frac{80}{q-2} - 28, 52\right\}$

Proof. Note that for all $x \in \mathbb{R}$ it is the case that:

$$(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) = (\mathfrak{I}_{\mathfrak{r}}((\mathsf{Aff}_{\alpha^{-2}} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}) \oplus (\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0})))(x)$$

$$= (\mathfrak{I}_{\mathfrak{r}}(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}))(x) + (\mathfrak{I}_{\mathfrak{r}}(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}))(x)$$

$$= \frac{1}{\alpha^2}(\mathfrak{I}_{\mathfrak{r}}(\Phi))(\alpha x) + \frac{1}{\alpha^2}(\mathfrak{I}_{\mathfrak{r}}(\Phi))(-\alpha x)$$

$$= \frac{1}{\left(\frac{\varepsilon}{2}\right)^{\frac{2}{q-2}}}\left[(\mathfrak{I}_{\mathfrak{r}}(\Phi))\left(\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x\right) + (\mathfrak{I}_{\mathfrak{r}}(\Phi))\left(-\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x\right)\right] \quad (5.1.36)$$

This and the assumption that $\Phi \in C(\mathbb{R}, \mathbb{R})$ along with the assumption that $\sup_{x \in \mathbb{R} \setminus [0,1]} |(\mathfrak{I}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ tells us that for all $x \in \mathbb{R}$ it holds that:

$$(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0) = \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}}[(\mathfrak{I}_{\mathfrak{r}}(\Phi))(0) + (\mathfrak{I}_{\mathfrak{r}}(\Phi))(0)]$$

$$= \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}}[\mathfrak{r}(0) + \mathfrak{r}(0)]$$

$$= 0 \quad (5.1.37)$$

This, in turn, establishes Item (i)--(ii). Observe next that from the assumption that $\mathfrak{I}_{\mathfrak{r}}(\Phi) \in C(\mathbb{R}, \mathbb{R})$ and the assumption that $\sup_{x \in \mathbb{R} \setminus [0,1]} |(\mathfrak{I}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ ensure that for all

128

$x \in \mathbb{R} \setminus [-1, 1]$ it holds that:

$$[\mathfrak{I}_{\mathfrak{r}}(\Phi)](x) + [\mathfrak{I}_{\mathfrak{r}}(\Phi)(-x)] = \mathfrak{r}(x) + \mathfrak{r}(-x) = \max\{x, 0\} + \max\{-x, 0\}$$

$$= |x| \tag{5.1.38}$$

The assumption that for all $\sup_{x \in \mathbb{R} \setminus [0,1]} |(\mathfrak{I}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ and the assumption that $\sup_{x \in [0,1]} |x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi))(x)| \leqslant \delta$ show that:

$$\sup_{x \in [-1,1]} \left| x^2 - ([\mathfrak{I}_{\mathfrak{r}}(\Phi)](x) + [\mathfrak{I}_{\mathfrak{r}}(\Phi)(x)]) \right|$$

$$= \max \left\{ \sup_{x \in [-1,0]} \left| x^2 - (\mathfrak{r}(x) + [\mathfrak{I}_{\mathfrak{r}}(\Phi)](-x)) \right|, \ \sup_{x \in [0,1]} \left| x^2 - ([\mathfrak{I}_{\mathfrak{r}}(\Phi)](x) + \mathfrak{r}(-x)) \right| \right\}$$

$$= \max \left\{ \sup_{x \in [-1,0]} \left| (-x)^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi))(-x) \right|, \ \sup_{x \in [0,1]} \left| x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi))(x) \right| \right\}$$

$$= \sup_{x \in [0,1]} \left| x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi))(x) \right| \leqslant \delta \tag{5.1.39}$$

Next observe that (5.1.36) and (5.1.38) show that for all $x \in \mathbb{R} \setminus \left[-\left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}}, \left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}} \right]$ it holds that:

$$0 \leqslant [\mathfrak{I}_{\mathfrak{r}}(\Psi)](x) = \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} \left([\mathfrak{I}_{\mathfrak{r}}(\Phi)] \left(\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x \right) + [\mathfrak{I}_{\mathfrak{r}}(\Phi)] \left(-\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x \right) \right)$$

$$= \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} \left| \left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x \right| = \left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}|x|} \leqslant |x|^2 \tag{5.1.40}$$

The triangle inequality then tells us that for all $x \in \mathbb{R} \setminus \left[- \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}}, \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}} \right]$ it holds that:

$$
\begin{aligned}
\left| x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \right| &= \left| x^2 - \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}} |x| \right| \leqslant \left(|x|^2 + \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}} |x| \right) \\
&= \left(|x|^q |x|^{-(q-2)} + \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}} |x|^q |x|^{-(q-1)} \right) \\
&\leqslant \left(|x|^q \left(\frac{\varepsilon}{2} \right)^{\frac{q-2}{q-2}} + \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}} |x|^q \left(\frac{\varepsilon}{2} \right)^{\frac{q-1}{q-2}} \right) \\
&= \left(\frac{\varepsilon}{2} + \frac{\varepsilon}{2} \right) |x|^q = \varepsilon |x|^q \leqslant \varepsilon \max \left\{ 1, |x|^q \right\}
\end{aligned}
\tag{5.1.41}
$$

Note that (5.1.39), (5.1.36) and the fact that $\delta = 2^{\frac{-2}{q-2}} \varepsilon^{\frac{q}{q-2}}$ then tell for all $x \in \left[- \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}}, \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}} \right]$ it holds that:

$$
\begin{aligned}
& \left| x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Phi))(x) \right| \\
&= \left(\frac{\varepsilon}{2} \right)^{\frac{-2}{q-2}} \left| \left(\left(\frac{\varepsilon}{2} \right)^{\frac{1}{q-2}} x \right)^2 - \left([\mathfrak{I}_{\mathfrak{r}}(\Phi)] \left(\left(\frac{\varepsilon}{2} \right)^{\frac{1}{q-2}} x \right) + [\mathfrak{I}_{\mathfrak{r}}(\Phi)](-y) \right) \right| \\
&\leqslant \left(\frac{\varepsilon}{2} \right)^{\frac{-2}{q-2}} \left[\sup_{y \in [-1,1]} \left| y^2 - [\mathfrak{I}_{\mathfrak{r}}(\Phi)](y) + [\mathfrak{I}_{\mathfrak{r}}(\Phi)](-y) \right| \right] \\
&\leqslant \left(\frac{\varepsilon}{2} \right)^{\frac{-2}{q-2}} \delta = \left(\frac{\varepsilon}{2} \right)^{\frac{-2}{q-2}} 2^{\frac{-2}{q-2}} \varepsilon^{\frac{q}{q-2}} = \varepsilon \leqslant \varepsilon \max \{ 1, |x|^q \}
\end{aligned}
\tag{5.1.42}
$$

Now note that this and (5.1.41) tells us that for all $x \in \mathbb{R}$ it is the case that:

$$
\left| x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \right| \leqslant \varepsilon \max \{ 1, |x|^q \}
\tag{5.1.43}
$$

This establishes Item (v). Note that, (5.1.42) tells that for all $x \in \left[- \left(\frac{\varepsilon}{2} \right)^{\frac{-1}{q-2}}, \left(\frac{\varepsilon}{2} \right)^{\frac{1}{q-2}} \right]$ it is the case that:

$$
\left| (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \right| \leqslant \left| x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \right| + |x|^2 \leqslant \varepsilon + |x|^2
\tag{5.1.44}
$$

This and (5.1.41) tells us that for all $x \in \mathbb{R}$:

$$|(\mathfrak{I}_{\mathfrak{r}})(x)| \leqslant \varepsilon + |x|^2 \tag{5.1.45}$$

This establishes Item (iv).

Note next that by Corollary 4.4.5.1, Remark 4.4.2, the hypothesis, and the fact that $\delta = 2^{\frac{-2}{q-2}}\varepsilon^{\frac{q}{q-2}}$ tells us that:

$$\begin{aligned}
\mathsf{D}(\Psi) = \mathsf{D}(\Phi) &\leqslant \max\left\{\frac{1}{2}\log_2(\delta^{-1}) + 1, 2\right\} \\
&= \max\left\{\frac{1}{q-2} + \left[\frac{q}{q-2}\right]\log_2(\varepsilon) + 1, 2\right\}
\end{aligned} \tag{5.1.46}$$

This establishes Item (v).

Notice next that the fact that $\delta = 2^{\frac{-2}{q-2}}\varepsilon^{\frac{q}{q-2}}$ tells us that:

$$\log_2\left(\delta^{-1}\right) = \log_2\left(2^{\frac{2}{q-2}}\varepsilon^{\frac{-q}{q-2}}\right) = \frac{2}{q-2} + \left[\left[\frac{q}{q-2}\right]\log_2\left(\varepsilon^{-1}\right)\right] \tag{5.1.47}$$

Note that by , Corollary 4.4.5.1 we have that:

$$\mathsf{P}\left(\Phi \bullet \mathsf{Aff}_{-\alpha,0}\right) \leqslant \left[\max\left\{1, \frac{\mathsf{I}\left(\mathsf{Aff}_{-\alpha,0}\right) + 1}{\mathsf{I}\left(\Phi\right) + 1}\right\}\right]\mathsf{P}\left(\Phi\right) = \mathsf{P}\left(\Phi\right) \tag{5.1.48}$$

and further that:

$$\begin{aligned}
\mathsf{P}\left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}\right) &= \left[\max\left\{1, \frac{\mathsf{O}\left(\mathsf{Aff}_{-\alpha^2,0}\right)}{\mathsf{O}\left(\Phi \bullet \mathsf{Aff}_{-\alpha,0}\right)}\right\}\right]\mathsf{P}\left(\Phi \bullet \mathsf{Aff}_{-\alpha,0}\right) \\
&\leqslant \mathsf{P}\left(\Phi\right)
\end{aligned} \tag{5.1.49}$$

By symmetry note also that $\mathsf{P}\left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}\right) = \mathsf{P}\left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}\right)$ and also that $\mathsf{L}\left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}\right) = \mathsf{L}\left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}\right)$. Thus Lemma 4.5.10, Corollary 4.3.5.1, and

131

the hypothesis tells us that:

$$P\left(\Psi\right) = P\left(\Phi \boxminus \Phi\right)$$

$$\leqslant 4\,P\left(\Phi\right)$$

$$= 4\max\left\{10\log_2\left(\delta^{-1}\right) - 7, 13\right\} \tag{5.1.50}$$

This, and the fact that $\delta = 2^{\frac{-2}{q-2}}\varepsilon^{\frac{q}{q-2}}$ renders (5.1.50) as:

$$4\max\left\{10\log_2\left(\delta^{-1}\right) - 7, 13\right\} = 4\max\left\{10\log_2\left(\delta^{-1}\right) - 7, 13\right\}$$

$$= 4\max\left\{10\left(\frac{2}{q-2} + \frac{q}{q-2}\log_2\left(\varepsilon^{-1}\right)\right) - 7, 13\right\}$$

$$= \max\left\{\left[\frac{40q}{q-2}\right]\log_2\left(\varepsilon^{-1}\right) + \frac{80}{q-2} - 28, 52\right\} \tag{5.1.51}$$

$\square$

Remark 5.1.8. *We will often find it helpful to refer to this network for fixed $\varepsilon \in (0,\infty)$ and $q \in (2,\infty)$ as the* $\mathsf{Sqr}^{q,\varepsilon}$ *network.*

Remark 5.1.9. *For an* R *implementation see Listing 9.17*

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Experimental $\lvert x^2 - \mathfrak{I}_{\mathfrak{r}}(\mathsf{Sqr}^{q,\varepsilon})(x)\rvert$	0.00000	0.08943	0.33787	3.14893	4.67465	20.00
Theoretical $\lvert x^2 - \mathfrak{I}_{\mathfrak{r}}(\mathsf{Sqr}^{q,\varepsilon})(x)\rvert$	0.010	1.715	10.402	48.063	45.538	1250.00
Forward Difference	0.01	1.6012	9.8655	44.9141	40.7102	1230

Table 5.1: Theoretical upper bounds for L^1 error, experimental L^1 error and their forward difference, with $q \in [2.1, 4]$, $\varepsilon \in (0.1, 2]$, and $x \in [-5, 5]$, all with 50 mesh-points for $\mathsf{Sqr}^{q,\varepsilon}$.
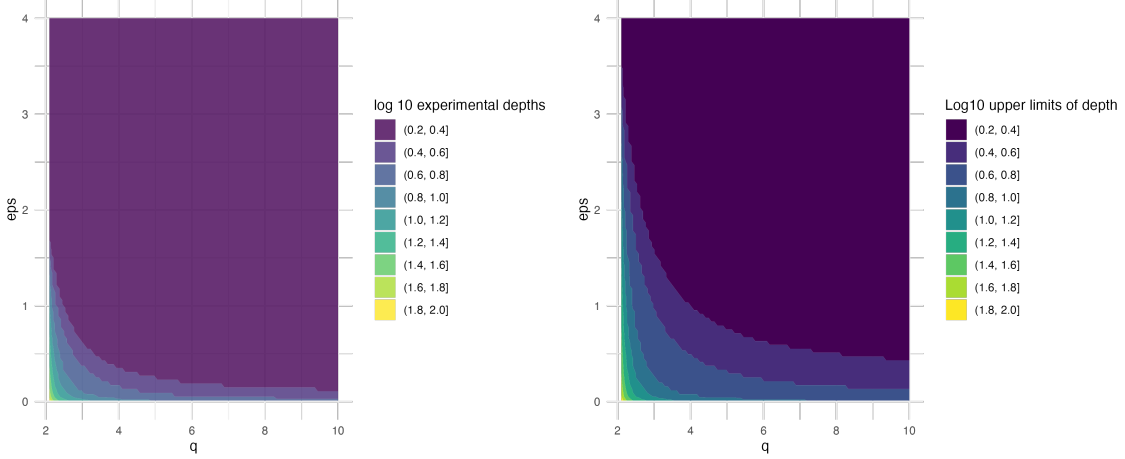
Figure 5.3: Left: $\log_{10}$ of depths for a simulation with $q \in [2.1, 4]$, $\varepsilon \in (0.1, 2]$, and $x \in [-5, 5]$, all with 50 mesh-points for $\mathsf{Sqr}^{q,\varepsilon}$. Right: The theoretical upper limits over the same range of values

5.1.3 The $\mathsf{Prd}^{q,\varepsilon}$ Neural Networks and Products of Two Real Numbers

We are finally ready to give neural network representations of arbitrary products of real numbers. However, this representation differs somewhat from those found in the literature, especially Grohs et al. (2023), where parallelization (stacking) is used instead of neural network sums. This will help us calculate W_1 and the width of the second to last layer for later neural network calculations.

Lemma 5.1.10. *Let* $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$, $A_1, A_2, A_3 \in \mathbb{R}^{1\times 2}$, $\Psi \in \mathsf{NN}$ *satisfy for all* $x \in \mathbb{R}$ *that* $\delta = \varepsilon \left(2^{q-1} + 1 \right)^{-1}$, $A_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $A_3 = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $\mathfrak{I}_{\mathfrak{r}} \in C(\mathbb{R},\mathbb{R})$, $(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0) = 0$, $0 \leqslant (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \leqslant \delta + |x|^2$, $|x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x)| \leqslant \delta \max\{1, |x|^q\}$, $\mathsf{D}(\Psi) \leqslant \max\{1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \log_2(\delta^{-1}), 2\}$, *and* $\mathsf{P}(\Psi) \leqslant \max\left\{ \left\lceil \frac{40q}{q-2} \right\rceil \log_2(\delta^{-1}) + \frac{80}{q-2} - 28, 52 \right\}$, *then:*

(i) there exists a unique $\Gamma \in \mathsf{NN}$ *satisfying:*

$$\Gamma = \left(\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_1,0}) \right) \bigoplus \left(\left(-\frac{1}{2} \right) \triangleright (\Psi \bullet \mathsf{Aff}_{A_2,0}) \right) \bigoplus \left(\left(-\frac{1}{2} \right) \triangleright (\Psi \bullet \mathsf{Aff}_{A_3,0}) \right)$$

$$(5.1.52)$$

<div align="center">133</div>
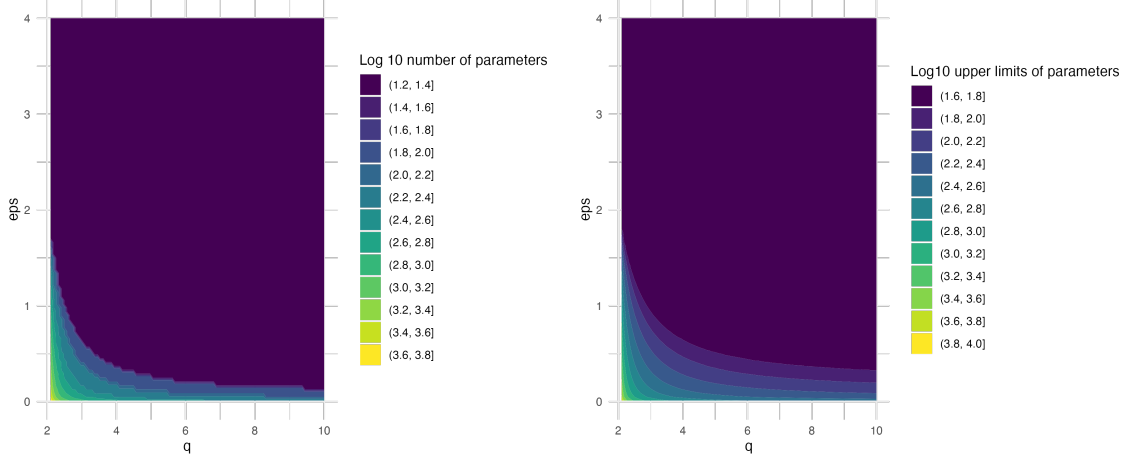
Figure 5.4: Left: $\log_{10}$ of params for a simulation with $q \in [2.1, 4]$, $\varepsilon \in (0.1, 2]$, and $x \in [-5, 5]$, all with 50 mesh-points for $\mathsf{Sqr}^{q,\varepsilon}$. Right: The theoretical upper limits over the same range of values

(ii) it that $\mathfrak{I}_{\mathfrak{r}}(\Gamma) \in C(\mathbb{R}^2, \mathbb{R})$

(iii) it holds for all $x \in \mathbb{R}$ that $(\mathfrak{I}_{\mathfrak{r}}(\Gamma))(x, 0) = (\mathfrak{I}_{\mathfrak{r}}(\Gamma))(0, y) = 0$

(iv) it holds for any $x, y \in \mathbb{R}$ that $\left| xy - (\mathfrak{I}_{\mathfrak{r}}(\Gamma)) \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) \right| \leqslant \varepsilon \max\{1, |x|^q, |y|^q\}$

(v) it holds that $\mathsf{P}(\Gamma) \leqslant \frac{360q}{q-2} [\log_2(\varepsilon^{-1}) + q + 1] - 252$

(vi) it holds that $\mathsf{D}(\Gamma) \leqslant \frac{q}{q-2} [\log_2(\varepsilon^{-1}) + q]$

(vii) it holds that $\mathsf{W}_1(\Gamma) = 24$

(viii) it holds that $\mathsf{W}_{\mathsf{H}(\Gamma)} = 24$

134

	Min	1st. Qu	Median	Mean	3rd Qu	Max.
Experimental $\left\|x^2 - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sqr}^{q,\varepsilon}\right)(x)\right\|$	0.0000	0.0894	0.3378	3.1489	4.6746	20.0000
Theoretical upper limits for $\left\|x^2 - \mathfrak{I}_{\mathfrak{r}}(\mathsf{Sqr}^{q,\varepsilon})(x)\right\|$	0.010	1.715	10.402	48.063	45.538	1250.000
Forward Difference	0.001	1.6012	9.8655	44.9141	40.7102	1230
Experimental depths	2	2	2	2.307	2	80
Theoretical upper bound on depths	2	2	2	2.73	2	91
Forward Difference	0	0	0	0.423	0	11
Experimental params	25	25	25	47.07	25	5641
Theoretical upper limit on params	52	52	52	82.22	52	6353
Forward Differnce	27	27	27	35.16	27	712

Table 5.2: Table showing the experimental and theoretical 1-norm difference, depths, and parameter counts respectively for $\mathsf{Sqr}^{q,\varepsilon}$ with $q \in [2.1, 4]$, $\varepsilon \in [0.01, 2]$, and $x \in [-5, 5]$ all with 50 mesh-points, and their forward differences.

Proof. Note that:

$$
(\mathfrak{I}_{\mathfrak{r}}(\Gamma))\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \mathfrak{I}_{\mathfrak{r}}\left(\left(\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_1,0})\right) \bigoplus \left(\left(-\frac{1}{2}\right) \triangleright (\Psi \bullet \mathsf{Aff}_{A_2,0})\right) \bigoplus \quad (5.1.53)
$$

$$
\left(\left(-\frac{1}{2}\right) \triangleright (\Psi \bullet \mathsf{Aff}_{A_3,0})\right)\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)
$$

$$
= \mathfrak{I}_{\mathfrak{r}}\left(\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_1,0})\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) + \mathfrak{I}_{\mathfrak{r}}\left(\left(-\frac{1}{2}\right) \triangleright (\Psi \bullet \mathsf{Aff}_{A_2,0})\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)
$$

$$
+ \mathfrak{I}_{\mathfrak{r}}\left(\left(-\frac{1}{2}\right) \triangleright (\Psi \bullet \mathsf{Aff}_{A_3,0})\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)
$$

$$
= \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))\left(\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right) - \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))\left(\begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right)
$$

$$
- \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))\left(\begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right)
$$

$$
= \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x+y) - \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) - \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(y) \quad (5.1.54)
$$

Note that this, and the assumption that $(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \in C(\mathbb{R}, \mathbb{R})$ and that $(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0) = 0$ ensures:

$$(\mathfrak{I}_{\mathfrak{r}}(\Gamma))\left(\begin{bmatrix} x \\ 0 \end{bmatrix}\right) = \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x+0) - \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) - \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0)$$

$$= 0$$

$$= \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0+y) - \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0) - \frac{1}{2}(\mathfrak{I}_{\mathfrak{r}}(\Psi))(y)$$

$$= (\mathfrak{I}_{\mathfrak{r}}(\Gamma))\left(\begin{bmatrix} 0 \\ y \end{bmatrix}\right) \tag{5.1.55}$$

Next, observe that since by assumption it is the case for all $x, y \in \mathbb{R}$ that $|x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x)| \leqslant \delta \max\{1, |x|^q\}$, $xy = \frac{1}{2}|x+y|^2 - \frac{1}{2}|x|^2 - \frac{1}{2}|y|^2$, triangle Inequality and from (5.1.54) we have that:

$$|(\mathfrak{I}_{\mathfrak{r}}(\Gamma)(x,y)) - xy|$$

$$= \left| \frac{1}{2}\left[(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x+y) - |x+y|^2\right] - \frac{1}{2}\left[(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) - |x|^2\right] - \frac{1}{2}\left[(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) - |y|^2\right]\right|$$

$$\leqslant \left| \frac{1}{2}\left[(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x+y) - |x+y|^2\right] + \frac{1}{2}\left[(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) - |x|^2\right] + \frac{1}{2}\left[(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) - |y|^2\right]\right|$$

$$\leqslant \frac{\delta}{2}\left[\max\{1, |x+y|^q\} + \max\{1, |x|^q\} + \max\{1, |y|^q\}\right]$$

Note also that since for all $\alpha, \beta \in \mathbb{R}$ and $p \in [1, \infty)$ we have that $|\alpha + \beta|^p \leqslant 2^{p-1}(|\alpha|^p + |\beta|^p)$ we have that:

$$|(\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) - xy|$$

$$\leqslant \frac{\delta}{2}\left[\max\{1, 2^{q-1}|x|^q + 2^{q-1}|y|^q\} + \max\{1, |x|^q\} + \max\{1, |y|^q\}\right]$$

$$\leqslant \frac{\delta}{2}\left[\max\{1, 2^{q-1}|x|^q\} + 2^{q-1}|y|^q + \max\{1, |x|^q\} + \max\{1, |y|^q\}\right]$$

$$\leqslant \frac{\delta}{2}\left[2^q + 2\right]\max\{1, |x|^q, |y|^q\} = \varepsilon \max\{1, |x|^q, |x|^q\}$$

This proves Item (iv).

By symmetry it holds that $\mathsf{P}\left(\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_1,0})\right)$
$= \mathsf{P}\left(-\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_2,0})\right) = \mathsf{P}\left(-\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_3,0})\right)$ and further that $\mathsf{L}\left(\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_1,0})\right) =$
$\mathsf{L}\left(-\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_2,0})\right) = \mathsf{L}\left(-\frac{1}{2} \triangleright (\Psi \bullet \mathsf{Aff}_{A_3,0})\right)$. Note also that Corollary 4.4.5.1 tells us that
for all $i \in \{1, 2, 3\}$ and $a \in \{\frac{1}{2}, -\frac{1}{2}\}$ it is the case that:

$$\mathsf{P}\left(a \triangleright (\Psi \bullet \mathsf{Aff}_{A_i,0})\right) = \mathsf{P}\left(\Psi\right) \tag{5.1.56}$$

This, together with Corollary 4.5.10.1 indicates that:

$$\mathsf{P}\left(\Gamma\right) \leqslant 9\,\mathsf{P}\left(\Psi\right)$$
$$\leqslant 9 \max\left\{\left[\frac{40q}{q-2}\right]\log_2\left(\delta^{-1}\right) + \frac{80}{q-2} - 28, 52\right\} \tag{5.1.57}$$

Combined with the fact that $\delta = \varepsilon \left(2^{q-1} + 1\right)^{-1}$, this is then rendered as:

$$9 \max\left\{\left[\frac{40q}{q-2}\right]\log_2\left(\delta^{-1}\right) + \frac{80}{q-2} - 28, 52\right\}$$
$$= 9 \max\left\{\left[\frac{40q}{q-2}\right]\left(\log_2\left(\varepsilon^{-1}\right) + \log_2\left(2^{q-1} + 1\right)\right) + \frac{80}{q-2} - 28, 52\right\} \tag{5.1.58}$$

Note that:

$$\log_2\left(2^{q-1} + 1\right) = \log_2\left(2^{q-1} + 1\right) - \log_2\left(2^q\right) + q$$
$$= \log_2\left(\frac{2^{q-1} + 1}{2^q}\right) + q = \log_2\left(2^{-1} + 2^{-q}\right) + q$$
$$\leqslant \log_2\left(2^{-1} + 2^{-2}\right) + q = \log_2\left(\frac{3}{4}\right) + q = \log_2\left(3\right) - 2 + q \tag{5.1.59}$$

Combine this with the fact that for all $q \in (2, \infty)$ it is the case that $\frac{q(q-1)}{q-2} \geqslant 2$ then gives us

that:

$$\left\lceil \frac{40q}{q-2} \right\rceil \log_2\left(2^{q-1}+1\right) - 28 \geqslant \left\lceil \frac{40q}{q-2} \right\rceil \log_2\left(2^{q-1}\right) - 28 = \frac{40q(q-1)}{q-2} - 28 \geqslant 52 \quad (5.1.60)$$

This then finally renders (5.1.58) as:

$$
\begin{aligned}
& 9\max\left\{ \left\lceil \frac{40q}{q-2} \right\rceil \left(\log_2\left(\varepsilon^{-1}\right) + \log_2\left(2^{q-1}+1\right)\right) + \frac{80}{q-2} - 28, 52 \right\} \\
& \leqslant 9 \left[\left\lceil \frac{40q}{q-2} \right\rceil \left(\log_2\left(\varepsilon^{-1}\right) + \log_2(3) - 2 + q\right) + \frac{80}{q-2} - 28 \right] \\
& = 9 \left[\left\lceil \frac{40q}{q-2} \right\rceil \left(\log_2\left(\varepsilon^{-1}\right) + \log_2(3) - 2 + \frac{2}{q}\right) - 28 \right] \\
& \leqslant 9 \left[\left\lceil \frac{40q}{q-2} \right\rceil \left(\log_2\left(\varepsilon^{-1}\right) + \log_2(3) - 1\right) - 28 \right] \\
& = \frac{360q}{q-2} \left[\log_2\left(\varepsilon^{-1}\right) + q + \log_2(3) - 1\right] - 252 \quad (5.1.61)
\end{aligned}
$$

Note that Lemma 4.5.11, Lemma 4.4.5, the hypothesis, and the fact that $\delta = \varepsilon \left(2^{q-1}+1\right)^{-1}$ tell us that:

$$
\begin{aligned}
\mathsf{D}\left(\Gamma\right) = \mathsf{D}\left(\Psi\right) & \leqslant \max\left\{ 1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \log_2\left(\delta^{-1}\right), 2 \right\} \\
& = \max\left\{ 1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \left[\log_2\left(\varepsilon^{-1}\right) + \log_2\left(2^{q-1}+1\right)\right], 2 \right\} \\
& = \max\left\{ 1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \left(\log_2\left(\varepsilon^{-1}\right) + q - 1\right), 2 \right\} \quad (5.1.62)
\end{aligned}
$$

Since it is the case that $\frac{q(q-1)}{2(q-2)} > 2$ for $q \in (2, \infty)$ we have that:

$$
\begin{aligned}
& \max\left\{ 1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \left(\log_2\left(\varepsilon^{-1}\right) + q - 1\right), 2 \right\} \\
& = 1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \left(\log_2\left(\varepsilon^{-1}\right) + q - 1\right) \\
& \leqslant \frac{q-1}{q-2} + \frac{q}{2(q-2)} \left(\log_2\left(\varepsilon^{-1}\right) + q\right)
\end{aligned}
$$

$$(5.1.63)$$

138

Observe next that for $q \in (0, \infty)$, $\varepsilon \in (0, \infty)$, Γ consists of, among other things, three stacked $(\Psi \bullet \mathsf{Aff}_{A_i,0})$ networks where $i \in \{1, 2, 3\}$. Corollary 4.4.5.1 tells us therefore, that $\mathsf{W}_1(\Gamma) = 3 \cdot \mathsf{W}_1(\Psi)$. On the other hand, note that each Ψ networks consist of, among other things, two stacked Φ networks, which by Corollary 4.4.5.1 and Lemma 5.1.7, yields that $\mathsf{W}_1(\Gamma) = 6 \cdot \mathsf{W}_1(\Phi)$. Finally from Corollary 5.1.4.1, and Corollary 4.4.5.1, we see that the only thing contributing to the $\mathsf{W}_1(\Phi)$ is $\mathsf{W}_1(\mathfrak{i}_4)$, which was established from Lemma 5.1.2 as 4. Whence we get that $\mathsf{W}_1(\Gamma) = 6 \cdot 4 = 24$, and that $\mathsf{W}_{\mathsf{H}(\Gamma)}(\Gamma) = 24$. This proves Item (vii)—(viii). This then completes the proof of the Lemma. $\qquad \square$

Corollary 5.1.10.1. *Let* $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$, $A_1, A_2, A_3 \in \mathbb{R}^{1 \times 2}$, $\Psi \in \mathbb{N}$ *satisfy for all* $x \in \mathbb{R}$ *that* $\delta = \varepsilon(2^{q-1} + 1)^{-1}$, $A_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $A_3 = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $\mathfrak{I}_{\mathfrak{r}} \in C(\mathbb{R}, \mathbb{R})$, $(\mathfrak{I}_{\mathfrak{r}}(\Psi))(0) = 0$, $0 \leqslant (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x) \leqslant \delta + |x|^2$, $|x^2 - (\mathfrak{I}_{\mathfrak{r}}(\Psi))(x)| \leqslant \delta \max\{1, |x|^q\}$, $\mathsf{D}(\Psi) \leqslant \max\{1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \log_2(\delta^{-1}), 2\}$, *and* $\mathsf{P}(\Psi) \leqslant \max\{\left\lceil \frac{40q}{q-2} \right\rceil \log_2(\delta^{-1}) + \frac{80}{q-2} - 28, 52\}$, *and finally let* Γ *be defined as in Lemma 5.1.10, i.e.:*

$$\Gamma = \left(\frac{1}{2} \circledast (\Psi \bullet \mathsf{Aff}_{A_1,0})\right) \bigoplus \left(\left(-\frac{1}{2}\right) \circledast (\Psi \bullet \mathsf{Aff}_{A_2,0})\right) \bigoplus \left(\left(-\frac{1}{2}\right) \circledast (\Psi \bullet \mathsf{Aff}_{A_3,0})\right)$$

(5.1.64)

It is then the case for all $x, y \in \mathbb{R}$ *that:*

$$\mathfrak{I}_{\mathfrak{r}}(\Gamma)(x, y) \leqslant \frac{3}{2}\left(\frac{\varepsilon}{3} + x^2 + y^2\right) \leqslant \varepsilon + 2x^2 + 2y^2 \tag{5.1.65}$$

Proof. Note that the triangle inequality, the fact that $\delta = \varepsilon(2^{q-1} + 1)^{-1}$, the fact that for

all $x, y \in \mathbb{R}$ it is the case that $|x + y|^2 \leqslant 2 \left(|x|^2 + |y|^2 \right)$ and (5.1.54) tell us that:

$$
\begin{aligned}
|\mathfrak{I}_{\mathfrak{r}} \left(\Gamma \right) \left(x, y \right)| &\leqslant \frac{1}{2} |\mathfrak{I}_{\mathfrak{r}} \left(\Psi \right) \left(x + y \right)| + \frac{1}{2} |\mathfrak{I}_{\mathfrak{r}} \left(\Psi \right) \left(x \right)| + \frac{1}{2} |\mathfrak{I}_{\mathfrak{r}} \left(\Psi \right) \left(y \right)| \\
&\leqslant \frac{1}{2} \left(\delta + |x + y|^2 \right) + \frac{1}{2} \left(\delta + |x|^2 \right) + \frac{1}{2} \left(\delta + |y|^2 \right) \\
&\leqslant \frac{3\delta}{2} + \frac{3}{2} \left(|x|^2 + |y|^2 \right) = \left(\frac{3\varepsilon}{2} \right) \left(2^{q-1} + 1 \right)^{-1} + \frac{3}{2} \left(|x|^2 + |y|^2 \right) \\
&= \frac{3}{2} \left(\frac{\varepsilon}{2^{q-1} + 1} + |x|^2 + |y|^2 \right) \leqslant \frac{3}{2} \left(\frac{\varepsilon}{3} + |x|^2 + |y|^2 \right) \\
&\leqslant \varepsilon + 2x^2 + 2y^2 \qquad\qquad\qquad\qquad\qquad (5.1.66)
\end{aligned}
$$

$\square$

Remark 5.1.11. *We shall refer to this neural network for a given $q \in (2, \infty)$ and given $\varepsilon \in (0, \infty)$ from now on as $\mathsf{Prd}^{q,\varepsilon}$.*

Remark 5.1.12. *For an R implementation see Listing 9.20*

Remark 5.1.13. *Diagrammatically, this can be represented as:*
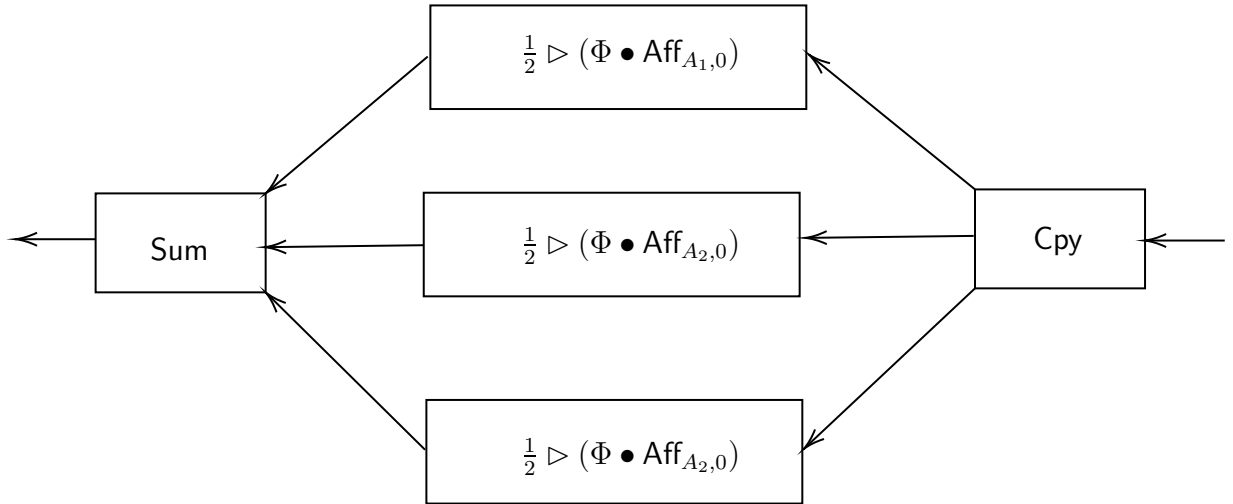
Figure 5.5: Neural network diagram of the $\mathsf{Prd}^{q,\varepsilon}$ network.
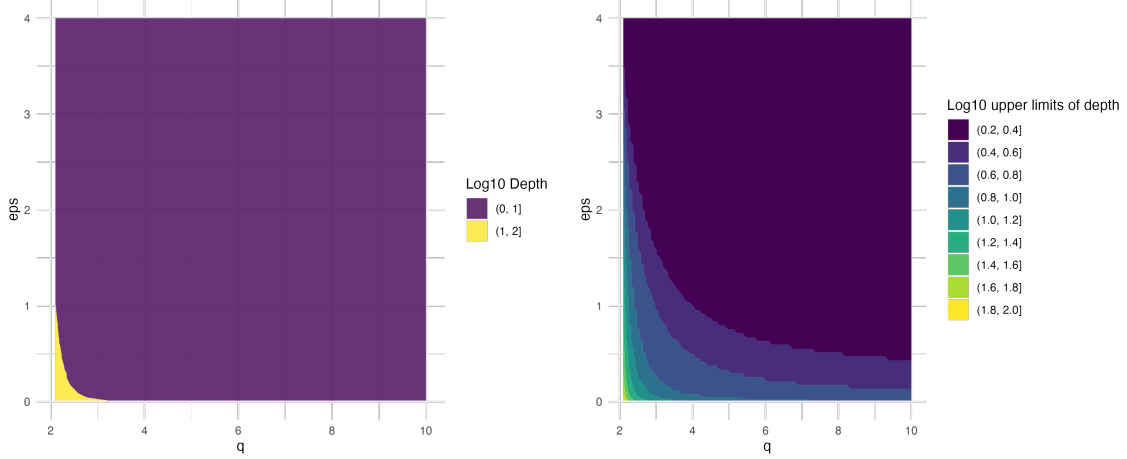
Figure 5.6: Left: $\log_{10}$ of deps for a simulation of $\mathsf{Prd}^{q,\varepsilon}$ with $q \in [2.1, 4]$, $\varepsilon \in (0.1, 2]$, and $x \in [-5, 5]$, all with 50 mesh-points. Right: The theoretical upper limits over the same range of values.

5.2 Higher Approximations

We take inspiration from the Sum neural network to create the Prd neural network. However, we first need to define a special neural network called *tunneling neural network* to stack two neural networks not of the same length effectively.

5.2.1 The Tun_n^d Neural Networks and Their Properties

Definition 5.2.1 (R—,2023, The Tunneling Neural Networks). *We define the tunneling neural network, denoted as Tun_n for $n \in \mathbb{N}$ by:*

$$\mathsf{Tun}_n = \begin{cases} \mathsf{Aff}_{1,0} & : n = 1 \\ \mathsf{Id}_1 & : n = 2 \\ \bullet^{n-2}\,\mathsf{Id}_1 & n \in \mathbb{N} \cap [3, \infty) \end{cases} \tag{5.2.1}$$

Where Id_1 is as in Definition 6.1.1.

Remark 5.2.2. *For an* R *implementation see Listing 9.12*
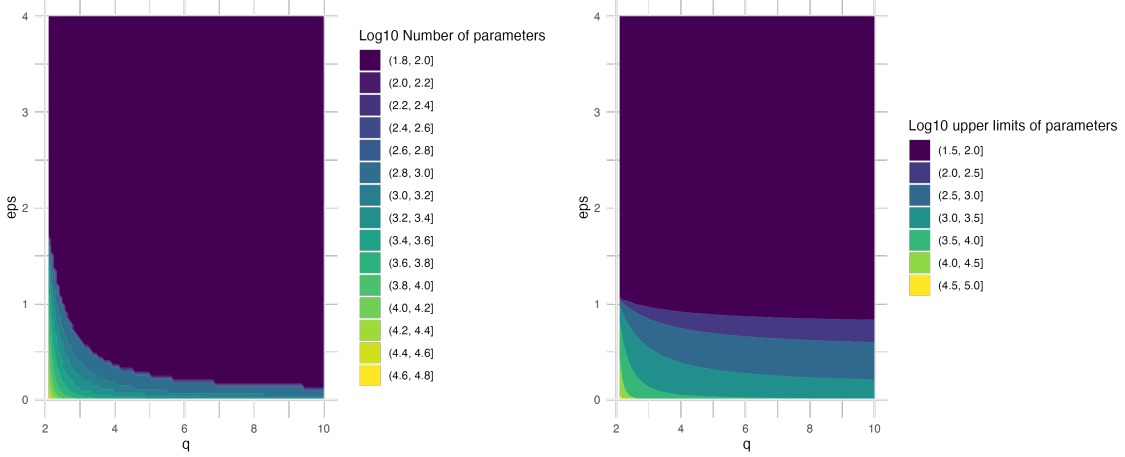
Figure 5.7: Left: $\log_{10}$ of params for a simulation of $\mathsf{Prd}^{q,\varepsilon}$ with $q \in [2.1, 4]$, $\varepsilon \in (0.1, 2]$, and $x \in [-5, 5]$, all with 50 mesh-points. Right: The theoretical upper limits over the same range of values.

Lemma 5.2.3. *Let $n \in \mathbb{N}$, $x \in \mathbb{R}$ and $\mathsf{Tun}_n \in \mathsf{NN}$. For all $n \in \mathbb{N}$ and $x \in \mathbb{R}$, it is then the case that:*

(i) $\mathfrak{I}_{\mathfrak{r}}(\mathsf{Tun}_n) \in C(\mathbb{R}, \mathbb{R})$

(ii) $\mathsf{D}(\mathsf{Tun}_n) = n$

(iii) $(\mathfrak{I}_{\mathfrak{r}}(\mathsf{Tun}_n))(x) = x$

(iv) $\mathsf{P}(\mathsf{Tun}_n) = \begin{cases} 2 & : n = 1 \\ 7 + 6(n-2) & : n \in \mathbb{N} \cap [2, \infty) \end{cases}$

(v) $\mathsf{L}(\mathsf{Tun}_n) = (l_0, l_1, ..., l_{L-1}, l_L) = (1, 2, ..., 2, 1)$

Proof. Note that $\mathsf{Aff}_{0,1} \in C(\mathbb{R}, \mathbb{R})$ and by Lemma 6.1.2 we have that $\mathsf{Id}_1 \in C(\mathbb{R}, \mathbb{R})$. Finally, the composition of continuous functions is continuous, hence $\mathsf{Tun}_n \in C(\mathbb{R}, \mathbb{R})$ for $n \in \mathbb{N} \cap [2, \infty)$. This proves Item (i).

Note that by Lemma 4.4.2 it is the case that $\mathsf{D}(\mathsf{Aff}_{1,0}) = 1$ and by Lemma 6.1.1 it is the case that $\mathsf{D}(\mathsf{Id}_1) = 2$. Assume now that for all $n \leqslant N$ that $\mathsf{D}(\mathsf{Tun}_n) = n$, then for the inductive
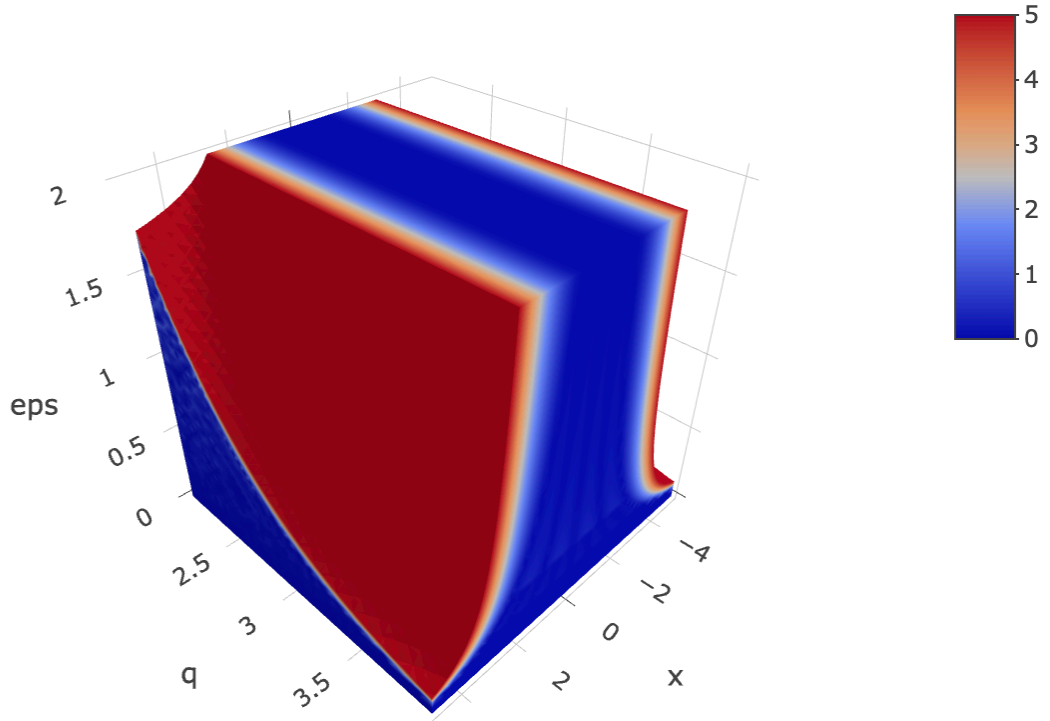
Figure 5.8: Isosurface plot showing $|x^2 - \mathsf{Sqr}^{q,\varepsilon}|$ for $q \in [2.1, 4]$, $\varepsilon \in [0.01, 2]$, and $x \in [-5, 5]$ with 50 mesh-points in each.

step, by Lemma 4.2.5 we have that:

$$\mathsf{D}\left(\mathsf{Tun}_{n+1}\right) = \mathsf{D}\left(\bullet^{n-1}\,\mathsf{Id}_1\right)$$

$$= \mathsf{D}\left(\left(\bullet^{n-2}\,\mathsf{Id}_1\right) \bullet \mathsf{Id}_1\right)$$

$$= n + 2 - 1 = n + 1 \tag{5.2.2}$$

This completes the induction and proves Item (i)—(iii). Note next that by (4.1.11) we have that:

$$\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Aff}_{1,0}\right)\right)(x) = x \tag{5.2.3}$$

143

Lemma 6.1.2, Item (iii) also tells us that:

$$\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_1\right)\right)(x) = \mathfrak{r}(x) - \mathfrak{r}(-x) = x \tag{5.2.4}$$

Assume now that for all $n \leqslant N$ that $\mathsf{Tun}_n(x) = x$. For the inductive step, by Lemma 6.1.2, Item (iii), and we then have that:

$$
\begin{aligned}
\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Tun}_{n+1}\right)\right)(x) &= \left(\mathfrak{I}_{\mathfrak{r}}\left(\bullet^{n-1}\,\mathsf{Id}_1\right)\right)(x)\,(x) \\
&= \left(\mathfrak{I}_{\mathfrak{r}}\left(\left(\bullet^{n-2}\,\mathsf{Id}_1\right)\bullet\mathsf{Id}_1\right)\right) \\
&= \left(\left(\mathfrak{I}_{\mathfrak{r}}\left(\bullet^{n-2}\,\mathsf{Id}_1\right)\right)\circ\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_1\right)\right)\right)(x) \\
&= \left(\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Tun}_n\right)\right)\circ\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_1\right)\right)\right)(x) \\
&= x \tag{5.2.5}
\end{aligned}
$$

This proves Item (ii). Next note that $\mathsf{P}\left(\mathsf{Tun}_1\right) = \mathsf{P}\left(\mathsf{Aff}_{1,0}\right) = 2$. Note also that:

$$\mathsf{P}\left(\mathsf{Tun}_2\right) = \mathsf{P}\left(\mathsf{Id}_1\right) = \mathsf{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1&-1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\right]$$
$$= 7$$

And that by definition of composition:

$$\mathsf{P}\left(\mathsf{Tun}_3\right) \tag{5.2.6}$$
$$= \mathsf{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1&-1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\bullet\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1&-1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\right]$$
$$= \mathsf{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1&-1\\-1&1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1&-1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\right]$$
$$= 13$$

Now for the inductive step assume that for all $n \leqslant N \in \mathbb{N}$, it is the case that $\mathsf{P}\left(\mathsf{Tun}_n\right) = 7 + 6(n-2)$. For the inductive step, we then have:

$$\mathsf{P}\left(\mathsf{Tun}_{n+1}\right) = \mathsf{P}\left(\mathsf{Tun}_n \bullet \mathsf{Id}_1\right) =$$

$$\mathsf{P}\left[\left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right), \cdots, \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix}\right)\right) \bullet \mathsf{Id}_1\right] =$$

$$\mathsf{P}\left[\left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right), \cdots, \left(\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix}\right)\right)\right]$$

$$= 7 + 6(n-2) + 6 = 7 + 6\left((n+1) - 2\right) \tag{5.2.7}$$

This proves Item (iv).

Note finally that Item (v) is a consequence of Lemma 6.1.2, Item (i), and Lemma 4.2.5 $\quad\square$

Definition 5.2.4 (R—, 2023, The Multi-dimensional Tunneling Network). *We define the multi-dimensional tunneling neural network, denoted as* Tun_n^d *for* $n \in \mathbb{N}$ *and* $d \in \mathbb{N}$ *by:*

$$\mathsf{Tun}_n^d = \begin{cases} \mathsf{Aff}_{\mathbb{I}_d, \mathbb{0}_d} & : n = 1 \\[2ex] \mathsf{Id}_d & : n = 2 \\[2ex] \bullet^{n-2} \, \mathsf{Id}_d & : n \in \mathbb{N} \cap [3, \infty) \end{cases} \tag{5.2.8}$$

Where Id_d *is as in Definition 6.1.1.*

Remark 5.2.5. *We may drop the requirement for a d and write* Tun_n *where* $d = 1$*, and it is evident from the context.*

Lemma 5.2.6. *Let* $n \in \mathbb{N}$*,* $d \in \mathbb{N}$*,* $x \in \mathbb{R}$ *and* $\mathsf{Tun}_n^d \in \mathsf{NN}$*. For all* $n \in \mathbb{N}$*,* $d \in \mathbb{N}$*, and* $x \in \mathbb{R}$*, it is then the case that:*

(i) $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Tun}_n^d\right) \in C\left(\mathbb{R}, \mathbb{R}\right)$

(ii) $\mathsf{D}\left(\mathsf{Tun}_n^d\right) = n$

(iii) $\left(\Im_{\mathfrak{r}} \left(\mathsf{Tun}_n^d \right) \right)(x) = x$

(iv) $\mathsf{P}\left(\mathsf{Tun}_n^d \right) = \begin{cases} 8d^2 + 5d & : n = 1 \\ 4d^2 + 3d + (n-1)\left(4d^2 + 2d\right) & : n \in \mathbb{N} \cap [2, \infty) \end{cases}$

(v) $\mathsf{L}\left(\mathsf{Tun}_n^d \right) = (l_0, l_1, ..., l_{L-1}, l_L) = (d, 2d, ..., 2d, d)$

Proof. Note that Items (i)–(iii) are consequences of Lemma 6.1.2 and Lemma 4.2.5 respectively. Note now that by observation $\mathsf{P}\left(\mathsf{Tun}_1^d \right) = d^2 + d$. Next Lemma 6.1.4 tells us that $\mathsf{P}\left(\mathsf{Tun}_2^d \right) = 4d^2 + 3d$ Note also that by definition of neural network composition, we have the

following:

$$\mathsf{P}\left(\mathsf{Tun}_3^d\right) \tag{5.2.9}$$

$$= \mathsf{P}\left[\left(\left(\begin{bmatrix} 1 & & & \\ -1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & -1 & & \\ & \ddots & & \\ & & 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\right)\right] \bullet \tag{5.2.10}$$

$$\left(\left(\begin{bmatrix} 1 & & & \\ -1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & -1 & & \\ & \ddots & & \\ & & 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\right)\right] =$$

$$\mathsf{P}\left[\left(\left(\begin{bmatrix} 1 & & & \\ -1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & -1 & & \\ -1 & 1 & & \\ & \ddots & & \\ & & 1 & -1 \\ & & -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right), \tag{5.2.11}$$

$$\left(\begin{bmatrix} 1 & -1 & & \\ & \ddots & & \\ & & 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\right)\right]$$

$$= 2d \times d + 2d + 2d \times 2d + 2d + 2d \times d + d$$

$$= 2d^2 + 2d + 4d^2 + 2d + 2d^2 + d$$

$$= 8d^2 + 5d \tag{5.2.12}$$

Suppose now that for all naturals up to and including n, it is the case that $\mathsf{P}\left(\mathsf{Tun}_n^d\right) =$

$4d^2 + 3d + (n-2)\left(4d^2 + 2d\right)$. For the inductive step, we have the following:

$$\mathsf{P}\left(\mathsf{Tun}_{n+1}^d\right) = \mathsf{P}\left(\mathsf{Tun}_n^d \bullet \mathsf{Id}_d\right)$$

$$= \mathsf{P}\left[\left(\begin{bmatrix} 1 & & & \\ -1 & & & \\ & \ddots & & \\ & & 1 & \\ & & -1 & \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right)\left(\begin{bmatrix} 1 & -1 & & \\ -1 & 1 & & \\ & & \ddots & \\ & & 1 & -1 \\ & & -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right), \ldots, \right.$$

$$\left.\left(\begin{bmatrix} 1 & -1 & & \\ & \ddots & & \\ & & 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\right) \bullet \mathsf{Id}_d\right] \qquad (5.2.13)$$

$$= \mathsf{P}\left[\left(\begin{bmatrix} 1 & & & \\ -1 & & & \\ & \ddots & & \\ & & 1 & \\ & & -1 & \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right)\left(\begin{bmatrix} 1 & -1 & & \\ -1 & 1 & & \\ & & \ddots & \\ & & 1 & -1 \\ & & -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right), \ldots, \right.$$

$$\left.\left(\begin{bmatrix} 1 & -1 & & \\ -1 & 1 & & \\ & \ddots & & \\ & & 1 & -1 \\ & & -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & -1 & \\ & \ddots & \\ & & 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\right)\right] \qquad (5.2.14)$$

$$= 4d^2 + 3d + (n-2)\left(4d^2 + 2d\right) + 4d^2 + 2d$$

$$= 4d^2 + 3d + (n-1)\left(4d^2 + 2d\right)$$

This proves Item (iv). Finally, Item (v) is a consequence of Lemma 4.4.2 $\qquad \square$
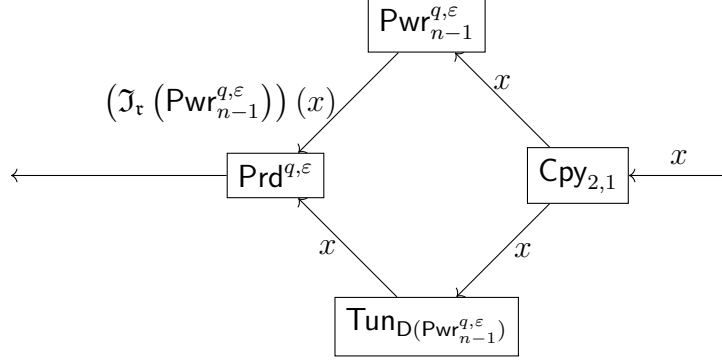
Figure 5.9: A representation of a typical $\mathsf{Pwr}_n^{q,\varepsilon}$ network.

5.2.2 The $\mathsf{Pwr}_n^{q,\varepsilon}$ Neural Networks and Their Properties

Definition 5.2.7 (R—, 2023, The Power Neural Network). *Let $n \in \mathbb{N}$. Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$, satisfy that $\delta = \varepsilon \left(2^{q-1} + 1\right)^{-1}$. We define the power neural networks $\mathsf{Pwr}_n^{q,\varepsilon} \in \mathsf{NN}$, denoted for $n \in \mathbb{N}_0$ as:*

$$\mathsf{Pwr}_n^{q,\varepsilon} = \begin{cases} \mathsf{Aff}_{0,1} & : n = 0 \\ \mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathsf{D}(\mathsf{Pwr}_{n-1}^{q,\varepsilon})} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right] \bullet \mathsf{Cpy}_{2,1} & : n \in \mathbb{N} \end{cases}$$

Diagrammatically, this can be represented as:

Remark 5.2.8. *For an* R *implementation see Listing 9.20*

Remark 5.2.9. *Note that for all $i \in \mathbb{N}$, $q \in (2, \infty)$, $\varepsilon \in (0, \infty)$, each $\mathsf{Pwr}_i^{q,\varepsilon}$ differs from $\mathsf{Pwr}_{i+1}^{q,\varepsilon}$ by atleast one $\mathsf{Prd}^{q,\varepsilon}$ network.*

Lemma 5.2.10. *Let $x, y \in \mathbb{R}$, $\varepsilon \in (0, \infty)$ and $q \in (2, \infty)$. It is then the case for all $x, y \in \mathbb{R}$ that:*

$$\varepsilon \max \left\{1, |x|^q, |y|^q\right\} \leqslant \varepsilon + \varepsilon |x|^q + \varepsilon |y|^q. \tag{5.2.15}$$

Proof. We will do this in the following cases:

For the case that $|x| \leqslant 1$ and $|y| \leqslant 1$ we then have:

$$\varepsilon \max \left\{1, |x|^q, |y|^q\right\} = \varepsilon \leqslant \varepsilon + \varepsilon |x|^q + \varepsilon |y|^q \tag{5.2.16}$$

For the case that $|x| \leqslant 1$ and $|y| \geqslant 1$, without loss of generality we have then:

$$\varepsilon \max \left\{1, |x|^q, |y|^q\right\} \leqslant \varepsilon |y|^q \leqslant \varepsilon + \varepsilon |x|^q + \varepsilon |y|^q : \tag{5.2.17}$$

For the case that $|x| \geqslant 1$ and $|y| \geqslant 1$, and without loss of generality that $|x| \geqslant |y|$ we have that:

$$\varepsilon \max\{1, |x|^q, |y|^q\} = \varepsilon |x|^q \leqslant \varepsilon + \varepsilon |x|^q + \varepsilon |y|^q \tag{5.2.18}$$

$\square$

Lemma 5.2.11. *Let $\mathfrak{p}_i$ for $i \in \{1, 2, ...\}$ be the set of functions defined for $\varepsilon \in (0, \infty)$, and $x \in \mathbb{R}$ as follows:*

$$\mathfrak{p}_1 = \varepsilon + 2 + 2|x|^2$$
$$\mathfrak{p}_i = \varepsilon + 2 \left(\mathfrak{p}_{i-1}\right)^2 + 2|x|^2 \text{ for } i \geqslant 2 \tag{5.2.19}$$

For all $n \in \mathbb{N}$ and $\varepsilon \in (0, \infty)$ and $q \in (2, \infty)$ it holds for all $x \in \mathbb{R}$ that:

$$\left|\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right| \leqslant \mathfrak{p}_n \tag{5.2.20}$$

Proof. Note that by Corollary 5.1.10.1 it is the case that:

$$\left|\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_1^{q,\varepsilon}\right)(x)\right| = \left|\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Prd}^{q,\varepsilon}\right)(1, x)\right| \leqslant \mathfrak{p}_1 \tag{5.2.21}$$

and applying (5.2.21) twice, it is the case that:

$$|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_2^{q,\varepsilon}\right)(x)| = |\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right),x\right)|$$

$$\leqslant \varepsilon + 2\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\right)(1,x)\right|^2 + 2|x|^2$$

$$\leqslant \varepsilon + 2\mathfrak{p}_1^2 + 2|x|^2 = \mathfrak{p}_2 \tag{5.2.22}$$

Let's assume this holds for all cases up to and including n. For the inductive step, Corollary 5.1.10.1 tells us that:

$$\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n+1}^{q,\varepsilon}\right)(x)\right| \leqslant |\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(\mathfrak{I}_{\mathfrak{r}}\cdots\left(1,x\right),x\right),x\right)\cdots\right)\right)|$$

$$\leqslant \mathfrak{I}_{\mathfrak{r}}\left[\mathsf{Prd}^{q,\varepsilon}\left(\mathsf{Pwr}_n^{q,\varepsilon}\left(x\right),x\right)\right]$$

$$\leqslant \varepsilon + 2\mathfrak{p}_n^2 + 2|x|^2 = \mathfrak{p}_{n+1} \tag{5.2.23}$$

This completes the proof of the lemma. $\square$

Remark 5.2.12. *Note that since any instance of $\mathfrak{p}_i$ contains an instance of $\mathfrak{p}_{i-1}$ for $i \in \mathbb{N} \cap [2,\infty)$, we have that $\mathfrak{p}_n \in \mathcal{O}\left(\varepsilon^{2(n-1)}\right)$*

Lemma 5.2.13. *For all $n \in \mathbb{N}$, $q \in (2,\infty)$, and $\varepsilon \in (0,\infty)$, it is the case that $\mathsf{P}\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)}\right) \leqslant \mathsf{P}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)$.*

Proof. Note that for all $n \in \mathbb{N}$ it is straightforwardly the case that $\mathsf{P}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) \geqslant \mathsf{P}\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right)$ because for all $n \in \mathbb{N}$, a $\mathsf{Pwr}_n^{q,\varepsilon}$ network contains a $\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}$ network. Note now that for all $i \in \mathbb{N}$ we have from Lemma 5.2.3 that $5 \leqslant \mathsf{P}\left(\mathsf{Tun}_{i+1}\right) - \mathsf{P}\left(\mathsf{Tun}_i\right) \leqslant 6$. Recall from Corollary 5.1.4.1 that every instance of the Φ network contains atleast one $\mathfrak{i}_4$ network, which by Lemma 5.1.2 has 40 parameters, whence the $\mathsf{Prd}^{q,\varepsilon}$ network has atleast 40 parameters for all $\varepsilon \in (0,\infty)$ and $q \in (2,\infty)$. Note now that for all $i \in \mathbb{N}$, $\mathsf{Pwr}_i^{q,\varepsilon}$ and $\mathsf{Pwr}_{i+1}^{q,\varepsilon}$ differ by atleast as many parameters as there are in $\mathsf{Prd}^{q,\varepsilon}$, since, indeed, they differ by atleast one more $\mathsf{Prd}^{q,\varepsilon}$. Thus for every increment in i, $\mathsf{Pwr}_i^{q,\varepsilon}$ outstrips Tun_i by at-least $40 - 6 = 34$

parameters. This is true for all $i \in \mathbb{N}$. Whence it is the case that for all $i \in \mathbb{N}$, it is the case that $\mathsf{P}\left(\mathsf{Tun}_i\right) \leqslant \mathsf{P}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)$. $\qquad\square$

Lemma 5.2.14 (R——,2023). *Let* $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$, *and* $\delta = \varepsilon \left(2^{q-1} + 1\right)^{-1}$. *Let* $n \in \mathbb{N}_0$, *and* $\mathsf{Pwr}_n \in \mathsf{NN}$. *It is then the case for all* $n \in \mathbb{N}_0$, *and* $x \in \mathbb{R}$ *that:*

(i) $\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)\right)(x) \in C\left(\mathbb{R}, \mathbb{R}\right)$

(ii) $\mathsf{D}(\mathsf{Pwr}_n^{q,\varepsilon}) \leqslant \begin{cases} 1 & : n = 0 \\[2mm] n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] - 1\right] + 1 & : n \in \mathbb{N} \end{cases}$

(iii) $\mathsf{W}_1\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) = \begin{cases} 1 & : n = 0 \\[2mm] 24 + 2\left(n - 1\right) & : n \in \mathbb{N} \end{cases}$

(iv) $\mathsf{P}(\mathsf{Pwr}_n^{q,\varepsilon}) \leqslant \begin{cases} 2 & : n = 0 \\[2mm] 4^{n+\frac{3}{2}} + \left(\frac{4^{n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right) & : n \in \mathbb{N} \end{cases}$

(v) $\left|x^n - \left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)\right)(x)\right| \leqslant \begin{cases} 0 & : n = 0 \\[2mm] \left|x\left(x^{n-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right)\right| + \varepsilon + |x|^q + \mathfrak{p}_{n-1}^q & : n \in \mathbb{N} \end{cases}$

Where we let $\mathfrak{p}_i$ *for* $i \in \{1, 2, ...\}$ *be the set of functions defined as follows:*

$$\mathfrak{p}_1 = \varepsilon + 2 + 2|x|^2$$

$$\mathfrak{p}_i = \varepsilon + 2\left(\mathfrak{p}_{i-1}\right)^2 + 2|x|^2 \tag{5.2.24}$$

And whence we get that:

$$\left|x^n - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right| \in \mathcal{O}\left(\varepsilon^{2q(n-1)}\right) \text{ for } n \geqslant 2 \tag{5.2.25}$$

$$(vi) \ \mathsf{W}_{\mathsf{H}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) = \begin{cases} 1 & n = 0 \\ \\ 24 & n \in \mathbb{N} \end{cases}$$

Proof. Note that Item (ii) of Lemma 4.4.2 ensures that $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_0\right) = \mathsf{Aff}_{1,0} \in C\left(\mathbb{R},\mathbb{R}\right)$. Note next that by Item (v) of Lemma 4.2.5, with $\Phi_1 \curvearrowright \nu_1, \Phi_2 \curvearrowright \nu_2, a \curvearrowright \mathfrak{r}$, we have that:

$$\left(\mathfrak{I}_{\mathfrak{r}}\left(\nu_1 \bullet \nu_2\right)\right)(x) = \left(\left(\mathfrak{I}_{\mathfrak{r}}\left(\nu_1\right)\right) \circ \left(\mathfrak{I}_{\mathfrak{r}}\left(\nu_2\right)\right)\right)(x) \tag{5.2.26}$$

This, with the fact that the composition of continuous functions is continuous, the fact the stacking of continuous instantiated neural networks is continuous tells us that $\left(\mathfrak{I}_{\mathfrak{r}}\mathsf{Pwr}_n\right) \in C\left(\mathbb{R},\mathbb{R}\right)$ for $n \in \mathbb{N} \cap [2,\infty)$. This establishes Item (i).

Note next that by observation $\mathsf{D}\left(\mathsf{Pwr}_0^{q,\varepsilon}\right) = 1$ and by Item (iv) of Lemma 6.1.2, it is the case that $\mathsf{D}\left(\mathsf{Id}_1\right) = 2$. By Lemmas 4.5.3 and 4.2.3 it is also the case that $\mathsf{D}\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right] \bullet \mathsf{Cpy}\right) = \mathsf{D}\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right]\right)$. Note also that by Lemma we have that $\mathsf{D}\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) = \mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)$. This with Lemma 4.2.5 then yields for $n \in \mathbb{N}$ that:

$$\begin{aligned} \mathsf{D}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) &= \mathsf{D}\left(\mathsf{Prd} \bullet \left[\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right] \bullet \mathsf{Cpy}_{2,1}\right) \\ &= \mathsf{D}\left(\mathsf{Prd} \bullet \left[\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right]\right) \\ &= \mathsf{D}\left(\mathsf{Prd}\right) + \mathsf{D}\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right) - 1 \\ &\leqslant \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] + \mathsf{D}\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right) - 1 \\ &= \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] + \mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) - 1 \tag{5.2.27} \end{aligned}$$

And hence for all $n \in \mathbb{N}$ it is the case that:

$$\mathsf{D}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) - \mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) \leqslant \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] - 1 \tag{5.2.28}$$

This, in turn, indicates that:

$$D\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) \leqslant n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q\right]-1\right]+1$$

$$\leqslant n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q\right]-1\right]+1 \qquad (5.2.29)$$

This proves Item (ii).

Note now that $\mathsf{W}_1\left(\mathsf{Pwr}_0^{q,\varepsilon}\right) = \mathsf{W}_1\left(\mathsf{Aff}_{0,1}\right) = 1$. Further Lemma 4.2.5, Remark 4.4.2, tells us that for all $i,k \in \mathbb{N}$ it is the case that $\mathsf{W}_i\left(\mathsf{Tun}_k\right) \leqslant 2$. Observe that since $\mathsf{Cpy}_{2,1}, \mathsf{Pwr}_0^{q,\varepsilon}$, and $\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_0^{q,\varepsilon}\right)}$ are all affine neural networks, Lemma 4.4.5, Corollary 4.4.5.1, and Lemma 5.1.10 tells us that:

$$\mathsf{W}_1\left(\mathsf{Pwr}_1^{q,\varepsilon}\right) = \mathsf{W}_1\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathsf{D}(\mathsf{Pwr}_0^{q,\varepsilon})} \boxminus \mathsf{Pwr}_0^{q,\varepsilon}\right] \bullet \mathsf{Cpy}_{2,1}\right)$$

$$= \mathsf{W}_1\left(\mathsf{Prd}^{q,\varepsilon}\right) = 24 \qquad (5.2.30)$$

And that:

$$\mathsf{W}_1\left(\mathsf{Pwr}_2^{q,\varepsilon}\right) = \mathsf{W}_1\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathsf{D}(\mathsf{Pwr}_1^{q,\varepsilon})} \boxminus \mathsf{Pwr}_1^{q,\varepsilon}\right] \bullet \mathsf{Cpy}_{2,1}\right)$$

$$= \mathsf{W}_1\left(\left[\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_1^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_1^{q,\varepsilon}\right]\right)$$

$$= 24 + 2 = 26$$

This completes the base case. For the inductive case, assume that for all i up to and including $k \in \mathbb{N}$ it is the case that $\mathsf{W}_1\left(\mathsf{Pwr}_i^{q,\varepsilon}\right) \leqslant \begin{cases} 1 & : i = 0 \\ 24 + 2(i-1) & : i \in \mathbb{N} \end{cases}$. For the case of $k+1$, we

get that:

$$
\begin{aligned}
\mathsf{W}_1\left(\mathsf{Pwr}^{q,\varepsilon}_{k+1}\right) &= \mathsf{W}_1\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathsf{D}(\mathsf{Pwr}^{q,\varepsilon}_k)} \boxminus \mathsf{Pwr}^{q,\varepsilon}_k\right] \bullet \mathsf{Cpy}_{2,1}\right) \\
&= \mathsf{W}_1\left(\left[\mathsf{Tun}_{\mathsf{D}(\mathsf{Pwr}^{q,\varepsilon}_k)} \boxminus \mathsf{Pwr}^{q,\varepsilon}_k\right]\right) \\
&= \mathsf{W}_1\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}^{q,\varepsilon}_k\right)}\right) + \mathsf{W}_1\left(\mathsf{Pwr}^{q,\varepsilon}_k\right) \\
&\leqslant
\begin{cases}
2 & : k = 0 \\[2mm]
24 + 2k & : k \in \mathbb{N}
\end{cases}
\end{aligned}
\tag{5.2.31}
$$

This establishes Item (iii).

For Item (iv), we will prove this in cases.

Case 1: $\mathsf{Pwr}^{q,\varepsilon}_0$:

Note that by Lemma 4.4.2 we have that:

$$
\mathsf{P}\left(\mathsf{Pwr}^{q,\varepsilon}_0\right) = \mathsf{P}\left(\mathsf{Aff}_{0,1}\right) = 2
\tag{5.2.32}
$$

This completes Case 1.

Case 2: $\mathsf{Pwr}^{q,\varepsilon}_n$ **where** $n \in \mathbb{N}$**:**

Note that Lemma 4.3.5, Lemma 5.2.13, Corollary 4.3.5.1, Lemma 4.3.6, and Corollary 4.3.6.1, tells us it is the case that:

$$
\begin{aligned}
\mathsf{P}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1} \boxminus \mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right)}\right) &\leqslant \mathsf{P}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1} \boxminus \mathsf{Pwr}^{q,\varepsilon}_{n-1}\right) \\
&\leqslant 4\,\mathsf{P}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right)
\end{aligned}
\tag{5.2.33}
$$

Then Lemma 4.2.5 and Corollary 4.4.5.1 tells us that:

$$\mathsf{P}\left(\left[\mathsf{Pwr}_{n-1}^{q,\varepsilon}\boxminus\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right]\bullet\mathsf{Cpy}_{2,1}\right)$$

$$=\mathsf{P}\left(\left[\mathsf{Pwr}_{n-1}^{q,\varepsilon}\boxminus\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right]\right)$$

$$\leqslant 4\,\mathsf{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) \tag{5.2.34}$$

Note next that by definition for all $q \in (2,\infty)$, and $\varepsilon \in (0,\infty)$ it is case that $\mathsf{W}_{\mathsf{H}\left(\mathsf{Pwr}_0^{q,\varepsilon}\right)}\mathsf{Pwr}_0^{q,\varepsilon} = \mathsf{W}_{\mathsf{H}(\mathsf{Aff}_{0,1})} = 1$. Now, by Lemma 5.1.10, and by construction of $\mathsf{Pwr}_i^{q,\varepsilon}$ we may say that for $i \in \mathbb{N}$ it is the case that:

$$\mathsf{W}_{\mathsf{H}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)} = \mathsf{W}_{\mathsf{H}(\mathsf{Prd}^{q,\varepsilon})} = 24 \tag{5.2.35}$$

Note also that by Lemma 5.2.3 it is the case that:

$$\mathsf{W}_{\mathsf{H}\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)}\right)}\left(\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)}\right) = 2 \tag{5.2.36}$$

Furthermore, note that for $n \in [2,\infty) \cap \mathbb{N}$ Lemma 5.1.10 tells us that:

$$\mathsf{W}_{\mathsf{H}\left(\left[\mathsf{Pwr}_{n-1}^{q,\varepsilon}\boxminus\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right]\right)}\left(\left[\mathsf{Pwr}_{n-1}^{q,\varepsilon}\boxminus\mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right]\right) = 24 + 2 = 26 \tag{5.2.37}$$

Finally Lemma 4.2.5, (5.2.34), a geometric series argument, and Corollary 4.3.5.1, also tells

us that:

$$P\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) \tag{5.2.38}$$

$$= P\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Pwr}_{n-1}^{q,\varepsilon} \boxminus \mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right] \bullet \mathsf{Cpy}_{2,1}\right)$$

$$= P\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Pwr}_{n-1}^{q,\varepsilon} \boxminus \mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right]\right)$$

$$\leqslant P\left(\mathsf{Prd}^{q,\varepsilon}\right) + 4\,P\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) +$$

$$+ W_1\left(\mathsf{Prd}^{q,\varepsilon}\right) \cdot W_{\mathsf{H}\left(\left[\mathsf{Pwr}_{n-1}^{q,\varepsilon} \boxminus \mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right]\right)} \left(\left[\mathsf{Pwr}_{n-1}^{q,\varepsilon} \boxminus \mathsf{Tun}_{\mathsf{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right]\right)$$

$$= P\left(\mathsf{Prd}^{q,\varepsilon}\right) + 4\,P\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 624$$

$$= 4^{n+1}\,P\left(\mathsf{Pwr}_0^{q,\varepsilon}\right) + \left(\frac{4^{n+1}-1}{3}\right)\left(P\left(\mathsf{Prd}^{q,\varepsilon}\right) + 624\right)$$

$$= 4^{n+\frac{3}{2}} + \left(\frac{4^{n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right) \tag{5.2.39}$$

Next note that $\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Pwr}_{0,1}\right)\right)(x)$ is exactly 1, which implies that for all $x \in \mathbb{R}$ we have that $|x^0 - \left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Pwr}_{0.1}\right)(x)\right)| = 0$. Note also that the instantiations of Tun_n and $\mathsf{Cpy}_{2,1}$ are exact. Note next that since Tun_n and $\mathsf{Cpy}_{2,1}$ are exact, the only sources of error for $\mathsf{Pwr}_n^{q,\varepsilon}$ are n compounding applications of $\mathsf{Prd}^{q,\varepsilon}$.

Note also that by definition, it is the case that:

$$\mathfrak{I}_\mathfrak{r}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) = \mathfrak{I}_\mathfrak{r}\left[\underbrace{\mathsf{Prd}^{q,\varepsilon}\left(\mathfrak{I}_\mathfrak{r}\left[\mathsf{Prd}^{q,\varepsilon}\left(\cdots \mathfrak{I}_\mathfrak{r}\left[\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right],\cdots x\right)\right],x\right)}_{n-copies}\right] \tag{5.2.40}$$

Lemma 5.1.10 tells us that:

$$|x - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right)| \leqslant \varepsilon \max\{1, |x|^q\} \leqslant \varepsilon + |x|^q \tag{5.2.41}$$

The triangle inequality, Lemma 5.2.10, Lemma 5.1.10, and Corollary 5.1.10.1 then tells us

that:

$$\left|x^2 - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_2^{q,\varepsilon}\right)(x)\right|$$

$$= \left|x \cdot x - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right),x\right)\right)\right|$$

$$\leqslant \left|x \cdot x - x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right)\right| + \left|x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right),x\right)\right)\right|$$

$$= \left|x\left(x - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right)\right)\right| + \varepsilon + \varepsilon\left|x\right|^q + \varepsilon\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(1,x\right)\right)\right|^q$$

$$\leqslant \left|x\varepsilon + x\varepsilon\left|x\right|^q\right| + \varepsilon + \varepsilon\left|x\right|^q + \varepsilon\left|\varepsilon + 2 + x^2\right|^q$$

$$= \left|x\varepsilon + x\varepsilon\left|x\right|^q\right| + \varepsilon + \varepsilon\left|x\right|^q + \varepsilon\mathfrak{p}_1^q \tag{5.2.42}$$

Note that this takes care of our base case. Assume now that for all integers up to and including n, it is the case that:

$$\left|x^n - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right| \leqslant \left|x \cdot x^{n-1} - x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right| + \left|x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right|$$

$$\leqslant \left|x\left(x^{n-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right)\right| + \varepsilon + \varepsilon\left|x\right|^q + \varepsilon\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right|^q$$

$$\leqslant \left|x\left(x^{n-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right)\right| + \varepsilon + \varepsilon\left|x\right|^q + \varepsilon\mathfrak{p}_{n-1}^q \tag{5.2.43}$$

For the inductive case, we see that:

$$\left|x^{n+1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n+1}^{q,\varepsilon}\right)(x)\right| \leqslant \left|x^{n+1} - x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right| + \left|x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n+1}^{q,\varepsilon}\right)\right|$$

$$\leqslant \left|x\left(x^n - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right)\right| + \varepsilon + \varepsilon\left|x\right|^q + \varepsilon\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right|^q$$

$$\leqslant \left|x\left(x^n - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right)\right| + \varepsilon + \varepsilon\left|x\right|^q + \varepsilon\mathfrak{p}_n^q \tag{5.2.44}$$

Note that since $\mathfrak{p}_n \in \mathcal{O}\left(\varepsilon^{2(n-1)}\right)$ for $n \in \mathbb{N} \cap [2,\infty)$, it is the case for all $x \in \mathbb{R}$ then that $\left|x^n - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)(x)\right| \in \mathcal{O}\left(\varepsilon^{2q(n-1)}\right)$ for $n \geqslant 2$.

Finally note that $\mathsf{W}_{\mathsf{H}\left(\mathsf{Pwr}_0^{q,\varepsilon}\right)}\left(\mathsf{Pwr}_0^{q,\varepsilon}\right) = 1$ from observation. For $n \in \mathbb{N}$, note that the second to last layer is the second to last layer of the $\mathsf{Prd}^{q,\varepsilon}$ network. Thus Lemma 5.1.10 tells us

that:

$$W_{\mathsf{H}\left(\mathsf{Pwr}_m^{q,\varepsilon}\right)}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) = \begin{cases} 1 & n = 0 \\ \\ 24 & n \in \mathbb{N} \end{cases} \qquad (5.2.45)$$

This completes the proof of the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 5.2.15. *Note each power network $\mathsf{Pwr}_n^{q,\varepsilon}$ is at least as deep and parameter-rich as the previous power network $\mathsf{Pwr}_{n-1}^{q,\varepsilon}$, one differs from the next by one $\mathsf{Prd}^{q,\varepsilon}$ network.*
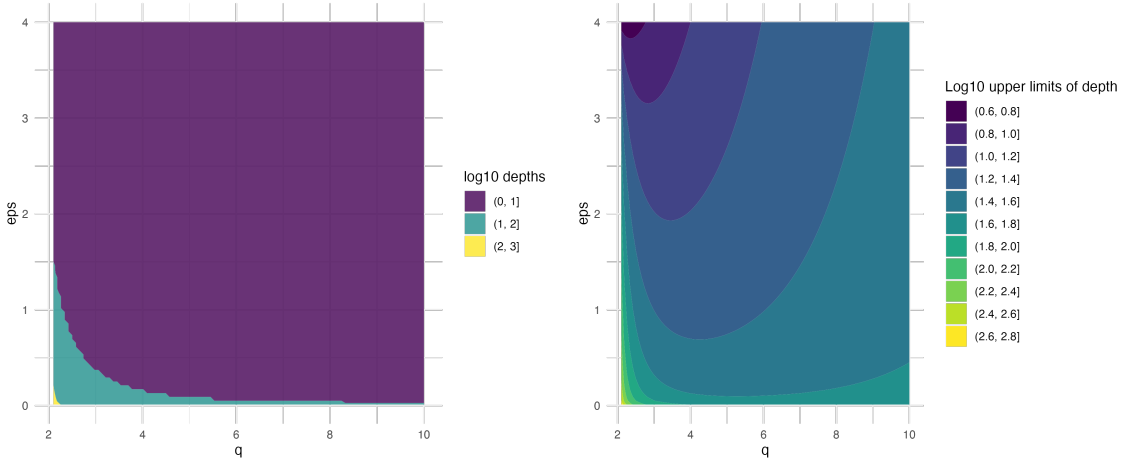
Figure 5.10: Left: $\log_{10}$ of depths for a simulation of $\mathsf{Pwr}_3^{q,\varepsilon}$ with $q \in [2.1,4]$, $\varepsilon \in (0.1,2]$, and $x \in [-5,5]$, all with 50 mesh-points. Right: The theoretical upper limits over the same range of values

5.2.3 $\mathsf{Pnm}_{n,C}^{q,\varepsilon}$ and Neural Network Polynomials.

Definition 5.2.16 (Neural Network Polynomials)**.** *Let $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$ and $\delta = \varepsilon\left(2^{q-1}+1\right)^{-1}$. For fixed q, ε, fixed $n \in \mathbb{N}_0$, and for $C = \{c_0, c_1, \ldots, c_n\} \in \mathbb{R}^{n+1}$ (the set of coefficients), we will define the following objects as neural network polynomials:*

$$\mathsf{Pnm}_{n,C}^{q,\varepsilon} := \bigoplus_{i=0}^{n} \left(c_i \triangleright \left[\mathsf{Tun}_{\max_i\left\{\mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\}+1-\mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon} \right] \right) \qquad (5.2.46)$$

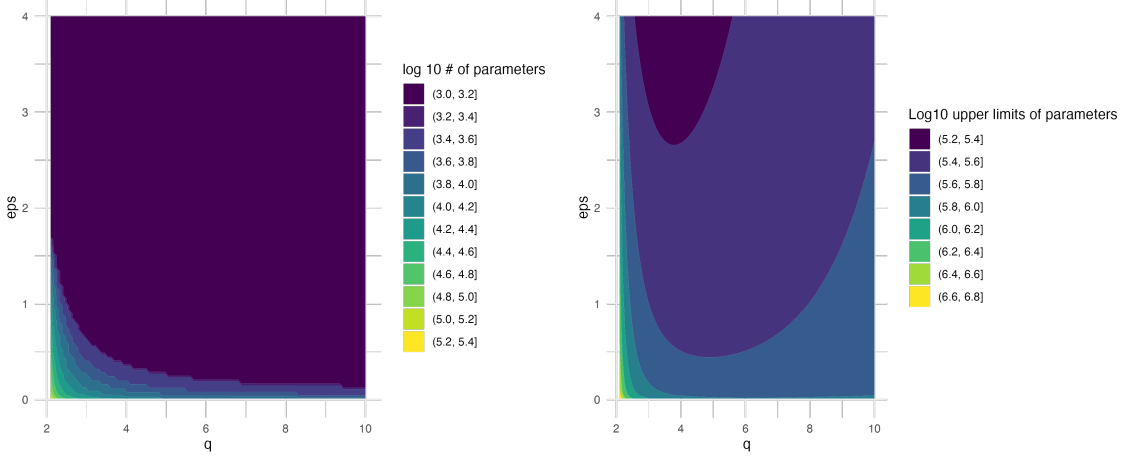**Remark 5.2.17.** *Diagrammatically, these can be represented as*

Figure 5.11: Left: $\log_{10}$ of params for a simulation of $\mathsf{Pwr}_3^{q,\varepsilon}$ with $q \in [2.1, 4]$, $\varepsilon \in (0.1, 2]$, and $x \in [-5, 5]$, all with 50 mesh-points. Right: The theoretical upper limits over the same range of values

Lemma 5.2.18 (R—,2023). *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon \left(2^{q-1} + 1 \right)^{-1}$ and let $C = \{ c_1, c_2, \ldots c_n \} \in \mathbb{R}^n$ be a set of real numbers, i.e. the set of coefficients. It is then the case for all $n \in \mathbb{N}_0$ and $x \in \mathbb{R}$ that:*

(i) $\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pnm}_{n,C}^{q,\varepsilon} \right) \in C \left(\mathbb{R}, \mathbb{R} \right)$

(ii) $\mathsf{D} \left(\mathsf{Pnm}_{n,C}^{q,\varepsilon} \right) \leqslant \begin{cases} 1 & : n = 0 \\ n \left[\frac{q}{q-2} \left[\log_2 \left(\varepsilon^{-1} \right) + q \right] - 1 \right] + 1 & : n \in \mathbb{N} \end{cases}$

(iii) $\mathsf{P} \left(\mathsf{Pnm}_{n,C}^{q,\varepsilon} \right) \leqslant \begin{cases} 2 & : n = 0 \\ (n+1) \left[4^{n+\frac{3}{2}} + \left(\frac{4^{n+1}-1}{3} \right) \left(\frac{360q}{q-2} \left[\log_2 \left(\varepsilon^{-1} \right) + q + 1 \right] + 372 \right) \right] & : n \in \mathbb{N} \end{cases}$

(iv) $\left| \sum_{i=0}^{n} c_i x^i - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pnm}_{n,C}^{q,\varepsilon} \right) (x) \right| \leqslant \sum_{i=1}^{n} c_i \left(\left| x \left(x^{i-1} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_{i-1}^{q,\varepsilon} \right) (x) \right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{i-1}^q \right)$
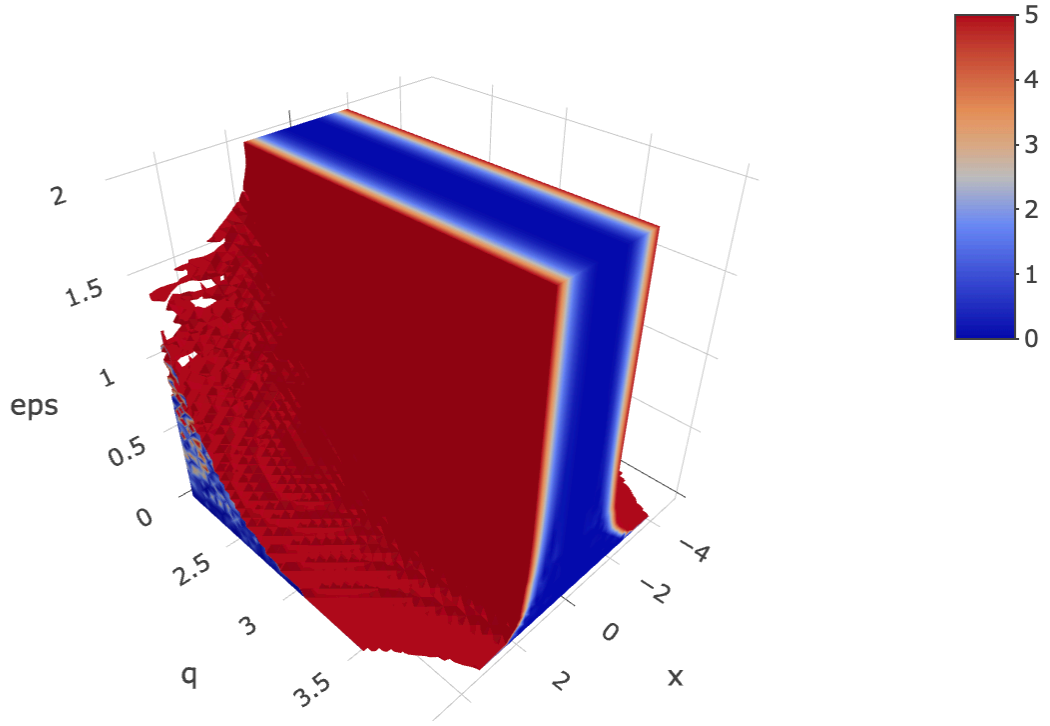
Figure 5.12: Isosurface plot showing $|x^3 - \mathfrak{J}_{\mathfrak{r}}(\mathsf{Pwr}_3^{q,\varepsilon})(x)|$ for $q \in [2.1, 4]$, $\varepsilon \in [0.01, 2]$, and $x \in [-5, 5]$ with 50 mesh-points in each.

Where $\mathfrak{p}_i$ are the set of functions defined for $i \in \mathbb{N}$ as such:

$$\mathfrak{p}_1 = \varepsilon + 1 + |x|^2$$

$$\mathfrak{p}_i = \varepsilon + (\mathfrak{p}_{i-1})^2 + |x|^2 \tag{5.2.47}$$

Whence it is the case that:

$$\left| \sum_{i=0}^{n} c_i x^i - \mathfrak{J}_{\mathfrak{r}} \left(\mathsf{Pnm}_{n,C}^{q,\varepsilon} \right)(x) \right| \in \mathcal{O}\left(\varepsilon^{2q(n-1)} \right) \tag{5.2.48}$$

(v) $\mathsf{W}_1 \left(\mathsf{Pnm}_{n,C}^{q,\varepsilon} \right) = 2 + 23n + n^2$

Figure 5.13: Neural network diagram for an elementary neural network polynomial, with all coefficients being uniformly 1.
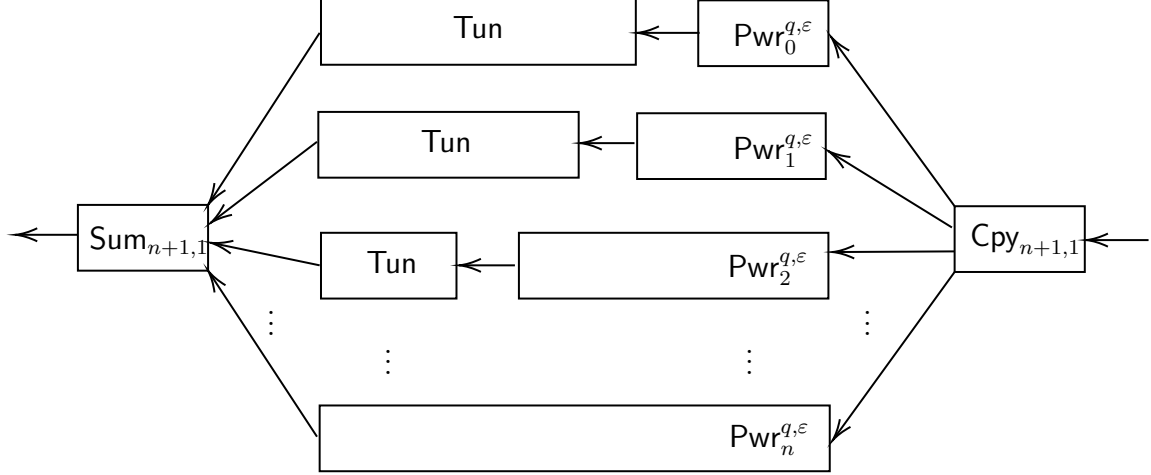
$$(vi) \ \mathsf{W}_{\mathsf{H}\left(\mathsf{Pnm}_{n,C}^{q,\varepsilon}\right)}\left(\mathsf{Pnm}_{n,C}^{q,\varepsilon}\right) \leqslant \begin{cases} 1 & : n = 0 \\ \\ 24 + 2n & : n \in \mathbb{N} \end{cases}$$

Proof. Note that by Lemma 4.6.5, Lemma 5.2.14, and Lemma 4.2.5 indicate for all $n \in \mathbb{N}_0$ it is the case that:

$$\begin{aligned} \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pnm}_{n,C}^{q,\varepsilon}\right) &= \mathfrak{I}_{\mathfrak{r}}\left(\bigoplus_{i=0}^{n}\left[c_i \triangleright \left[\mathsf{Tun}_{\max_i\left\{\mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\}+1-\mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon}\right]\right]\right) \\ &= \sum_{i=1}^{n} c_i\, \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Tun}_{\max_i\left\{\mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\}+1-\mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon}\right) \\ &= \sum_{i=1}^{n} c_i\, \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right) \end{aligned}$$

Since Lemma 5.2.14 tells us that $\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)\right)(x) \in C\left(\mathbb{R}, \mathbb{R}\right)$, for all $n \in \mathbb{N}_0$ and since the finite sum of continuous functions is continuous, this proves Item (i).

Note that $\mathsf{Pnm}_n^{q,\varepsilon}$ is only as deep as the deepest of the $\mathsf{Pwr}_i^{q,\varepsilon}$ networks, which from the definition is $\mathsf{Pwr}_n^{q,\varepsilon}$, which in turn also has the largest bound. Therefore, by Lemma 4.2.5,

	Min	1st. Qu	Median	Mean	3rd Qu	Max.
Experimental $\left\|x^3 - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}^{q,\varepsilon}\right)(x)\right\|$	0.0000	0.2053	7.2873	26.7903	45.4275	125.00
Experimental depths	4	4	4	4.92	4	238
Theoretical upper bound on depths	4.30	17.82	23.91	25.80	29.63	548.86
Forward Difference	0.30	13.82	19.91	20.88	25.63	310.86
Experimental params	1483	1483	1483	1546	1483	5711
Theoretical upper limit on params	9993	9993	9993	11589	9993	126843
Forward Differnce	8510	8510	8510	10043	8510	121132

Table 5.3: Table showing the experimental and theoretical 1-norm difference, depths, and parameter counts respectively for $\mathsf{Pwr}_3^{q,\varepsilon}$ with $q \in [2.1, 4]$, $\varepsilon \in [0.01, 2]$, and $x \in [-5, 5]$ all with 50 mesh-points, and their forward differences.

Lemma 4.4.5, Lemma 4.5.11, and Lemma 5.2.14, we have that:

$$\mathsf{D}\left(\mathsf{Pnm}_{n,C}^{q,\varepsilon}\right) \leqslant \mathsf{D}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)$$

$$\leqslant \begin{cases} 1 & : n = 0 \\ n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] - 1\right] + 1 & : n \in \mathbb{N} \end{cases}$$

This proves Item (ii).

Note next that for the case of $n = 0$, we have that:

$$\mathsf{Pnm}_n^{q,\varepsilon} = c_i \triangleright \mathsf{Pwr}_0^{q,\varepsilon} \tag{5.2.49}$$

This then yields us 2 parameters.

Note that each neural network summand in $\mathsf{Pnm}_n^{q,\varepsilon}$ consists of a combination of Tun_k and Pwr_k for some $k \in \mathbb{N}$. Each Pwr_k has at least as many parameters as a tunneling neural network of that depth, as Lemma 5.2.13 tells us. This, finally, with Lemma 4.4.5, Corollary 4.4.5.1, and Lemma 5.2.14 then implies that:

$$P\left(\mathsf{Pnm}_{n,C}^{q,\varepsilon}\right) = P\left(\bigoplus_{i=0}^{n}\left[c_i \triangleright \left[\mathsf{Tun}_{\max_i\left\{D\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\}+1-D\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon}\right]\right]\right)$$

$$\leqslant (n+1) \cdot P\left(c_i \triangleright \left[\mathsf{Tun}_1 \bullet \mathsf{Pwr}_n^{q,\varepsilon}\right]\right)$$

$$\leqslant (n+1) \cdot P\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)$$

$$\leqslant \begin{cases} 2 & : n = 0 \\ (n+1)\left[4^{n+\frac{3}{2}} + \left(\frac{4^{n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right)\right] & : n \in \mathbb{N} \end{cases}$$

This proves Item (iii).

Finally, note that for all $i \in \mathbb{N}$, Lemma 5.2.14, and the triangle inequality then tells us that it is the case for all $i \in \mathbb{N}$ that:

$$\left|x^i - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)(x)\right| \leqslant \left|x^i - x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)(x)\right| + \left|x \cdot \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)(x) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)(x)\right|$$

$$(5.2.50)$$

This, Lemma 5.2.29, and the fact that instantiation of the tunneling neural network leads to the identity function (Lemma 5.2.3 and Lemma 4.2.5), together with Lemma 4.6.8, and the absolute homogeneity condition of norms, then tells us that for all $x \in \mathbb{R}$, and $c_0, c_1, \ldots, c_n \in$

$\mathbb{R}$ it is the case that:

$$\left| \sum_{i=0}^{n} c_i x^i - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pnm}_{n,C}^{q,\varepsilon}(x) \right) \right|$$

$$= \left| \sum_{i=0}^{n} c_i x^i - \mathfrak{I}_{\mathfrak{r}} \left[\bigoplus_{i=0}^{n} \left[c_i \triangleright \mathsf{Tun}_{\max_i \left\{ \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon} \right] \right] (x) \right|$$

$$= \left| \sum_{i=1}^{n} c_i x^i - \sum_{i=0}^{n} c_i \left(\mathfrak{I}_{\mathfrak{r}} \left[\mathsf{Tun}_{\max_i \left\{ \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon} \right] (x) \right) \right|$$

$$\leqslant \sum_{i=1}^{n} |c_i| \cdot \left| x^i - \mathfrak{I}_{\mathfrak{r}} \left[\mathsf{Tun}_{\max_i \left\{ \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon} \right] (x) \right|$$

$$\leqslant \sum_{i=1}^{n} |c_i| \cdot \left(\left| x \left(x^{i-1} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_{i-1}^{q,\varepsilon} \right) (x) \right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{i-1}^q \right)$$

Note however that since for all $x \in \mathbb{R}$ and $i \in \mathbb{N} \cap [2, \infty)$, Lemma 5.1.10 tells us that $|x^i - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_i^{q,\varepsilon} \right) (x)| \in \mathcal{O} \left(\varepsilon^{2q(i-1)} \right)$, this, and the fact that $f + g \in \mathcal{O}(x^a)$ if $f \in \mathcal{O}(x^a)$, $g \in \mathcal{O}\left(x^b\right)$, and $a \geqslant b$, then implies that:

$$\sum_{i=1}^{n} |c_i| \cdot \left(\left| x \left(x^{i-1} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_{i-1}^{q,\varepsilon} \right) (x) \right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{i-1}^q \right) \in \mathcal{O} \left(\varepsilon^{2q(n-1)} \right) \tag{5.2.51}$$

This proves Item (iv).

Note next in our construction $\mathsf{Aff}_{0,1}$ will require tunneling whenever $i \in \mathbb{N}$ in $\mathsf{Pwr}_i^{q,\varepsilon}$. Lemma 4.4.5 and Corollary 4.4.5.1 then tell us that:

$$\mathsf{W}_1 \left(\mathsf{Pnm}_n^{q,\varepsilon} \right) = \mathsf{W}_1 \left(\bigoplus_{i=0}^{n} \left[c_i \triangleright \left[\mathsf{Tun}_{\max_i \left\{ \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon} \right] \right] \right)$$

$$= \mathsf{W}_1 \left(\bigoplus_{i=0}^{n} \mathsf{Pwr}_i^{q,\varepsilon} \right)$$

$$\leqslant \sum_{i=0}^{n} \mathsf{W}_1 \left(\mathsf{Pwr}_i^{q,\varepsilon} \right) = 2 + \frac{n}{2} \left(24 + 24 + 2 \left(n - 1 \right) \right) = 2 + 23n + n^2$$

$$\tag{5.2.52}$$

This proves Item (v).

Finally note that from the definition of the $\mathsf{Pnm}_{n,C}^{q,\varepsilon}$, it is evident that $\mathsf{W}_{\mathsf{H}\left(\mathsf{Pwr}_{0,C}^{q,\varepsilon}\right)}\left(\mathsf{Pwr}_{0,C}^{q,\varepsilon}\right) = 1$ since $\mathsf{Pwr}_{0,C}^{q,\varepsilon} = \mathsf{Aff}_{0,1}$. Other than this network, for all $i \in \mathbb{N}$, $\mathsf{Pwr}_{i,C}^{q,\varepsilon}$ end in the $\mathsf{Prd}^{q,\varepsilon}$ network, and the deepest of the $\mathsf{Pwr}_i^{q,\varepsilon}$ networks is $\mathsf{Pwr}_n^{q,\varepsilon}$ inside $\mathsf{Pnm}_{n,C}^{q,\varepsilon}$. All other $\mathsf{Pwr}_i^{q,\varepsilon}$ must end in tunnels. Whence in the second to last layer, Lemma 5.1.10 tells us that:

$$\mathsf{W}_{\mathsf{H}\left(\mathsf{Pnm}_{n,C}^{q,\varepsilon}\right)} \leqslant \begin{cases} 1 & : n = 0 \\ 24 + 2n & : n \in \mathbb{N} \end{cases} \tag{5.2.53}$$

This completes the proof of the Lemma. $\qquad\square$

Remark 5.2.19. *Note that we will implement this in R as the so-called* `Tay` *function. Our implementations of neural network exponentials, cosines, and sines will be instantiations of this* `Tay` *function with the appropriate coefficients and exponents being replaced to give the appropriate Taylor expansions.*

5.2.4 $\mathsf{Xpn}_n^{q,\varepsilon}$, $\mathsf{Csn}_n^{q,\varepsilon}$, $\mathsf{Sne}_n^{q,\varepsilon}$, and ANN Approximations of e^x, $\cos(x)$, and $\sin(x)$.

Once we have neural network polynomials, we may take the next leap to transcendental functions. For approximating them we will use Taylor expansions which will swiftly give us our approximations for our desired functions. Here, we will explore neural network approximations for three common transcendental functions: e^x, $\cos(x)$, and $\sin(x)$.

Lemma 5.2.20. *Let* $\nu_1, \nu_2 \in \mathsf{NN}$, $f, g \in C(\mathbb{R}, \mathbb{R})$, *and* $\varepsilon_1, \varepsilon_2 \in (0, \infty)$ *such that for all* $x \in \mathbb{R}$ *it holds that* $|f(x) - \mathfrak{I}_{\mathfrak{r}}(\nu_1)| \leqslant \varepsilon_1$ *and* $|g(x) - \mathfrak{I}_{\mathfrak{r}}(\nu_2)| \leqslant \varepsilon_2$. *It is then the case for all* $x \in \mathbb{R}$ *that:*

$$|[f + g](x) - \mathfrak{I}_{\mathfrak{r}}([\nu_1 \oplus \nu_2])(x)| \leqslant \varepsilon_1 + \varepsilon_2 \tag{5.2.54}$$

Proof. Note that the triangle inequality tells us:

$$\left|[f+g](x) - \mathfrak{I}_{\mathfrak{r}}[\nu_1 \oplus \nu_2](x)\right| = \left|f(x) + g(x) - \mathfrak{I}_{\mathfrak{r}}(\nu_1)(x) - \mathfrak{I}_{\mathfrak{r}}(\nu_2)(x)\right|$$

$$\leqslant \left|f(x) - \mathfrak{I}_{\mathfrak{r}}(\nu_1)(x)\right| + \left|g(x) - \mathfrak{I}_{\mathfrak{r}}(\nu_2)(x)\right|$$

$$\leqslant \varepsilon_1 + \varepsilon_2$$

$\square$

Lemma 5.2.21. *Let $n \in \mathbb{N}$. Let $\nu_1, \nu_2, ..., \nu_n \in \mathsf{NN}$, $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n \in (0, \infty)$ and $f_1, f_2, ..., f_n \in C(\mathbb{R}, \mathbb{R})$ such that for all $i \in \{1, 2, ..., n\}$, and for all $x \in \mathbb{R}$, it is the case that, $|f_i(x) - \mathfrak{I}_{\mathfrak{r}}(\nu_i)(x)| \leqslant \varepsilon_i$. It is then the case for all $x \in \mathbb{R}$, that:*

$$\left|\sum_{i=1}^{n} f_i(x) - \bigoplus_{i=1}^{n} (\mathfrak{I}_{\mathfrak{r}}(\nu_i))(x)\right| \leqslant \sum_{i=1}^{n} \varepsilon_i \tag{5.2.55}$$

Proof. This is a consequence of a finite number of applications of (5.2.54). $\square$

Definition 5.2.22 (R—2023, $\mathsf{Xpn}_n^{q,\varepsilon}$ and the Neural Network Taylor Approximations for e^x around $x = 0$). *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon (2^{q-1} + 1)^{-1}$, and let $\mathsf{Pwr}_n^{q,\varepsilon} \subsetneq \mathsf{NN}$ be as in Lemma 5.2.14. We define, for all $n \in \mathbb{N}_0$, the family of neural networks $\mathsf{Xpn}_n^{q,\varepsilon}$ as:*

$$\mathsf{Xpn}_n^{q,\varepsilon} := \bigoplus_{i=0}^{n} \left[\frac{1}{i!} \triangleright \left[\mathsf{Tun}_{\max_i\{\mathsf{D}(\mathsf{Pwr}_i^{q,\varepsilon})\}+1-\mathsf{D}(\mathsf{Pwr}_i^{q,\varepsilon})} \bullet \mathsf{Pwr}_i^{q,\varepsilon}\right]\right] \tag{5.2.56}$$

Lemma 5.2.23 (R—,2023). *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon (2^{q-1} + 1)^{-1}$. It is then the case for all $n \in \mathbb{N}_0$ and $x \in \mathbb{R}$ that:*

(i) $\mathfrak{I}_{\mathfrak{r}}(\mathsf{Xpn}_n^{q,\varepsilon})(x) \in C(\mathbb{R}, \mathbb{R})$

(ii) $\mathsf{D}(\mathsf{Xpn}_n^{q,\varepsilon}) \leqslant \begin{cases} 1 & : n = 0 \\ n\left[\frac{q}{q-2}\left[\log_2(\varepsilon^{-1}) + q\right] - 1\right] + 1 & : n \in \mathbb{N} \end{cases}$

(iii) $\mathsf{P}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) \leqslant \begin{cases} 2 & : n = 0 \\ \\ (n+1)\left[4^{n+\frac{3}{2}} + \left(\frac{4^{n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right)\right] & : n \in \mathbb{N} \end{cases}$

(iv)

$$\left|\sum_{i=0}^{n}\left[\frac{x^i}{i!}\right] - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right)(x)\right| \leqslant \sum_{i=1}^{n}\frac{1}{i!}\left(\left|x\left(x^{i-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)(x)\right)\right| + \varepsilon + |x|^q + \mathfrak{p}_{i-1}^q\right)$$

Where $\mathfrak{p}_i$ are the set of functions defined for $i \in \mathbb{N}$ as such:

$$\mathfrak{p}_1 = \varepsilon + 1 + |x|^2$$
$$\mathfrak{p}_i = \varepsilon + (\mathfrak{p}_{i-1})^2 + |x|^2 \tag{5.2.57}$$

Whence it is the case that:

$$\left|\sum_{i=0}^{n}\left[\frac{x^i}{i!}\right] - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right)(x)\right| \in \mathcal{O}\left(\varepsilon^{2q(n-1)}\right) \tag{5.2.58}$$

(v) $\mathsf{W}_1\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) = 2 + 23n + n^2$

(vi) $\mathsf{W}_{\mathsf{H}\left(\mathsf{Xpn}_{q,\varepsilon}^n\right)}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) \leqslant 24 + 2n$

Proof. This follows straightforwardly from Lemma 5.2.18 with $c_i \curvearrowleft \frac{1}{i!}$ for all $n \in \mathbb{N}$ and $i \in \{0, 1, \ldots, n\}$. In particular, Item (iv) benefits from the fact that for all $i \in \mathbb{N}_0$, it is the case that $\frac{1}{i!} \geqslant 0$. $\square$

Lemma 5.2.24 (R—, 2023)**.** *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon\left(2^{q-1} + 1\right)^{-1}$. It is then*

168

the case for all $n \in \mathbb{N}_0$ *and* $x \in [a, b] \subsetneq \mathbb{R}$, *where* $0 \in [a, b] \subsetneq \mathbb{R}$ *that:*

$$\left| e^x - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Xpn}_n^{q,\varepsilon} \right) (x) \right| \leqslant \sum_{i=0}^{n} \frac{1}{i!} \left(\left| x \left(x^{n-1} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_{n-1}^{q,\varepsilon} \right) (x) \right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{n-1}^q \right) + \frac{e^b \cdot |x|^{n+1}}{(n+1)!}$$

$$(5.2.59)$$

Proof. Note that Taylor's theorem states that for $x \in [a, b] \subsetneq \mathbb{R}$ it is the case that:

$$e^x = \sum_{i=0}^{n} \left[\frac{x^i}{i!} \right] + \frac{e^\xi \cdot x^{n+1}}{(n+1)!} \tag{5.2.60}$$

Where ξ is between 0 and x in the Lagrange form of the remainder. Note then, for all $n \in \mathbb{N}_0$, $x \in [a, b] \subsetneq \mathbb{R}$, and ξ between 0 and x, it is the case, by monotonicity of e^x that the second summand is bounded by:

$$\frac{e^\xi \cdot x^{n+1}}{(n+1)!} \leqslant \frac{e^b \cdot |x|^{n+1}}{(n+1)!} \tag{5.2.61}$$

This, and the triangle inequality, then indicates that for all $x \in [a, b] \subsetneq \mathbb{R}$, and ξ between 0 and x that:

$$
\begin{aligned}
\left| e^x - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Xpn}_n^{q,\varepsilon} \right) (x) \right| &= \left| \sum_{i=0}^{n} \left[\frac{x^i}{i!} \right] + \frac{e^\xi \cdot x^{n+1}}{(n+1)!} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Xpn}_n^{q,\varepsilon} \right) (x) \right| \\
&\leqslant \left| \sum_{i=0}^{n} \left[\frac{x^i}{i!} \right] - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Xpn}_n^{q,\varepsilon} \right) (x) \right| + \frac{e^b \cdot |x|^{n+1}}{(n+1)!} \\
&\leqslant \sum_{i=1}^{n} \frac{1}{i!} \left(\left| x \left(x^{n-1} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_{n-1}^{q,\varepsilon} \right) (x) \right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{n-1}^q \right) + \frac{e^b \cdot |x|^{n+1}}{(n+1)!}
\end{aligned}
$$

Whence we have that for fixed $n \in \mathbb{N}_0$ and $b \in [0, \infty)$, the last summand is constant, whence it is the case that:

$$\left| e^x - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Xpn}_n^{q,\varepsilon} \right) (x) \right| \in \mathcal{O} \left(\varepsilon^{2q(n-1)} \right) \tag{5.2.62}$$

169

$\square$

Definition 5.2.25 (The $\mathsf{Csn}_n^{q,\varepsilon}$ Networks, and Neural Network Cosines). *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon \left(2^{q-1} + 1\right)^{-1}$. Let $\mathsf{Pwr}_n^{q,\varepsilon}$ be a neural networks as defined in Definition 5.2.7. We will define the neural networks $\mathsf{Csn}_n^{q,\varepsilon}$ as:*

$$\mathsf{Csn}_n^{q,\varepsilon} := \bigoplus_{i=0}^{n} \left[\frac{(-1)^i}{2i!} \triangleright \left[\mathsf{Tun}_{\max_i \left\{ \mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\}+1-\mathsf{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)} \bullet \mathsf{Pwr}_{2i}^{q,\varepsilon} \right] \right] \tag{5.2.63}$$

Lemma 5.2.26 (R—, 2023). *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon \left(2^{q-1} + 1\right)^{-1}$. It is then the case for all $n \in \mathbb{N}_0$ and $x \in \mathbb{R}$ that:*

(i) $\mathfrak{I}_\mathfrak{r} \left(\mathsf{Csn}_n^{q,\varepsilon}\right)(x) \in C\left(\mathbb{R}, \mathbb{R}\right)$

(ii) $\mathsf{D}\left(\mathsf{Csn}_n^{q,\varepsilon}\right) \leqslant \begin{cases} 1 & : n = 0 \\ 2n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] - 1\right] + 1 & : n \in \mathbb{N} \end{cases}$

(iii) $\mathsf{P}\left(\mathsf{Csn}_n^{q,\varepsilon}\right) \leqslant \begin{cases} 2 & : n = 0 \\ (2n+1)\left[4^{2n+\frac{3}{2}} + \left(\frac{4^{2n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right)\right] & : n \in \mathbb{N} \end{cases}$

(iv) $\left| \sum_{i=0}^{n} \frac{(-1)^i}{2i!} x^{2i} - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Csn}_n^{q,\varepsilon}\right)(x) \right|$
$\leqslant \sum_{i=1}^{n} \left| \frac{(-1)^i}{2i!} \right| \left(\left| x\left(x^{2i-1} - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Pwr}_{2i-1}^{q,\varepsilon}\right)(x)\right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{2i-1}^q \right)$

Where $\mathfrak{p}_i$ are the set of functions defined for $i \in \mathbb{N}$ as such:

$$\mathfrak{p}_1 = \varepsilon + 1 + |x|^2$$
$$\mathfrak{p}_i = \varepsilon + \left(\mathfrak{p}_{i-1}\right)^2 + |x|^2 \tag{5.2.64}$$

Whence it is the case that:

$$\left| \sum_{i=0}^{n} \frac{(-1)^i}{2i!} x^{2i} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Csn}_n^{q,\varepsilon} \right)(x) \right| \in \mathcal{O} \left(\varepsilon^{2q(2n-1)} \right) \tag{5.2.65}$$

Proof. Item (i) derives straightforwardly from Lemma 5.2.18. This proves Item (i).

Next, observe that since $\mathsf{Csn}_n^{q,\varepsilon}$ will contain, as the deepest network in the summand, $\mathsf{Pwr}_{2n}^{q,\varepsilon}$, we may then conclude that

$$\mathsf{D} \left(\mathsf{Csn}_n^{q,\varepsilon} \right) \leqslant \mathsf{D} \left(\mathsf{Pwr}_{2n}^{q,\varepsilon} \right)$$

$$\leqslant \begin{cases} 1 & : n = 0 \\ 2n \left[\frac{q}{q-2} \left[\log_2 \left(\varepsilon^{-1} \right) + q \right] - 1 \right] + 1 & : n \in \mathbb{N} \end{cases}$$

This proves Item (ii).

A similar argument to the above, Lemma 4.4.5, and Corollary 4.4.5.1 reveals that:

$$\mathsf{P} \left(\mathsf{Csn}_n^{q,\varepsilon} \right) = \mathsf{P} \left(\bigoplus_{i=0}^{n} \left[\frac{(-1)^i}{2i!} \triangleright \left[\mathsf{Tun}_{\max_i \left\{ \mathsf{D} \left(\mathsf{Pwr}_i^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D} \left(\mathsf{Pwr}_i^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_i^{q,\varepsilon} \right] \right] \right)$$

$$\leqslant (n+1) \cdot \mathsf{P} \left(c_i \triangleright \left[\mathsf{Tun}_1 \bullet \mathsf{Pwr}_{2n}^{q,\varepsilon} \right] \right)$$

$$\leqslant (n+1) \cdot \mathsf{P} \left(\mathsf{Pwr}_{2n}^{q,\varepsilon} \right)$$

$$\leqslant \begin{cases} 2 & : n = 0 \\ (n+1) \left[4^{2n+\frac{3}{2}} + \left(\frac{4^{2n+1}-1}{3} \right) \left(\frac{360q}{q-2} \left[\log_2 \left(\varepsilon^{-1} \right) + q + 1 \right] + 372 \right) \right] & : n \in \mathbb{N} \end{cases}$$

This proves Item (iii).

In a similar vein, we may argue from Lemma 5.2.18 and from the absolute homogeneity

property of norms that:

$$
\left| \sum_{i=0}^{n} \frac{(-1)^i}{2i!} x^{2i} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Csn}_n^{q,\varepsilon}(x) \right) \right|
$$

$$
= \left| \sum_{i=0}^{n} \frac{(-1)^i}{2i!} x^{2i} - \mathfrak{I}_{\mathfrak{r}} \left[\bigoplus_{i=0}^{n} \left[\frac{(-1)^i}{2i!} \triangleright \mathsf{Tun}_{\max_{2i} \left\{ \mathsf{D} \left(\mathsf{Pwr}_{2i}^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D} \left(\mathsf{Pwr}_{2i}^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_{2i}^{q,\varepsilon} \right] \right] (x) \right|
$$

$$
= \left| \sum_{i=1}^{n} \frac{(-1)^i}{2i!} x^{2i} - \sum_{i=0}^{n} \frac{(-1)^i}{2i!} \left(\mathfrak{I}_{\mathfrak{r}} \left[\mathsf{Tun}_{\max_{2i} \left\{ \mathsf{D} \left(\mathsf{Pwr}_{2i}^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D} \left(\mathsf{Pwr}_{2i}^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_{2i}^{q,\varepsilon} \right] (x) \right) \right|
$$

$$
\leqslant \sum_{i=1}^{n} \left| \frac{(-1)^i}{2i!} \right| \cdot \left| x^{2i} - \mathfrak{I}_{\mathfrak{r}} \left[\mathsf{Tun}_{\max_{2i} \left\{ \mathsf{D} \left(\mathsf{Pwr}_{2i}^{q,\varepsilon} \right) \right\} + 1 - \mathsf{D} \left(\mathsf{Pwr}_{2i}^{q,\varepsilon} \right)} \bullet \mathsf{Pwr}_{2i}^{q,\varepsilon} \right] (x) \right|
$$

$$
\leqslant \sum_{i=1}^{n} \left| \frac{(-1)^i}{2i!} \right| \cdot \left| \left(\left| x \left(x^{2i-1} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_{2i-1}^{q,\varepsilon} \right)(x) \right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{2i-1}^q \right) \right|
$$

Whence we have that:

$$
\left| \sum_{i=0}^{n} \left[\frac{(-1)^i x^{2i}}{2i!} \right] - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Csn}_n^{q,\varepsilon} \right)(x) \right| \in \mathcal{O} \left(\varepsilon^{2q(2n-1)} \right) \tag{5.2.66}
$$

This proves Item (iv). $\qquad\qquad\square$

Lemma 5.2.27 (R—, 2023). *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon \left(2^{q-1} + 1 \right)^{-1}$. It is then the case for all $n \in \mathbb{N}_0$ and $x \in [a, b] \subseteq [0, \infty)$ that:*

$$
\left| \cos(x) - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Csn}_n^{q,\varepsilon} \right)(x) \right| \tag{5.2.67}
$$

$$
\leqslant \sum_{i=0}^{n} \frac{(-1)^i}{2i!} \left(\left| x \left(x^{n-1} - \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Pwr}_{n-1}^{q,\varepsilon} \right)(x) \right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{n-1}^q \right) + + \frac{|x|^{n+1}}{(n+1)!}
$$

Proof. Note that Taylor's theorem states that for all $x \in [a, b] \subsetneq \mathbb{R}$, where $0 \in [a, b]$, it is the case that:

$$
\cos(x) = \sum_{i=0}^{n} \frac{(-1)^i}{2i!} x^i + \frac{\cos^{(n+1)}(\xi) \cdot x^{n+1}}{(n+1)!} \tag{5.2.68}
$$

Note further that for all $n \in \mathbb{N}_0$, and $x \in \mathbb{R}$, it is the case that $\cos^{(n)}(x) \leqslant 1$. Whence we

may conclude that for all $n \in \mathbb{N}_0$, $x \in [a,b] \subseteq \mathbb{R}$, where $0 \in [a,b]$ and ξ between 0 and x, we may bound the second summand by:

$$\frac{\cos^{(n+1)}(\xi) \cdot x^{n+1}}{(n+1)!} \leqslant \frac{|x|^{n+1}}{(n+1)!} \tag{5.2.69}$$

This, and the triangle inequality, then indicates that for all $x \in [a,b] \subsetneq [0,\infty)$ and $\xi \in [0,x]$:

$$
\begin{aligned}
\left|\cos(x) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Csn}_n^{q,\varepsilon}\right)(x)\right| &= \left| \sum_{i=0}^{n} \frac{(-1)^i}{2i!} x^i + \frac{\cos^{(n+1)}(\xi) \cdot x^{n+1}}{(n+1)!} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Csn}_n^{q,\varepsilon}\right)(x) \right| \\
&\leqslant \left| \sum_{i=0}^{n} \frac{(-1)^i}{2i!} x^i - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Csn}_n^{q,\varepsilon}\right)(x) \right| + \frac{|x|^{n+1}}{(n+1)!} \\
&\leqslant \sum_{i=1}^{n} \left| \frac{(-1)^i}{2i!} \right| \cdot \left| \left(\left| x\left(x^{2i-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{2i-1}^{q,\varepsilon}\right)(x)\right) \right| + \varepsilon + |x|^q + \mathfrak{p}_{2i-1}^q \right) \right| \\
&\quad + \frac{|x|^{n+1}}{(n+1)!}
\end{aligned}
$$

This completes the proof of the Lemma. $\qquad\square$

Definition 5.2.28 (R—, 2023, The $\mathsf{Sne}_n^{q,\varepsilon}$ Newtorks and Neural Network Sines.). *Let $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$ and $\delta = \varepsilon\left(2^{q-1}+1\right)^{-1}$. Let $\mathsf{Pwr}^{q,\varepsilon}$ be a neural network defined in Definition 5.2.7. We will define the neural network $\mathsf{Csn}_{n,q,\varepsilon}$ as:*

$$\mathsf{Sne}_n^{q,\varepsilon} \coloneqq \mathsf{Csn}^{q,\varepsilon} \bullet \mathsf{Aff}_{1,-\frac{\pi}{2}} \tag{5.2.70}$$

Lemma 5.2.29 (R—, 2023). *Let $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$ and $\delta = \varepsilon\left(2^{q-1}+1\right)^{-1}$. It is then the case for all $n \in \mathbb{N}_0$ and $x \in \mathbb{R}$ that:*

(i) $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sne}_n^{q,\varepsilon}\right) \in C\left(\mathbb{R},\mathbb{R}\right)$

(ii) $\mathsf{D}\left(\mathsf{Sne}_n^{q,\varepsilon}\right) \leqslant \begin{cases} 1 & : n = 0 \\ 2n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] - 1\right] + 1 & : n \in \mathbb{N} \end{cases}$

173

(iii) $\mathsf{P}\left(\mathsf{Sne}_n^{q,\varepsilon}\right) \leqslant \begin{cases} 2 & : n = 0 \\ (2n+1)\left[4^{2n+\frac{3}{2}} + \left(\frac{4^{2n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right)\right] & : n \in \mathbb{N} \end{cases}$

(iv)

$$\left|\sum_{i=0}^n \frac{(-1)^i}{2i!}\left(x - \frac{\pi}{2}\right)^{2i} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sne}_n^{q,\varepsilon}\right)(x)\right|$$

$$= \left|\sum_{i=0}^n \frac{(-1)^i}{2i!}\left(x - \frac{\pi}{2}\right)^{2i} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Csn}_n^{q,\varepsilon} \bullet \mathsf{Aff}_{1,-\frac{\pi}{2}}\right)(x)\right|$$

$$\leqslant \sum_{i=1}^n \left|\frac{(-1)^i}{2i!}\right|\left(\left|\left(x - \frac{\pi}{2}\right)\left(\left(x - \frac{\pi}{2}\right)^{2i-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)\left(x - \frac{\pi}{2}\right)\right)\right| + \varepsilon + |x|^q + \mathfrak{p}_{i-1}^q\right)$$

Where $\mathfrak{p}_i$ are the set of functions defined for $i \in \mathbb{N}$ as such:

$$\mathfrak{p}_1 = \varepsilon + 1 + |x|^2$$

$$\mathfrak{p}_i = \varepsilon + (\mathfrak{p}_{i-1})^2 + |x|^2 \tag{5.2.71}$$

Whence it is the case that:

$$\left|\sum_{i=0}^n \frac{(-1)^i}{2i!}\left(x - \frac{\pi}{2}\right)^{2i} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sne}_n^{q,\varepsilon}\right)(x)\right| \in \mathcal{O}\left(\varepsilon^{2q(2n-1)}\right) \tag{5.2.72}$$

Proof. This follows straightforwardly from Lemma 5.2.26, and the fact that by Corollary 4.4.5.1, there is not a change to the parameter count, by Lemma 4.2.11, there is no change in depth, by Lemma 4.4.2, and Lemma 5.2.26, continuity is preserved, and the fact that $\mathsf{Aff}_{1,-\frac{\pi}{2}}$ is exact and hence contributes nothing to the error, and finally by the fact that $\mathsf{Aff}_{1,-\frac{\pi}{2}} \to (\cdot) - \frac{\pi}{2}$ under instantiation, assures us that the $\mathsf{Sne}_n^{q,\varepsilon}$ has the same error bounds

as $\mathsf{Csn}_n^{q,\varepsilon}$. $\qquad\square$

Lemma 5.2.30 (R—, 2023). *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon \left(2^{q-1} + 1\right)^{-1}$. It is then the case for all $n \in \mathbb{N}_0$ and $x \in [a, b] \subseteq [0, \infty)$ that:*

$$\left|\sin(x) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Sne}_n^{q,\varepsilon}\right)(x)\right|$$

$$\leqslant \sum_{i=1}^{n} \left|\frac{(-1)^i}{2i!}\right| \left(\left|\left(x - \frac{\pi}{2}\right)\left(\left(x - \frac{\pi}{2}\right)^{2i-1} - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)\left(x - \frac{\pi}{2}\right)\right)\right| + \varepsilon + |x|^q + \mathfrak{p}_{i-1}^q\right)$$

$$+ \frac{|x|^{n+1}}{(n+1)!} \tag{5.2.73}$$

Proof. Note that the fact that $\sin(x) = \cos\left(x - \frac{\pi}{2}\right)$, Lemma 4.2.5, and Lemma 4.4.2 then renders (5.2.73) as:

$$\left|\sin(x) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Sne}_n^{q,\varepsilon}\right)\right|$$

$$= \left|\cos\left(x - \frac{\pi}{2}\right) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Csn}_n^{q,\varepsilon} \bullet \mathsf{Aff}_{1, -\frac{\pi}{2}}\right)(x)\right|$$

$$= \left|\cos\left(x - \frac{x}{2}\right) - \mathfrak{I}_\mathfrak{r}\, \mathsf{Csn}_n^{q,\varepsilon}\left(x - \frac{\pi}{2}\right)\right|$$

$$\leqslant \sum_{i=1}^{n} \left|\frac{(-1)^i}{2i!}\right| \left(\left|\left(x - \frac{\pi}{2}\right)\left(\left(x - \frac{\pi}{2}\right)^{2i-1} - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)\left(x - \frac{\pi}{2}\right)\right)\right| + \varepsilon + |x|^q + \mathfrak{p}_{i-1}^q\right)$$

$$+ \frac{|x|^{n+1}}{(n+1)!}$$

$\qquad\square$

Remark 5.2.31. *Note that under these neural network architectures the famous Pythagorean identity $\sin^2(x) + \cos^2(x) = 1$, may be rendered approximately, for appropriately fixed n, q, ε as: $\left[\mathsf{Sqr}^{q,\varepsilon} \bullet \mathsf{Csn}_n^{q,\varepsilon}\right] \oplus \left[\mathsf{Sqr}^{q,\varepsilon} \bullet \mathsf{Sne}_n^{q,\varepsilon}\right] \approx 1$. On a similar note, it is the case, with appropriate n, q, ε that $\mathfrak{I}_\mathfrak{r}\left(\mathsf{Xpn}_n^{q,\varepsilon} \triangleleft i\right)(\pi) \approx -1$*

A full discussion of the associated parameter, depth, and accuracy bounds are beyond the scope of this dissertation, and may be appropriate for future work.

Chapter 6

ANN first approximations

We will give here a few ANN representations of common functions. Specifically we will posit the existence of a 1-dimensional identity and show that it acts as a compositional idenity for neural networks with fixed end-widths. Thus under composition neural networks with fixed end-widths under composition act as a monoid with Id_d as the compositional identity.

We will also posit two new neural networks Trp^h, and $\mathsf{Etr}^{n,h}$ neural networks for approximating trapezoidal rule integration.

We will then go on to posit the Nrm_1^d and Mxm^d network, taken mainly from (Jentzen et al., 2023, Chapter 3), our contribution will be to add parameter estimates.

We will finally go on to show the $\mathsf{MC}_{x,y}^{N,d}$ neural network which will perform the maximum convolution approximation for functions $f : \mathbb{R}^d \to \mathbb{R}$. Our contribution will be to show that the parameter counts are polynomial on dimension, d.

6.1 ANN Representations for One-Dimensional Identity

Definition 6.1.1 (One Dimensional Identity Neural Network)**.** *We will denote by* $\mathsf{Id}_d \in \mathsf{NN}$ *the neural network satisfying for all* $d \in \mathbb{N}$ *that:*

(i)

$$\mathsf{Id}_1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \in \left(\left(\mathbb{R}^{2\times 1} \times \mathbb{R}^2 \right) \times \left(\mathbb{R}^{1\times 2} \times \mathbb{R}^1 \right) \right)$$

$$(6.1.1)$$

(ii)

$$\mathsf{Id}_d = \boxminus_{i=1}^{d} \mathsf{Id}_1 \qquad (6.1.2)$$

For $d > 1$.

Lemma 6.1.2. *Let $d \in \mathbb{N}$, it is then the case that:*

(i) $\mathsf{L}(\mathsf{Id}_d) = (d, 2d, d) \in \mathbb{N}^3$.

(ii) $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_d \right) \in C\left(\mathbb{R}^d, \mathbb{R}^d \right)$.

(iii) *For all $x \in \mathbb{R}^d$ that:*

$$\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_d \right) \right)(x) = x$$

(iv) *For $d \in \mathbb{N}$ it is the case that $\mathsf{D}\left(\mathsf{Id}_d \right) = 2$*

Proof. Note that (6.1.1) ensure that $\mathsf{L}(\mathsf{Id}_d) = (1, 2, 1)$. Furthermore, (6.1.2) and Remark 4.3.12 prove that $\mathsf{L}(\mathsf{Id}_d) = (d, 2d, d)$ which in turn proves Item (i). Note now that Remark 4.3.12 tells us that:

$$\mathsf{Id}_d = \boxminus_{i=1}^{d}\left(\mathsf{Id}_1 \right) \in \left(\bigtimes_{i=1}^{L} \left[\mathbb{R}^{dl_i \times dl_{i-1}} \times \mathbb{R}^{dl_i} \right] \right) = \left(\left(\mathbb{R}^{2d \times d} \times \mathbb{R}^{2d} \right) \times \left(\mathbb{R}^{d \times 2d} \times \mathbb{R}^d \right) \right) \qquad (6.1.3)$$

Note that 6.1.1 ensures that for all $x \in \mathbb{R}$ it is the case that:

$$\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_1\right)\right)(x) = \mathfrak{r}(x) - \mathfrak{r}(-x) = \max\{x,0\} - \max\{-x,0\} = x \qquad (6.1.4)$$

And Lemma 4.3.8 shows us that for all $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ it is the case that $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_d\right) \in C\left(\mathbb{R}^d, \mathbb{R}^d\right)$ and that:

$$\begin{aligned}
\left(\mathfrak{I}_{\mathfrak{a}}\left(\mathsf{Id}_d\right)\right)(x) &= \left(\mathfrak{I}_{\mathfrak{a}}\left(\boxminus_{i=1}^d \left(\mathsf{Id}_1\right)\right)\right)(x_1, x_2, ..., x_d) \\
&= \left(\left(\mathfrak{I}_{\mathfrak{a}}\left(\mathsf{Id}_1\right)\right)(x_1), \left(\mathfrak{I}_{\mathfrak{a}}\left(\mathsf{Id}_1\right)\right)(x_1), ..., \left(\mathfrak{I}_{\mathfrak{a}}\left(\mathsf{Id}_1\right)\right)(x_d)\right) \\
&= (x_1, x_2, ..., x_d) = x \qquad (6.1.5)
\end{aligned}$$

This proves Item (ii)—(iii). Item (iv) follows straightforwardly from Item (i). This establishes the lemma. $\square$

Remark 6.1.3. *Note here the difference between Definition 5.1.1 and Definition 6.1.1.*

Lemma 6.1.4 (R—, 2023)**.** *Let $d \in \mathbb{N}$. It then the case that for all $d \in \mathbb{N}$ we have that $\mathsf{P}\left(\mathsf{Id}_d\right) = 4d^2 + 3d$*

Proof. By observation we have that $\mathsf{P}\left(\mathsf{Id}_1\right) = 4(1)^2 + 3(1) = 7$. By induction, suppose that this holds for all natural numbers up to and including n, i.e., for all naturals up to and including n; it is the case that $\mathsf{P}\left(\mathsf{Id}_n\right) = 4n^2 + 3n$. Note then that $\mathsf{Id}_{n+1} = \mathsf{Id}_n \boxminus \mathsf{Id}_1$. For W_1 and W_2 of this new network, this adds a combined extra $8n + 4$ parameters. For b_1 and b_2 of this new network, this adds a combined extra 3 parameters. Thus, we have the following:

$$4n^2 + 3n + 8n + 4 + 3 = 4(n+1)^2 + 3(n+1) \qquad (6.1.6)$$

This completes the induction and hence proves the Lemma. $\square$

Lemma 6.1.5. *Let $\nu \in \mathsf{NN}$ with end-widths d. It is then the case that $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Id}_d \bullet \nu\right)(x) = \mathfrak{I}_{\mathfrak{r}}\left(\nu \bullet \mathsf{Id}_d\right) = \mathfrak{I}_{\mathfrak{r}}\left(\nu\right)$, i.e. Id_d acts as a compositional identity.*

Proof. From (4.2.1) and Definition 6.1.1 we have eight cases.

Case 1 where $d = 1$ and subcases:

(1.i) $\mathsf{Id}_d \bullet \nu$ where $\mathsf{D}(\nu) = 1$

(1.ii) $\mathsf{Id}_d \bullet \nu$ where $\mathsf{D}(\nu) > 1$

(1.iii) $\nu \bullet \mathsf{Id}_d$ where $\mathsf{D}(\nu) = 1$

(1.iv) $\nu \bullet \mathsf{Id}_d$ where $\mathsf{D}(\nu) > 1$

Case 2 where $d > 1$ and subcases:

(2.i) $\mathsf{Id}_d \bullet \nu$ where $\mathsf{D}(\nu) = 1$

(2.ii) $\mathsf{Id}_d \bullet \nu$ where $\mathsf{D}(\nu) > 1$

(2.iii) $\nu \bullet \mathsf{Id}_d$ where $\mathsf{D}(\nu) = 1$

(2.iv) $\nu \bullet \mathsf{Id}_d$ where $\mathsf{D}(\nu) > 1$

Case 1.i: Let $\nu = ((W_1, b_1))$. Deriving from Definitions 6.1.1 and 4.2.1 we have that:

$$\mathsf{Id}_1 \bullet \nu = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} W_1, \begin{bmatrix} 1 \\ -1 \end{bmatrix} b_1 + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \tag{6.1.7}$$

$$= \left(\left(\begin{bmatrix} W_1 \\ -W_1 \end{bmatrix}, \begin{bmatrix} b_1 \\ -b_1 \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \tag{6.1.8}$$

Let $x \in \mathbb{R}$. Upon instantiation with $\mathfrak{r}$ and $d = 1$ we have:

$$\left(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Id}_1 \bullet \nu \right) \right)(x) = \mathfrak{r}(W_1 x + b_1) - \mathfrak{r}(-W_1 x - b_1)$$

$$= \max\{W_1 x + b_1, 0\} - \max\{-W_1 x - b_1, 0\}$$

$$= W_1 x + b_1$$

$$= \mathfrak{I}_{\mathfrak{r}}(\nu)$$

179

Case 1.ii: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definition 6.1.1 and 4.2.1 we have that:

$\mathsf{Id}_1 \bullet \nu$

$$= \left((W_1, b_1), (W_2, b_2), ..., (W_{L-1}, b_{L-1}), \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} W_L, \begin{bmatrix} 1 \\ -1 \end{bmatrix} b_L + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right)$$

$$= \left((W_1, b_1), (W_2, b_2), ..., (W_{L-1}, b_{L-1}), \left(\begin{bmatrix} W_L \\ -W_L \end{bmatrix}, \begin{bmatrix} b_L \\ -b_L \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right)$$

Let $x \in \mathbb{R}$. Note that upon instantiation with $\mathfrak{r}$, the last two layers are:

$$\mathfrak{r}(W_L x + b_L) - \mathfrak{r}(-W_L x - b_L, 0)$$

$$= \max\{W_L x + b_L, 0\} - \max\{-W_L x - b_L, 0\}$$

$$= W_L x + b_L \tag{6.1.9}$$

This, along with Case 1. i, implies that the uninstantiated last layer is equivalent to (W_L, b_L) whence $\mathsf{Id}_1 \bullet \nu = \nu$.

Case 1.iii: Let $\nu = ((W_1, b_1))$. Deriving from Definition 6.1.1 and 4.2.1 we have:

$$\nu \bullet \mathsf{Id}_1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(W_1 \begin{bmatrix} 1 & -1 \end{bmatrix}, W_1 \begin{bmatrix} 0 \\ 0 \end{bmatrix} + b_1 \right) \right)$$

$$= \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} W_1 & -W_1 \end{bmatrix}, b_1 \right) \right)$$

Let $x \in \mathbb{R}$. Upon instantiation with $\mathfrak{r}$ we have that:

$$(\mathfrak{I}_{\mathfrak{r}} (\nu \bullet \mathsf{Id}_1)) (x) = \begin{bmatrix} W_1 & -W_1 \end{bmatrix} \mathfrak{r} \left(\begin{bmatrix} x \\ -x \end{bmatrix} \right) + b_1$$

$$= W_1 \, \mathfrak{r}(x) - W_1 \, \mathfrak{r}(-x) + b_1$$

$$= W_1 \left(\mathfrak{r}(x) - \mathfrak{r}(-x) \right) + b_1$$

$$= W_x + b_1 = \mathfrak{I}_{\mathfrak{r}} (\nu) \tag{6.1.10}$$

Case 1.iv: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definitions 6.1.1 and 4.2.1 we have that:

$$\nu \bullet \mathsf{Id}_1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} W_1 & -W_1 \end{bmatrix}, b_1 \right), (W_2, b_2), ..., (W_L, b_L) \right) \tag{6.1.11}$$

Let $x \in \mathbb{R}$. Upon instantiation with $\mathfrak{r}$, we have that the first two layers are:

$$\begin{bmatrix} W_1 & -W_1 \end{bmatrix} \mathfrak{r} \left(\begin{bmatrix} x \\ -x \end{bmatrix} \right) + b_1$$

$$= W_1 \, \mathfrak{r}(x) - W_1 \, \mathfrak{r}(-x) + b_1$$

$$= W_1 \left(\mathfrak{r}(x) - \mathfrak{r}(-x) \right) + b_1$$

$$= W_1 x + b_1 = \mathfrak{I}_{\mathfrak{r}} (\nu) \tag{6.1.12}$$

This, along with Case 1. iii, implies that the uninstantiated first layer is equivalent (W_1, b_1) whence we have that $\nu \bullet \mathsf{Id}_1 = \nu$.

Observe that Definitions 4.3.1 and 6.1.1 tells us that:

181

$$\boxdot_{i=1}^{d} \mathsf{Id}_i \tag{6.1.13}$$

$$= \left(\left(\left[\overbrace{\begin{array}{ccc} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{array}}^{d-many} \right], \mathbb{0}_{2d} \right), \left(\left[\overbrace{\begin{array}{ccc} \text{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,2} \end{array}}^{d-many} \right], \mathbb{0}_{d} \right) \right)$$

Case 2.i Let $d \in \mathbb{N} \cap [1,\infty)$. Let $\nu \in \mathsf{NN}$ be $\nu = (W_1, b_1)$ with end-widths d. Deriving from Definitions 4.2.1 and 6.1.1 we have:

$$\mathsf{Id}_d \bullet \nu = \left(\left(\left[\begin{array}{ccc} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{array} \right] W_1, \left[\begin{array}{ccc} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{array} \right] b_1 \right), \right.$$

$$\left. \left(\left[\begin{array}{ccc} \text{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,2} \end{array} \right], \mathbb{0}_{d} \right) \right)$$

$$= \left(\left(\left[\begin{array}{c} [W_1]_{1,*} \\ -[W_1]_{1,*} \\ \vdots \\ [W_1]_{d,*} \\ -[W_1]_{d,*} \end{array} \right], \left[\begin{array}{c} [b_1]_1 \\ -[b_1]_1 \\ \vdots \\ [b_1]_d \\ -[b_1]_d \end{array} \right] \right), \left(\left[\begin{array}{ccc} \text{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,2} \end{array} \right], \mathbb{0}_{d} \right) \right)$$

Let $x \in \mathbb{R}^d$. Upon instantiation with $\mathfrak{r}$ we have that:

$$(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Id}_d \bullet \nu\right)) (x)$$

$$= \mathfrak{r}([W_1]_{1,*} \cdot x + [b_1]_1) - \mathfrak{r}(-[W_1]_{1,*} \cdot x - [b_1]_1) + \cdots$$

$$+ \mathfrak{r}([W_1]_{d,*} \cdot x + [b_1]_d) - \mathfrak{r}(-[W_1]_{d,*} \cdot x - [b_1]_d)$$

$$= [W_1]_{1,*} \cdot x + [b_1]_1 + \cdots + [W_1]_{d,*} \cdot x + [b_1]_d$$

$$= W_1 x + b_1 = \mathfrak{I}_{\mathfrak{r}} (\nu)$$

Case 2.ii: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definition 6.1.1 and 4.2.1 we have that:

$\mathsf{Id}_d \bullet \nu$

$$= \left((W_1, b_1), (W_2, b_2), ..., (W_{L-1}, b_{L-1}), \left(\begin{bmatrix} [W_L]_{1,*} \\ -[W_L]_{1,*} \\ \vdots \\ [W_L]_{d,*} \\ -[W_L]_{d,*} \end{bmatrix}, \begin{bmatrix} [b_L]_1 \\ -[b_L]_1 \\ \vdots \\ [b_L]_d \\ -[b_L]_d \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right)$$

Note that upon instantiation with $\mathfrak{r}$, the last two layers become:

$$\mathfrak{r}([W_L]_{1,*} \cdot x + [b_L]_1) - \mathfrak{r}(-[W_L]_{1,*} \cdot x - [b_L]_1) + \cdots$$

$$+ \mathfrak{r}([W_L]_{d,*} \cdot x + [b_L]_d) - \mathfrak{r}(-[W_L]_{d,*} \cdot x - [b_L]_d)$$

$$= [W_L]_{1,*} \cdot x + [b_L]_1 + \cdots + [W_L]_{d,*} \cdot x + [b_L]_d$$

$$= W_L x + b_L \tag{6.1.14}$$

This, along with Case 2.i implies that the uninstantiated last layer is equivalent to (W_L, b_L) whence $\mathsf{Id}_d \bullet \nu = \nu$.

Case 2.iii: Let $\nu = ((W_1, b_1))$. Deriving from Definition 6.1.1 and 4.2.1 we have:

$$\nu \bullet \mathsf{Id}_d =$$

$$\left(\left(\left[\begin{matrix} \mathrm{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \mathrm{Weight}_{\mathsf{Id}_1,1} \end{matrix} \right], \mathbb{0}_{2d} \right), \left(W_1 \left[\begin{matrix} \mathrm{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \mathrm{Weight}_{\mathsf{Id}_1,2} \end{matrix} \right], b_1 \right) \right)$$

Upon instantiation with $\mathfrak{r}$ we have that:

$$(\mathfrak{I}_{\mathfrak{r}}(\nu))(x) \tag{6.1.15}$$

$$= \left[[W_1]_{*,1} \ -[W_1]_{*,1} \ \cdots \ [W_1]_{*,d} \ -[W_1]_{*,d} \right] \mathfrak{r} \left(\left[\begin{matrix} [x]_1 \\ -[x]_1 \\ \vdots \\ [x]_d \\ -[x]_d \end{matrix} \right] \right) + b_1$$

$$= [W_1]_{*,1}\,\mathfrak{r}([x]_1) - [W_1]_{*,1}\,\mathfrak{r}(-[x]_1) + \cdots + [W_1]_{*,d}\,\mathfrak{r}([x]_d) - [W_1]_{*,d}\,\mathfrak{r}(-[x]_d) + b_1$$

$$= [W_1]_{*,1} \cdot [x]_1 + \cdots + [W_1]_{*,d} \cdot [x]_d$$

$$= W_1 x + b_1 = \mathfrak{I}_{\mathfrak{r}}(\nu) \tag{6.1.16}$$

Case 2.iv: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definitions 6.1.1 and 4.2.1 we have:

$$\nu \bullet \mathsf{Id}_d$$

$$= \left(\left(\left[\begin{matrix} \mathrm{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \mathrm{Weight}_{\mathsf{Id}_1,1} \end{matrix} \right], \mathbb{0}_{2d} \right), \left(\left[[W_1]_{*,1} \ -[W_1]_{*,1} \ \cdots \ [W_1]_{*,d} \ -[W_1]_{*,d} \right], b_1 \right), ... \right.$$

$$(W_2, b_2), ..., (W_L, b_L))$$

Upon instantiation with $\mathfrak{r}$, we have that the first two layers are:

$$(\mathfrak{I}_{\mathfrak{r}}(\nu))(x) \tag{6.1.17}$$

$$= \begin{bmatrix} [W_1]_{*,1} & -[W_1]_{*,1} & \cdots & [W_1]_{*,d} & -[W_1]_{*,d} \end{bmatrix} \mathfrak{r} \left(\begin{bmatrix} [x]_1 \\ -[x]_1 \\ \vdots \\ [x]_d \\ -[x]_d \end{bmatrix} \right) + b_1$$

$$= [W_1]_{*,1}\,\mathfrak{r}([x]_1) - [W_1]_{*,1}\,\mathfrak{r}(-[x]_1) + \cdots + [W_1]_{*,d}\,\mathfrak{r}([x]_d) - [W_1]_{*,d}\,\mathfrak{r}(-[x]_d) + b_1$$

$$= [W_1]_{*,1} \cdot [x]_1 + \cdots + [W_1]_{*,d} \cdot [x]_d$$

$$= W_1 x + b_1 \tag{6.1.18}$$

This, along with Case 2. iii, implies that the uninstantiated first layer is equivalent to (W_L, b_L) whence $\mathsf{Id}_d \bullet \nu = \nu$.

This completes the proof. $\qquad\square$

Definition 6.1.6 (Monoid). *Given a set X with binary operation $*$, we say that X is a monoid under the operation $*$ if:*

(i) For all $x, y \in X$ it is the case that $x * y \in X$

(ii) For all $x, y, z \in X$ it is the case that $(x * y) * z = x * (y * z)$

(iii) The exists a unique element $e \in X$ such that $e * x = x * e = x$

Theorem 6.1.7. *Let $d \in \mathbb{N}$. For a fixed d, the set of all neural networks $\nu \in \mathsf{NN}$ with instantiations in $\mathfrak{r}$ and end-widths d form a monoid under the operation of $\bullet$.*

Proof. This is a consequence of Lemma 6.1.5 and Lemma 4.2.4. $\qquad\square$

Remark 6.1.8. *By analogy with matrices, we may find it helpful to refer to neural networks of end-widths d as ``square neural networks of size d''.*

6.2 Trp^h, $\mathsf{Etr}^{n,h}$ and Neural Network Approximations For the Trapezoidal Rule.

Definition 6.2.1 (The Trp neural network). *Let $h \in [0, \infty)$. We define the $\mathsf{Trp}^h \in \mathsf{NN}$ neural network as:*

$$\mathsf{Trp}^h := \mathsf{Aff}_{\left[\frac{h}{2} \ \frac{h}{2}\right], 0} \tag{6.2.1}$$

Lemma 6.2.2. *Let $h \in (-\infty, \infty)$. It is then the case that:*

(i) for $x = \{x_1, x_2\} \in \mathbb{R}^2$ that $\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Trp}^h\right)\right)(x) \in C\left(\mathbb{R}^2, \mathbb{R}\right)$

(ii) for $x = \{x_1, x_2\} \in \mathbb{R}^2$ that $\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Trp}^h\right)\right)(x) = \frac{1}{2}h\left(x_1 + x_2\right)$

(iii) $\mathsf{D}\left(\mathsf{Trp}^h\right) = 1$

(iv) $\mathsf{P}\left(\mathsf{Trp}^h\right) = 3$

(v) $\mathsf{L}\left(\mathsf{Trp}^h\right) = (2, 1)$

Proof. This a straight-forward consequence of Lemma 4.4.1 $\qquad \square$

Definition 6.2.3 (The Etr neural network). *Let $n \in \mathbb{N}$ and $h \in [0, \infty)$. We define the neural network $\mathsf{Etr}^{n,h} \in \mathsf{NN}$ as:*

$$\mathsf{Etr}^{n,h} := \mathsf{Aff}_{\underbrace{\left[\frac{h}{2} \ h \ h \ \dots \ h \ \frac{h}{2}\right]}_{n+1-many}, 0} \tag{6.2.2}$$

Lemma 6.2.4. *Let $n \in \mathbb{N}$. Let $x_0 \in (-\infty, \infty)$, and $x_n \in [x_0, \infty)$. Let $x = [x_0 \ x_1 \ \dots \ x_n] \in \mathbb{R}^{n+1}$ and $h \in (-\infty, \infty)$ such that for all $i \in \{0, 1, \dots, n\}$ it is the case that $x_i = x_0 + i \cdot h$. Then:*

(i) for all $x \in \mathbb{R}^{n+1}$ it is the case that $\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Etr}^{n,h}\right)\right)(x) \in C\left(\mathbb{R}^{n+1}, \mathbb{R}\right)$

(ii) for all $n \in \mathbb{N}$, and $h \in (0, \infty)$ it is the case that $\left(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Etr}^{n,h} \right) \right)(x) = \frac{h}{2} \cdot x_0 + h \cdot x_1 + \cdots + h \cdot x_{n-1} + \frac{h}{2} \cdot x_n$

(iii) for all $n \in \mathbb{N}$, and $h \in (0, \infty)$ it is the case that $\mathsf{D}\left(\mathsf{Etr}^{n,h} \right) = 1$

(iv) for all $n \in \mathbb{N}$ and $h \in (0, \infty)$ it is the case that $\mathsf{P}\left(\mathsf{Etr}^{n,h} \right) = n + 2$

(v) for all $n \in \mathbb{N}$ and $h \in (0, \infty)$ it is the case that $\mathsf{L}\left(\mathsf{Etr}^{n,h} \right) = (n+1, 1)$

Proof. This a straightforward consequence of Lemma 4.4.1. $\qquad\square$

Remark 6.2.5. *Let $h \in (0, \infty)$. Note then that Trp^h is simply $\mathsf{Etr}^{2,h}$.*

Remark 6.2.6. *For an R implementation, see Listing 9.27*

6.3 Maximum Convolution Approximations for Multi--Dimensional Functions

We will present here an approximation scheme for continuous functions called maximum convolution approximation. This derives mainly from Chapter 4 of Jentzen et al. (2023), and our contribution is mainly to show parameter bounds, and convergence in the case of 1-D approximation.

6.3.1 The Nrm_1^d Neural Networks

Definition 6.3.1 (The Nrm_1^d neural network)**.** *We denote by $\left(\mathsf{Nrm}_1^d \right)_{d \in \mathbb{N}} \subseteq \mathsf{NN}$ the family of neural networks that satisfy:*

(i) for $d = 1$:

$$
\mathsf{Nrm}_1^1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \in \left(\mathbb{R}^{2\times1} \times \mathbb{R}^2 \right) \times \left(\mathbb{R}^{1\times2} \times \mathbb{R}^1 \right)
$$

$$(6.3.1)$$

(ii) for $d \in \{2, 3, ...\}$:

$$\mathsf{Nrm}_1^d = \mathsf{Sum}_{d,1} \bullet \left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1 \right] \tag{6.3.2}$$

Lemma 6.3.2. *Let $d \in \mathbb{N}$. It is then the case that:*

(i) $\mathsf{L}\left(\mathsf{Nrm}_1^d\right) = (d, 2d, 1)$

(ii) $\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Nrm}_1^d\right)\right)(x) \in C\left(\mathbb{R}^d, \mathbb{R}\right)$

(iii) that for all $x \in \mathbb{R}^d$ that $\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Nrm}_1^d\right)\right)(x) = \|x\|_1$

(iv) it holds $\mathsf{H}\left(\mathsf{Nrm}_1^d\right) = 1$

(v) it holds that $\mathsf{P}\left(\mathsf{Nrm}_1^d\right) \leqslant 7d^2$

(vi) it holds that $\mathsf{D}\left(\mathsf{Nrm}_1^d\right) = 2$

Proof. Note that by observation, it is the case that $\mathsf{L}\left(\mathsf{Nrm}_1^d\right) = (1, 2, 1)$. This and Remark 4.4.2 tells us that for all $d \in \{2, 3, ...\}$ it is the case that $\mathsf{L}\left(\boxminus_{i=1}^d \mathsf{Nrm}_1^d\right) = (d, 2d, d)$. This, Lemma 4.2.5, and Lemma 4.4.2 ensure that for all $d \in \{2, 3, 4, ...\}$ it is the case that $\mathsf{L}\left(\mathsf{Nrm}_1^d\right) = (d, 2d, 1)$, which in turn establishes Item (i).

Notice now that (6.3.1) ensures that:

$$\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Nrm}_1^d\right)\right)(x) = \mathfrak{r}(x) + \mathfrak{r}(-x) = \max\{x, 0\} + \max\{-x, 0\} = |x| = \|x\|_1 \tag{6.3.3}$$

This along with (Grohs et al., 2023, Proposition 2.19) tells us that for all $d \in \{2, 3, 4, ...\}$ and $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ it is the case that:

$$\left(\mathfrak{I}_\mathfrak{r}\left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right]\right)(x) = (|x_1|, |x_2|, ..., |x_d|) \tag{6.3.4}$$

This together with Lemma 4.2.3 tells us that:

$$\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Nrm}_1^d\right)\right) = \left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sum}_{d,1} \bullet \left[\boxminus_{i=1}^d \mathsf{Nrm}_1^d\right]\right)\right)(x)$$

$$= \left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sum}_{d,1}\right)\right)(|x_1|, |x_2|, ..., |x_d|) = \sum_{i=1}^d |x_i| = \|x\|_1 \qquad (6.3.5)$$

Note next that by observation $\mathsf{H}\left(\mathsf{Nrm}_1^1\right) = 1$. Remark 4.4.2 then tells us that since the number of layers remains unchanged under stacking, it is then the case that $\mathsf{H}\left(\mathsf{Nrm}_1^1\right) = \mathsf{H}\left(\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right) = 1$. Note next that Lemma 4.2.4 then tells us that $\mathsf{H}\left(\mathsf{Sum}_{d,1}\right) = 0$ whence Lemma 4.2.5 tells us that:

$$\mathsf{H}\left(\mathsf{Nrm}_1^d\right) = \mathsf{H}\left(\mathsf{Sum}_{d,1} \bullet \left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right]\right)$$

$$= \mathsf{H}\left(\mathsf{Sum}_{d,1}\right) + \mathsf{H}\left(\left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right]\right) = 0 + 1 = 1 \qquad (6.3.6)$$

Note next that:

$$\mathsf{Nrm}_1^1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix}\right)\right) \in \left(\mathbb{R}^{2\times 1} \times \mathbb{R}^2\right) \times \left(\mathbb{R}^{1\times 2} \times \mathbb{R}^1\right) \qquad (6.3.7)$$

and as such $\mathsf{P}\left(\mathsf{Nrm}_1^1\right) = 7$. This, combined with Cor 4.3.5.1, and the fact that we are stacking identical neural networks then tells us that:

$$\mathsf{P}\left(\left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right]\right) \leqslant 7d^2 \qquad (6.3.8)$$

Then Lemma Corollary 4.4.5.1, Lemma 4.5.6, and Lemma 4.2.5 tells us that:

$$\mathsf{P}\left(\mathsf{Nrm}_1^d\right) = \mathsf{P}\left(\mathsf{Sum}_{d,1} \bullet \left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right]\right)$$

$$\leqslant \mathsf{P}\left(\left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right]\right) \leqslant 7d^2 \qquad (6.3.9)$$

189

This establishes Item (v).

Finally, by observation $\mathsf{D}\left(\mathsf{Nrm}_1^1\right) = 2$, we are stacking the same neural network when we have Nrm_1^d. Stacking has no effect on depth from Definition 4.3.1, and by Lemma 4.2.5, $\mathsf{D}\left(\mathsf{Sum}_{d,1} \bullet \left[\boxminus_{i=1}^d \mathsf{Nrm}_1^1\right]\right) = \mathsf{D}\left(\boxminus \mathsf{Nrm}_1^1\right)$. Thus we may conclude that $\mathsf{D}\left(\mathsf{Nrm}_1^d\right) = \mathsf{D}\left(\mathsf{Nrm}_1^1\right) = 2$.

This concludes the proof of the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Remark 6.3.3. *For an R implementation, see Listing 9.22*

6.3.2 The Mxm^d Neural Networks

Given $x \in \mathbb{R}$, it is straightforward to find the maximum; x is the maximum. For $x \in \mathbb{R}^d$ we may find the maximum via network (6.3.11.1), i.e. Mxm^2. The strategy is to find maxima for half our entries and half repeatedly until we have one maximum. For $x \in \mathbb{R}^d$ where d is even we may stack d copies of Mxm^2 to halve, and for $x \in \mathbb{R}^d$ where d is odd and greater than 3 we may introduce ``padding'' via the Id_1 network and thus require $\frac{d-1}{2}$ copies of Mxm^2 to halve.

Definition 6.3.4 (Maxima ANN representations)**.** *Let* $\left(\mathsf{Mxm}^d\right)_{d \in \mathbb{N}} \subseteq \mathsf{NN}$ *represent the neural networks that satisfy:*

(i) for all $d \in \mathbb{N}$ that $\mathsf{I}\left(\mathsf{Mxm}^d\right) = d$

(ii) for all $d \in \mathbb{N}$ that $\mathsf{O}\left(\mathsf{Mxm}^d\right) = 1$

(iii) that $\mathsf{Mxm}^1 = \mathsf{Aff}_{1,0} \in \mathbb{R}^{1\times 1} \times \mathbb{R}^1$

(iv) that:

$$\mathsf{Mxm}^2 = \left(\left(\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right), \left(\begin{bmatrix} 1 & 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix}\right)\right) \qquad (6.3.10)$$

(v) it holds for all $d \in \{2, 3, ...\}$ that $\mathsf{Mxm}^{2d} = \mathsf{Mxm}^d \bullet \left[\boxminus_{i=1}^d \mathsf{Mxm}^2 \right]$, and

(vi) it holds for all $d \in \{2, 3, ...\}$ that $\mathsf{Mxm}^{2d-1} = \mathsf{Mxm}^d \bullet \left[\left(\boxminus_{i=1}^d \mathsf{Mxm}^2 \right) \boxminus \mathsf{Id}_1 \right]$.

Remark 6.3.5. *Diagrammatically, this can be represented as:*
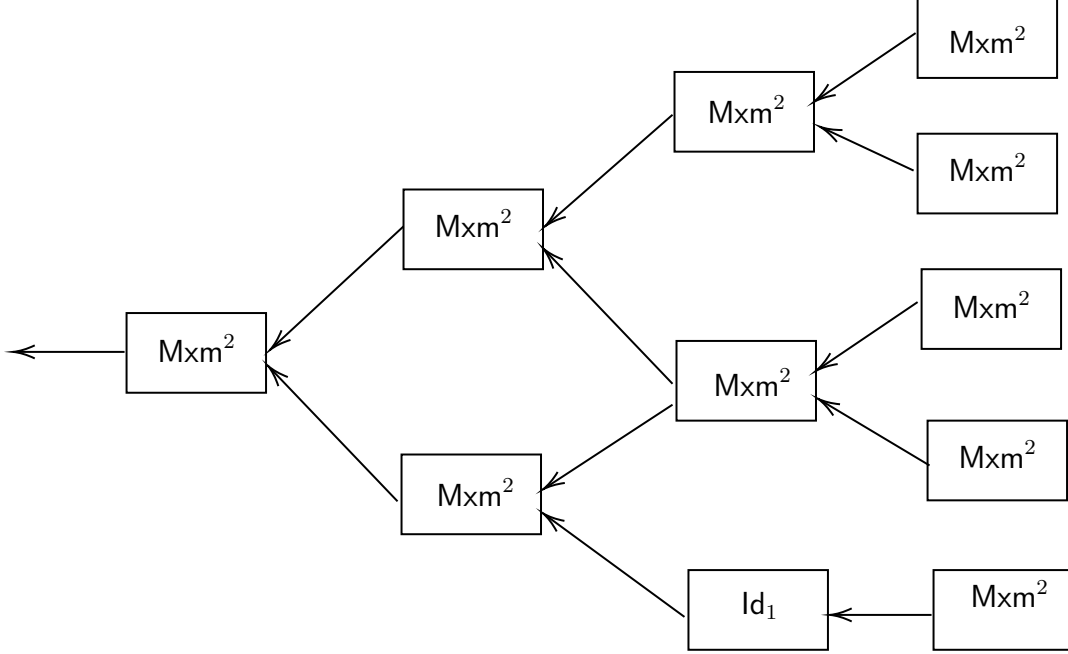
Figure 6.1: Neural network diagram for Mxm^5.

Lemma 6.3.6. *Let $d \in \mathbb{N}$, it is then the case that:*

(i) $\mathsf{H}\left(\mathsf{Mxm}^d\right) = \lceil \log_2(d) \rceil$

(ii) *for all $i \in \mathbb{N}$ that* $\mathsf{W}_i\left(\mathsf{Mxm}^d\right) \leqslant 3 \left\lceil \frac{d}{2^i} \right\rceil$

(iii) $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Mxm}^d\right) \in C\left(\mathbb{R}^d, \mathbb{R}\right)$ *and*

(iv) *for all $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ we have that* $\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Mxm}^d\right)\right)(x) = \max\{x_1, x_2, ..., x_d\}$.

(v) $\mathsf{P}\left(\mathsf{Mxm}^d\right) \leqslant \left(\frac{4}{3}d^2 + 3d\right)\left(1 + \frac{1}{2}^{\lceil \log_2(d) \rceil + 1}\right)$

(vi) $\mathsf{D}\left(\mathsf{Mxm}^d\right) = \lceil \log_2(d) \rceil + 1$

Proof. Assume w.l.o.g. that $d > 1$. Note that (6.3.10) ensures that $\mathsf{H}\left(\mathsf{Mxm}^d\right) = 1$. This and (4.3.1) then tell us that for all $d \in \{2, 3, 4, ...\}$ it is the case that:

$$\mathsf{H}\left(\boxminus_{i=1}^d \mathsf{Mxm}^2\right) = \mathsf{H}\left(\left[\boxminus_{i=1}^d \mathsf{Mxm}^2\right] \boxminus \mathsf{Id}_1\right) = \mathsf{H}\left(\mathsf{Mxm}^2\right) = 1$$

This and Lemma 4.2.5 tells us that for all $d \in \{3, 4, 5, ...\}$ it holds that:

$$\mathsf{H}\left(\mathsf{Mxm}^d\right) = \mathsf{H}\left(\mathsf{Mxm}^{\lceil \frac{d}{2} \rceil}\right) + 1 \tag{6.3.11}$$

And for $d \in \{4, 6, 8, ...\}$ with $\mathsf{H}\left(\mathsf{Mxm}^{\lceil \frac{d}{2} \rceil}\right) = \left\lceil \log_2\left(\frac{d}{2}\right)\right\rceil$ it holds that:

$$\mathsf{H}\left(\mathsf{Mxm}^d\right) = \left\lceil \log_2\left(\frac{d}{2}\right)\right\rceil + 1 = \lceil \log_2(d) - 1 \rceil + 1 = \lceil \log_2(d)\rceil \tag{6.3.12}$$

Moreover (6.3.11) and the fact that for all $d \in \{3, 5, 7, ...\}$ it holds that $\lceil \log_2(d+1)\rceil = \lceil \log_2(d)\rceil$ ensures that for all $d \in \{3, 5, 7, ...\}$ with $\mathsf{H}\left(\mathsf{Mxm}^{\lceil \frac{d}{2} \rceil}\right) = \left\lceil \log_2\left(\lceil \frac{d}{2}\rceil\right)\right\rceil$ it holds that:

$$\mathsf{H}\left(\mathsf{Mxm}^d\right) = \left\lceil \log_2\left(\left\lceil \frac{d}{2}\right\rceil\right)\right\rceil + 1 = \left\lceil \log_2\left(\left\lceil \frac{d+1}{2}\right\rceil\right)\right\rceil + 1$$
$$= \lceil \log_2(d+1) - 1\rceil + 1 = \lceil \log_2(d+1)\rceil = \lceil \log_2(d)\rceil \tag{6.3.13}$$

This and (6.3.12) demonstrate that for all $d \in \{3, 4, 5, ...\}$ with $\forall k \in \{2, 3, ..., d-1\}$: $\mathsf{H}\left(\mathsf{Mxm}^d\right) = \lceil \log_2(k)\rceil$ it holds htat $\mathsf{H}\left(\mathsf{Mxm}^d\right) = \lceil \log_2(d)\rceil$. The fact that $\mathsf{H}\left(\mathsf{Mxm}^2\right) = 1$ and induction establish Item (i).

We next note that $\mathsf{L}\left(\mathsf{Mxm}^2\right) = (2, 3, 1)$. This then indicates that for all $i \in \mathbb{N}$ that:

$$\mathsf{W}_i\left(\mathsf{Mxm}^2\right) \leqslant 3 = 3\left\lceil \frac{2}{2^i}\right\rceil. \tag{6.3.14}$$

Note then that Lemma 4.2.5 and Remark 4.4.2 tells us that:

$$W_i \left(\mathsf{Mxm}^{2d} \right) = \begin{cases} 3d & : i = 1 \\ W_{i-1} \left(\mathsf{Mxm}^d \right) & : i \geqslant 2 \end{cases} \tag{6.3.15}$$

And:

$$W_i \left(\mathsf{Mxm}^{2d-1} \right) = \begin{cases} 3d - 1 & : i = 1 \\ W_{i-1} \left(\mathsf{Mxm}^d \right) & : i \geqslant 2 \end{cases} \tag{6.3.16}$$

This in turn assures us that for all $d \in \{2, 4, 6, ...,\}$ it holds that:

$$W_1 \left(\mathsf{Mxm}^d \right) = 3 \left(\frac{d}{2} \right) \leqslant 3 \left\lceil \frac{d}{2} \right\rceil \tag{6.3.17}$$

Moreover, note that (6.3.16) tells us that for all $d \in \{3, 5, 7, ...\}$ it holds that:

$$W_1 \left(\mathsf{Mxm}^d \right) = 3 \left\lceil \frac{d}{2} \right\rceil - 1 \leqslant 3 \left\lceil \frac{d}{2} \right\rceil \tag{6.3.18}$$

This and (6.3.17) shows that for all $d \in \{2, 3, ...\}$ it holds that:

$$W_1 \left(\mathsf{Mxm}^d \right) \leqslant 3 \left\lceil \frac{d}{2} \right\rceil \tag{6.3.19}$$

Additionally note that (6.3.15) demonstrates that for all $d \in \{4, 6, 8, ...\}$, $i \in \{2, 3, ...\}$ with $W_{i-1} \left(\mathsf{Mxm}^{\frac{d}{2}} \right) \leqslant 3 \left\lceil \left(\frac{d}{2} \right) \frac{1}{2^{i-1}} \right\rceil$ it holds that:

$$W_i \left(\mathsf{Mxm}^d \right) = W_{i-1} \left(\mathsf{Mxm}^{\frac{d}{2}} \right) \leqslant 3 \left\lceil \left(\frac{d}{2} \right) \frac{1}{2^{i-1}} \right\rceil = 3 \left\lceil \frac{d}{2^i} \right\rceil \tag{6.3.20}$$

Furthermore note also the fact that for all $d \in \{3, 5, 7, ...\}$, $i \in \mathbb{N}$ it holds that $\left\lceil \frac{d+1}{2^i} \right\rceil = \left\lceil \frac{d}{2^i} \right\rceil$ and (6.3.16) assure that for all $d \in \{3, 5, 7, ...\}$, $i \in \{2, 3, ...\}$ with $W_{i-1} \left(\mathsf{Mxm}^{\left\lceil \frac{d}{2} \right\rceil} \right) \leqslant$

193

$3 \left\lceil \left\lceil \frac{d}{2} \right\rceil \frac{1}{2^{i-1}} \right\rceil$ it holds that:

$$\mathsf{W}_i \left(\mathsf{Mxm}^d \right) = \mathsf{W}_{i-1} \left(\mathsf{Mxm}^{\left\lceil \frac{d}{2} \right\rceil} \right) \leqslant 3 \left\lceil \left\lceil \frac{d}{2} \right\rceil \frac{1}{2^{i-1}} \right\rceil = 3 \left\lceil \frac{d+1}{2^i} \right\rceil = 3 \left\lceil \frac{d}{2^i} \right\rceil \tag{6.3.21}$$

This and (6.3.20) tells us that for all $d \in \{3, 4, ...\}$, $i \in \{2, 3, ...\}$ with $\forall k \in \{2, 3, ..., d-1\}$, $j \in \{1, 2, ..., i-1\} : \mathsf{W}_j \left(\mathsf{Mxm}^k \right) \leqslant 3 \left\lceil \frac{k}{2^j} \right\rceil$ it holds that:

$$\mathsf{W}_i \left(\mathsf{Mxm}^d \right) \leqslant 3 \left\lceil \frac{d}{2^i} \right\rceil \tag{6.3.22}$$

This, combined with (6.3.14), (6.3.19), with induction establishes Item (ii).

Next observe that (6.3.10) tells that for $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ it becomes the case that:

$$\left(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Mxm}^2 \right) \right)(x) = \max\{x_1 - x_2, 0\} + \max\{x_2, 0\} - \max\{-x_2, 0\}$$
$$= \max\{x_1 - x_2, 0\} + x_2 = \max\{x_1, x_2\} \tag{6.3.23}$$

Note next that Lemma 6.1.2, Lemma 4.2.5, and (Grohs et al., 2023, Proposition 2.19) then imply for all $d \in \{2, 3, 4, ...\}$, $x = \{x_1, x_2, ..., x_d\} \in \mathbb{R}^d$ it holds that $\left(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Mxm}^d \right) \right)(x) \in C \left(\mathbb{R}^d, \mathbb{R} \right)$. and $\left(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{Mxm}^d \right) \right)(x) = \max\{x_1, x_2, ..., x_d\}$. This establishes Items (iii)—(iv).

Consider now the fact that Item (ii) implies that the layer architecture forms a geometric series whence we have that the number of bias parameters is bounded by:

$$\frac{\frac{3d}{2} \left(1 - \left(\frac{1}{2} \right)^{\lceil \log_2(d) \rceil + 1} \right)}{\frac{1}{2}} = 3d \left(1 - \frac{1}{2}^{\lceil \log_2(d) \rceil + 1} \right)$$
$$\leqslant \left\lceil 3d \left(1 - \frac{1}{2}^{\lceil \log_2(d) \rceil + 1} \right) \right\rceil \tag{6.3.24}$$

For the weight parameters, consider the fact that our widths follow a geometric series with ratio $\frac{1}{2}$, and considering that we have an upper bound for the number of hidden layers, and the fact that $\mathsf{W}_0 \left(\mathsf{Mxm}^d \right) = d$, would then tell us that the number of weight parameters is

bounded by:

$$\sum_{i=0}^{\lceil \log_2(d) \rceil} \left[\left(\frac{1}{2} \right)^i \cdot \mathsf{W}_0 \left(\mathsf{Mxm}^d \right) \cdot \left(\frac{1}{2} \right)^{i+1} \cdot \mathsf{W}_0 \left(\mathsf{Mxm}^d \right) \right]$$

$$= \sum_{i=0}^{\lceil \log_2(d) \rceil} \left[\left(\frac{1}{2} \right)^{2i+1} \left(\mathsf{W}_0 \left(\mathsf{Mxm}^d \right) \right)^2 \right]$$

$$= \frac{1}{2} \sum_{i=0}^{\lceil \log_2(d) \rceil} \left[\left(\left(\frac{1}{2} \right)^i \mathsf{W}_0 \left(\mathsf{Mxm}^d \right) \right)^2 \right] = \frac{1}{2} \sum_{i=0}^{\lceil \log_2(d) \rceil} \left[\left(\frac{1}{4} \right)^i d^2 \right] \qquad (6.3.25)$$

Notice that this is a geometric series with ratio $\frac{1}{4}$, which would then reveal that:

$$\frac{1}{2} \sum_{i=0}^{\lceil \log_2(d) \rceil} \left[\left(\frac{1}{4} \right)^i d^2 \right] \leqslant \frac{2}{3} d^2 \left(1 - \frac{1}{2}^{2(\lceil \log_2(d) \rceil + 1)} \right) \qquad (6.3.26)$$

Thus, we get that:

$$\mathsf{P} \left(\mathsf{Mxm}^d \right) \leqslant \frac{2}{3} d^2 \left(1 - \frac{1}{2}^{2(\lceil \log_2(d) \rceil)+1} \right) + \left\lceil 3d \left(1 - \frac{1}{2}^{\lceil \log_2(d) \rceil + 1} \right) \right\rceil$$

$$\leqslant \frac{2}{3} d^2 \left(1 - \frac{1}{2}^{2(\lceil \log_2(d) \rceil)+1} \right) + \left\lceil 3d \left(1 - \frac{1}{2}^{2(\lceil \log_2(d) \rceil + 1)} \right) \right\rceil \qquad (6.3.27)$$

$$\leqslant \left\lceil \left(\frac{2}{3} d^2 + 3d \right) \left(1 + \frac{1}{2}^{2(\lceil \log_2(d) \rceil + 1)} \right) + 1 \right\rceil \qquad (6.3.28)$$

This proves Item (v).

Item (vi) is a straightforward consequence of Item (i). This completes the proof of the lemma. $\qquad\square$

Remark 6.3.7. *For an R implementation, see Listing 9.26*

6.3.3 The $\mathsf{MC}_{x,y}^{N,d}$ Neural Networks

Let $f : [a,b] \to \mathbb{R}$ be a continuous bounded function with Lipschitz constant L. Let $x_0 \leqslant x_1 \leqslant \cdots \leqslant x_N$ be a set of sample points within $[a,b]$, with it being possibly the case that for all $i \in \{0, 1, \ldots, N\}$, $x_i \sim \mathrm{Unif}([a,b])$. For all $i \in \{0, 1, \ldots, N\}$, define a series of

functions $f_0, f_1, \ldots f_N : [a, b] \to \mathbb{R}$, as such:

$$f_i = f(x_i) - L \cdot |x - x_i| \tag{6.3.29}$$

We will call the approximant $\max_{i \in \{0,1,\ldots,N\}} \{f_i\}$, the *maximum convolution approximation*. This converges to f, as shown in

Lemma 6.3.8. *Let* $d, N \in \mathbb{N}$, $L \in [0, \infty)$, $x_1, x_2, \ldots, x_N \in \mathbb{R}^d$, $y = (y_1, y_2, \ldots, y_N) \in \mathbb{R}^N$ *and* $\mathsf{MC} \in \mathsf{NN}$ *satisfy that:*

$$\mathsf{MC}_{x,y}^{N,d} = \mathsf{Mxm}^N \bullet \mathsf{Aff}_{-L\mathbb{I}_N, y} \bullet \left(\boxminus_{i=1}^N \left[\mathsf{Nrm}_1^d \bullet \mathsf{Aff}_{\mathbb{I}_d, -x_i} \right] \right) \bullet \mathsf{Cpy}_{N,d} \tag{6.3.30}$$

It is then the case that:

(i) $\mathsf{I}\left(\mathsf{MC}_{x,y}^{N,d}\right) = d$

(ii) $\mathsf{O}\left(\mathsf{MC}_{x,y}^{N,d}\right) = 1$

(iii) $\mathsf{H}\left(\mathsf{MC}_{x,y}^{N,d}\right) = \lceil \log_2(N) \rceil + 1$

(iv) $\mathsf{W}_1\left(\mathsf{MC}_{x,y}^{N,d}\right) = 2dN$

(v) for all $i \in \{2, 3, \ldots\}$ *we have* $\mathsf{W}_i\left(\mathsf{MC}_{x,y}^{N,d}\right) \leqslant 3 \left\lceil \frac{N}{2^{i-1}} \right\rceil$

(vi) it holds for all $x \in \mathbb{R}^d$ *that* $\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{MC}_{x,y}^{N,d}\right)\right)(x) = \max_{i \in \{1,2,\ldots,N\}}\left(y_i - L\|x - x_i\|_1\right)$

(vii) it holds that $\mathsf{P}\left(\mathsf{MC}_{x,y}^{N,d}\right) \leqslant \left\lceil \left(\frac{2}{3}d^2 + 3d\right)\left(1 + \frac{1}{2}2^{(\lceil \log_2(d) \rceil + 1)}\right) + 1 \right\rceil + 7N^2d^2 + 3\left\lceil \frac{N}{2} \right\rceil \cdot 2dN$

Proof. Throughout this proof let $\mathsf{S}_i \in \mathsf{NN}$ satisfy for all $i \in \{1, 2, \ldots, N\}$ that $\mathsf{S}_i = \mathsf{Nrm}_1^d \bullet \mathsf{Aff}_{\mathbb{I}_d, -x_i}$ and let $\mathsf{X} \in \mathsf{NN}$ satisfy:

$$\mathsf{X} = \mathsf{Aff}_{-L\mathbb{I}_N, y} \bullet \left(\left[\boxminus_{i=1}^N \mathsf{S}_i \right] \right) \bullet \mathsf{Cpy}_{N,d} \tag{6.3.31}$$

Note that (6.3.30) and Lemma 4.2.5 tells us that $O(\mathbb{R}) = O\left(\mathsf{Mxm}^N\right) = 1$ and $\mathsf{I}\left(\mathsf{MC}_{x,y}^{N,d}\right) = \mathsf{I}\left(\mathsf{Cpy}_{N,d}\right) = d$. This proves Items (i)--(ii). Next observe that since it is the case that $\mathsf{H}\left(\mathsf{Cpy}_{N,d}\right)$ and $\mathsf{H}\left(\mathsf{Nrm}_1^d\right) = 1$, Lemma 4.2.5 then tells us that:

$$\mathsf{H}(\mathsf{X}) = \mathsf{H}\left(\mathsf{Aff}_{-L\mathbb{I}_N,y}\right) + \mathsf{H}\left(\boxminus_{i=1}^N \mathsf{S}_i\right) + \mathsf{H}\left(\mathsf{Cpy}_{N,d}\right) = 1 \tag{6.3.32}$$

Thus Lemma 4.2.5 and Lemma 6.3.6 then tell us that:

$$\mathsf{H}(\mathsf{MC}) = \mathsf{H}\left(\mathsf{Mxm}^N \bullet \mathsf{X}\right) = \mathsf{H}\left(\mathsf{Mxm}^N\right) + \mathsf{H}(\mathsf{X}) = \lceil \log_2(N)\rceil + 1 \tag{6.3.33}$$

Which in turn establishes Item (iii).

Note next that Lemma 4.2.5 and (Grohs et al., 2023, Proposition 2.20) tells us that:

$$\mathsf{W}_1\left(\mathsf{MC}_{x,y}^{N,d}\right) = \mathsf{W}_1(\mathsf{X}) = \mathsf{W}_1\left(\boxminus_{i=1}^N \mathsf{S}_i\right) = \sum_{i=1}^N \mathsf{W}_1(\mathsf{S}_i) = \sum_{i=1}^N \mathsf{W}_1\left(\mathsf{Nrm}_1^d\right) = 2dN \tag{6.3.34}$$

This establishes Item (iv).

Next observe that the fact that $\mathsf{H}(\mathsf{X}) = 1$, Lemma 4.2.5 and Lemma 6.3.6 tells us that for all $i \in \{2,3,...\}$ it is the case that:

$$\mathsf{W}_i\left(\mathsf{MC}_{x,y}^{N,d}\right) = \mathsf{W}_{i-1}\left(\mathsf{Mxm}^N\right) \leqslant 3\left\lceil \frac{N}{2^{i-1}}\right\rceil \tag{6.3.35}$$

This establishes Item (v).

Next observe that Lemma 6.3.2 and Lemma 4.4.5 tells us that for all $x \in \mathbb{R}^d$, $i \in \{1,2,...,N\}$ it holds that:

$$\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{MC}_{x,y}^{N,d}\right)\right)(x) - \left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Nrm}_1^d\right) \circ \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Aff}_{\mathbb{I}_d,-x_i}\right)\right)(x) = \|x - x_i\|_1 \tag{6.3.36}$$

This an (Grohs et al., 2023, Proposition 2.20) combined establishes that for all $x \in \mathbb{R}^d$ it

holds that:

$$\left(\mathfrak{I}_\mathfrak{r}\left(\left[\boxminus_{i=1}^N \mathsf{S}_i\right]\bullet\mathsf{Cpy}_{N,d}\right)\right)(x)=\left(\|x-x_1\|_1,\|x-x_2\|_1,...,\|x-x_N\|_1\right)$$

$$(6.3.37)$$

This and Lemma 4.4.5 establishes that for all $x\in\mathbb{R}^d$ it holds that:

$$\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{X}\right)\right)(x)=\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Aff}_{-L\mathbb{I}_N,y}\right)\right)\circ\left(\mathfrak{I}_\mathfrak{r}\left(\left[\boxminus_{i=1}^N \mathsf{S}_i\right]\bullet\mathsf{Cpy}_{N,d}\right)\right)(x)$$

$$=\left(y_1-L\|x-x_1\|,y_2-L\|x-x_2\|,...,y_N-L\|x-x_N\|_1\right)\qquad(6.3.38)$$

Then Lemma 4.2.5 and Lemma 6.3.6 tells us that for all $x\in\mathbb{R}^d$ it holds that:

$$\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{MC}_{x,y}^{N,d}\right)\right)(x)=\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Mxm}^N\right)\circ\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{X}\right)\right)\right)(x)$$

$$=\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{Mxm}^N\right)\right)\left(y_1-L\|x-x_1\|_1,y_2-L\|x-x_2\|_1,...,y_N-L\|x-x_N\|_1\right)$$

$$=\max_{i\in\{1,2,...,N\}}\left(y_i-L\|x-x_i\|_1\right)\qquad(6.3.39)$$

This establishes Item (vi).

For Item (vii) note that Lemma 6.3.2, Remark 4.3.4, Lemma 6.3.2, and Corollary 4.4.5.1 tells us that for all $d\in\mathbb{N}$ and $x\in\mathbb{R}^d$ it is the case that:

$$\mathsf{P}\left(\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right)\leqslant\mathsf{P}\left(\mathsf{Nrm}_1^d\right)\leqslant 7d^2\qquad(6.3.40)$$

This, along with Corollary 4.3.5.1, and because we are stacking identical neural networks, then tells us that for all $N\in\mathbb{N}$, it is the case that:

$$\mathsf{P}\left(\boxminus_{i=1}^N\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)\leqslant 7N^2d^2\qquad(6.3.41)$$

Observe next that Corollary 4.4.5.1 tells us that for all $d,N\in\mathbb{N}$ and $x\in\mathbb{R}^d$ it is the case

that:

$$\mathsf{P}\left(\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)\bullet\mathsf{Cpy}_{N,d}\right)\leqslant\mathsf{P}\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)\leqslant 7N^2d^2 \qquad (6.3.42)$$

Now, let $d, N \in \mathbb{N}$, $L \in [0, \infty)$, let $x_1, x_2, \ldots, x_N \in \mathbb{R}^d$ and let $y = \{y_1, y_2, \ldots, y_N\} \in \mathbb{R}^N$. Observe that again, Corollary 4.4.5.1, and (6.3.42) tells us that:

$$\mathsf{P}\left(\mathsf{Aff}_{-L\mathbb{I}_N,y}\bullet\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x_i}\right]\right)\bullet\mathsf{Cpy}_{N,d}\right)$$
$$\leqslant\mathsf{P}\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)\leqslant 7N^2d^2$$

Finally Lemma 4.2.5, (6.3.34), and Lemma 6.3.6 yields that:

$$\mathsf{P}(\mathsf{MC}_{x,y}^{N,d})=\mathsf{P}\left(\mathsf{Mxm}^N\bullet\mathsf{Aff}_{-L\mathbb{I}_N,y}\bullet\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x_i}\right]\right)\bullet\mathsf{Cpy}_{N,d}\right)$$
$$\leqslant\mathsf{P}\left(\mathsf{Mxm}^N\bullet\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)\right)$$
$$\leqslant\mathsf{P}\left(\mathsf{Mxm}^N\right)+\mathsf{P}\left(\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)\right)+$$
$$\mathsf{W}_1\left(\mathsf{Mxm}^N\right)\cdot\mathsf{W}_{\mathsf{H}\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)}\left(\boxminus_{i=1}^{N}\left[\mathsf{Nrm}_1^d\bullet\mathsf{Aff}_{\mathbb{I}_d,-x}\right]\right)$$
$$\leqslant\left\lceil\left(\frac{2}{3}d^2+3d\right)\left(1+\frac{1}{2}^{2(\lceil\log_2(d)\rceil+1)}\right)+1\right\rceil+7N^2d^2+3\left\lceil\frac{N}{2}\right\rceil\cdot 2dN \qquad (6.3.43)$$

$\square$

Remark 6.3.9. *We may represent the neural network diagram for* Mxm^d *below.*

Remark 6.3.10. *For an R implementation, see Listing 9.28.*

6.3.4 Lipschitz Function Approximations

Lemma 6.3.11. *Let* (E, d) *be a metric space. Let* $L \in [0, \infty)$, $D \subseteq E$, $\varnothing \neq C \subseteq D$. *Let* $f : D \to \mathbb{R}$ *satisfy for all* $x \in D$, $y \in C$ *that* $|f(x) - f(y)| \leqslant Ld(x, y)$, *and let*
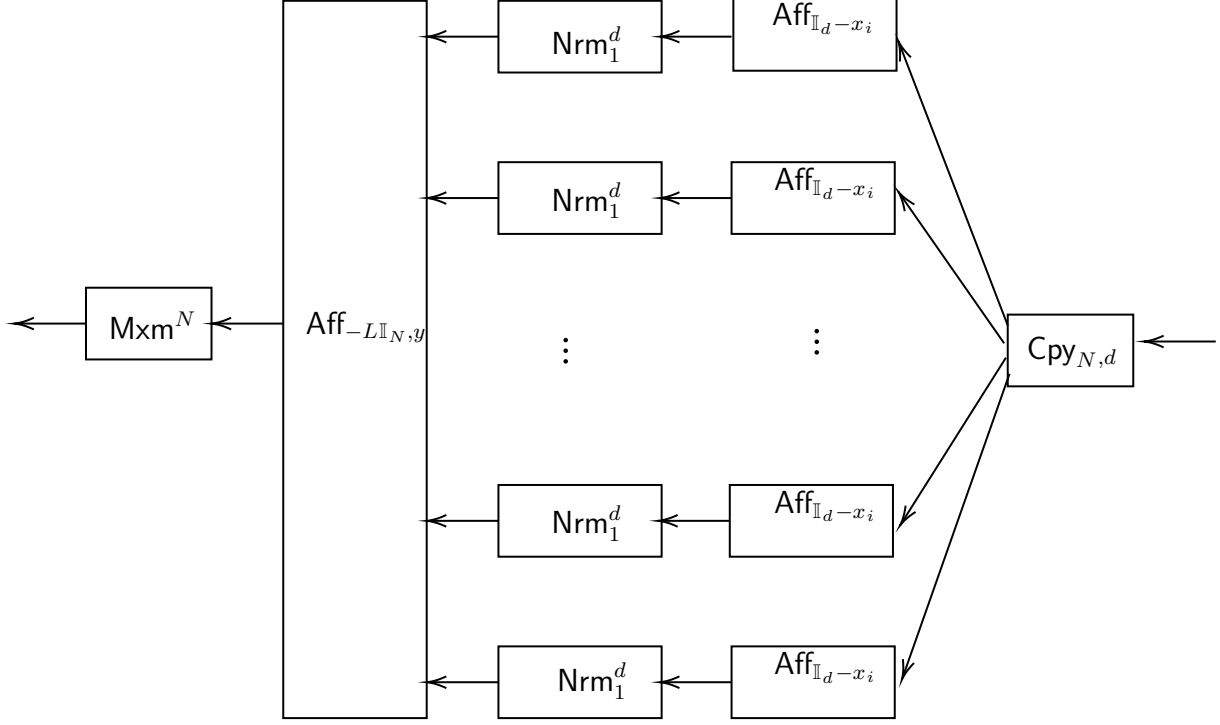
Figure 6.2: Neural network diagram for the $\mathsf{MC}^{N,d}_{x,y}$ network

$F : E \to \mathbb{R} \cup \{\infty\}$ *satisfy for all* $x \in E$ *that:*

$$F(x) = \sup_{y \in C} \left[f(y) - L d(x,y) \right] \qquad (6.3.44)$$

It is then the case that:

   (i) for all $x \in C$ *that* $F(x) = f(x)$

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

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

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

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

$$|F(x) - f(x)| \leqslant 2L \left[\inf_{y \in C} d(x,y) \right] \qquad (6.3.45)$$

200

Proof. The assumption that $\forall x \in D, y \in C : |f(x) - f(y)| \leqslant Ld(x,y)$ ensures that:

$$f(y) - Ld(x,y) \leqslant f(x) \leqslant f(y) + Ld(x,y) \tag{6.3.46}$$

For $x \in D$, it then renders as:

$$f(x) \geqslant \sup_{y \in C} [f(y) - Ld(x,y)] \tag{6.3.47}$$

This establishes Item (i). Note that (6.3.45) then tells us that for all $x \in C$ it holds that:

$$F(x) \geqslant f(x) - Ld(x,y) = f(x) \tag{6.3.48}$$

This with (6.3.47) then yields Item (i).

Note next that (6.3.46, with $x \frown y$ and $y \frown z$) and the triangle inequality ensure that for all $x \in E$, $y, z \in C$ it holds that:

$$f(y) - Ld(x,y) \leqslant f(z) + Ld(y,z) - Ld(x,y) \leqslant f(z) + Ld(x,z) \tag{6.3.49}$$

We then obtain for all $x \in E, z \in C$ it holds that:

$$F(x) = \sup_{y \in C} [f(y) - Ld(x,y)] \leqslant f(x) + Ld(x,z) < \infty \tag{6.3.50}$$

This proves Item (iii). Item (iii), (6.3.44), and the triangle inequality then shows that for all

$x, y \in E$, it holds that:

$$F(x) - F(y) = \left[\sup_{v \in C} \left(f(v) - Ld(x, v) \right) \right] - \left[\sup_{w \in C} \left(f(w) - Ld(y, w) \right) \right]$$

$$= \sup_{v \in C} \left[f(v) - Ld(x, v) - \sup_{w \in C} \left(f(w) - Ld(y, w) \right) \right]$$

$$\leqslant \sup_{v \in C} \left[f(v) - Ld(x, v) - \left(f(v) - Ld(y, w) \right) \right]$$

$$= \sup_{v \in C} \left(Ld(y, v) + Ld(x, v) - Ld(x, v) \right) = Ld(x, y) \qquad (6.3.51)$$

This establishes Item (v). Finally, note that Items (i) and (iv), the triangle inequality, and the assumption that $\forall x \in D, y \in C : |f(x) - f(y)| \leqslant Ld(x, y)$ ensure that for all $x \in D$ it holds that:

$$|F(x) - f(x)| = \inf_{y \in C} |F(x) - F(y) + f(y) - f(x)|$$

$$\leqslant \inf_{y \in C} \left(|F(x) - F(y)| + |f(y) - f(x)| \right)$$

$$\leqslant \inf_{y \in C} \left(2Ld(x, y) \right) = 2L \left[\inf_{y \in C} d(x, y) \right] \qquad (6.3.52)$$

This establishes Item (v) and hence establishes the Lemma. $\qquad\square$

Corollary 6.3.11.1. *Let (E, d) be a metric space, let $L \in [0, \infty)$, $\varnothing \neq C \subseteq E$, let $f : E \to \mathbb{R}$ satisfy for all $x \in E$, $y \in C$ that $\|f(x) - f(y)\| \leqslant Ld(x, y)$, and let $F : E \to \mathbb{R} \cup \{\infty\}$ satisfy for all $x \in E$ that:*

$$F(x) = \sup_{y \in C} \left[f(y) - Ld(x, y) \right] \qquad (6.3.53)$$

It is then the case that:

(i) for all $x \in C$ that $F(x) = f(x)$

(ii) for all $x \in E$ that $F(x) \leqslant f(x)$

(iii) for all $x, y \in E$ that $|F(x) - f(y)| \leqslant Ld(x, y)$ and

(iv) for all $x \in E$ that:

$$|F(x) - f(x)| \leqslant 2L \left[\inf_{y \in C} d(x, y) \right] \tag{6.3.54}$$

Proof. Note that Lemma 6.3.4 establishes Items (i)—(iv). $\qquad\square$

6.3.5 Explicit ANN Approximations

Lemma 6.3.12. *Let $d, N \in \mathbb{N}$, $L \in [0, \infty)$. Let $E \subseteq \mathbb{R}^d$. Let $x_1, x_2, ..., x_N \in E$, let $f : E \to \mathbb{R}$ satisfy for all $x_1, y_1 \in E$ that $|f(x_1) - f(y_1)| \leqslant L \|x_1 - x_2\|_1$ and let $\mathsf{MC} \in \mathsf{NN}$ and $y = (f(x_1), f(x_2), ..., f(x_N))$ satisfy:*

$$\mathsf{MC}^{N,d}_{x,y} = \mathsf{Mxm}^N \bullet \mathsf{Aff}_{-L\mathbb{I}_N, y} \bullet \left[\boxminus_{i=1}^N \mathsf{Nrm}_1^d \bullet \mathsf{Aff}_{\mathbb{I}_d, -x_i} \right] \bullet \mathsf{Cpy}_{N,d} \tag{6.3.55}$$

It is then the case that:

$$\sup_{x \in E} \left| \left(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{MC}^{N,d}_{x,y} \right) \right)(x) - f(x) \right| \leqslant 2L \left[\sup_{x \in E} \left(\min_{i \in \{1,2,...,N\}} \|x - x_i\|_1 \right) \right] \tag{6.3.56}$$

Proof. Throughout this proof let $F : \mathbb{R}^d \to \mathbb{R}$ satisfy that:

$$F(x) = \max_{i \in \{1,2,...,N\}} \left(f(x_i) - L \|x - x_i\|_1 \right) \tag{6.3.57}$$

Note then that Corollary 6.3.11.1, (6.3.57), and the assumption that for all $x, y \in E$ it holds that $|f(x) - f(y)| \leqslant L \|x - y\|_1$ assures that:

$$\sup_{x \in E} |F(x) - f(x)| \leqslant 2L \left[\sup_{x \in E} \left(\min_{i \in \{1,2,...,N\}} \|x - x_i\|_1 \right) \right] \tag{6.3.58}$$

Then Lemma 6.3.8 tells us that for all $x \in E$ it holds that $F(x) = (\mathfrak{I}_{\mathfrak{r}}(\mathsf{MC}))(x)$. This

combined with (6.3.58) establishes (6.3.56). $\square$

Lemma 6.3.13. *Let* $d, N \in \mathbb{N}$, $L \in [0, \infty)$. *Let* $[a, b] \subsetneq \mathbb{R}^d$. *Let* $x_1, x_2, ..., x_N \in [a, b]$, *let* $f : [a, b] \to \mathbb{R}$ *satisfy for all* $x_1, x_2 \in [a, b]$ *that* $|f(x_1) - f(x_2)| \leqslant L |x_1 - x_2|$ *and let* $\mathsf{MC}_{x,y}^{N,1} \in \mathsf{NN}$ *and* $y = f([x]_*)$ *satisfy:*

$$\mathsf{MC}_{x,y}^{N,1} = \mathsf{Mxm}^N \bullet \mathsf{Aff}_{-L\mathbb{I}_N,y} \bullet \left[\boxminus_{i=1}^N \mathsf{Nrm}_1^1 \bullet \mathsf{Aff}_{1,-x_i} \right] \bullet \mathsf{Cpy}_{N,1} \qquad (6.3.59)$$

It is then the case that for approximant $\mathsf{MC}_{x,y}^{N,1}$ *that:*

 (i) $\mathsf{I}\left(\mathsf{MC}_{x,y}^{N,1}\right) = 1$

 (ii) $\mathsf{O}\left(\mathsf{MC}_{x,y}^{N,1}\right) = 1$

 (iii) $\mathsf{H}\left(\mathsf{MC}_{x,y}^{N,1}\right) = \lceil \log_2(N) \rceil + 1$

 (iv) $\mathsf{W}_1\left(\mathsf{MC}_{x,y}^{N,1}\right) = 2N$

 (v) *for all* $i \in \{2, 3, ...\}$ *we have* $\mathsf{W}_1\left(\mathsf{MC}_{x,y}^{N,1}\right) \leqslant 3 \left\lceil \frac{N}{2^{i-1}} \right\rceil$

 (vi) *it holds for all* $x \in \mathbb{R}^d$ *that* $\left(\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{MC}_{x,y}^{N,1} \right) \right)(x) = \max_{i \in \{1,2,...,N\}} (y_i - L|x - x_i|)$

 (vii) *it holds that* $\mathsf{P}\left(\mathsf{MC}_{x,y}^{N,1}\right) \leqslant 6 + 7N^2 + 3 \left\lceil \frac{N}{2} \right\rceil \cdot 2N$

(viii) $\sup_{x \in [a,b]} |F(x) - f(x)| \leqslant 2L \frac{|a-b|}{N}$

Proof. Items (i)—(vii) is an assertion of Lemma 6.3.8. Item (viii) is an assertion of Lemma 6.3.12 with $d \curvearrowleft 1$. $\square$

Part III

Artificial Neural Networks for u and Brownian motions

Chapter 7

ANN representations of Brownian Motion Monte Carlo

We will now take the modified and simplified version of Multi-level Picard introduced in Chapter 2 and show a neural network representation and associated, parameters, depth, and accuracy bounds. However we will also try a different approach in that we will also give a direct neural network representation of the expectation of the stochastic process that Feynman-Kac asserts in Lemma 7.4.4, and to build up to it we must build the requisite technology in Lemma 7.1.1, Lemma 7.2.1, Lemma 7.3.1.

Lemma 7.0.1 (R—,2023). *Let $d, M \in \mathbb{N}$, $T \in (0, \infty)$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $\Gamma \in \mathsf{NN}$, satisfy that $\mathfrak{I}_\mathfrak{a}(\mathsf{G}_d) \in C\left(\mathbb{R}^d, \mathbb{R}\right)$, for every $\theta \in \Theta$, let $\mathcal{U}^\theta : [0, T] \to [0, T]$ and $\mathcal{W}^\theta : [0, T] \to \mathbb{R}^d$ be functions , for every $\theta \in \Theta$, let $U^\theta : [0, T] \to \mathbb{R}^d \to \mathbb{R}$ satisfy satisfy for all $t \in [0, T]$, $x \in \mathbb{R}^d$ that:*

$$U^\theta(t, x) = \frac{1}{M} \left[\sum_{k=1}^M \left(\mathfrak{I}_\mathfrak{a}\left(\mathsf{G}_d\right)\right) \left(x + \mathcal{W}^{(\theta, 0, -k)}\right) \right] \tag{7.0.1}$$

Let $\mathsf{U}_t^\theta \in \mathsf{NN}$, $\theta \in \Theta$ *satisfy for all* $\theta \in \Theta$, $t \in [0,T]$ *that:*

$$\mathsf{U}_t^\theta = \left[\bigoplus_{k=1}^{M} \left(\frac{1}{M} \triangleright \left(\mathsf{G}_d \bullet \mathsf{Aff}_{\mathbb{I}_d, \mathcal{W}_{T-t}^{(\theta,0,-k)}} \right) \right) \right] \tag{7.0.2}$$

It is then the case that:

(i) *for all* $\theta_1, \theta_2 \in \Theta$, $t_1, t_2 \in [0,T]$ *that* $\mathsf{L}\left(\mathsf{U}_{t_1}^{\theta_1} \right) = \mathsf{L}\left(\mathsf{U}_{t_2}^{\theta_2} \right)$.

(ii) *for all* $\theta \in \Theta$, $t \in [0,T]$, *that* $\mathsf{D}\left(\mathsf{U}_t^\theta \right) \leqslant \mathsf{D}(\mathsf{G}_d)$

(iii) *for all* $\theta \in \Theta$, $t \in [0,T]$ *that:*

$$\left\| \mathsf{L}\left(\mathsf{U}_t^\theta \right) \right\|_\infty \leqslant \left\| \mathsf{L}\left(\mathsf{G}_d \right) \right\|_\infty \left(1 + \sqrt{2} \right) M \tag{7.0.3}$$

(iv) *for all* $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ *that* $U^\theta(t,x) = \left(\mathfrak{I}_\mathfrak{a}\left(\mathbf{U}_t^\theta \right) \right)(x)$ *and*

(v) *for all* $\theta \in \Theta$, $t \in [0,T]$ *that:*

$$\mathsf{P}\left(\mathsf{U}_t^\theta \right) \leqslant 2\,\mathsf{D}\left(\mathsf{G}_d \right) \left[\left(1 + \sqrt{2} \right) M \left\| \mathsf{L}\left(\mathsf{G}_d \right) \right\|_\infty \right]^2 \tag{7.0.4}$$

Proof. Throughout the proof let $\mathsf{P}_t^\theta \in \mathsf{NN}$, $\theta \in \Theta$, $t \in [0,T]$ satisfy for all $\theta \in \Theta$, $t \in [0,T]$ that:

$$\mathsf{P}_t^\theta = \bigoplus_{k=1}^{M} \left[\frac{1}{M} \triangleright \left(\mathsf{G}_d \bullet \mathsf{Aff}_{\mathbb{I}_d, \mathcal{W}_{T-t}^\theta} \right) \right] \tag{7.0.5}$$

Note the hypothesis that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that $\mathcal{W}_t^\theta \in \mathbb{R}^d$ and Lemma 4.6.9 applied for every $\theta \in \Theta$ $t \in [0,T]$ with $v \curvearrowleft M$, $c_{i \in \{u,u+1,\dots,v\}} \curvearrowleft \left(\frac{1}{M} \right)_{i \in \{u,u+1,\dots,v\}}$, $(B_i)_{i \in \{u,u+1,\dots,v\}} \curvearrowleft \left(\mathcal{W}_{T-t}^{(\theta,0,-k)} \right)_{k \in \{1,2,\dots,M\}}$, $(\nu_i)_{i \in \{u,u+1,\dots,v\}} \curvearrowleft (\mathsf{G}_d)_{i \in \{u,u+1,\dots,v\}}$, $\mu \curvearrowleft \Phi_t^\theta$ and with the notation of Lemma 4.6.9 tells us that for all $\theta \in \Theta$, $t \in [0,T]$, and $x \in \mathbb{R}^d$ it holds

that:

$$L\left(P_t^\theta\right) = \left(d, M\, W_1\left(G\right), M\, W_2\left(G\right), ..., M\, W_{D(G)-1}\left(G\right), 1\right) = L\left(P_0^0\right) \in \mathbb{N}^{D(G)+1} \qquad (7.0.6)$$

and that:

$$\left(\mathfrak{I}_\mathfrak{a}\left(P_t^\theta\right)\right)(x) = \frac{1}{M}\left[\sum_{k=1}^{M}\left(\mathfrak{I}_\mathfrak{a}\left(G\right)\right)\left(x + \mathcal{W}_{T-t}^{(\theta,0,-k)}\right)\right]$$

$$= U^\theta\left(t, x\right) \qquad (7.0.7)$$

This proves Item (i).

Note that (7.0.6), and (7.0.7) also implies that:

$$L\left(U_t^\theta\right) = L\left(P_t^\theta\right)$$

$$= \left(d, W_1\left(P_t^\theta\right), W_2\left(P_t^\theta\right), ..., W_{D(G)}\left(P_t^\theta\right), t\right)$$

$$= L\left(U_0^0\right) \in \mathbb{N}^{D(G)+1} \qquad (7.0.8)$$

This indicates that for all $\theta \in \Theta$, $t \in [0, T]$ it is the case that:

$$\left\|L\left(U_t^\theta\right)\right\|_\infty = \left\|L\left(U_0^0\right)\right\|_\infty$$

$$= \max_{k\in\{1,2,...,D(G)\}}\left(W_k\left(P_0^0\right)\right)$$

This, (7.0.6), and Lemma 4.2.5 ensure that for all $\theta \in \Theta$, $t \in [0, T]$ it is the case that:

$$\left\|L\left(U_t^\theta\right)\right\|_\infty = \left\|L\left(U_0^0\right)\right\|_\infty \leqslant \left\|L\left(P_0^0\right)\right\|_\infty \leqslant M\left\|L\left(G\right)\right\|_\infty$$

$$\leqslant M\left\|L\left(G\right)\right\|_\infty + M\left[\left\|L\left(U_0^0\right)\right\|_\infty\right] \qquad (7.0.9)$$

Then (Hutzenthaler et al., 2021, Corollary 4.3), with $\gamma \curvearrowleft 0$, $\beta \curvearrowleft M$, $k \curvearrowleft 1$, $\alpha_0 \curvearrowleft \left\|L\left(G\right)\right\|_\infty$, $\alpha_1 \curvearrowleft 0$, $(x_i)_{i\in\{0,1,...,k\}} \curvearrowleft \left(\left\|L\left(U_0^0\right)\right\|_\infty\right)_{i\in\{0,1,...,n\}}$ in the notation of (Hutzenthaler et al., 2021,

Corollary 4.3) yields for all $\theta \in \Theta$, $t \in [0, T]$ that:

$$\left\| \mathsf{L}\left(\mathsf{U}_t^\theta\right) \right\|_\infty \leqslant \frac{1}{2} \left(\left\| \mathsf{L}\left(\mathsf{G}\right) \right\|_\infty \right) \left(1 + \sqrt{2}\right) M$$
$$\leqslant \left(\left\| \mathsf{L}\left(\mathsf{G}\right) \right\|_\infty \right) \left(1 + \sqrt{2}\right) M$$

Note that Lemma 4.2.5, Item (iii), proves that for all $\theta \in \Theta$, $t \in [0, T]$ it is the case that:

$$\mathsf{D}\left(\mathsf{U}_t^\theta\right) = \mathsf{D}\left(\mathsf{U}_0^0\right) = \mathsf{D}\left(\mathsf{G}\right) \tag{7.0.10}$$

This proves Items (ii)--(iii) and (7.0.7) proves Item (iv).

Items (ii)--(iii) together shows that for all $\theta \in \Theta$, $t \in [0, T]$ it is the case that:

$$\mathsf{P}\left(\mathsf{U}_t^\theta\right) \leqslant \sum_{k=1}^{\mathsf{D}\left(\mathsf{U}_t^\theta\right)} \left\| \mathsf{L}\left(\mathsf{U}_t^\theta\right) \right\|_{\max}$$
$$= \mathsf{D}\left(\mathsf{U}_t^\theta\right) \left\| \mathsf{L}\left(\mathsf{U}_t^\theta\right) \right\|_\infty$$
$$\leqslant \mathsf{D}\left(\mathsf{U}_t^\theta\right) \left(\left\| \mathsf{L}\left(\mathsf{G}\right) \right\|_\infty \right) \left(1 + \sqrt{2}\right) M$$
$$= \mathsf{D}\left(\mathsf{G}\right) \left(\left\| \mathsf{L}\left(\mathsf{G}\right) \right\|_\infty \right) \left(1 + \sqrt{2}\right) M$$

This proves Item (v) and hence the whole lemma. $\qquad\square$

While we realize that the modified Multi-Level Picard may approximate solutions to non--linear PDEs we may chose a more circuitous route. It is quite possible, now that we have networks $\mathsf{Pwr}_n^{q,\varepsilon}$, to approximate polynomials using these networks. Once we have polynomials we may approximate more sophisticated PDEs.

7.1   The $\mathsf{E}_n^{N,h,q,\varepsilon}$ Neural Networks

Lemma 7.1.1 (R—, 2023). *Let $n, N \in \mathbb{N}$ and $h \in (0, \infty)$. Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$, satisfy that $\delta = \varepsilon \left(2^{q-1} + 1\right)^{-1}$. Let $a \in (-\infty, \infty)$, $b \in [a, \infty)$. Let $f : [a, b] \to \mathbb{R}$ be continuous*

and have second derivatives almost everywhere in $[a, b]$. Let $a = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_{N-1} \leqslant x_N = b$ such that for all $i \in \{0, 1, ..., N\}$ it is the case that $h = \frac{b-a}{N}$, and $x_i = x_0 + i \cdot h$. Let $x = [x_0 \ x_1 \ \cdots \ x_N]$ and as such let $f\left([x]_{*,*}\right) = [f(x_0) \ f(x_1) \ \cdots \ f(x_N)]$. Let $\mathsf{E}_n^{N,h,q,\varepsilon} \in \mathsf{NN}$ be the neural network given by:

$$\mathsf{E}_n^{N,h,q,\varepsilon} = \mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Etr}^{N,h} \tag{7.1.1}$$

It is then the case that:

(i) for all $x \in \mathbb{R}^{N+1}$ we have that $\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)\right)(x) \in C\left(\mathbb{R}^{N+1}, \mathbb{R}\right)$

(ii) $\mathsf{D}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right) \leqslant \begin{cases} 1 & : n = 0 \\ n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] - 1\right] + 1 & : n \geqslant 1 \end{cases}$

(iii)

$\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)$

$\leqslant \begin{cases} N + 2 & : n = 0 \\ \left(\frac{1}{2}N + 1\right)(n+1)\left[4^{n+\frac{3}{2}} + \left(\frac{4^{n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right)\right] & : n \in \mathbb{N} \end{cases}$

(iv) for all $x = \{x_0, x_1, \ldots, x_N\} \in \mathbb{R}^{N+1}$, where $a = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_{N-1} \leqslant x_N = b$ we have that:

$$\left|\exp\left[\int_a^b f\,dx\right] - \mathfrak{I}_\mathfrak{r}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)\left(f\left([x]_{*,*}\right)\right)\right|$$

$$\leqslant \frac{(b-a)^3}{12N^2}f''(\xi) \cdot n^2 \cdot \left[\Xi + \frac{(b-a)^3}{12N^2}f''(\xi)\right]^{n-1}$$

$$+ \sum_{i=1}^n \frac{1}{i!}\left(\left|\Xi\left(\Xi^{i-1} - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)(\Xi)\right)\right| + \varepsilon + |\Xi|^q + \mathfrak{p}_{i-1}^q\right) + \frac{e^\xi \cdot \left|\int_a^b f\,dx\right|^{n+1}}{(n+1)!} \tag{7.1.2}$$

Where $\Xi = \mathfrak{I}_\mathfrak{r}\left(\mathsf{Etr}^{N,h}\right)\left(f\left([x]_{*,*}\right)\right)$.

(v) it is the case that $\mathsf{W}_{\mathsf{H}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right) \leqslant 24 + 2n$

Proof. Note that Lemma 6.2.4 tells us that $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Etr}^{N,h}\right) \in C\left(\mathbb{R}^{N+1}, \mathbb{R}\right)$, and Lemma 5.2.23 tells us that $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right)(x) \in C\left(\mathbb{R}, \mathbb{R}\right)$. Next, note that Lemma 4.2.5 and the fact that the composition of continuous functions is continuous yields that:

$$\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right) = \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Aff}_{\left[\frac{h}{2} \ h \ \dots \ h \ \frac{h}{2}\right],0}\right)$$
$$= \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) \circ \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Aff}_{\left[\frac{h}{2} \ h \ \dots \ h \ \frac{h}{2}\right],0}\right) \in C\left(\mathbb{R}^{N+1}, \mathbb{R}\right)$$

Since both component neural networks are continuous, and the composition of continuous functions is continuous, so is $\mathsf{E}_n^{N,h,q,\varepsilon}$. This proves Item (i).

Next note that $\mathsf{D}\left(\mathsf{Aff}_{\left[\frac{h}{2} \ h \ \dots \ h \ \frac{h}{2}\right]}\right) = 1$, and thus Lemma 4.2.5 and Lemma 5.2.23 tells us that:

$$\mathsf{D}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right) = \mathsf{D}\left(\mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Aff}_{\left[\frac{h}{2} \ h \ \dots \ h \ \frac{h}{2}\right],0}\right)$$
$$= \mathsf{D}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) + \mathsf{D}\left(\mathsf{Aff}_{\left[\frac{h}{2} \ h \ \dots \ h \ \frac{h}{2}\right],0}\right) - 1$$
$$= \mathsf{D}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right)$$
$$\leqslant \begin{cases} 1 & : n = 0 \\ n\left[\frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] - 1\right] + 1 & : n \in \mathbb{N} \end{cases}$$

This proves Item (ii).

Next note that by Corollary 4.4.5.1, Lemma 5.2.23, Lemma 6.2.4, and the fact that $\mathsf{I}\left(\mathsf{Etr}^{N,h}\right) =$

$N + 1$, and $\mathsf{I}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) = 1$, tells us that, for all $N \in \mathbb{N}$ it is the case that:

$$
\begin{aligned}
&\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right) \\
&\leqslant \left[\max\left\{1, \frac{\mathsf{I}\left(\mathsf{Etr}^{N,h}\right) + 1}{\mathsf{I}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) + 1}\right\}\right] \cdot \mathsf{P}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) \\
&= \left(\frac{1}{2}N + 1\right) \cdot \mathsf{P}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right) \\
&\leqslant \begin{cases} N + 2 & : n = 0 \\ \left(\frac{1}{2}N + 1\right)(n + 1)\left[4^{n + \frac{3}{2}} + \left(\frac{4^{n+1}-1}{3}\right)\left(\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 372\right)\right] & : n \in \mathbb{N} \end{cases}
\end{aligned}
$$

This proves Item (iii).

Note next that:

$$
\mathsf{Aff}_{\left[\frac{h}{2} \, h \, \ldots \, h \, \frac{h}{2}\right],0} = \mathsf{Etr}^{N,h} \tag{7.1.3}
$$

Thus the well-known error term of the trapezoidal rule tells us that for $[a, b] \subsetneq \mathbb{R}$, and for $\xi \in [a, b]$ it is the case that:

$$
\left|\int_a^b f(x)\, dx - \left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Etr}^{N,h}\right)\right)\left(f\left([x]_{*,*}\right)\right)\right| \leqslant \frac{(b - a)^3}{12N^2} f''(\xi) \tag{7.1.4}
$$

and for $n \in \mathbb{N}_0$, $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon\left(2^{q-1} + 1\right)^{-1}$, and for $x \in [\mathfrak{a}, \mathfrak{b}] \subsetneq \mathbb{R}$, with $0 \in [\mathfrak{a}, \mathfrak{b}]$ it is the case, according to Lemma 7.1.1, that:

$$
\left|e^x - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right)(x)\right| \leqslant \sum_{i=1}^n \frac{1}{i!}\left(\left|x\left(x^{n-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right)\right| + \varepsilon + |x|^q + \mathfrak{p}_{n-1}^q\right) + \frac{e^{\mathfrak{b}} \cdot |x|^{n+1}}{(n + 1)!}
$$

$$
\tag{7.1.5}
$$

Note now that for $f \in C_{ae}(\mathbb{R}, \mathbb{R})$, $\int_a^b f\, dx \in [\mathfrak{a}, \mathfrak{b}] \subsetneq \mathbb{R}$, $0 \in [\mathfrak{a}, \mathfrak{b}]$, and ξ between 0 and $\int_a^b f\, dx$

it is the case that:

$$\exp\left[\int_a^b f dx\right] = \sum_{i=1}^n \left[\frac{1}{i!}\left(\int_a^b f dx\right)^i\right] + \frac{e^\xi \cdot \left|\int_a^b f dx\right|^{n+1}}{(n+1)!} \tag{7.1.6}$$

And thus the triangle inequality, Lemma 4.2.5, and Lemma 5.2.23, tells us that for $x = \{x_0, x_1, \ldots, x_N\}$, $a = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_N = b$ and $[a,b] \subsetneq \mathbb{R}$ that:

$$\left|\exp\left[\int_a^b f dx\right] - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)\left(f\left([x]_{*,*}\right)\right)\right|$$

$$= \left|\sum_{i=1}^n \left[\frac{1}{i!}\left(\int_a^b f dx\right)^i\right] + \frac{e^\xi \cdot \left(\int_a^b f dx\right)^{n+1}}{(n+1)!} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Etr}^{N,h}\right)\left(f\left([x]_{*,*}\right)\right)\right|$$

$$\leqslant \left|\sum_{i=1}^n \left[\frac{1}{i!}\left(\int_a^b f dx\right)^i\right] - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right)(x) \circ \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Etr}^{N,h}\right)\left(f\left([x]_{*,*}\right)\right)\right| + \frac{e^\xi \cdot \left|\int_a^b f dx\right|^{n+1}}{(n+1)!}$$

$$\tag{7.1.7}$$

Note that the instantiation of $\mathsf{Etr}^{N,h}$ is exact as it is the instantiation of an affine neural network. For notational simplicity let $\Xi = \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Etr}^{N,h}\right)\left(f\left([x]_{*,*}\right)\right)$. Then Lemma 5.2.23 tells us that:

$$\left|\sum_{i=0}^n \left[\frac{\Xi^i}{i!}\right] - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Xpn}_n^{q,\varepsilon}\right)(\Xi)\right| \leqslant \sum_{i=1}^n \frac{1}{i!}\left(\left|\Xi\left(\Xi^{i-1} - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)(\Xi)\right)\right| + \varepsilon + |\Xi|^q + \left(\mathfrak{p}_{i-1}^\Xi\right)^q\right) \tag{7.1.8}$$

Where for $i \in \mathbb{N}$, let $\mathfrak{p}_{i-1}^\Xi$ be the family of functions defined as such:

$$\mathfrak{p}_1^\Xi = \varepsilon + 1 + |\Xi|^2$$

$$\mathfrak{p}_i^\Xi = \varepsilon + \left(\mathfrak{p}_{i-1}\right)^2 + |\Xi|^2 \tag{7.1.9}$$

This then leaves us with:

$$\left| \sum_{i=0}^{n} \left[\frac{1}{i!} \left(\int_a^b f dx \right)^i \right] - \sum_{i=0}^{n} \left[\frac{\Xi^i}{i!} \right] \right| \leqslant \sum_{i=0}^{n} \left| \left[\frac{1}{i!} \left(\int_a^b f dx \right)^i - \frac{\Xi^i}{i!} \right] \right|$$

$$\leqslant (n+1) \max_{i \in \{0,1,\ldots,n\}} \left| \left[\frac{1}{i!} \left(\int_a^b f dx \right)^i - \frac{\Xi^i}{i!} \right] \right|$$

$$\leqslant n \cdot \max_{i \in \{1,\ldots,n\}} \frac{1}{i!} \left| \left[\left(\int_a^b f dx \right)^i - \Xi^i \right] \right| \qquad (7.1.10)$$

Note that for each $i \in \{1, \ldots, n\}$ it holds that:

$$\left(\int_a^b f dx \right)^i - \Xi^i = \left(\int_a^b f dx - \Xi \right) \left[\left(\int_a^b f dx \right)^{i-1} + \left(\int_a^b f dx \right)^{i-2} \cdot \Xi + \cdots + \Xi^{i-1} \right]$$

$$(7.1.11)$$

Note that the well-known error bounds for the trapezoidal rule tell us that Ξ and $\int_a^b f dx$ differ by at most $\frac{(b-a)^3}{12N^2} f''(\xi)$ in absolute terms, and thus:

$$\max \left\{ \Xi, \int_a^b f dx \right\} \leqslant \Xi + \frac{(b-a)^3}{12N^2} f''(\xi) \qquad (7.1.12)$$

This then renders (7.1.11) as:

$$\left(\int_a^b f dx \right)^i - \Xi^i \leqslant \frac{(b-a)^3}{12N^2} f''(\xi) \cdot i \cdot \left[\Xi + \frac{(b-a)^3}{12N^2} f''(\xi) \right]^{i-1} \qquad (7.1.13)$$

Note that this also renders (7.1.10) as:

$$\left| \sum_{i=0}^{n} \left[\frac{1}{i!} \left(\int_a^b f dx \right)^i \right] - \sum_{i=0}^{n} \left[\frac{\Xi^i}{i!} \right] \right| \leqslant \frac{(b-a)^3}{12N^2} f''(\xi) \cdot n^2 \cdot \left[\Xi + \frac{(b-a)^3}{12N^2} f''(\xi) \right]^{n-1} \qquad (7.1.14)$$

214

This, the triangle inequality and (7.1.8), then tell us for all $x \in [a, b] \subseteq [0, \infty)$ that:

$$
\left| \sum_{i=0}^{n} \left[\frac{1}{i!} \left(\int_{a}^{b} f dx \right)^{i} \right] - \Im_{\mathfrak{r}} \left(\mathsf{Xpn}_{n}^{q, \varepsilon} \right) (x) \circ \Xi \right|
$$

$$
\leqslant \left| \sum_{i=0}^{n} \left[\frac{1}{i!} \left(\int_{a}^{b} f dx \right)^{i} \right] - \sum_{i=0}^{n} \left[\frac{\Xi^{i}}{i!} \right] \right| + \left| \sum_{i=0}^{n} \left[\frac{\Xi^{i}}{i!} \right] - \Im_{\mathfrak{r}} \left(\mathsf{Xpn}_{n}^{q, \varepsilon} \right) (x) \circ \Xi \right|
$$

$$
\leqslant \frac{(b-a)^{3}}{12 N^{2}} f''(\xi) \cdot n^{2} \cdot \left[\Xi + \frac{(b-a)^{3}}{12 N^{2}} f''(\xi) \right]^{n-1} +
$$

$$
\sum_{i=1}^{n} \frac{1}{i!} \left(\left| \Xi \left(\Xi^{i-1} - \Im_{\mathfrak{r}} \left(\mathsf{Pwr}_{i-1}^{q, \varepsilon} \right) (\Xi) \right) \right| + \varepsilon + |\Xi|^{q} + \left(\mathfrak{p}_{i-1}^{\Xi} \right)^{q} \right) \tag{7.1.15}
$$

This, applied to (7.1.7) then gives us that:

$$
\left| \exp \left[\int_{a}^{b} f dx \right] - \Im_{\mathfrak{r}} \left(\mathsf{E}_{n}^{N, h, q, \varepsilon} \right) \left(f \left([x]_{*, *} \right) \right) \right|
$$

$$
\leqslant \left| \sum_{i=1}^{n} \left[\frac{1}{i!} \left(\int_{a}^{b} f dx \right)^{i} \right] - \Im_{\mathfrak{r}} \left(\mathsf{Xpn}_{n}^{q, \varepsilon} \right) (x) \circ \Im_{\mathfrak{r}} \left(\mathsf{Etr}^{N, h} \right) \left(f \left([x]_{*, *} \right) \right) \right| + \frac{e^{\xi} \cdot \left| \int_{a}^{b} f dx \right|^{n+1}}{(n+1)!}
$$

$$
\leqslant \frac{(b-a)^{3}}{12 N^{2}} f''(\xi) \cdot n^{2} \cdot \left[\Xi + \frac{(b-a)^{3}}{12 N^{2}} f''(\xi) \right]^{n-1} +
$$

$$
\sum_{i=1}^{n} \frac{1}{i!} \left(\left| \Xi \left(\Xi^{i-1} - \Im_{\mathfrak{r}} \left(\mathsf{Pwr}_{i-1}^{q, \varepsilon} \right) (\Xi) \right) \right| + \varepsilon + |\Xi|^{q} + \left(\mathfrak{p}_{i-1}^{\Xi} \right)^{q} \right) + \frac{e^{\xi} \cdot \left| \int_{a}^{b} f dx \right|^{n+1}}{(n+1)!} \tag{7.1.16}
$$

This proves Item (iv).

Finally note that Lemma 5.2.23 tells us that:

$$
\mathsf{W}_{\mathsf{H} \left(\mathsf{E}_{n}^{N, h, q, \varepsilon} \right)} \left(\mathsf{E}_{n}^{N, h, q, \varepsilon} \right) = \mathsf{W}_{\mathsf{H} \left(\mathsf{Xpn}_{n}^{q, \varepsilon} \right)} \left(\mathsf{Xpn}_{n}^{q, \varepsilon} \right)
$$

$$
\leqslant 24 + 2n \tag{7.1.17}
$$

$\square$

Remark 7.1.2. *We may represent the* $\mathsf{E}_{n}^{N, h, q, \varepsilon}$ *diagrammatically as follows:*
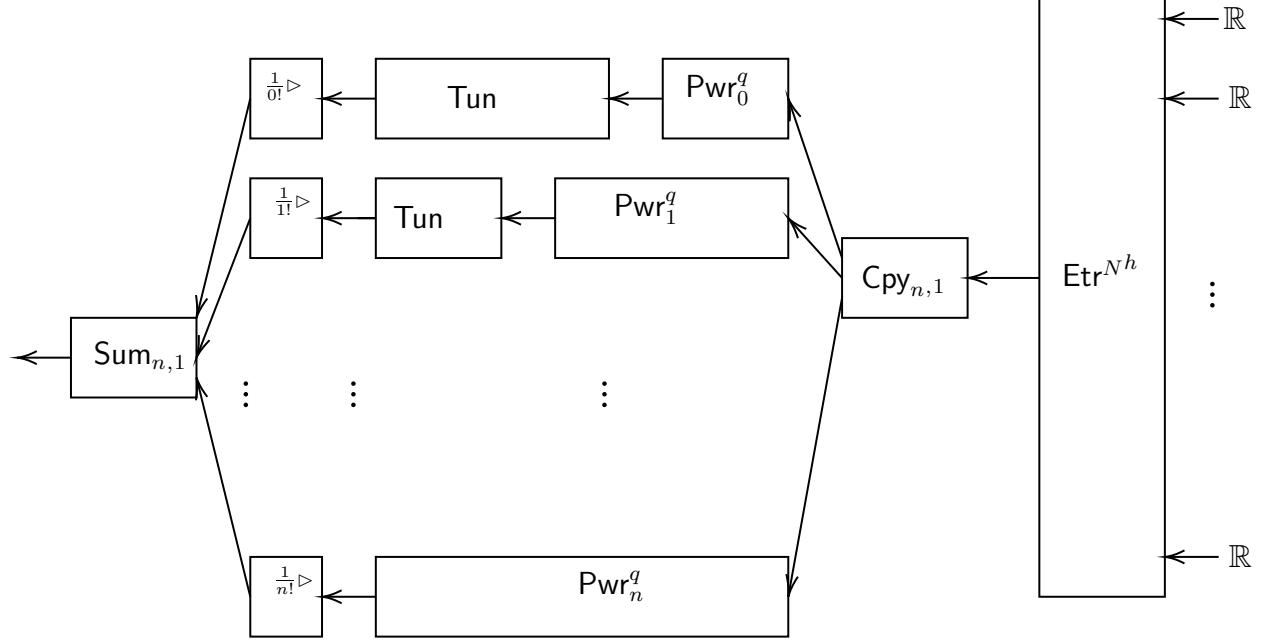
Figure 7.1: Diagram of $\mathsf{E}_n^{N,h,q,\varepsilon}$.

7.2 The $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}$ Neural Networks

Lemma 7.2.1 (R—,2023)**.** *Let $n, N \in \mathbb{N}$, $h \in (0,\infty)$. Let $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$, satisfy that $\delta = \varepsilon \left(2^{q-1}+1\right)^{-1}$. Let $a \in (-\infty, \infty)$, $b \in [a, \infty)$. Let $f : [a,b] \to \mathbb{R}$ be continuous and have second derivatives almost everywhere in $[a,b]$. Let $a = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_{N-1} \leqslant x_N = b$ such that for all $i \in \{0,1,...,N\}$ it is the case that $h = \frac{b-a}{N}$, and $x_i = x_0 + i \cdot h$. Let $x = [x_0 \, x_1 \, \cdots x_N]$ and as such let $f\left([x]_{*,*}\right) = [f(x_0) \, f(x_1) \, \cdots \, f(x_N)]$. Let $\mathsf{E}_{n,h,q,\varepsilon}^{\exp} \in \mathsf{NN}$ be the neural network given by:*

$$\mathsf{E}_n^{N,h,q,\varepsilon} = \mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Etr}^{N,h} \qquad (7.2.1)$$

Let $\mathsf{G}_d \in \mathsf{NN}$ be the neural network which instantiates as $\mathfrak{u}_d = \mathfrak{I}_\mathfrak{r}\left(\mathsf{G}_d\right)(x) \in C\left(\mathbb{R}^d, \mathbb{R}\right)$ for all $x \in \mathbb{R}^d$.

Let $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}$ be the neural network given as:

$$\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} = \mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{E}_n^{N,h,q,\varepsilon} \diamondsuit \mathsf{G}_d\right] \tag{7.2.2}$$

It is then the case that for all $\mathfrak{x} = \{x_0, x_1, \ldots, x_N\} \in \mathbb{R}^{N+1}$ and $x \in \mathbb{R}^d$ that:

(i) $\mathfrak{I}_\mathfrak{r}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right)(f([\mathfrak{x}]_*) \frown x) \in C\left(\mathbb{R}^{N+1} \times \mathbb{R}^d, \mathbb{R}\right)$

(ii) $\mathsf{D}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right) \leqslant \begin{cases} \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] + \mathsf{D}\left(\mathsf{G}_d\right) - 1 & : n = 0 \\ \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] + \max\left\{\mathsf{D}\left(\mathsf{E}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right), \mathsf{D}\left(\mathsf{G}_d\right)\right\} - 1 & : n \geqslant 1 \end{cases}$

(iii) It is also the case that:

$$\mathsf{P}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right) \leqslant \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 324 + 48n$$
$$+ 24\,\mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d\right) + 4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathsf{P}\left(\mathsf{G}_d\right)\right\} \tag{7.2.3}$$

(iv) It is also the case that:

$$\left|\exp\left(\int_a^b f\,dx\right)\mathfrak{u}_d(x) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right)(f([\mathfrak{x}]_* \frown x)\right|$$
$$\leqslant 3\varepsilon + 2\varepsilon\,|\mathfrak{u}(x)|^q + 2\varepsilon\left|\exp\left(\int_a^b f\,dx\right)\right|^q + \varepsilon\left|\exp\left(\int_a^b f\,dx\right) - \mathfrak{e}\right|^q - \mathfrak{e}\mathfrak{u}(x)$$

Where, as per Lemma 7.1.1, $\mathfrak{e}$ is defined as:

$$\left|\mathsf{E}_n^{N,h,q,\varepsilon}\left(f([\mathfrak{x}]_*)\right) - \exp\left(\int_a^b f\,dx\right)\right| \leqslant \mathfrak{e} \tag{7.2.4}$$

Remark 7.2.2. *Diagrammatically* $\mathsf{UE}_n^{N,h,q,\varepsilon}$ *can be represented as:*

Proof. Note that from Lemma 4.2.5, and Lemma 4.3.3, we have that for $\mathfrak{x} \in \mathbb{R}^{N+1}$, and $x \in \mathbb{R}^d$ it is the case that $\mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{E}_n^{N,h,q,\varepsilon} \diamondsuit \mathsf{G}_d\right]\right)(f([\mathfrak{x}]_* \frown x)) = \mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon}\right) \circ \mathfrak{I}_\mathfrak{r}\left(\left[\mathsf{E}_n^{N,h,q,\varepsilon} \diamondsuit \mathsf{G}_d\right]\right)$
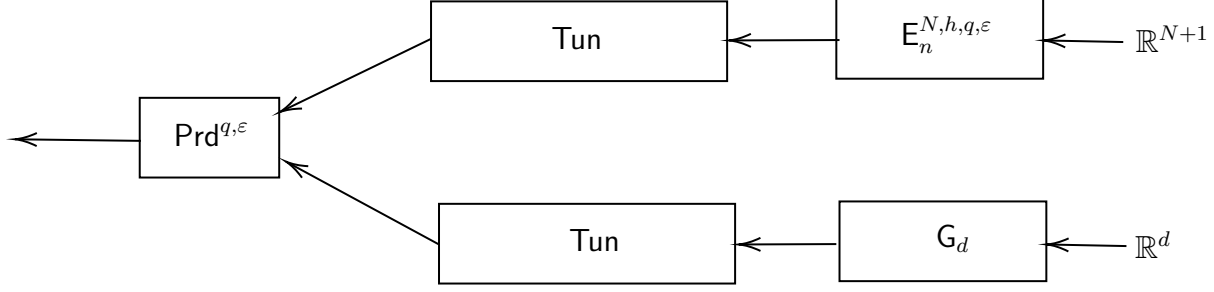
Figure 7.2: Neural network diagram for $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}$

$(f([\mathfrak{x}]_*) \frown x)$. Then Lemma 5.1.10 tells us that $\mathfrak{I}_\mathfrak{r}(\mathsf{Prd}^{q,\varepsilon}) \in C(\mathbb{R}^2, \mathbb{R})$. Lemma 7.1.1 tells us that $\mathfrak{I}_\mathfrak{r}(\mathsf{E}_n^{N,h,q,\varepsilon}) \in C(\mathbb{R}^{N+1}, \mathbb{R})$ and by hypothesis it is the case that $\mathfrak{I}_\mathfrak{r}(\mathsf{G}_d) \in C(\mathbb{R}^d, \mathbb{R})$. Thus, by the stacking properties of continuous instantiated networks and the fact that the composition of continuous functions is continuous, we have that $\mathfrak{I}_\mathfrak{r}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right) \in C(\mathbb{R}^{N+1} \times \mathbb{R}^d, \mathbb{R})$.

Note that by Lemma 4.2.5 it is the case that:

$$\mathsf{D}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right) = \mathsf{D}(\mathsf{Prd}^{q,\varepsilon}) + \mathsf{D}\left(\mathsf{E}_n^{N,h,q,\varepsilon} \ominus \mathsf{G}_d\right) - 1 \tag{7.2.5}$$

Lemma 7.1.1 and Lemma 5.1.10 then tell us that:

$$\mathsf{D}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right)$$
$$\leqslant \begin{cases} \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] + \mathsf{D}(\mathsf{G}_d) - 1 & : n = 0 \\ \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] + \max\left\{\mathsf{D}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathsf{D}(\mathsf{G}_d)\right\} - 1 & : n \geqslant 1 \end{cases} \tag{7.2.6}$$

Note that then Lemma 4.2.5, Lemma 5.2.23, and Lemma 7.1.1 tell us that:

$$\mathsf{P}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right) \leqslant \mathsf{P}\left(\mathsf{Prd}^{q,\varepsilon}\right) + 4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathsf{P}\left(\mathsf{G}_d\right)\right\}$$

$$+ \mathsf{W}_1\left(\mathsf{Prd}^{q,\varepsilon}\right) \cdot \mathsf{W}_{\mathsf{H}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\diamondsuit\mathsf{G}_d\right)}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\diamondsuit\mathsf{G}_d\right)$$

$$\leqslant \mathsf{P}\left(\mathsf{Prd}^{q,\varepsilon}\right) + 4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathsf{P}\left(\mathsf{G}_d\right)\right\}$$

$$+ 24\left[(24 + 2n) + \mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d\right)\right]$$

$$= \mathsf{P}\left(\mathsf{Prd}^{q,\varepsilon}\right) + 4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathsf{P}\left(\mathsf{G}_d\right)\right\}$$

$$+ 576 + 48n + 24 \cdot \mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d\right)$$

$$\leqslant \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 324 + 48n$$

$$+ 24\,\mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d\right) + 4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathsf{P}\left(\mathsf{G}_d\right)\right\} \qquad (7.2.7)$$

Now note that Lemma 4.2.5, and Lemma 4.3.3 tells us that for all $\mathfrak{x} = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n$ and $x \in \mathbb{R}^d$:

$$\mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{E}_n^{N,h,q,\varepsilon}\diamondsuit\mathsf{G}_d\right]\right)\left(f\left([\mathfrak{x}]_*\right) \frown x\right) = \mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathfrak{I}_\mathfrak{r}\left(\mathsf{G}_d\right)\right)\left(f\left([\mathfrak{x}]_*\right) \frown x\right).$$
$$(7.2.8)$$

Note then that the triangle inequality tells us that:

$$\left|\exp\left(\int_a^b f dx\right)\mathfrak{u}_d\left(x\right) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right)\left(f\left([\mathfrak{x}]_*\right) \frown x\right)\right|$$

$$= \left|\exp\left(\int_a^b f dx\right) \cdot \mathfrak{u}_d\left(x\right) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathfrak{I}_\mathfrak{r}\left(\mathsf{G}_d\right)\right)\left(f\left([\mathfrak{x}]_*\right) \frown x\right)\right|$$

$$\leqslant \left|\exp\left(\int_a^b f dx\right) \cdot \mathfrak{u}_d\left(x\right) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\exp\left(\int_a^b f dx\right), \mathfrak{u}_d\left(x\right)\right)\right|$$

$$+ \left|\mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\exp\left(\int_a^b f dx\right), \mathfrak{u}_d\left(x\right)\right) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\mathfrak{I}_\mathfrak{r}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathfrak{I}_\mathfrak{r}\left(\mathsf{G}_d\right)\right)\left(f\left([\mathfrak{x}]\right) \frown x\right)\right|$$
$$(7.2.9)$$

Note that Lemma 5.1.10 bounds the first summand. Note that by hypothesis $\mathfrak{I}_\mathfrak{r}\left(\mathsf{G}_d\right)$ is

exactly $\mathfrak{u}_d(x)$. Note also that by Lemma 7.1.1, Lemma 5.1.10, we realize that the second summand can be bounded as such:

$$\left| \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\exp\left(\int_a^b f\,dx\right),\mathfrak{u}_d(x)\right) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\right)\left(\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right),\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{G}_d\right)\right)\left(f\left([\mathfrak{r}]_*\right) \frown x\right) \right|$$

$$\leqslant \exp\left(\int_a^b f\,dx\right)\mathfrak{u}_d(x) + \varepsilon + \varepsilon\left|\exp\left(\int_a^b f\,dx\right)\right|^q + \varepsilon\left|\mathfrak{u}_d(x)\right|^q$$

$$- \left[\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)\left(f\left([\mathfrak{r}]_*\right)\right)\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{G}_d\right)(x) - \varepsilon - \varepsilon\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)\left(f\left([\mathfrak{r}]_*\right)\right)\right|^q - \varepsilon\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{G}_d\right)(x)\right|^q\right]$$

$$\tag{7.2.10}$$

Per Lemma 7.1.1, let $\mathfrak{e}$ represent the error in approximation of $\mathsf{E}_n^{N,h,q,\varepsilon}$, that is to say for all $\mathfrak{r}\in\mathbb{R}^{N+1}$ and corresponding $f\left([\mathfrak{r}]_*\right)$, let it be the case that:

$$\left|\mathsf{E}_n^{N,h,q,\varepsilon}\left(f\left([\mathfrak{r}]_*\right)\right) - \exp\left(\int_a^b f\,dx\right)\right| \leqslant \mathfrak{e} \tag{7.2.11}$$

Thus $\mathsf{E}_n^{N,h,q,\varepsilon}\left(f\left([\mathfrak{r}]_*\right)\right)$ is maximally $\mathfrak{e}+\exp\left(\int_a^b f\,dx\right)$ and minimally $\exp\left(\int_a^b f\,dx\right)-\mathfrak{e}$. Thus (7.2.10) is rendered as:

$$\exp\left(\int_a^b f\,dx\right)\mathfrak{u}_d(x) + \varepsilon + \varepsilon\left|\exp\left(\int_a^b f\,dx\right)\right|^q + \varepsilon\left|\mathfrak{u}_d(x)\right|^q$$

$$- \left[\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{G}_d\right)\left(f\left([\mathfrak{r}]_*\right) \frown x\right) - \varepsilon - \varepsilon\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right)\left(f\left([\mathfrak{r}]_*\right)\right)\right|^q - \varepsilon\left|\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{G}_d\right)(x)\right|^q\right]$$

$$\leqslant \exp\left(\int_a^b f\,dx\right)\mathfrak{u}_d(x) + \varepsilon + \varepsilon\left|\exp\left(\int_a^b f\,dx\right)\right|^q + \varepsilon\left|\mathfrak{u}_d(x)\right|^q$$

$$- \left[\left(\mathfrak{e}+\exp\left(\int_a^b f\,dx\right)\right)\mathfrak{u}_d(x) - \varepsilon - \varepsilon\left|\exp\left(\int_a^b f\,dx\right) - \mathfrak{e}\right|^q - \varepsilon\left|\mathfrak{u}_d(x)\right|^q\right]$$

$$= \exp\left(\int_a^b f\,dx\right)\mathfrak{u}_d(x) + \varepsilon + \varepsilon\left|\exp\left(\int_a^b f\,dx\right)\right|^q + \varepsilon\left|\mathfrak{u}_d(x)\right|^q$$

$$- \mathfrak{e}u(t,x) - \exp\left(\int_a^b f\,dx\right)\mathfrak{u}_d(x) + \varepsilon + \varepsilon\left|\exp\left(\int_a^b f\,dx\right) - \mathfrak{e}\right|^q + \varepsilon\left|\mathfrak{u}_d(x)\right|^q$$

$$= 2\varepsilon + 2\varepsilon\left|\mathfrak{u}_d(x)\right|^q + \varepsilon\left|\exp\left(\int_a^b f\,dx\right) - \mathfrak{e}\right|^q + \varepsilon\left|\exp\left(\int_a^b f\,dx\right)\right|^q - \mathfrak{e}\mathfrak{u}_d(x) \tag{7.2.12}$$

This, together with (7.2.9), then tells us that:

$$\left| \exp \left(\int_a^b f dx \right) \mathfrak{u}_d(x) - \mathfrak{I}_\mathfrak{r} \left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \right) (f([\mathfrak{r}]_*) \frown x) \right|$$

$$\leqslant \left| \exp \left(\int_a^b f dx \right) \mathfrak{u}_d(x) - \mathfrak{I}_\mathfrak{r} (\mathsf{Prd}^{q,\varepsilon}) \left(\exp \left(\int_a^b f dx \right), \mathfrak{u}_d(x) \right) \right|$$

$$+ \left| \mathfrak{I}_\mathfrak{r} (\mathsf{Prd}^{q,\varepsilon}) \left(\exp \left(\int_a^b f dx \right), \mathfrak{u}_d(x) \right) - \mathfrak{I}_\mathfrak{r} (\mathsf{Prd}^{q,\varepsilon}) \left(\mathfrak{I}_\mathfrak{r} \left(\mathsf{E}_n^{N,h,q,\varepsilon} \right) (f([\mathfrak{r}]_*)), \mathfrak{I}_\mathfrak{r} (\mathsf{G}_d)(x) \right) \right|$$

$$\leqslant \varepsilon + \varepsilon \left| \exp \left(\int_a^b f dx \right) \right|^q + \varepsilon \left| \mathfrak{u}_d(x) \right|^q$$

$$+ 2\varepsilon + 2\varepsilon \left| \mathfrak{u}_d(x) \right|^q + \varepsilon \left| \exp \left(\int_a^b f dx \right) - \mathfrak{e} \right|^q + \varepsilon \left| \exp \left(\int_a^b f dx \right) \right|^q - \mathfrak{e} \mathfrak{u}_d(x)$$

$$= 3\varepsilon + 2\varepsilon \left| \mathfrak{u}_d(x) \right|^q + 2\varepsilon \left| \exp \left(\int_a^b f dx \right) \right|^q + \varepsilon \left| \exp \left(\int_a^b f dx \right) - \mathfrak{e} \right|^q - \mathfrak{e} \mathfrak{u}_d(x)$$

$\square$

7.3   The $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$ Neural Networks

Lemma 7.3.1 (R—,2023)**.** *Let* $n, N \in \mathbb{N}$, $h \in (0,\infty)$. *Let* $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$, *satisfy that* $\delta = \varepsilon \left(2^{q-1} + 1 \right)^{-1}$. *Let* $a \in (-\infty, \infty)$, $b \in [a, \infty)$. *Let* $f : [a,b] \to \mathbb{R}$ *be continuous and have second derivatives almost everywhere in* $[a,b]$. *Let* $a = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_{N-1} \leqslant x_N = b$ *such that for all* $i \in \{0, 1, ..., N\}$ *it is the case that* $h = \frac{b-a}{N}$, *and* $x_i = x_0 + i \cdot h$. *Let* $x = [x_0 \ x_1 \ \cdots \ x_N]$ *and as such let* $f \left([x]_{*,*} \right) = [f(x_0) \ f(x_1) \ \cdots \ f(x_N)]$. *Let* $\mathsf{E}_{n,h,q,\varepsilon}^{\exp} \in \mathsf{NN}$ *be the neural network given by:*

$$\mathsf{E}_n^{N,h,q,\varepsilon} = \mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Etr}^{N,h} \tag{7.3.1}$$

Let $\mathsf{G}_d \subsetneq \mathsf{NN}$ *be the neural networks which, for* $d \in \mathbb{N}$, *instantiate as* $\mathfrak{u}_d = \mathfrak{I}_\mathfrak{r} (\mathsf{G}_d)(x) \in C \left(\mathbb{R}^d, \mathbb{R} \right)$ *for all* $x \in \mathbb{R}^d$.

Let $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \subsetneq \mathsf{NN}$ *be the neural networks given as:*

$$\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} = \mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{E}_n^{N,h,q,\varepsilon} \diamondsuit \mathsf{G}_d \right] \tag{7.3.2}$$

Finally let $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \subsetneq \mathsf{NN}$ *be given the neural networks given by:*

$$\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} = \mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \bullet \left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}} \right] \tag{7.3.3}$$

It is then the case that for all $\mathfrak{x} = \{x_0, x_1, \ldots, x_N\} \in \mathbb{R}^{N+1}$ *and* $x \in \mathbb{R}^d$ *that:*

(i) $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right) \in C\left(\mathbb{R}^{N+1} \times \mathbb{R}^d, \mathbb{R} \right)$

(ii)

$$\mathsf{D}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right) = \mathsf{D}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,n,h,q,\varepsilon} \right)$$

$$\leqslant \begin{cases} \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1} \right) + q \right] + \mathsf{D}\left(\mathsf{G}_d \right) - 1 & : n = 0 \\[2mm] \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1} \right) + q \right] + \max\left\{ \mathsf{D}\left(\mathsf{E}_{N,n,h,q,\varepsilon}^{\exp,f} \right), \mathsf{D}\left(\mathsf{G}_d \right) \right\} - 1 & : n \in \mathbb{N} \end{cases}$$

(iii) It is also the case that:

$$\mathsf{P}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right) = \mathsf{P}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \right) \leqslant \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1} \right) + q + 1 \right] + 324 + 48n$$

$$+ 24\,\mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d \right) + 4\max\left\{ \mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon} \right), \mathsf{P}\left(\mathsf{G}_d \right) \right\} \tag{7.3.4}$$

(iv) It is also the case that:

$$\left| \exp\left(\int_t^T f\,ds \right) \mathfrak{u}_d^T\left(x \right) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right)\left(f\left([\mathfrak{r}]_* \right) \frown x \right) \right|$$

$$\leqslant 3\varepsilon + 2\varepsilon\left| \mathfrak{u}_d^T\left(t, x \right) \right|^q + 2\varepsilon\left| \exp\left(\int_a^b f\,dx \right) \right|^q + \varepsilon\left| \exp\left(\int_a^b f\,dx \right) - \mathfrak{e} \right|^q - \mathfrak{e}\mathfrak{u}_d^T\left(x \right)$$

Where, as per Lemma 7.1.1, $\mathfrak{e}$ is defined as:

$$\left| \mathsf{E}_n^{N,h,q,\varepsilon} \left(f\left([\mathfrak{r}]_* \right) \right) - \exp\left(\int_a^b f dx \right) \right| \leqslant \mathfrak{e} \tag{7.3.5}$$

Proof. Note that (7.4.9) is an assertion of Feynman-Kac. Now notice that for $x \in \mathbb{R}^{N+1} \times \mathbb{R}^d$ it is the case that:

$$\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right)(x) = \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \bullet \left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}} \right] \right)(x)$$

$$= \mathfrak{I}_{\mathfrak{r}} \left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \right) \circ \mathfrak{I}_{\mathfrak{r}} \left(\left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}} \right] \right)(x)$$

Note that by Lemma 7.2.1 it holds that $\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \right) \in C\left(\mathbb{R}^{N+1} \times \mathbb{R}^d, \mathbb{R} \right)$. Note also that by Lemma 5.2.6, Tun_1^{N+1} is continuous and by Lemma 4.4.2, $\mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}}$ is continuous, and whence by Lemma 5.2.6 and Lemma 4.4.5 it is the case that $\mathfrak{I}_{\mathfrak{r}} \left(\left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}} \right] \right)(x) \in C\left(\mathbb{R}^{N+1} \times \mathbb{R}^d, \mathbb{R}^{N+1} \times \mathbb{R}^d \right)$. Finally, since the composition of continuous functions is continuous, and since we have functions $\left(\mathbb{R}^{N+1} \times \mathbb{R}^d \right) \mapsto \left(\mathbb{R}^{N+1} \times \mathbb{R}^d \right) \mapsto \mathbb{R}$ we have that $\mathfrak{I}_{\mathfrak{r}} \left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right) \in C\left(\mathbb{R}^{N+1} \times \mathbb{R}^d, \mathbb{R} \right)$. This proves Item (i).

Note next that by Lemma 5.2.6, it is the case that $\mathsf{D}\left(\mathsf{Tun}_1^{N+1} \right) = \mathsf{D}\left(\mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}} \right) = 1$. Thus by Lemma 4.2.5 it is the case that $\mathsf{D}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right) = \mathsf{D}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \right)$. This proves Item (ii)

Next note that:

$$\mathsf{P}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right) = \mathsf{P}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \bullet \left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}} \right] \right) \tag{7.3.6}$$

Note carefully that Definition 5.2.4 tells us that $\mathsf{Tun}_1^{N+1} = \mathsf{Aff}_{\mathbb{I}_{N+1,N+1},\mathbb{0}_{N+1}}$, and so by Lemma 4.4.6, it must be the case that $\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}}$ is also an affine neural network. Whence,

Corollary 4.4.5.1 tells us that:

$$\mathsf{P}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right) = \mathsf{P}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \bullet \left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}}\right]\right)$$

$$\leqslant \left[\max\left\{1, \frac{\mathsf{I}\left(\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}}\right)+1}{\mathsf{I}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right)+1}\right\}\right] \cdot \mathsf{P}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right)$$

$$= \mathsf{P}\left(\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}\right)$$

$$\leqslant \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 324 + 48n$$

$$+ 24\,\mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d\right) + 4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right), \mathsf{P}\left(\mathsf{G}_d\right)\right\} \tag{7.3.7}$$

Finally, note that both $\mathsf{Aff}_{W,b}$ and Tun_n^d are exact and contribute nothing to the uncertainty. Thus $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$ has the same error bounds as $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}$. That is to say that:

$$\left|\exp\left(\int_t^T f\,ds\right)\mathfrak{u}_d^T\left(x\right) - \mathfrak{I}_\mathfrak{r}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)\left(f\left(\left[\mathfrak{x}\right]_*\right) \frown x\right)\right|$$

$$\leqslant 3\varepsilon + 2\varepsilon\left|\mathfrak{u}_d^T\left(t,x\right)\right|^q + 2\varepsilon\left|\exp\left(\int_a^b f\,dx\right)\right|^q + \varepsilon\left|\exp\left(\int_a^b f\,dx\right) - \mathfrak{e}\right|^q - \mathfrak{e}\mathfrak{u}_d^T\left(x\right)$$

$$\square$$

Corollary 7.3.1.1 (R—, 2024, Approximants for Brownian Motion)**.** *Let $t \in (0,\infty)$ and $T \in (t,\infty)$. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. Let $n, N \in \mathbb{N}$, and $h \in (0,\infty)$. Let $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$, satisfy that $\delta = \varepsilon\left(2^{q-1}+1\right)^{-1}$. Let $f: [t,T] \to \mathbb{R}$ be continuous almost everywhere in $[t,T]$. Let it also be the case that $f = g \circ \mathfrak{h}$, where $\mathfrak{h}: [t,T] \to \mathbb{R}^d$, and $g: \mathbb{R}^d \to \mathbb{R}$. Let $t = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_{N-1} \leqslant t_N = T$ such that for all $i \in \{0,1,...,N\}$ it is the case that $h = \frac{T-t}{N}$, and $t_i = t_0 + i \cdot h$. Let $\mathbf{t} = [t_0\ t_1\ \cdots t_N]$ and as such let $f\left(\left[\mathbf{t}\right]_{*,*}\right) = [f(t_0)\ f(t_1)\ \cdots\ f(t_N)]$. Let $\mathsf{E}_n^{N,h,q,\varepsilon} \in \mathsf{NN}$ be the neural network given by:*

$$\mathsf{E}_n^{N,h,q,\varepsilon} = \mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Etr}^{N,h} \tag{7.3.8}$$

Let $u_d \in C^{1,2}\left([0,T] \times \mathbb{R}^d, \mathbb{R}\right)$ satisfy for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that:

$$\left(\frac{\partial}{\partial t}u_d\right)(t,x) + \left(\nabla_x^2 u_d\right)(t,x) + \alpha_d(x)\,u_d(t,x) = 0 \tag{7.3.9}$$

Furthermore, let $\mathfrak{u}_d^T(x) = u_d(T,x)$. Let $\mathsf{G}_d \subsetneq \mathsf{NN}$ be the neural networks which instantiate as $\mathfrak{u}_d^T = \mathfrak{I}_{\mathfrak{r}}(\mathsf{G}_d) \in C\left(\mathbb{R}^d, \mathbb{R}\right)$.

Let $\mathcal{W}^d : [0,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$ be standard Brownian motions, and let $\mathcal{X}^{d,t,x} : [t,T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ be stochastic processes with continuous sample paths satisfying that for all $d \in \mathbb{N}$, $t \in [0,T]$, $s \in [t,T]$, $x \in \mathbb{R}^d$ we have $\mathbb{P}$-a.s. that:

$$\mathcal{X}_s^{d,t,x} = x + \int_s^t \sqrt{2}\, d\mathcal{W}_r^d \tag{7.3.10}$$

It is then the case that for all $d \in \mathbb{N}$, $t \in [0,T]$, $x \in \mathbb{R}^d$ it holds that:

$$u_d(t,x) = \mathbb{E}\left[\exp\left(\int_t^T \left(\alpha_d \circ \mathcal{X}_r^{d,t,x}\right) dr\right) u_d\left(T, \mathcal{X}_T^{d,t,x}\right)\right] \tag{7.3.11}$$

Let $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon}$ be the neural network given as:

$$\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} = \mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{E}_n^{N,h,q,\varepsilon} \diamond \mathsf{G}_d\right] \tag{7.3.12}$$

Finally let $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$ be given by:

$$\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} = \mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \bullet \left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{0_{d,d},\mathcal{X}_{\omega_i}}\right] \tag{7.3.13}$$

It is then the case that for all $\mathfrak{x} = \{x_0, x_1, \ldots, x_N\} \in \mathbb{R}^{N+1}$ and $x \in \mathbb{R}^d$ that:

$$\left|\exp\left(\int_t^T f\,ds\right)\mathfrak{u}_d^T(x) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)(f([\mathfrak{x}]_*) \frown x)\right|$$
$$\leqslant 3\varepsilon + 2\varepsilon \left|\mathfrak{u}_d^T(t,x)\right|^q + 2\varepsilon \left|\exp\left(\int_a^b f\,dx\right)\right|^q + \varepsilon \left|\exp\left(\int_a^b f\,dx\right) - \mathfrak{e}\right|^q - \mathfrak{e}\mathfrak{u}_d^T(x)$$

Where, as per Lemma 7.1.1, $\mathfrak{e}$ is defined as:

$$\left| \mathsf{E}_n^{N,h,q,\varepsilon} \left(f\left([\mathfrak{x}]_* \right) \right) - \exp\left(\int_a^b f\,dx \right) \right| \leqslant \mathfrak{e} \tag{7.3.14}$$

Proof. Note that for a fixed $T \in (0,\infty)$ it is the case that $u_d\left(t,x\right) \in C^{1,2}\left([0,T] \times \mathbb{R}^d, \mathbb{R}\right)$ projects down to a function $\mathfrak{u}_d^T\left(x\right) \in C^2\left(\mathbb{R}^d, \mathbb{R}\right)$. Furthermore given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $\mathcal{X}^{d,t,x} : [t,T] \times \Omega \to \mathbb{R}^d$, for a fixed outcome space $\omega_i \in \Omega$ it is the case that $\mathcal{X}^{d,t,x}$ projects down to $\mathcal{X}_{\omega_i}^{d,t,x} : [t,T] \to \mathbb{R}^d$. Thus given $\alpha_d : \mathbb{R}^d \to \mathbb{R}$ that is infinitely often differentiable, we get that $\alpha_d \circ \mathcal{X}_{\omega_i}^{d,t,x} : [t,T] \to \mathbb{R}$.

Taken together with Lemma 7.2.1 with $x \curvearrowright \mathcal{X}_{r,\omega}^{d,t,x}$, $f \curvearrowright \alpha_d \circ \mathcal{X}_{\omega_i}^{d,t,x}$, $b \curvearrowright T$, $a \curvearrowright t$, and $\mathfrak{u}_d^T\left(x\right) \curvearrowright u_d\left(T, \mathcal{X}_{\omega_i}^{d,t,x}\right)$, our error term is rendered as:

$$\left| \exp\left(\int_t^T \left(\alpha_d \circ \mathcal{X}_{r,\omega_i}^{d,t,x} ds \right) \right) u_d\left(T, \mathcal{X}_{\omega_i}^{d,t,x}\right) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \right) \right|$$

$$\leqslant 3\varepsilon + 2\varepsilon \left| u_d\left(T, \mathcal{X}_{r,\omega_i}^{d,t,x}\right) \right|^q + 2\varepsilon \left| \exp\left(\int_a^b f\,dx \right) \right|^q + \varepsilon \left| \exp\left(\int_a^b f\,dx \right) - \mathfrak{e} \right|^q - \mathfrak{e} u_d\left(T, \mathcal{X}_{r,\omega_i}^{d,t,x}\right)$$

This completes the proof of the Lemma. $\qquad \square$

Remark 7.3.2. *Diagrammatically, this can be represented as:*
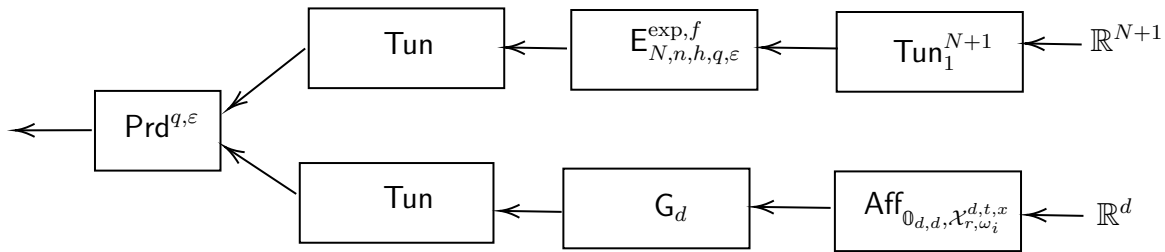
Figure 7.3: Neural network diagram for $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$

7.4 The $\text{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}$ Neural Networks

Lemma 7.4.1. *Let $\nu_1, \nu_2, \ldots, \nu_n \in \mathsf{NN}$ such that for all $i \in \{1, 2, \ldots, n\}$ it is the cast that $\mathsf{O}(\nu_i) = 1$, and it is also the case that $\mathsf{D}(\nu_1) = \mathsf{D}(\nu_2) = \cdots = \mathsf{D}(\nu_n)$. Let $x_1 \in \mathbb{R}^{\mathsf{I}(\nu_1)}, x_2 \in \mathbb{R}^{\mathsf{I}(\nu_2)}, \ldots, x_n \in \mathbb{R}^{\mathsf{I}(\nu_n)}$ and $\mathfrak{x} \in \mathbb{R}^{\sum_{i=1}^{n} \mathsf{I}(\nu_i)}$. It is then the case that we have that:*

$$\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sum}_{n,1} \bullet [\boxminus_{i=1}^{n} \nu_i]\right)(\mathfrak{x}) = \sum_{i=1}^{n} \mathfrak{I}_{\mathfrak{r}}(\nu_i)(x_i) \tag{7.4.1}$$

Proof. Throughout the proof let $x_1 \in \mathbb{R}^{\mathsf{I}(\nu_1)}, x_2 \in \mathbb{R}^{\mathsf{I}(\nu_2)}, \ldots, x_n \in \mathbb{R}^{\mathsf{I}(\nu_n)}$ and $\mathfrak{x} \in \mathbb{R}^{\sum_{i=1}^{n} \mathsf{I}(\nu_i)}$ such that $\mathfrak{x} = x_1 \frown x_2 \frown \cdots \frown x_n$. Observe that by Lemma 4.2.5 we have that:

$$\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sum}_{n,1} \bullet [\boxminus_{i=1}^{n} \nu_i]\right)(\mathfrak{x}) = \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sum}_{n,1}\right) \circ \mathfrak{I}_{\mathfrak{r}}\left(\boxminus_{i=1}^{n} \nu_i\right)(\mathfrak{x}) \tag{7.4.2}$$

Note however that Defiition 4.1.5 yields that $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sum}_{n,1}\right)(x) = [1\,1\,\cdots\,1] \cdot x + 0$ for $x \in \mathbb{R}^n$. On the other hand $\mathsf{O}\left(\boxminus_{i=1}^{n} \nu_i\right) = n$ and furthermore by Lemma 4.3.3 it is the case for $\mathfrak{x} \in \mathbb{R}^{\sum_{i=1}^{n} \mathsf{I}(\nu_i)}$ that $\mathfrak{I}_{\mathfrak{r}}\left(\boxminus_{i=1}^{n} \nu_i\right)(\mathfrak{x}) = \mathfrak{I}_{\mathfrak{r}}(\nu_1)(x_1) \frown \mathfrak{I}_{\mathfrak{r}}(\nu_2)(x_2) \frown \cdots \frown \mathfrak{I}_{\mathfrak{r}}(\nu_n)(x_n)$. Thus $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{Sum}_{n,1} \bullet [\boxminus_{i=1}^{n} \nu_i]\right)(\mathfrak{x})$ is rendered as:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mathfrak{I}_{\mathfrak{r}}(\nu_1)(x_1) \\ \mathfrak{I}_{\mathfrak{r}}(\nu_2)(x_2) \\ \vdots \\ \mathfrak{I}_{\mathfrak{r}}(\nu_n)(x_n) \end{bmatrix} + 0 = \sum_{i=1}^{n} \mathfrak{I}_{\mathfrak{r}}(\nu_i)(x_i) \tag{7.4.3}$$

This completes the proof of the lemma. $\qquad\square$

Lemma 7.4.2. *Let $\nu_1, \nu_2, \ldots, \nu_n \in \mathsf{NN}$ with $\mathsf{I}(\nu_1) = \mathsf{I}(\nu_2) = \ldots = \mathsf{I}(\nu_n)$ and $\mathsf{O}(\nu_1) = \mathsf{O}(\nu_2) = \ldots = \mathsf{O}(\nu_n) = 1$ such that for all $i \in \{1, 2, \ldots, n\}$ it is the case that there exists $f_i \in C\left(\mathbb{R}^{\mathsf{I}(\nu_1)}, \mathbb{R}\right)$, and $\varepsilon_i \in (0, \infty)$, where for all $x_i \in \mathbb{R}^{\mathsf{I}(\nu_1)}$, it is the case that $|\mathfrak{I}_{\mathfrak{r}}(\nu_i)(x_i) - f(x_i)| \leqslant \varepsilon_i$. It is then the case that for all $\mathfrak{x} \in \mathbb{R}^{n \cdot \mathsf{I}(\nu_1)}$ and $x_i \in \mathbb{R}^{\mathsf{I}(\nu_i)}$ with*

$\mathfrak{x} = x_1 \frown x_2 \frown \cdots \frown x_n$ *that:*

$$\left\| \mathfrak{I}_{\mathfrak{r}} \left(\boxminus_i^n \nu_i \right) (\mathfrak{x}) - \left[\frown_{i=1}^n f_i \right] (\mathfrak{x}) \right\|_1 \leqslant \sum_{i=1}^n \varepsilon_i \qquad (7.4.4)$$

Proof. We will prove this with induction. This is straight-forward for the case where we have just one neural network where for all $x \in \mathbb{R}^{\mathfrak{l}(\nu_1)}$ it is the case that $\left\| \mathfrak{I}_{\mathfrak{r}} (\nu_1) (x) - f(x) \right\|_1 \leqslant \varepsilon_1 = \sum_{i=1}^1 \varepsilon_i$. Suppose now, that, $\left\| \mathfrak{I}_{\mathfrak{r}} \left(\boxminus_i^n \nu_i \right) (\mathfrak{x}) - \left[\frown_{i=1}^n f_i \right] (\mathfrak{x}) \right\|_1 \leqslant \sum_{i=1}^n \varepsilon_i$ holds true for all cases upto and including n. Consider what happens when we have a triple, a function f_{n+1}, a neural network ν_{n+1}, and $\varepsilon_{n+1} \in (0, \infty)$ with a maximum error over all $x \in \mathbb{R}^{\mathfrak{l}(\nu_1)}$ of $\left| \mathfrak{I}_{\mathfrak{r}} (\nu_{n+1}) (x) - f_{n+1} (x) \right| \leqslant \varepsilon_{n+1}$. Then Lemma 4.3.3, Corollary 1.2.35.1, and the triangle inequality tells us that:

$$\begin{aligned}
& \left\| \mathfrak{I}_{\mathfrak{r}} \left(\boxminus_i^{n+1} \nu_i \right) (\mathfrak{x}) - \left[\frown_{i=1}^{n+1} f_i \right] (\mathfrak{x}) \right\|_1 \\
& \leqslant \left\| \mathfrak{I}_{\mathfrak{r}} \left(\boxminus_i^n \nu_i \right) (\mathfrak{x}) - \left[\frown_{i=1}^n f_i \right] (\mathfrak{x}) \right\|_1 + \left| \mathfrak{I}_{\mathfrak{r}} (\nu_{n+1}) (x) - f_{n+1} (x) \right| \\
& \leqslant \sum_{i=1}^{n+1} \varepsilon_i \qquad\qquad (7.4.5)
\end{aligned}$$

This proves the inductive case and hence the Lemma. $\qquad\square$

Lemma 7.4.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X_d : \Omega \to \mathbb{R}_d$ be a random variable. Let $f : \mathbb{R}_d \to \mathbb{R}$ be a function such that for all $x, \mathfrak{x} \in \mathbb{R}^d$ it is the case that $\left\| f(x) - f(\mathfrak{x}) \right\|_E \leqslant \mathfrak{L} |x - \mathfrak{x}|$. It is then the case that $\mathbb{V} [f(X_d)] \leqslant 2\mathfrak{L}^2 \mathbb{V} [X_d]$.*

Proof. Let $\mathfrak{X}_d$ be an i.i.d. copy of X_d. As such it is the case that $\mathrm{Cov} (X_d, \mathfrak{X}_d) = 0$, whence it is the case that $\mathbb{V} [X_d, \mathfrak{X}_d] = \mathbb{V} [X_d] + \mathbb{V} [\mathfrak{X}_d] = \mathbb{V}[X_d] + \mathbb{V} [-\mathfrak{X}_d] = \mathbb{V} [X_d - \mathfrak{X}_d] = 2 \mathbb{V} [X_d]$. Note that $f(X_d)$ and $f(\mathfrak{X}_d)$ are also indepentend and thus $\mathrm{Cov} (f(X_d), f(\mathfrak{X}_d)) = 0$, and

whence we get that $\mathbb{V}[f(X_d) - f(\mathfrak{X}_d)] = 2\,\mathbb{V}[\mathfrak{X}_d]$. This then yields that:

$$
\begin{aligned}
2\,\mathbb{V}[f(X_d)] &= \mathbb{V}[f(X_d) - f(\mathfrak{X}_d)] \\
&= \mathbb{E}\left[(f(X_d) - f(\mathfrak{X}_d))^2\right] - \left(\mathbb{E}[f(X_d) - f(\mathfrak{X}_d)]\right)^2 \\
&= \mathbb{E}\left[(f(X_d) - f(\mathfrak{X}_d))^2\right] \\
&= \mathfrak{L}^2 \cdot \mathbb{E}\left[(X_d - \mathfrak{X}_d)^2\right] \\
&= \mathfrak{L}^2 \cdot 2\,\mathbb{V}[X_d] \\
\implies \mathbb{V}[f(X_d)] &= \mathfrak{L}^2 \cdot \mathbb{V}[X_d]
\end{aligned}
\tag{7.4.6}
$$

This proves the Lemma. $\square$

Lemma 7.4.4 (R—, 2024, Approximants for Brownian Motion)**.** *Let $t \in (0, \infty)$ and $T \in (t, \infty)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $n, N \in \mathbb{N}$, and $h \in (0, \infty)$. Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$, satisfy that $\delta = \varepsilon(2^{q-1} + 1)^{-1}$. Let $f : [t, T] \to \mathbb{R}$ be continuous almost everywhere in $[t, T]$. Let it also be the case that $f = g \circ \mathfrak{h}$, where $\mathfrak{h} : [t, T] \to \mathbb{R}^d$, and $g : \mathbb{R}^d \to \mathbb{R}$. Let $t = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_{N-1} \leqslant t_N = T$ such that for all $i \in \{0, 1, ..., N\}$ it is the case that $h = \frac{T-t}{N}$, and $t_i = t_0 + i \cdot h$. Let $\mathbf{t} = [t_0\, t_1 \cdots t_N]$ and as such let $f\left([\mathbf{t}]_{*,*}\right) = [f(t_0)\, f(t_1)\, \cdots\, f(t_N)]$. Let $u_d \in C\left(\mathbb{R}^d, \mathbb{R}\right)$, bounded by $\mathfrak{M}_{u,d}$ satisfy for all $d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ that:*

$$
\left(\frac{\partial}{\partial t} u_d\right)(t, x) + \left(\nabla_x^2 u_d\right)(t, x) + \alpha_d(x)\, u_d(t, x) = 0
\tag{7.4.7}
$$

Furthermore, let $\mathfrak{u}_d^T(x) = u_d(T, x)$. Let $\mathsf{G}_d \subsetneq \mathsf{NN}$ be the neural network which instantiates as $\mathfrak{u}_d^T = \mathfrak{I}_\mathfrak{r}(\mathsf{G}_d) \in C\left(\mathbb{R}^d, \mathbb{R}\right)$.

Let $\mathcal{W}^d : [0, T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$ be standard Brownian motions, and let $\mathcal{X}^{d,t,x} : [t, T] \times \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ be stochastic processes with continuous sample paths satisfying

229

that for all $d \in \mathbb{N}$, $t \in [0, T]$, $s \in [t, T]$, $x \in \mathbb{R}^d$ we have $\mathbb{P}$-a.s, that:

$$\mathcal{X}_s^{d,t,x} = x + \int_s^t \sqrt{2}d\mathcal{W}_r^d \tag{7.4.8}$$

It is then the case that for all $d \in \mathbb{N}$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that:

$$u_d(t, x) = \mathbb{E}\left[\exp\left(\int_t^T \left(\alpha_d \circ \mathcal{X}_r^{d,t,x}\right) dr\right) u_d\left(T, \mathcal{X}_T^{d,t,x}\right)\right] \tag{7.4.9}$$

Let $\mathsf{E}_n^{N,h,q,\varepsilon} \subsetneq \mathsf{NN}$ be neural networks given by:

$$\mathsf{E}_n^{N,h,q,\varepsilon} = \mathsf{Xpn}_n^{q,\varepsilon} \bullet \mathsf{Etr}^{N,h} \tag{7.4.10}$$

Furthermore, let $\mathsf{G}_d \in \mathsf{NN} \subsetneq \mathsf{NN}$ be neural networks which instantiate as $u_d = \mathfrak{I}_{\mathfrak{r}}(\mathsf{G}_d) \in C\left(\mathbb{R}^d, \mathbb{R}\right)$.

Furthermore, let $\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \subsetneq \mathsf{NN}$ be neural networks given by:

$$\mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} = \mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{E}_n^{N,h,q,\varepsilon} \diamondsuit \mathsf{G}_d\right] \tag{7.4.11}$$

Futhermore, let $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} \subsetneq \mathsf{NN}$ be neural networks given by:

$$\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon} = \mathsf{UE}_{n,\mathsf{G}_d}^{N,h,q,\varepsilon} \bullet \left[\mathsf{Tun}_1^{N+1} \boxminus \mathsf{Aff}_{\mathbb{0}_{d,d},\mathcal{X}_{\omega_i}}\right] \tag{7.4.12}$$

Finally let $\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon} \subsetneq \mathsf{NN}$ be neural networks which, for $\omega_i \in \Omega$, is given as:

$$\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon} = \frac{1}{\mathfrak{n}} \triangleright \left(\mathsf{Sum}_{\mathfrak{n},1} \bullet \left[\boxminus_{i=1}^{\mathfrak{n}} \mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right]\right) \tag{7.4.13}$$

It is then the case that for all $\mathfrak{X} \in \mathbb{R}^{\mathfrak{n}(N+1)} \times \mathbb{R}^{\mathfrak{n}d}$:

(i) $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right)(\mathfrak{X}) \in C\left(\mathbb{R}^{\mathfrak{n}(N+1)} \times \mathbb{R}^{\mathfrak{n}d}, \mathbb{R}\right)$

(ii) $\mathsf{D}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right) \leqslant \begin{cases} \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q\right]+\mathsf{D}\left(\mathsf{G}_d\right)-1 & : n=0 \\ \\ \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q\right]+\max\left\{\mathsf{D}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right),\mathsf{D}\left(\mathsf{G}_d\right)\right\}-1 & : n\in\mathbb{N} \end{cases}$

(iii) It is also the case that:

$$\mathsf{P}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right) \leqslant \mathfrak{n}^2\cdot\left[\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q+1\right]+324+48n\right.$$

$$\left.+24\,\mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d\right)+4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right),\mathsf{P}\left(\mathsf{G}_d\right)\right\}\right] \qquad (7.4.14)$$

(iv) It is also the case that:

$$\left|\frac{1}{\mathfrak{n}}\left[\sum_{i=1}^{\mathfrak{n}}\left[\exp\left(\int_t^T f\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)ds\cdot u_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right)\right]\right]-\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right)(\mathfrak{X})\right|$$

$$\leqslant 3\varepsilon+2\varepsilon\left|\mathfrak{u}_d^T(x)\right|^q+2\varepsilon\left|\exp\left(\int_a^b fdx\right)\right|^q+\varepsilon\left|\exp\left(\int_a^b fdx\right)-\mathfrak{e}\right|^q-\mathfrak{e}\mathfrak{u}_d^T(x)$$

Where, as per Lemma 7.1.1, $\mathfrak{e}$ is defined as:

$$\left|\mathsf{E}_n^{N,h,q,\varepsilon}-\exp\left(\int_a^b fdx\right)\right|\leqslant\mathfrak{e} \qquad (7.4.15)$$

Proof. Note that for all $i\in\{1,2,\ldots,\mathfrak{n}\}$, Lemma 7.3.1 tells us that $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)$

$\in C\left(\mathbb{R}^{N+1}\times\mathbb{R}^d,\mathbb{R}\right)$. Lemma 4.5.19 and Lemma 4.5.18, thus tells us that

$\mathfrak{I}_{\mathfrak{r}}\left(\left(\bigoplus_{i=1}^{\mathfrak{n}}\left[\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right]\right)\right)=\sum_{i=1}^{\mathfrak{n}}\left[\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)\right]$. The sum of continuous functions is

continuous. Note next that $\frac{1}{\mathfrak{n}}\triangleright$ is an affine neural network, and hence, by Lemma 4.4.2, must

be continuous.

Then Lemmas 4.2.5, 4.3.8, and the fact that by Lemma 7.3.1 each of the individual stacked

$\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$ neural networks is continuous then ensures us that it must therefore be the case

that: $\mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right)\in C\left(\mathbb{R}^{\mathfrak{n}(N+1)}\times\mathbb{R}^{\mathfrak{n}d},\mathbb{R}\right)$. This proves Item (i).

Next note that by construction each $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$ has the same depth, indeed for each i the

only thing different for each of the $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$ is the parameters themselves and not the count

or depth or layer architecture. Note that $\mathsf{D}\left(\mathsf{Sum}_{\mathfrak{n},1}\right) = \mathsf{D}\left(\frac{1}{\mathfrak{n}}\triangleright\right) = \mathsf{D}\left(\mathsf{Aff}_{\frac{1}{\mathfrak{n}},0}\right) = 1$.

Whence by Lemma 4.2.5 it is the case that $\mathsf{D}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right) = \mathsf{D}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)$. This then proves Item (ii).

Next, observe that each of the $\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}$ networks has the same architecture for all $\omega_i \in \Omega$ by construction. Corollary 4.3.5.1 then yields that:

$$\mathsf{P}\left(\boxminus_{i=1}^{\mathfrak{n}}\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right) \leqslant \mathfrak{n}^2 \cdot \mathsf{P}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right) \tag{7.4.16}$$

Note for instance also that by Remark 4.4.3, it is the case that $\mathsf{P}\left(\mathsf{Sum}_{\mathfrak{n},1}\right) = \mathfrak{n} + 1$. Furthermore, since the output of the Sum neural network has length one, by Definition 4.6.1 it is the case that $\mathsf{P}\left(\frac{1}{\mathfrak{n}}\triangleright\right) = 2$. Then Corollary 4.4.5.1 leads us to conclude that:

$$\mathsf{P}\left(\mathsf{Sum}_{\mathfrak{n},1} \bullet \left[\boxminus_{i=1}^{\mathfrak{n}}\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right]\right) \leqslant \mathsf{P}\left(\boxminus_{i=1}^{\mathfrak{n}}\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)$$
$$\leqslant \mathfrak{n}^2 \cdot \mathsf{P}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)$$

$$\tag{7.4.17}$$

and therefore that:

$$\mathsf{P}\left(\frac{1}{\mathfrak{n}}\triangleright\left(\mathsf{Sum}_{\mathfrak{n},1} \bullet \left[\boxminus_{i=1}^{\mathfrak{n}}\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right]\right)\right)$$
$$\leqslant \mathsf{P}\left(\mathsf{Sum}_{\mathfrak{n},1} \bullet \left[\boxminus_{i=1}^{\mathfrak{n}}\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right]\right)$$
$$\leqslant \mathsf{P}\left(\boxminus_{i=1}^{\mathfrak{n}}\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)$$
$$\leqslant \mathfrak{n}^2 \cdot \mathsf{P}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)$$
$$\leqslant \mathfrak{n}^2 \cdot \left[\frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 324 + 48n\right.$$
$$\left. + 24\,\mathsf{W}_{\mathsf{H}(\mathsf{G}_d)}\left(\mathsf{G}_d\right) + 4\max\left\{\mathsf{P}\left(\mathsf{E}_n^{N,h,q,\varepsilon}\right),\mathsf{P}\left(\mathsf{G}_d\right)\right\}\right] \tag{7.4.18}$$

Observe that the absolute homogeneity condition for norms, the fact that the Brownian

motions are independent of each other, Lemma 7.4.1, the fact that $\mathfrak{n} \in \mathbb{N}$, the fact that the upper limit of error remains bounded by the same bound for all $\omega_i \in \Omega$, and Lemma 7.4.2, then yields us:

$$
\left| \frac{1}{\mathfrak{n}} \left[\sum_{i=1}^{\mathfrak{n}} \left[\exp\left(\int_t^T f\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right) ds \cdot u_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right)\right]\right] - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right)\right|
$$

$$
= \left| \frac{1}{\mathfrak{n}} \left[\sum_{i=1}^{\mathfrak{n}} \left[\exp\left(\int_t^T f\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right) ds \cdot u_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right)\right]\right] - \mathfrak{I}_{\mathfrak{r}}\left[\frac{1}{\mathfrak{n}} \triangleright \left(\mathsf{Sum}_{\mathfrak{n},1} \bullet \left[\boxminus_{i=1}^{\mathfrak{n}} \mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right]\right)\right]\right|
$$

$$
\leqslant \left| \frac{1}{\mathfrak{n}} \left[\sum_{i=1}^{\mathfrak{n}} \exp\left(\int_t^T f\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right) ds \cdot u_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right)\right] - \frac{1}{\mathfrak{n}} \left[\sum_{i=1}^{\mathfrak{n}} \left(\mathfrak{I}_{\mathfrak{r}}\left[\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right]\right)\right]\right|
$$

$$
\leqslant \frac{1}{\mathfrak{n}} \sum_{i=1}^{\mathfrak{n}} \left| \exp\left(\int_t^T f\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right) ds \cdot u_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)\right|
$$

$$
= \left| \exp\left(\int_t^T f\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right) ds \cdot u_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right) - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UEX}_{n,\mathsf{G}_d,\omega_i}^{N,h,q,\varepsilon}\right)\right|
$$

$$
\leqslant 3\varepsilon + 2\varepsilon \left|\mathfrak{u}_d^T(t,x)\right|^q + 2\varepsilon \left|\exp\left(\int_a^b f dx\right)\right|^q + \varepsilon \left|\exp\left(\int_a^b f dx\right) - \mathfrak{e}\right|^q - \mathfrak{e}\mathfrak{u}_d^T(x)
$$

$\square$

Corollary 7.4.4.1. *Let* $N, n, \mathfrak{n} \in \mathbb{N}$, $h, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$, *given* $\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon} \subsetneq \mathsf{NN}$, *it is then the case that:*

$$
\left(\mathbb{E}\left[\left| \mathbb{E}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right]\right.\right.\right.
$$

$$
\left.\left.\left. - \frac{1}{\mathfrak{n}} \left(\sum_{i=1}^{\mathfrak{n}} \left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\omega_i}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right]\right) \right|^2 \right]\right)^{\frac{1}{2}}
$$

$$
\leqslant \frac{\mathfrak{k}_2}{\mathfrak{n}^{\frac{1}{2}}} \left(\mathbb{E}\left[\left| \exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right|^2\right]\right)^{\frac{1}{2}} \tag{7.4.19}
$$

Proof. Note that $\mathbb{E}\left[\mathcal{X}_{r,\Omega}^{d,tx}\right] < \infty$, and $\mathfrak{u}^T$ being bounded yields that $\mathbb{E}\left[\mathfrak{u}^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right] < \infty$, and also that $\mathbb{E}\left[\alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x}\right] < \infty$. Thus we also see that $\mathbb{E}\left[\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right] < \infty$, and thus $\mathbb{E}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right)\right] < \infty$. Thus together these two facts, along with the fact that the two factors are independent by (Hutzenthaler et al., 2020a, Corollary 2.5), then assert that

$$\mathbb{E}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x}\right) \cdot \mathfrak{u}^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right] < \infty.$$

Note that (Hutzenthaler et al., 2021, Corollary 3.8) tells us that:

$$\left(\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right]\right.\right.\right.$$
$$\left.\left.\left. -\frac{1}{\mathfrak{n}}\left(\sum_{i=1}^{\mathfrak{n}}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\omega_i}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right]\right)\right|^2\right]\right)^{\frac{1}{2}}$$
$$\leqslant \frac{\mathfrak{k}_2}{\mathfrak{n}^{\frac{1}{2}}}\left(\mathbb{E}\left[\left|\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right|^2\right]\right)^{\frac{1}{2}}$$
$$\leqslant \frac{\mathfrak{k}}{\mathfrak{n}^{\frac{1}{2}}}\left(\exp\left((T-t)\mathfrak{M}_{\alpha,d}\right)\mathfrak{M}_{u,d}\right) \tag{7.4.20}$$

This, combined with Lyapunov's Inequality for Expectation and the Triangle Inequality yields that:

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right]\right.\right.$$
$$\left.\left. -\frac{1}{\mathfrak{n}}\left(\sum_{i=1}^{\mathfrak{n}}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\omega_i}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\omega_i}^{d,t,x}\right)\right]\right)\right|\right]$$
$$\leqslant \frac{\mathfrak{k}}{\mathfrak{n}^{\frac{1}{2}}}\left(\exp\left((T-t)\mathfrak{M}_{\alpha,d}\right)\mathfrak{M}_{u,d}\right) \tag{7.4.21}$$

Finally, combined with, the linearity of expectation, and the fact that the expectation of a deterministic constant and a deterministic function is, respectively, the constant and function itself, and the triangle inequality then yields that:

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left(\int_t^T \alpha_d \circ \mathcal{X}_{r,\Omega}^{d,t,x} ds\right) \cdot \mathfrak{u}_d^T\left(\mathcal{X}_{r,\Omega}^{d,t,x}\right)\right] - \mathfrak{I}_{\mathfrak{r}}\left(\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}\right)\right|\right]$$
$$\leqslant \frac{\mathfrak{k}}{\mathfrak{n}^{\frac{1}{2}}}\left(\exp\left((T-t)\mathfrak{M}_{\alpha,d}\right)\mathfrak{M}_{u,d}\right)$$
$$+ 3\varepsilon + 2\varepsilon\left|\mathfrak{u}_d^T\left(t,x\right)\right|^q + 2\varepsilon\left|\exp\left(\int_a^b f dx\right)\right|^q + \varepsilon\left|\exp\left(\int_a^b f dx\right) - \mathfrak{e}\right|^q - \mathfrak{e}\mathfrak{u}_d^T\left(x\right)$$

This completes the proof of the corollary. $\qquad\square$

Remark 7.4.5. *Note that diagrammatically, this can be represented as in figure below.*
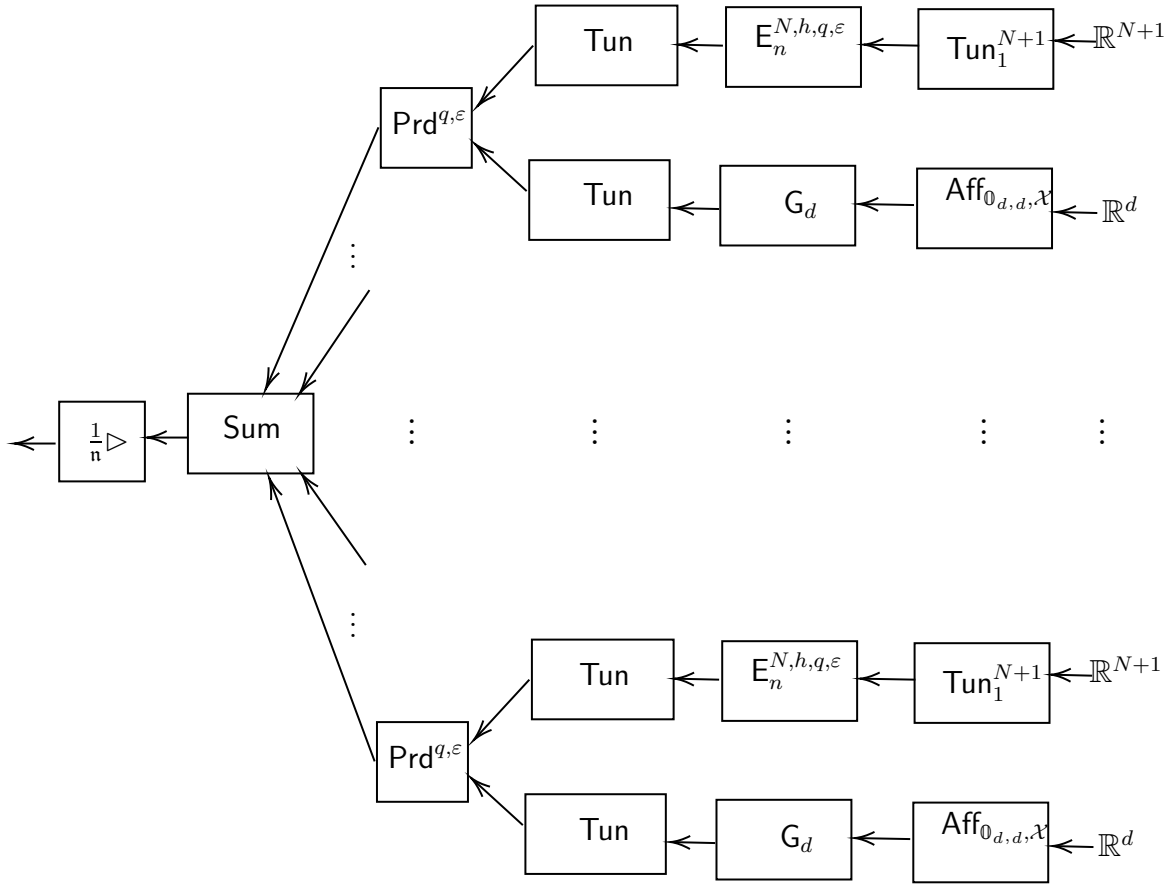
Figure 7.4: Neural network diagram for the $\mathsf{UES}_{n,\mathsf{G}_d,\Omega,\mathfrak{n}}^{N,h,q,\varepsilon}$ network.

Chapter 8

Conclusions and Further Research

We will present three avenues of further research and related work on parameter estimates here. We will present these as a series of recommendations and conjectures to further extend this framework for understanding neural networks.

8.1 Further operations

Note, for instance, that several classical operations are done on neural networks that have yet to be accounted for in this framework and talked about in the literature. We will discuss two of them *dropout* and *merger* and discuss how they may be brought into this framework. Overfitting presents an important challenge for all machine learning models, including deep learning. There ex

Definition 8.1.1 (Hadamard Product)**.** *Let* $m, n \in \mathbb{N}$*. Let* $A, B \in \mathbb{R}^{m \times n}$*. For all* $i \in \{1, 2, \ldots, m\}$ *and* $j \in \{1, 2, \ldots, n\}$ *define the Hadamard product* $\odot : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ *as:*

$$A \odot B := [A \odot B]_{i,j} = [A]_{i,j} \times [B]_{i,j} \quad \forall i, j \tag{8.1.1}$$

We will also define the dropout operator introduced in Srivastava et al. (2014), and explained

further in Goodfellow et al. (2016).

Definition 8.1.2 (Instantiation with dropout). *Let $\nu \in \mathsf{NN}$, $L, n \in \mathbb{N}$, $p \in (0,1)$, $\mathsf{L}(\nu) = (l_0, l_1, \ldots, B_L)$, and that $\nu = ((W_1, b_1), (W_2, b_2), \ldots, (W_L, b_L))$. Let it be the case that for each $n \in \mathbb{N}$, $\rho_n = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n$ where for each $i \in \{1, 2, \ldots, n\}$ it is the case that $x_i \sim \mathrm{Bern}(p)$. We will then denote $\mathfrak{I}_{\mathfrak{a}}^{D,p}(\nu) \in C\left(\mathbb{R}^{\mathsf{I}(\nu)}, \mathbb{R}^{\mathsf{O}(\nu)}\right)$, the continuous function given by:*

$$\mathfrak{I}_{\mathfrak{a}}^{D,p}(\nu) = \rho_{l_L} \odot \mathfrak{a}\left(W_l\left(\rho_{l_{L-1}} \odot \mathfrak{a}\left(W_{L-1}\left(\ldots\right) + b_{L-1}\right)\right) + b_L\right) \qquad (8.1.2)$$

Dropout is an example of *ensemble learning*, a form of learning where versions of our model (e.g. random forests or neural networks) are made (e.g. by dropout for neural networks or by enforcing a maximum depth to the trees in our forest), and a weighted average of the predictions of our different models is taken to be the predictive model. That such a model can work, and indeed work well, is the subject of Schapire (1990).

8.2 Further Approximants

In theory the approximation schemes given in the case of $\mathsf{Xpn}_n^{q,\varepsilon}$, $\mathsf{Csn}_n^{q,\varepsilon}$, and $\mathsf{Sne}_n^{q,\varepsilon}$ given in the previous sections, could be used to approximate more transcendental functions, and identities such as alluded to in Remark 5.2.31. Indeed, recent attempts have been made to approximate backwards and forward Euler methods as in Grohs et al. (2023). In fact, this architecture was originally envisioned to approximate, Multi-Level Picard iterations, as seen in Ackermann et al. (2023). These neural network methods have been proven to beat the curse of dimensionality in the sense that the size of these networks (parameter and depth counts) grow only polynomially with respect to the desired accuracy. In practice, it remains to be seen whether for larger dimensions, the increased number of operations and architectures to contend with do not make up for the polynomial increase in parameter and depths, especially when it comes to computaiton time.

In a similar vein, these architectures have so far lacked a consistent implementation in a widely available programming language. Part of the dissertation work has been focused on implementing these architectures as an R package, available at CRAN.

8.3   Algebraic Properties of this Framework

It is quite straightforward to see that the instantiation operation has sufficiently functorial properties, at the very least, when instantiating with the identity function. More specifically consider the category Mat whose objects are natural numbers, m, n, and whose arrows $m \xleftarrow{A} n$ are matrices $A \in \mathbb{R}^{m \times n}$, i.e. a continuous function between vector spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Consider as well the set of neural networks $\nu \subsetneq$ NN where $\mathsf{I}(\nu) = n$ and $\mathsf{O}(\nu) = m$.

In such a case, note that the instantiation operation preserves the axiom of functoriality, namely that composition is respected under instantiation. Note also that we have alluded to the fact that under neural network composition, with Id (the appropriate one for our dimension) behaves like a monoid under instantiation.

Note for example that a neural network analog for derivatives, one that respects the chain rule under instantiation already exist in the literature, e.g. Berner et al. (2019). Thus there is a growing and rather rich and growing set of algebraic operations that are and have been proposed for neural networks.

Taken together, these facts seem to imply that a further exploration of the algebraic properties of this artificial neural network framework could present a fruitful avenue of future study. Much remains to be studied.

This completes this Dissertation.

Chapter 9

Bibliography and Code Listings

Bibliography

Ackermann, J., Jentzen, A., Kruse, T., Kuckuck, B., and Padgett, J. L. (2023). Deep neural networks with relu, leaky relu, and softplus activation provably overcome the curse of dimensionality for kolmogorov partial differential equations with lipschitz nonlinearities in the l^p-sense.

Arik, S. n and Pfister, T. (2021). TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8):6679--6687. Number: 8.

Beck, C., Gonon, L., Hutzenthaler, M., and Jentzen, A. (2021a). On existence and uniqueness properties for solutions of stochastic fixed point equations. *Discrete & Continuous Dynamical Systems - B*, 26(9):4927.

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

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

Berner, J., Elbrächter, D., Grohs, P., and Jentzen, A. (2019). Towards a regularity theory for relu networks – chain rule and global error estimates. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, page 1—5.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 1800–1807, Los Alamitos, CA, USA. IEEE Computer Society.

Crandall, M. G., Ishii, H., and Lions, P.-L. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc.*, 27(1):1–67.

Da Prato, G. and Zabczyk, J. (2002). *Second Order Partial Differential Equations in Hilbert Spaces*. London Mathematical Society Lecture Note Series. Cambridge University Press.

Davies, A., Juhasz, A., Lackenby, M., and Tomasev, N. (2022). The signature and cusp geometry of hyperbolic knots. *Geometry and Topology*. Publisher: Mathematical Sciences Publishers.

Durrett, R. (2019). *Probability: Theory and Examples.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. (2019). On Multilevel Picard Numerical Approximations for High-Dimensional Nonlinear Parabolic Partial Differential Equations and High-Dimensional Nonlinear Backward Stochastic Differential Equations. *J Sci Comput*, 79(3):1534--1571.

E, W., Hutzenthaler, M., Jentzen, A., and Kruse, T. (2021). Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. *Partial Differ. Equ. Appl.*, 2(6):80.

Golub, G. and Van Loan, C. (2013). *Matrix Computations.* Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning.* MIT Press. `http://www.deeplearningbook.org`.

Graham, R., Knuth, D., and Patashnik, O. (1994). *Concrete Mathematics: A Foundation for Computer Science.* Addison-Wesley Professional, Upper Saddle River, NJ, 2nd edition edition.

Grohs, P., Hornung, F., Jentzen, A., and von Wurstemberger, P. (2018). A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. Papers 1809.02362, arXiv.org.

Grohs, P., Hornung, F., Jentzen, A., and Zimmermann, P. (2023). Space-time error estimates for deep neural network approximations for differential equations. *Advances in Computational Mathematics*, 49(1):4.

Grohs, P., Jentzen, A., and Salimova, D. (2022). Deep neural network approximations for solutions of PDEs based on monte carlo algorithms. *Partial Differential Equations and Applications*, 3(4).

Hutzenthaler, M., Jentzen, A., Kruse, T., Anh Nguyen, T., and von Wurstemberger, P. (2020a). Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2244):20190630.

Hutzenthaler, M., Jentzen, A., Kuckuck, B., and Padgett, J. L. (2021). Strong L^p-error analysis of nonlinear Monte Carlo approximations for high-dimensional semilinear partial differential equations. Technical Report arXiv:2110.08297, arXiv. arXiv:2110.08297 [cs, math] type: article.

Hutzenthaler, M., Jentzen, A., and von Wurstemberger (2020b). Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *Electronic Journal of Probability*, 25(none):1–73.

Itô, K. (1942a). Differential equations determining Markov processes (original in Japanese). *Zenkoku Shijo Sugaku Danwakai*, 244(1077):1352–1400.

Itô, K. (1942b). On a stochastic integral equation. *Proc. Imperial Acad. Tokyo*, 244(1077):1352–1400.

Jentzen, A., Kuckuck, B., and von Wurstemberger, P. (2023). Mathematical introduction to deep learning: Methods, implementations, and theory.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus.* Graduate Texts in Mathematics (113) (Book 113). Springer New York.

Petersen, P. and Voigtlaender, F. (2018). Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Netw*, 108:296–330.

Rafi, S. (2024). nnR.

Rafi, S. and Padgett, J. L. (2024). *nnR: Neural Networks Made Algebraic.* R package version 0.1.0.

Rio, E. (2009). Moment Inequalities for Sums of Dependent Random Variables under Projective Conditions. *J Theor Probab*, 22(1):146–163.

Schapire, R. E. (1990). The strength of weak learnability. *Mach Learn*, 5(2):197–227.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929--1958.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in nlp.

Tsaban, T., Varga, J. K., Avraham, O., Ben-Aharon, Z., Khramushin, A., and Schueler-Furman, O. (2022). Harnessing protein folding neural networks for peptide–protein docking. *Nat Commun*, 13(1):176. Number: 1 Publisher: Nature Publishing Group.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, î, and Polosukhin, I. (2017). Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Behram, F. A., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H. S., Akyildiz, B., Balandat, M., Spisak, J., Jain, R., Rabbat, M., and Hazelwood, K. (2022). Sustainable ai: Environmental implications, challenges and opportunities.

Yarotsky, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94:103--114.

Zhao, T., Lyu, R., Wang, H., Cao, Z., and Ren, Z. (2023). Space-based gravitational wave signal detection and extraction with deep neural network. *Commun Phys*, 6(1):1–12. Number: 1 Publisher: Nature Publishing Group.

9.1 Code Listings

Parts of this code have been released on CRAN under the package name nnR, and can be found in Rafi and Padgett (2024), with the corresponding repository being found at Rafi (2024):

Listing 9.1: R code for neural network generation

```r
#' Function to generate a random matrix with specified dimensions.
#'
#' @param rows number of rows.
#' @param cols number of columns.
#'
#' @return a random matrix of dimension rows times columns with elements from
#' a standard normal distribution

generate_random_matrix <- function(rows, cols) {
  (rows * cols) |>
    rnorm() |>
    matrix(rows, cols) -> result
  return(result)
}

#' @title create_neural_network
#' @description Function to create a list of lists for neural network layers
#'
#' @param layer_architecture a list specifying the width of each layer
#'
#' @return An ordered list of ordered pairs of \eqn{(W,b)}. Where \eqn{W} is
#'    the matrix
#' representing the weight matrix at that layer and \eqn{b} the bias vector.
#'    Entries
#' on the matrix come from a standard normal distribution. Neural networks
#' are defined to be elements belonging to the following set:
#' \deqn{
#'
#' \mathsf{NN} = \bigcup_{L\in \N} \bigcup_{l_0,l_1,...,l_L \in \N}
#' \left( \times^L_{k=1} \left[ \mathbb{R}^{l_k \times l_{k-1}} \times \R^{l_k
#'    }\right]  \right)
#'
#'
#' }
#'
#'
#'
#'
#' We will use the definition of neural networks as found in:
#'
#' @references Grohs, P., Hornung, F., Jentzen, A. et al.
#' Space-time error estimates for deep neural network approximations
#' for differential equations. Adv Comput Math 49, 4 (2023).
#' \url{https://doi.org/10.1007/s10444-022-09970-2}.
#' @export

```

```r
create_neural_network <- function(layer_architecture) {
  if (all(sapply(layer_architecture, function(x) is.numeric(x) && x %% 1 == 0
      && x > 0)) == FALSE) {
    stop("Non_integer_or_negative_neural_network_width_specified.")
  } else if (layer_architecture |> length() < 2) {
    stop("Neural_network_must_have_atleast_two_layers.")
  } else {
    layer_architecture |> length() -> L

    # Initialize the list of lists
    neural_network <- list()

    # Generate matrices W and vectors b for each layer
    for (i in 1:(L - 1)) {
      # Set dimensions for W and b
      layer_architecture[i] -> input_size
      layer_architecture[i + 1] -> output_size

      # Create matrix W
      generate_random_matrix(output_size, input_size) -> W

      # Create vector b
      output_size |>
        rnorm() |>
        matrix(output_size, 1) -> b

      # Add W and b to the list
      list(W = W, b = b) -> neural_network[[i]]
    }

    return(neural_network)
  }
}
```

Listing 9.2: R code for auxilliary functions

```r
source("R/is_nn.R")
#' @title hid
#'
#' @description The function that returns the number of hidden layers of a
#' neural network. Denoted \eqn{\mathsf{H}}
#'
#' @param nu a neural network of the type generated by create_neural_network()
#'
#' By definition \eqn{\mathsf{H}(\nu) = \mathsf{D}(\nu) - 1}
#'
#' @references Definition 1.3.1. Jentzen, A., Kuckuck, B., and von
#'     Wurstemberger, P. (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}.
#'
#' @return Integer representing the number of hidden layers.
#' @export

```

```r
hid <- function(nu) {
  if (nu |> is_nn() == TRUE) {
    return(length(nu) - 1)
  } else {
    stop("Only_neural_networks_can_have_hidden_layers")
  }
}

#' @title dep
#' @description The function that returns the depth of a neural network.
    Denoted
#' \eqn{\mathsf{D}}.
#'
#' @param nu a neural network of the type generated by
#' create_neural_network(). Very straightforwardly it is the
#' length of the list where neural networks are defined as an odered list of
#' lists.
#' @references Definition 1.3.1. Jentzen, A., Kuckuck, B., and von
    Wurstemberger, P. (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}.
#'
#' @return Integer representing the depth of the neural network.
#' @export

dep <- function(nu) {
  if (nu |> is_nn() == TRUE) {
    return(length(nu))
  } else {
    stop("Only_neural_networks_can_have_depth")
  }
}

#' @title inn
#' @description The function that returns the input layer size of a neural
#' network. Denoted \eqn{\mathsf{I}}
#'
#' @param nu A neural network of the type generated by
#' create_neural_network().
#'
#' @references Definition 1.3.1. Jentzen, A., Kuckuck, B., and von
    Wurstemberger, P. (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}.
#'
#' @return An integer representing the input width of the neural network.
#' @export

inn <- function(nu) {
  if (nu |> is_nn() == TRUE) {
    return(dim(nu[[1]]$W)[2])
  } else {
    stop("Only_neural_networks_can_have_size_of_input_layers")
```

```
68      }
69 }
70
71
72 #' @title out
73 #' @description The function that returns the output layer size of a neural
74 #' network. Denoted \eqn{\mathsf{O}}.
75 #'
76 #' @param nu A neural network of the type generated by create_neural_network()
        .
77 #'
78 #' @references Definition 1.3.1. Jentzen, A., Kuckuck, B., and von
        Wurstemberger, P. (2023).
79 #' Mathematical introduction to deep learning: Methods, implementations,
80 #' and theory. \url{https://arxiv.org/abs/2310.20360}.
81 #'
82 #' @return An integer representing the output width of the neural network.
83 #' @export
84
85 out <- function(nu) {
86    if (nu |> is_nn() == TRUE) {
87       return(dim(nu[[length(nu)]]$W)[1])
88    } else {
89       stop("Ony_neural_networks_can_have_size_of_output_layers")
90    }
91 }
92
93
94 #' @title lay
95 #' @description The function that returns the layer architecture of a neural
96 #' network.
97 #'
98 #' @param nu A neural network of the type generated by
99 #' create_neural_network(). Denoted \eqn{\mathsf{L}}.
100 #'
101 #' @references Definition 1.3.1. Jentzen, A., Kuckuck, B., and von
        Wurstemberger, P. (2023).
102 #' Mathematical introduction to deep learning: Methods, implementations,
103 #' and theory. \url{https://arxiv.org/abs/2310.20360}.
104 #'
105 #' @return A tuple representing the layer architecture of our neural network.
106 #' @export
107
108 lay <- function(nu) {
109    if (nu |> is_nn() == TRUE) {
110       layer_architecture <- list()
111       for (i in 1:length(nu)) {
112          layer_architecture |> append(dim(nu[[i]]$W)[1]) -> layer_architecture
113       }
114       inn(nu) |> append(layer_architecture) -> layer_architecture
115       return(layer_architecture)
116    } else {
117       stop("Only_neural_networks_can_have_layer_architectures")
```

```
118      }
119 }
120
121
122 #' @title param
123 #' @description The function that returns the numbe of parameters of a neural
124 #' network.
125 #'
126 #' @param nu A neural network of the type generated by
127 #' create_neural_network(). Denoted \eqn{\mathsf{P}}.
128 #'
129 #' @references Definition 1.3.1. Jentzen, A., Kuckuck, B., and von
        Wurstemberger, P. (2023).
130 #' Mathematical introduction to deep learning: Methods, implementations,
131 #' and theory. \url{https://arxiv.org/abs/2310.20360}.
132 #'
133 #' @return An integer representing the parameter count of our neural network.
134 #' @export
135
136 param <- function(nu) {
137    if (nu |> is_nn() == TRUE) {
138      0 -> param_count
139      for (i in 1:length(nu)) {
140        param_count + length(nu[[i]]$W) + length(nu[[i]]$b) -> param_count
141      }
142      return(param_count)
143    } else {
144      stop("Only_neural_networks_can_have_parameters")
145    }
146 }
```

Listing 9.3: R code for activation functions ReLU and Sigmoid

```
1  #' @title: ReLU
2  #' @description: The ReLU activation function
3  #'
4  #' @param x A real number that is the input to our ReLU function.
5  #'
6  #' @return The output of the standard ReLU function, i.e. \eqn{\max\{0,x\}}.
      See also \code{\link{Sigmoid}}.
7  #' and \code{\link{Tanh}}.
8  #' @export
9
10 ReLU <- function(x) {
11    if (x |> is.numeric() && x |> length() == 1 && x |> is.finite()) {
12      return(x |> max(0))
13    } else {
14      stop("x_must_be_a_real_number")
15    }
16 }
17
18 #' @title: Sigmoid
19 #' @description The Sigmoid activation function.
20 #'
```

```
21  #' @param x a real number that is the input to our Sigmoid function
22  #'
23  #' @return The output of a standard Sigmoid function,
24  #' i,e. \eqn{\frac{1}{1 + \exp(-x)}}.
25  #' See also \code{\link{Tanh}}.and \code{\link{ReLU}}.
26  #' @export
27
28  Sigmoid <- function(x) {
29    if (x |> is.numeric() && x |> length() == 1 && x |> is.finite()) {
30      return(1 / (1 + exp(-x)))
31    } else {
32      stop("x_must_be_a_real_number")
33    }
34  }
35
36  #' @title Tanh
37  #' @description The tanh activation function
38  #'
39  #' @param x a real number
40  #'
41  #' @return the \eqn{tanh} of x. See also \code{\link{Sigmoid}} and
42  #' \code{\link{ReLU}}.
43  #' @export
44
45  Tanh <- function(x) {
46    if (x |> is.numeric() && x |> length() == 1 && x |> is.finite()) {
47      return(x |> tanh())
48    } else {
49      stop("x_must_be_a_real_number")
50    }
51  }
```

Listing 9.4: R code for intanitation

```
1   source("R/aux_fun.R")
2   source("R/is_nn.R")
3
4   #' @title inst
5   #' @description The function that instantiates a neural network as created
6   #' by create_neural_network().
7   #'
8   #'
9   #' @param neural_network An ordered list of lists, of the type generated by
10  #' create_neural_network() where each element in the
11  #' list of lists is a pair \eqn{(W,b)} representing the weights and biases of
12  #' that layer.
13  #'
14  #' \emph{NOTE:} We will call istantiation what Grohs et. al. call "realization
        ".
15  #'
16  #' @references  Grohs, P., Hornung, F., Jentzen, A. et al. Space-time error
        estimates for deep
17  #' neural network approximations for differential equations. Adv Comput Math
        49, 4 (2023).
```

```r
#' https://doi.org/10.1007/s10444-022-09970-2.
#'
#' Definition 1.3.4. Jentzen, A., Kuckuck, B., and von Wurstemberger, P.
#'   (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}
#'
#' @param activation_function A continuous function applied to the output of
#'   each layer
#' @param x our input to the continuous function formed from activation. Our
#'   input will
#' be an element in \eqn{\mathbb{R}^d} for some appropriate \eqn{d}.
#'
#' @return The output of the function that is the instantiation of the given
#' neural network with the given activation function at the given \eqn{x}.
#'
#' @export
#'


inst <- function(neural_network, activation_function, x) {
  if (neural_network |> is_nn() == FALSE) {
    stop("Only_neural_networks_can_be_instantiated")
  } else if (neural_network |> inn() != x |>
    matrix() |>
    nrow()) {
    stop("x_does_not_match_input_size_required_by_neural_network")
  } else {
    if (dep(neural_network) == 1) {
      neural_network[[1]]$W %*% x + neural_network[[1]]$b -> output
      return(output)
    }

    x |> matrix() -> output

    for (i in 1:(length(neural_network) - 1)) {
      neural_network[[i]]$W %*% output + neural_network[[i]]$b -> linear_
          transform
      apply(linear_transform,
        MARGIN = 1,
        FUN = activation_function
      ) -> output
    }
    neural_network[[length(neural_network)]]$W %*% output +
      neural_network[[length(neural_network)]]$b -> output
    return(output)
  }
}
```

Listing 9.5: R code for parallelizing two neural networks

```r
source("R/aux_fun.R")
source("R/Tun.R")
source("R/is_nn.R")
```

```r
#' Function for creating a block diagonal given two matrices.
#'
#' @param matrix1 A matrix.
#' @param matrix2 A matrix
#'
#' @return A block diagonal matrix with matrix1 on top left
#' and matrix2 on bottom right.

create_block_diagonal <- function(matrix1, matrix2) {
  nrow(matrix1) -> m1
  nrow(matrix2) -> m2
  ncol(matrix1) -> n1
  ncol(matrix2) -> n2

  # Create a block diagonal matrix
  0 |> matrix(m1 + m2, n1 + n2) -> block_diagonal_matrix
  block_diagonal_matrix[1:m1, 1:n1] <- matrix1
  block_diagonal_matrix[(m1 + 1):(m1 + m2), (n1 + 1):(n1 + n2)] <-
    matrix2

  return(block_diagonal_matrix)
}

#' @title stk
#' @description A function that stacks neural networks.
#'
#' @param nu neural network.
#' @param mu neural network.
#'
#' @return A stacked neural network of \eqn{\nu} and \eqn{\mu}, i.e. \eqn{\nu
#'   \boxminus \mu}
#'
#'
#' \strong{NOTE:} This is different than the one given in Grohs, et. al. 2023.
#' While we use padding to equalize neural networks being parallelized our
#' padding is via the Tun network whereas Grohs et. al. uses repetitive
#' composition of the i network. We use repetitive composition of the \eqn{\
#'   mathsf{Id_1}}
#' network. See \code{\link{Id}} \code{\link{comp}}
#'
#' \strong{NOTE:} The terminology is also different from Grohs et. al. 2023.
#' We call stacking what they call parallelization. This terminology change
#'   was
#' inspired by the fact that parallelization implies commutativity but this
#' operation is not quite commutative. It is commutative up to transposition
#' of our input x under instantiation with a continuous activation function.
#'
#' Also the work parallelization has a lot of baggage when it comes to
#' artificial neural networks in that it often means many different CPUs
#'   working
#' together.
```

250

```
53  #'
54  #' \emph{Remark:} We will use only one symbol for stacking equal and unequal
       depth
55  #' neural networks, namely "stk". This is for usability but also that
56  #' for all practical purposes only the general stacking of neural networks
57  #' of different sizes is what is needed.
58  #'
59  #' \emph{Remark:} We have two versions, a prefix and an infix version.
60  #'
61  #' This operation on neural networks, called "parallelization" is found in:
62  #' @references  Grohs, P., Hornung, F., Jentzen, A. et al. Space-time error
       estimates for deep
63  #' neural network approximations for differential equations. Adv Comput Math
       49, 4 (2023).
64  #' https://doi.org/10.1007/s10444-022-09970-2
65  #'
66  #' @export
67
68  stk <- function(nu, mu) {
69    if (nu |> is_nn() && mu |> is_nn()) {
70      if (dep(nu) == dep(mu)) {
71        list() -> parallelized_network
72        for (i in 1:length(nu)) {
73          create_block_diagonal(nu[[i]]$W, mu[[i]]$W) -> parallelized_W
74          rbind(nu[[i]]$b, mu[[i]]$b) -> parallelized_b
75          list(W = parallelized_W, b = parallelized_b) -> parallelized_network[[
              i]]
76        }
77        return(parallelized_network)
78      }
79
80      if (dep(nu) > dep(mu)) {
81        (dep(nu) - dep(mu) + 1) |> Tun(d = out(mu)) -> padding
82        padding |> comp(mu) -> padded_network
83        nu |> stk(padded_network) -> parallelized_network
84        return(parallelized_network)
85      }
86
87      if (dep(nu) < dep(mu)) {
88        (dep(mu) - dep(nu) + 1) |> Tun(d = out(nu)) -> padding
89        padding |> comp(nu) -> padded_network
90        padded_network |> stk(mu) -> parallelized_network
91        return(parallelized_network)
92      }
93    } else {
94      stop("Please_try_stacking_neural_networks")
95    }
96  }
97
98  #' The stk function.
99  #'
100 #' @param nu neural network.
101 #' @param mu neural network.
```

```r
102  #'
103  #' @return A stacked neural network of nu and mu.
104  #' @export
105
106
107  `%stk%` <- function(nu, mu) {
108    if (nu |> is_nn() && mu |> is_nn()) {
109      if (dep(nu) == dep(mu)) {
110        list() -> parallelized_network
111        for (i in 1:length(nu)) {
112          create_block_diagonal(nu[[i]]$W, mu[[i]]$W) -> parallelized_W
113          rbind(nu[[i]]$b, mu[[i]]$b) -> parallelized_b
114          list(W = parallelized_W, b = parallelized_b) -> parallelized_network[[
                  i]]
115        }
116        return(parallelized_network)
117      }
118
119      if (dep(nu) > dep(mu)) {
120        (dep(nu) - dep(mu) + 1) |> Tun(d = out(mu)) -> padding
121        padding |> comp(mu) -> padded_network
122        nu |> stk(padded_network) -> parallelized_network
123        return(parallelized_network)
124      }
125
126      if (dep(nu) < dep(mu)) {
127        (dep(mu) - dep(nu) + 1) |> Tun(d = out(nu)) -> padding
128        padding |> comp(nu) -> padded_network
129        padded_network |> stk(mu) -> parallelized_network
130        return(parallelized_network)
131      }
132    } else {
133      stop("Please_try_stacking_neural_networks")
134    }
135  }
```

Listing 9.6: R code for affine neural networks

```r
1   #' @title Aff
2   #' @description The function that returns \eqn{\mathsf{Aff}} neural networks.
3   #'
4   #' @param W An \eqn{m \times n} matrix representing the weight of the affine
5   #' neural network
6   #' @param b An \eqn{m \times 1} vector representing the bias of the affine
7   #' neural network
8   #'
9   #' @references  Grohs, P., Hornung, F., Jentzen, A. et al. Space-time error
        estimates for deep
10  #' neural network approximations for differential equations. Adv Comput Math
        49, 4 (2023).
11  #' https://doi.org/10.1007/s10444-022-09970-2
12  #'
13  #' Definition 2.3.1. Jentzen, A., Kuckuck, B., and von Wurstemberger, P.
        (2023).
```

```r
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}
#'
#' @return Returns the network \eqn{((W,b))} representing an affine neural
#'  network. Also
#' denoted as \eqn{\mathsf{Aff}_{W,b}}
#' See also \code{\link{Cpy}} and \code{\link{Sum}}.
#' @export

Aff <- function(W, b) {
  if (W |> is.matrix() == FALSE) (W |> matrix() -> W)
  if (b |> is.matrix() == FALSE) (b |> matrix() -> b)
  list(list(W = W, b = b)) -> return_network
  return(return_network)
}

#' @title Cpy
#' @description The function that returns \eqn{\mathsf{Cpy}} neural networks.
#' These are neural networks defined as such
#' \deqn{
#' \mathsf{Aff}_{\left[ \mathbb{I}_k \: \mathbb{I}_k \: \cdots \: \mathbb{I}_k
#'  \right]^T,0_{k}}
#' }
#'
#' @param n number of copies to make.
#' @param k the size of the input vector.
#'
#' @return Returns an affine network that makes a concatenated vector that is
#'  \eqn{n}
#' copies of the input vector of size \eqn{k}. See \code{\link{Aff}} and \code
#'  {\link{Sum}}.
#'
#' @references Definition 2.4.6. Jentzen, A., Kuckuck, B., and von
#'  Wurstemberger, P. (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}
#'
#'
#' @export

Cpy <- function(n, k) {
  if (n %% 1 != 0 ||
    n < 1 ||
    k %% 1 != 0 ||
    k < 1) {
    stop("n_and_k_must_be_natural_numbers")
  } else {
    k |> diag() -> W
    for (i in 2:n) {
      W |> rbind(k |> diag()) -> W
    }
    0 |> matrix(n * k) -> b
    list(list(W = W, b = b)) -> return_network
```

```
62      return(return_network)
63    }
64 }
65
66 #' @title Sum
67 #' @description The function that returns \eqn{\mathsf{Sum}} neural networks.
68 #'
69 #' These are neural networks defined as such
70 #' \deqn{
71 #' \mathsf{Aff}_{\left[ \mathbb{I}_k \: \mathbb{I}_k \: \cdots \: \mathbb{I}_k
         \right],0_{k}}
72 #' }
73 #'
74 #' @param n number of copies of a certain vector to be summed.
75 #' @param k the size of the summation vector.
76 #'
77 #' @return An affine neural network that will take a vector of size
78 #' \eqn{n \times k} and return the summation vector that is of length
79 #' \eqn{k}. See also \code{\link{Aff}} and \code{\link{Cpy}}.
80 #'
81 #' @references Definition 2.4.1. Jentzen, A., Kuckuck, B., and von
         Wurstemberger, P. (2023).
82 #' Mathematical introduction to deep learning: Methods, implementations,
83 #' and theory. \url{https://arxiv.org/abs/2310.20360}
84 #'
85 #'
86 #' @export
87 #'
88
89
90 Sum <- function(n, k) {
91   if (n %% 1 != 0 ||
92     n < 1 ||
93     k %% 1 != 0 ||
94     k < 1) {
95     stop("n_and_k_must_be_natural_numbers")
96   } else {
97     k |> diag() -> W
98     for (i in 2:n) {
99       W |> cbind(k |> diag()) -> W
100     }
101     0 |> matrix(k) -> b
102     list(list(W = W, b = b)) -> return_network
103
104     return(return_network)
105   }
106 }
```

Listing 9.7: R code for composition of two neural networks

```
1 source("R/aux_fun.R")
2 source("R/is_nn.R")
3
4
```

```r
#' @title comp
#' @description The function that takes the composition of two neural
#' networks assuming they are compatible, i.e., given
#' \eqn{\nu_1, \nu_2 \in \mathsf{NN}}, it must be the case that
#' \eqn{\mathsf{I}(\nu)_1 = \mathsf{O}(\nu_2)}.
#'
#' @param phi_1 first neural network to be composed, goes on the left
#' @param phi_2 second neural network to be composed, goes on right
#'
#' @return The composed neural network. See also \code{\link{dep}}.
#' Composition of neural networks is the operation defined for \eqn{\nu_1 \in
#'     \mathsf{NN}}
#' and \eqn{\nu_2 \in \mathsf{NN}} as:
#'
#' \deqn{
#' \nu_1 \bullet \nu_2 = \begin{cases} (( W'_1,b'_1 ),
#' ( W'_2,b'_2 ), ...,( W'_{M-1}, b'_{M-1}), ( W_1W'_M, W_1b'_{M} + b_1), (W_
#'     2, b_2 ),\\...,
#' ( W_L,b_L )) & :( L> 1 ) \land ( M > 1 ) \\((W_1W'_1,W_1b'_1+b_1),(W_2,b_2)
#'     , (W_3,b_3),...,
#' (W_Lb_L)) & :(L>1) \land (M=1) \\(((W'_1, b'_1),(W'_2,b'_2), ...,
#' (W'_{M-1}, b'_{M-1})(W_1, b'_M + b_1)) &:(L=1) \land (M>1) \\ ((W_1W'_1, W_
#'     1b'_1+b_1)) &:(L=1)
#' \land (M=1)\end{cases}
#'
#'
#' }
#'
#'
#' @references Grohs, P., Hornung, F., Jentzen, A. et al.
#' Space-time error estimates for deep neural network approximations
#' for differential equations. Adv Comput Math 49, 4 (2023).
#' \url{https://doi.org/10.1007/s10444-022-09970-2}.
#'
#' @references Definition 2.1.1. Jentzen, A., Kuckuck, B., and von
#'     Wurstemberger, P. (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}
#'
#' \emph{Remark:} We have two versions of this function, an
#' infix version for close resemblance to mathematical notation and
#' prefix version.

#' @encoding utf8
#' @export
#'


comp <- function(phi_1, phi_2) {
  if (phi_1 |> is_nn() && phi_2 |> is_nn()) {
    dep(phi_1) -> L
    dep(phi_2) -> L_

```

```r
    if (L > 1 & L_ > 1) {
      phi_2[-L_] -> beginning
      phi_1[-1] -> end
      phi_1[[1]]$W %*% phi_2[[L_]]$W -> mid_W
      phi_1[[1]]$W %*% phi_2[[L_]]$b + phi_1[[1]]$b -> mid_b
      list(W = mid_W, b = mid_b) -> mid
      c(
        beginning,
        list(mid),
        end
      ) -> composed_network
      return(composed_network)
    } else if (L > 1 & L_ == 1) {
      phi_1[[1]]$W %*% phi_2[[1]]$W -> beginning_W
      phi_1[[1]]$W %*% phi_2[[1]]$b + phi_1[[1]]$b -> beginning_b
      list(
        W = beginning_W,
        b = beginning_b
      ) -> beginning
      phi_1[-1] -> end
      c(
        list(beginning),
        end
      ) -> composed_network
      return(composed_network)
    } else if (L == 1 & L_ > 1) {
      phi_2[-L_] -> beginning
      phi_1[[1]]$W %*% phi_2[[L_]]$W -> end_W
      phi_1[[1]]$W %*% phi_2[[L_]]$b + phi_1[[1]]$b -> end_b
      list(
        W = end_W,
        b = end_b
      ) -> end
      c(
        beginning,
        list(end)
      ) -> composed_network
      return(composed_network)
    } else if (L == 1 & L_ == 1) {
      list() -> composed_network
      phi_1[[1]]$W %*% phi_2[[1]]$W -> W
      phi_1[[1]]$W %*% phi_2[[1]]$b + phi_1[[1]]$b -> b
      list(
        W = W,
        b = b
      ) -> composed_network[[1]]
      return(composed_network)
    } else {
      stop("Dimensionality_mismatch")
    }
  } else {
    stop("Only_neural_networks_can_be_composed.")
  }
```

```r
}

#' The `infix version of comp function
#'
#' @param phi_1 first neural network to be composed, goes on the left
#' @param phi_2 second neural network to be composed, goes on right
#'
#' @rdname comp
#' @export


`%comp%` <- function(phi_1, phi_2) {
  if (phi_1 |> is_nn() && phi_2 |> is_nn()) {
    dep(phi_1) -> L
    dep(phi_2) -> L_

    if (L > 1 & L_ > 1) {
      phi_2[-L_] -> beginning
      phi_1[-1] -> end
      phi_1[[1]]$W %*% phi_2[[L_]]$W -> mid_W
      phi_1[[1]]$W %*% phi_2[[L_]]$b + phi_1[[1]]$b -> mid_b
      list(W = mid_W, b = mid_b) -> mid
      c(
        beginning,
        list(mid),
        end
      ) -> composed_network
      return(composed_network)
    } else if (L > 1 & L_ == 1) {
      phi_1[[1]]$W %*% phi_2[[1]]$W -> beginning_W
      phi_1[[1]]$W %*% phi_2[[1]]$b + phi_1[[1]]$b -> beginning_b
      list(
        W = beginning_W,
        b = beginning_b
      ) -> beginning
      phi_1[-1] -> end
      c(
        list(beginning),
        end
      ) -> composed_network
      return(composed_network)
    } else if (L == 1 & L_ > 1) {
      phi_2[-L_] -> beginning
      phi_1[[1]]$W %*% phi_2[[L_]]$W -> end_W
      phi_1[[1]]$W %*% phi_2[[L_]]$b + phi_1[[1]]$b -> end_b
      list(
        W = end_W,
        b = end_b
      ) -> end
      c(
        beginning,
        list(end)
      ) -> composed_network
```

257

```
159    return(composed_network)
160  } else if (L == 1 & L_ == 1) {
161    list() -> composed_network
162    phi_1[[1]]$W %*% phi_2[[1]]$W -> W
163    phi_1[[1]]$W %*% phi_2[[1]]$b + phi_1[[1]]$b -> b
164    list(
165      W = W,
166      b = b
167    ) -> composed_network[[1]]
168    return(composed_network)
169  } else {
170    stop("Dimensionality_mismatch")
171  }
172  } else {
173    stop("Only_neural_networks_can_be_composed.")
174  }
175 }
```

Listing 9.8: R code for scalar multiplication

```
1  source("R/comp.R")
2  source("R/aux_fun.R")
3  source("R/is_nn.R")
4
5  #' @title slm
6  #'
7  #' @description The function that returns the left  scalar multiplication
8  #' neural network
9  #'
10 #' @param a A real number.
11 #' @param nu A neural network of the kind created by create_neural_network.
12 #'
13 #' @return Returns a neural network that is \eqn{a \triangleright \nu}. This
14 #' instantiates as \eqn{a \cdot f(x)} under continuous function activation.
       More specifically
15 #' we define operation as:
16 #'
17 #' Let \eqn{\lambda \in \mathbb{R}}. We will denote by \eqn{(\cdot) \
       triangleright (\cdot):
18 #' \mathbb{R} \times \mathsf{NN} \rightarrow \mathsf{NN}} the function
       satisfying for all
19 #' \eqn{\nu \in \mathsf{NN}} and \eqn{\lambda \in \mathbb{R}} that \eqn{\
       lambda \triangleright \nu =
20 #' \mathsf{Aff}_{\lambda \mathbb{I}_{\mathsf{I}(\nu)},0} \bullet \nu}.
21
22 #' @references Definition 2.3.4. Jentzen, A., Kuckuck, B., and von
       Wurstemberger, P. (2023).
23 #' Mathematical introduction to deep learning: Methods, implementations,
24 #' and theory. \url{https://arxiv.org/abs/2310.20360}.
25 #'
26 #' \emph{Note:} We will have two versions of this operation, a prefix and an
27 #' infix version.
28 #' @export
29
```

```r
slm <- function(a, nu) {
  if (a |> is.numeric() &&
    length(a) == 1 &&
    a |> is.finite() &&
    nu |> is_nn()) {
    nu |> out() -> constant_matrix_size
    list() -> multiplier_network
    a |> diag(constant_matrix_size) -> W
    0 |> matrix(constant_matrix_size) -> b
    list(W = W, b = b) -> multiplier_network[[1]]
    multiplier_network |> comp(nu) -> return_network
    return(return_network)
  } else {
    stop("a_must_be_a_real_number_and_nu_must_be_a_neural_network")
  }
}

#' @title srm
#' @description The function that returns the right scalar multiplication
#' neural network
#'
#' @param nu A neural network of the type generated by create_neural_network()
#'   .
#' @param a A real number.
#'
#' @return Returns a neural network that is \eqn{\nu \triangleleft a}. This
#' instantiates as \eqn{f(a \cdot x)}.under continuous function activation.
#'   More
#' specifically we will define this operation as:
#'
#' Let \eqn{\lambda \in \mathbb{R}}. We will denote by \eqn{(\cdot) \
#'   triangleleft (\cdot):
#' \mathsf{NN} \times \mathbb{R} \rightarrow \mathsf{NN}} the function
#'   satisfying for all
#' \eqn{\nu \in \mathsf{NN}} and \eqn{\lambda \in \mathbb{R}} that \eqn{\nu \
#'   triangleleft \lambda =
#' \nu \bullet \mathsf{Aff}_{\lambda \mathbb{I}_{\mathsf{I}(\nu)},0}}.
#'
#' @references Definition 2.3.4. Jentzen, A., Kuckuck, B., and von
#'   Wurstemberger, P. (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}.
#'
#' \emph{Note:} We will have two versions of this operation, a prefix
#' and an infix version.
#' @export

srm <- function(nu, a) {
  if (a |> is.numeric() &&
    length(a) == 1 &&
    a |> is.finite() &&
    nu |> is_nn()) {
```

```r
    nu |> inn() -> constant_matrix_size
    list() -> multiplier_network
    a |> diag(constant_matrix_size) -> W
    0 |> matrix(constant_matrix_size) -> b
    list(W = W, b = b) -> multiplier_network[[1]]
    nu |> comp(multiplier_network) -> return_network
    return(return_network)
  } else {
    stop("a_must_be_a_real_number_and_nu_must_be_a_neural_network")
  }
}


#'
#' @param a A real number.
#' @param nu A neural network of the type generated by create_neural_network()
    .
#'
#' @rdname slm
#' @export

`%slm%` <- function(a, nu) {
  if (a |> is.numeric() &&
      length(a) == 1 &&
      a |> is.finite() &&
      nu |> is_nn()) {
    nu |> out() -> constant_matrix_size
    list() -> multiplier_network
    a |> diag(constant_matrix_size) -> W
    0 |> matrix(constant_matrix_size) -> b
    list(W = W, b = b) -> multiplier_network[[1]]
    multiplier_network |> comp(nu) -> return_network
    return(return_network)
  } else {
    stop("a_must_be_a_real_number_and_nu_must_be_a_neural_network")
  }
}

#' @param nu A neural network
#' @param a A real number.
#'
#' @rdname srm
#' @export

`%srm%` <- function(nu, a) {
  if (a |> is.numeric() &&
      length(a) == 1 &&
      a |> is.finite() &&
      nu |> is_nn()) {
    nu |> inn() -> constant_matrix_size
    list() -> multiplier_network
    a |> diag(constant_matrix_size) -> W
    0 |> matrix(constant_matrix_size) -> b
```

```
129    list(W = W, b = b) -> multiplier_network[[1]]
130    nu |> comp(multiplier_network) -> return_network
131    return(return_network)
132  } else {
133    stop("a_must_be_a_real_number_and_nu_must_be_a_neural_network")
134  }
135 }
```

Listing 9.9: R code for sum of two neural networks

```
1 source("R/comp.R")
2 source("R/stacking.R")
3 source("R/aux_fun.R")
4 source("R/Aff.R")
5 source("R/is_nn.R")
6
7 #' @title nn_sum
8 #' @description A function that performs the neural network sum for two
9 #' neural networks of the type generated by
10 #' create_neural_network(). Neural network sums are defined for
11 #' \eqn{\nu_1 \in \mathsf{NN}} and \eqn{\nu_2 \in \mathsf{NN}} as:
12 #' \deqn{
13 #' \oplus^v_{i=u}\nu_i \coloneqq \left( \mathsf{Sum}_{v-u+1,\mathsf{O}(\nu_2)}
       \bullet \left[ \boxminus^v_{i=u}\nu_i \right] \bullet \mathsf{Cpy}_{(v-u
     +1),\mathsf{I}(\nu_1)} \right)
14 #' }
15 #'
16 #' @references Grohs, P., Hornung, F., Jentzen, A. et al.
17 #' Space-time error estimates for deep neural network approximations
18 #' for differential equations. Adv Comput Math 49, 4 (2023).
19 #' \url{https://doi.org/10.1007/s10444-022-09970-2}.
20 #'
21 #' @param nu_1 A neural network.
22 #' @param nu_2 A neural network.
23 #'
24 #' @return A neural network that is the neural network sum of \eqn{\nu_1} and
       \eqn{\nu_2}
25 #' i.e. \eqn{\nu_1 \oplus \nu_2}.
26 #'
27 #' \emph{Note:} We have two versions, an infix version and a prefix version.
28 #' @export
29
30 nn_sum <- function(nu_1, nu_2) {
31   if (nu_1 |> is_nn() &&
32     nu_2 |> is_nn() &&
33     inn(nu_1) == inn(nu_2) &&
34     out(nu_1) == out(nu_2)) {
35     Cpy(2, inn(nu_1)) -> first_third
36     nu_1 |> stk(nu_2) -> mid_third
37     Sum(2, out(nu_1)) -> last_third
38
39     last_third |>
40       comp(mid_third) |>
41       comp(first_third) -> return_network
```

```
42      return(return_network)
43    } else {
44      stop("Only_neural_networks_with_same_end-widths_may_be_summed")
45    }
46 }
47
48 #' Function for calculating neural network sums
49 #'
50 #' @param nu_1 A neural network.
51 #' @param nu_2 A neural network.
52 #'
53 #' @rdname nn_sum
54 #' @export
55 #'
56 `%nn_sum%` <- function(nu_1, nu_2) {
57    if (nu_1 |> is_nn() &&
58      nu_2 |> is_nn() &&
59      inn(nu_1) == inn(nu_2) &&
60      out(nu_1) == out(nu_2)) {
61      Cpy(2, inn(nu_1)) -> first_third
62      nu_1 |> stk(nu_2) -> mid_third
63      Sum(2, out(nu_1)) -> last_third
64
65      last_third |>
66        comp(mid_third) |>
67        comp(first_third) -> return_network
68      return(return_network)
69    } else {
70      stop("Only_neural_networks_of_same_end_widths_may_be_summed")
71    }
72 }
```

Listing 9.10: R code for $\mathfrak{i}$

```
1 #' @title: i
2 #' @description The function that returns the \eqn{\mathbb{i}} network.
3 #'
4 #' @param d the size of the \eqn{\mathsf{i}} network
5 #'
6 #' @return returns the i_d network
7
8 i <- function(d) {
9    list() -> return_network
10   d |> diag() -> W
11   0 |> matrix(d, 1) -> b
12   list(W = W, b = b) -> return_network[[1]]
13   list(W = W, b = b) -> return_network[[2]]
14   return(return_network)
15 }
```

Listing 9.11: R code for Id neural networks

```
1 #' @title: Id
2 #' @description The function that returns the \eqn{\mathsf{Id_1}} networks.
```

```r
#' @param d the dimension of the \eqn{Id} network, by default it is \eqn{1}.
#'
#' @return Returns the \eqn{\mathsf{Id_1}} network.
#' @export

Id <- function(d = 1) {
  if (d %% 1 != 0 ||
    d < 1
  ) {
    stop("d_must_be_natural_numbers")
  } else if (d == 1) {
    W_1 <- c(1, -1) |> matrix()
    b_1 <- c(0, 0) |> matrix()
    layer_1 <- list(W = W_1, b = b_1)
    W_2 <- c(1, -1) |> matrix(1, 2)
    b_2 <- 0 |> matrix()
    layer_2 <- list(W = W_2, b = b_2)
    result <- list(layer_1, layer_2)
    return(result)
  } else if (d > 1) {
    Id() -> return_network
    for (j in 2:d) {
      return_network |> stk(Id()) -> return_network
    }
    return(return_network)
  } else {
    stop("Unknown_error")
  }
}
```

Listing 9.12: R code for Tun

```r
source("R/comp.R")
source("R/Id.R")

#' Tun: The function that returns tunneling neural networks
#'
#' @param n The depth of the tunnel network where \eqn{n \in \mathbb{N} \cap
#'   [1,\infty)}.
#' @param d The dimension of the tunneling network. By default it is assumed
#'   to be \eqn{1}.
#'
#' @return A tunnel neural network of depth n. A tunneling neural
#' network is defined as the neural network \eqn{\mathsf{Aff}_{1,0}} for \eqn{
#'   n=1},
#' the neural network \eqn{\mathsf{Id}_1} for \eqn{n=1} and the neural network
#' \eqn{\bullet^{n-2}\mathsf{Id}_1} for \eqn{n >2}. For this to work we
#' must provide an appropriate \eqn{n} and instantiate with ReLU at some
#' real number \eqn{x}.
#' @export
#'
Tun <- function(n, d = 1) {
  if (n %% 1 != 0 ||
    n < 1 ||
```

```r
     d %% 1 != 0 ||
     d < 1
  ) {
    stop("n_and_d_must_be_natural_numbers")
  }
  if (d == 1) {
    if (n == 1) {
      return(Aff(1, 0))
    } else if (n == 2) {
      return(Id())
    } else if (n > 2) {
      Id() -> return_network
      for (i in 3:n) {
        return_network |> comp(Id()) -> return_network
      }
      return(return_network)
    }
  } else if (d > 1) {
    if (n == 1) {
      return(Aff(diag(d), 0 |> matrix()))
    } else if (n == 1) {
      return(Id(d))
    } else if (n == 2) {
      return(Id(d))
    } else if (n > 2) {
      Id(d) -> return_network
      for (i in 3:n) {
        return_network |> comp(Id(d)) -> return_network
      }
      return(return_network)
    }
  } else {
    stop("Unknown_error")
  }
}
```

Listing 9.13: R code for Φ_k

```r
source("R/comp.R")
source("R/Aff.R")
source("R/i.R")
source("R/aux_fun.R")
source("R/activations.R")

#' The c_k function
#'
#' @param k an integer in \eqn{[1,\infty)}
#'
#' @return the real number \eqn{2^{1-2k}}
#' @references Grohs, P., Hornung, F., Jentzen, A. et al.
#' Space-time error estimates for deep neural network approximations
#' for differential equations. Adv Comput Math 49, 4 (2023).
#' https://doi.org/10.1007/s10444-022-09970-2
```

```r
#'
#' @references Definition 2.3.4. Jentzen, A., Kuckuck, B., and von
#'    Wurstemberger, P. (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}.

c_k <- function(k) {
  2^{
    1 - 2 * k
  } -> result
  return(result)
}

#' This is an intermediate variable, see reference.
c(0, -1 / 2, -1, 0) |> matrix() -> B


#' C_k: The function that returns the C_k matrix
#'
#' @param k Natural number, the precision with which to approximate squares
#' within \eqn{[0,1]}
#'
#' @return A neural network that approximates the square of any real within
#' \eqn{[0,1]}

C_k <- function(k) {
  c(-c_k(k), 2 * c_k(k), -c_k(k), 1) |> matrix(1, 4) -> result
  return(result)
}


#' A_k: The function that returns the matrix A_k
#'
#' @param k Natural number, the precision with which to approximate squares
#' within \eqn{[0,1]}
#'
#' @return A neural network that approximates the square of any real within
#' \eqn{[0,1]}
#'
A_k <- function(k) {
  c(2, 2, 2, -c_k(k)) |>
    c(-4, -4, -4, 2 * c_k(k)) |>
    c(2, 2, 2, -c_k(k)) |>
    c(0, 0, 0, 1) |>
    matrix(4, 4) -> result
  return(result)
}

#' This is an intermediate variable. See the reference
#'
c(1, 1, 1, 1) |> matrix(4, 1) -> A
```

```r
#' The Phi_k function
#'
#' @param k an integer \eqn{k \in (2,\infty)}
#'
#' @return The Phi_k neural network
#' @references Grohs, P., Hornung, F., Jentzen, A. et al.
#' Space-time error estimates for deep neural network approximations
#' for differential equations. Adv Comput Math 49, 4 (2023).
#' https://doi.org/10.1007/s10444-022-09970-2
#'
Phi_k <- function(k) {
  if (k |> is.numeric() &&
    k |> length() == 1 &&
    k >= 1 &&
    k |> is.finite() &&
    k %% 1 == 0) {
    if (k == 1) {
      C_k(1) |>
        Aff(0) |>
        comp(i(4)) |>
        comp(Aff(A, B)) -> return_network
      return(return_network)
    }
    if (k >= 2) {
      C_k(k) |>
        Aff(0) |>
        comp(i(4)) -> return_network
      for (j in (k - 1):1) {
        A_k(j) |>
          Aff(B) |>
          comp(i(4)) -> intermediate_network
        return_network |> comp(intermediate_network) -> return_network
      }
      return_network |> comp(A |> Aff(B)) -> return_network
      return(return_network)
    }
  } else {
    stop("k_must_a_natural_number")
  }
}
```

Listing 9.14: R code for simulations involving Φ_k

```r
source("Phi_k.R")
source("aux_fun")
source("realization.R")
source("activations.R")

library(ggplot2)

#' The Phi_k_diff function
#'
#' @param x the number to be squared in [0,1]
#' @param k a parameter for Phi_k in [0, \infty)]
#'
#' @return the 1-norm error between x^2 and Phi_k approximation

Phi_k_diff <- function(x, k) {
  return <- (k |> Phi_k() |> rlz(ReLU, x) - x^2) |>
    abs() -> result
  return(result)
}

k_values <- c(1, 2, 5, 10, 15, 20)
```

```r
22  x_values <- seq(-2, 2, length.out = 200)
23  Phi_k_diff_v <- Vectorize(Phi_k_diff)
24
25  Phi_k_diff_data <- expand.grid(k = k_values, x = x_values)
26  Phi_k_diff_data$diff <- Phi_k_diff_v(Phi_k_diff_data$x, Phi_k_diff_data$k)
27
28  library(ggplot2)
29  ggplot(Phi_k_diff_data, aes(x = x, y = diff, color = factor(k))) +
30    scale_y_log10() +
31    geom_line() +
32    geom_line(aes(y = 2^(-2 * k - 2))) +
33    labs(
34      x = "x",
35      y = "log10_of_the_1-norm_error_over_domain_[0,1]"
36    ) -> Phi_k_diff_plot
37  ggsave("Phi_k_properties/diff.png", plot = Phi_k_diff_plot, width = 6, height
        = 5, units = "in")
38
39  vectorized_Phi_k <- Vectorize(Phi_k)
40  vectorized_param <- Vectorize(param)
41
42  param_data <- data.frame(x = 1:100, y = vectorized_param(vectorized_Phi_k
        (1:100)))
43
44  ggplot(param_data, aes(x = x, y = y)) +
45    geom_line() +
46    theme_minimal() +
47    xlab("Size_of_k") +
48    ylab("Number_of_parameters") +
49    ggtitle("Plot_of_the_number_of_parameters_of_ϕ(k)_against_k") +
50    geom_smooth(method = "lm", se = FALSE, color = "blue")
51
52  vectorized_dep <- Vectorize(dep)
53
54  dep_data <- data.frame(x = 1:100, y = vectorized_dep(vectorized_Phi_k(1:100)))
55
56  ggplot(dep_data, aes(x = x, y = y)) +
57    geom_line() +
58    theme_minimal() +
59    xlab("Size_of_k") +
60    ylab("Depth_of_network") +
61    ggtitle("Plot_of_the_depth_of_ϕ(k)_against_k") +
62    geom_smooth(method = "lm", se = FALSE, color = "blue")
```

Listing 9.15: R code for Φ

```r
1  source("R/Phi_k.R")
2  source("R/i.R")
3  source("R/Aff.R")
4
5
6  #' The Phi function
7  #'
```

```
 8 #' @param eps parameter for Phi
 9 #' @references Grohs, P., Hornung, F., Jentzen, A. et al.
10 #' Space-time error estimates for deep neural network approximations
11 #' for differential equations. Adv Comput Math 49, 4 (2023).
12 #' https://doi.org/10.1007/s10444-022-09970-2
13 #'
14 #' @return neural network Phi that approximately squares a number between
15 #' 0 and 1.
16
17 Phi <- function(eps) {
18   if (eps |> is.numeric() &&
19     eps |> length() == 1 &&
20     eps |> is.finite() &&
21     eps > 0) {
22     (0.5 * log2(1 / eps) - 1) |> ceiling() -> M
23
24     if (M <= 0) 1 -> M
25
26     if (M == 1) {
27       C_k(1) |>
28         Aff(0) |>
29         comp(i(4)) |>
30         comp(Aff(A, B)) -> return_network
31       return(return_network)
32     }
33
34     if (M >= 2) {
35       C_k(M) |>
36         Aff(0) |>
37         comp(i(4)) -> return_network
38       for (j in (M - 1):1) {
39         A_k(j) |>
40           Aff(B) |>
41           comp(i(4)) -> intermediate_network
42         return_network |> comp(intermediate_network) -> return_network
43       }
44       return_network |> comp(A |> Aff(B)) -> return_network
45       return(return_network)
46     }
47   } else {
48     stop("eps_must_be_a_positive_real_number")
49   }
50 }
```

Listing 9.16: R code for simulations involving Φ

```
1 source("Phi.R")
2 source("aux_fun.R")
3 source("realization.R")
4 source("activations.R")
5
6 #' The Phi diff function
7 #'
8 #' @param eps parameter for Phi
```

```r
#' @param x number to be squared
#'
#' @return the 1-norm error between the result
#' and x^2

diff <- function(eps, x) {
  (x^2 - eps |> Phi() |> rlz(ReLU, x)) |>
    abs() -> result
  return(result)
}

eps_values <- c(1, 0.5, 0.1, 0.01, 0.001, 0.0001)
x_values <- seq(0, 1, length.out = 100)
vectorized_diff <- Vectorize(diff)

diff_data <- expand.grid(eps = eps_values, x = x_values)
diff_data$Phi_diff <- vectorized_diff(diff_data$eps, diff_data$x)

library(ggplot2)

ggplot(diff_data, aes(x = x, y = eps, z = Phi_diff)) +
  geom_contour_filled() +
  ggtitle("Contour_plot_of_the_1-norm_difference_for_values_of_x_and_eps") +
  theme_minimal() -> Phi_diff_contour_plot

ggsave("Phi_properties/Phi_diff_contour.png", plot = Phi_diff_contour_plot,
    width = 6, height = 5, units = "in")

vectorized_Phi_k <- Vectorize(Phi_k)
vectorized_param <- Vectorize(param)

param_data <- data.frame(x = 1:100, y = vectorized_param(vectorized_Phi_k
    (1:100)))

ggplot(param_data, aes(x, y)) +
  geom_line() +
  theme_minimal()


vectorized_dep <- Vectorize(dep)

dep_data <- data.frame(x = 1:100, y = vectorized_dep(vectorized_Phi_k(1:100)))

ggplot(dep_data, aes(x = x, y = y)) +
  geom_line() +
  theme_minimal() +
  xlab("Size_of_k") +
  ylab("Depth_of_network") +
  ggtitle("Plot_of_the_depth_of_ɸ(k)_against_k") +
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```

Listing 9.17: R code for Sqr

```r
source("R/comp.R")
source("R/Aff.R")
source("R/nn_sum.R")
source("R/Phi.R")
source("R/aux_fun.R")

#' @title Sqr
#' @description A function that returns the \eqn{\mathsf{Sqr}} neural networks
    .
#'
#' @param q parameter for the Sqr network. \eqn{2 \in (2,\infty)}
#' @param eps parameter for the Sqr network. \eqn{eps \in (0,1]}. You may
#' choose epsilon to be greater than 1 but that leads to large errors
#'
#' @return A neural network that approximates the square of a real number.when
#' provided appropriate \eqn{q,\varepsilon} and upon instantiation with ReLU,
#' and a real number \eqn{x}
#' @references  Grohs, P., Hornung, F., Jentzen, A. et al. Space-time error
    estimates for deep
#' neural network approximations for differential equations. Adv Comput Math
    49, 4 (2023).
#' https://doi.org/10.1007/s10444-022-09970-2
#'
#'
#' @export


Sqr <- function(q, eps) {
  if (q <= 2 || eps <= 0) {
    stop("q_must_be_>_2_and_eps_must_be_>_0")
  } else {
    2^(-2 / (q - 2)) * eps^(q / (q - 2)) -> delta
    (eps / 2)^(1 / (q - 2)) -> alpha

    (0.5 * log2(1 / eps) - 1) |> ceiling() -> M

    if (M <= 0) 1 else M -> M

    (Aff(alpha^(-2), 0) |> comp(Phi(delta))) |>
      comp(Aff(alpha, 0)) -> first_summand

    (Aff(alpha^(-2), 0) |> comp(Phi(delta))) |>
      comp(Aff(-alpha, 0)) -> second_summand

    first_summand |>
      nn_sum(second_summand) -> return_network

    return(return_network)
  }
}
```

271

Listing 9.18: R code simulations involving Sqr

```r
source("aux_fun.R")
source("Sqr.R")
source("instantiation.R")
source("activations.R")
library("tidyverse")

#' Sqr_diff function
#'
#' @param q parameter for the Sqr network
#' @param eps parameter for the Sqr network
#' @param x the number to be squered
#'
#' @return a neural network that approximately squares x.

Sqr_diff <- function(q, eps, x) {
  return <- (Sqr(q, eps) |> rlz(ReLU, x) - x^2) |> abs()
  return(return)
}

Sqr_diff_v <- Vectorize(Sqr_diff)

Sqr_data <- expand.grid(
  q = seq(2.1, 4, length.out = 50),
  eps = seq(0.01, 2, length.out = 50),
  x = seq(-5, 5, length.out = 50)
)


Sqr_data$diff <- Sqr_diff_v(Sqr_data$q, Sqr_data$eps, Sqr_data$x)

#' Function to calculate the theoretical upper bounds of the 1-norm error
#' over \mathbb{R}
#'
#' @param q parameter for the Sqr network
#' @param eps parameter for the Sqr network
#' @param x the number to be squered
#'
#' @return the maximum 1-norm error over \mathbb{R}

diff_upper_limit <- function(q, eps, x) {
  eps * max(1, abs(x)^q)
}

diff_upper_limit_v <- Vectorize(diff_upper_limit)

Sqr_data$diff_upper_limit <- diff_upper_limit_v(Sqr_data$q, Sqr_data$eps, Sqr_
    data$x)

write_csv(Sqr_data, "Sqr_properties/Sqr_data.csv")

library(plotly)
```

```r
fig <- plot_ly(
  type = "isosurface",
  x = Sqr_data$x,
  y = Sqr_data$q,
  z = Sqr_data$eps,
  value = Sqr_data$diff,
  isomin = 0.0001,
  isomax = 5,
  colorscale = "RdBu"
) |>
  layout(scene = list(
    xaxis = list(title = "x"),
    yaxis = list(title = "q"),
    zaxis = list(title = "eps")
  )) |>
  layout(scene = list(legend = list(title = "Diff_from_x^2")))

fig

library(ggplot2)

Sqr_data_aux <- expand.grid(
  q = seq(2.1, 10, length.out = 100),
  eps = seq(0.01, 4, length.out = 100)
)

Sqr_data_aux$param <- 0

for (k in 1:10000) {
  Sqr_data_aux$param[k] <- Sqr(Sqr_data_aux$q[k], Sqr_data_aux$eps[k]) |>
      param()
}

experimental_params <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z = log10(
    param))) +
  geom_contour_filled() +
  theme_minimal() +
  labs(fill = "Log_10_number_of_parameters")

Sqr_data_aux$dep <- 0

for (k in 1:10000) {
  Sqr_data_aux$dep[k] <- Sqr(Sqr_data_aux[k, ]$q, Sqr_data_aux[k, ]$eps) |>
      dep()
}

experimental_deps <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z = log10(dep)))
      +
  geom_contour_filled(alpha = 0.8) +
  # scale_fill_continuous(breaks = seq(0, max(Sqr_data_aux$dep), by = 1)) +
  theme_minimal() +
  labs(fill = "log_10_experimental_depths")
```

```r
param_upper_limit <- function(q, eps) {
  (((40 * q) / (q - 2)) * ((1 / eps) |> log(2)) + 80 / (q - 2) - 28) |> max
    (52)
}

dep_upper_limit <- function(q, eps) {
  ((q / (2 * q - 4)) * log2(1 / eps) + 1 / (q - 2) + 1 / (q - 2) + 1) |> max
    (2)
}

Sqr_data_aux$param_upper_limit <- 0

for (k in 1:10000) {
  Sqr_data_aux$param_upper_limit[k] <- param_upper_limit(Sqr_data_aux[k, ]$q,
    Sqr_data_aux[k, ]$eps) |>
    ceiling()
}

param_theoretical_upper_limits <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z =
    log10(param_upper_limit))) +
  geom_contour_filled() +
  theme_minimal() +
  labs(fill = "Log10_upper_limits_of_parameters")

Sqr_data_aux$dep_upper_limit <- 0

for (k in 1:10000) {
  Sqr_data_aux$dep_upper_limit[k] <- dep_upper_limit(Sqr_data_aux[k, ]$q, Sqr_
    data_aux[k, ]$eps) |>
    ceiling()
}

dep_theoretical_upper_limits <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z =
    log10(dep_upper_limit))) +
  geom_contour_filled() +
  theme_minimal() +
  labs(fill = "Log10_upper_limits_of_depth")

ggsave("Sqr_properties/param_theoretical_upper_limits.png", plot = param_
    theoretical_upper_limits, width = 6, height = 5, units = "in")

ggsave("Sqr_properties/dep_theoretical_upper_limits.png", plot = dep_
    theoretical_upper_limits, width = 6, height = 5, units = "in")

ggsave("Sqr_properties/experimental_deps.png", plot = experimental_deps, width
    = 6, height = 5, units = "in")

ggsave("Sqr_properties/experimental_params.png", plot = experimental_params,
    width = 6, height = 5, units = "in")
```

274

Listing 9.19: R code simulations involving Sqr

```r
source("aux_fun.R")
source("Sqr.R")
source("instantiation.R")
source("activations.R")
library("tidyverse")

#' Sqr_diff function
#'
#' @param q parameter for the Sqr network
#' @param eps parameter for the Sqr network
#' @param x the number to be squered
#'
#' @return a neural network that approximately squares x.

Sqr_diff <- function(q, eps, x) {
  return <- (Sqr(q, eps) |> rlz(ReLU, x) - x^2) |> abs()
  return(return)
}

Sqr_diff_v <- Vectorize(Sqr_diff)

Sqr_data <- expand.grid(
  q = seq(2.1, 4, length.out = 50),
  eps = seq(0.01, 2, length.out = 50),
  x = seq(-5, 5, length.out = 50)
)



Sqr_data$diff <- Sqr_diff_v(Sqr_data$q, Sqr_data$eps, Sqr_data$x)

#' Function to calculate the theoretical upper bounds of the 1-norm error
#' over \mathbb{R}
#'
#' @param q parameter for the Sqr network
#' @param eps parameter for the Sqr network
#' @param x the number to be squered
#'
#' @return the maximum 1-norm error over \mathbb{R}

diff_upper_limit <- function(q, eps, x) {
  eps * max(1, abs(x)^q)
}

diff_upper_limit_v <- Vectorize(diff_upper_limit)

Sqr_data$diff_upper_limit <- diff_upper_limit_v(Sqr_data$q, Sqr_data$eps, Sqr_
    data$x)

write_csv(Sqr_data, "Sqr_properties/Sqr_data.csv")

library(plotly)
```

```r
fig <- plot_ly(
  type = "isosurface",
  x = Sqr_data$x,
  y = Sqr_data$q,
  z = Sqr_data$eps,
  value = Sqr_data$diff,
  isomin = 0.0001,
  isomax = 5,
  colorscale = "RdBu"
) |>
  layout(scene = list(
    xaxis = list(title = "x"),
    yaxis = list(title = "q"),
    zaxis = list(title = "eps")
  )) |>
  layout(scene = list(legend = list(title = "Diff_from_x^2")))

fig

library(ggplot2)

Sqr_data_aux <- expand.grid(
  q = seq(2.1, 10, length.out = 100),
  eps = seq(0.01, 4, length.out = 100)
)

Sqr_data_aux$param <- 0

for (k in 1:10000) {
  Sqr_data_aux$param[k] <- Sqr(Sqr_data_aux$q[k], Sqr_data_aux$eps[k]) |>
      param()
}

experimental_params <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z = log10(
    param))) +
  geom_contour_filled() +
  theme_minimal() +
  labs(fill = "Log_10_number_of_parameters")

Sqr_data_aux$dep <- 0

for (k in 1:10000) {
  Sqr_data_aux$dep[k] <- Sqr(Sqr_data_aux[k, ]$q, Sqr_data_aux[k, ]$eps) |>
      dep()
}

experimental_deps <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z = log10(dep)))
      +
  geom_contour_filled(alpha = 0.8) +
  # scale_fill_continuous(breaks = seq(0, max(Sqr_data_aux$dep), by = 1)) +
  theme_minimal() +
  labs(fill = "log_10_experimental_depths")
```

```r
param_upper_limit <- function(q, eps) {
  (((40 * q) / (q - 2)) * ((1 / eps) |> log(2)) + 80 / (q - 2) - 28) |> max
      (52)
}

dep_upper_limit <- function(q, eps) {
  ((q / (2 * q - 4)) * log2(1 / eps) + 1 / (q - 2) + 1 / (q - 2) + 1) |> max
      (2)
}

Sqr_data_aux$param_upper_limit <- 0

for (k in 1:10000) {
  Sqr_data_aux$param_upper_limit[k] <- param_upper_limit(Sqr_data_aux[k, ]$q,
      Sqr_data_aux[k, ]$eps) |>
    ceiling()
}

param_theoretical_upper_limits <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z =
    log10(param_upper_limit))) +
  geom_contour_filled() +
  theme_minimal() +
  labs(fill = "Log10_upper_limits_of_parameters")

Sqr_data_aux$dep_upper_limit <- 0

for (k in 1:10000) {
  Sqr_data_aux$dep_upper_limit[k] <- dep_upper_limit(Sqr_data_aux[k, ]$q, Sqr_
      data_aux[k, ]$eps) |>
    ceiling()
}

dep_theoretical_upper_limits <- ggplot(Sqr_data_aux, aes(x = q, y = eps, z =
    log10(dep_upper_limit))) +
  geom_contour_filled() +
  theme_minimal() +
  labs(fill = "Log10_upper_limits_of_depth")

ggsave("Sqr_properties/param_theoretical_upper_limits.png", plot = param_
    theoretical_upper_limits, width = 6, height = 5, units = "in")

ggsave("Sqr_properties/dep_theoretical_upper_limits.png", plot = dep_
    theoretical_upper_limits, width = 6, height = 5, units = "in")

ggsave("Sqr_properties/experimental_deps.png", plot = experimental_deps, width
    = 6, height = 5, units = "in")

ggsave("Sqr_properties/experimental_params.png", plot = experimental_params,
    width = 6, height = 5, units = "in")
```

Listing 9.20: R code for $\mathsf{Pwr}^{q,\varepsilon}$ networks

```r
source("R/Prd.R")
source("R/Aff.R")
source("R/stacking.R")
source("R/Tun.R")
source("R/aux_fun.R")

#' @title Pwr
#' @description
#' A function that returns the \eqn{\mathsf{Pwr}} neural networks.
#'
#'
#' @param q inside \eqn{(2,\infty)}.
#' @param eps inside \eqn{(0,\infty)}.
#' @param exponent the exponent which the Pwr network will approximate. Must
     be
#' a non-negative integer.
#'
#' @return A neural network that approximates raising a number to exponent,
     when
#' given appropriate \eqn{q,\varepsilon} and exponent when isntanatiated
#' under ReLU activation at \eqn{x}.
#' @export


Pwr <- function(q, eps, exponent) {
  if (q <= 2) {
    stop("Too_small_q,_q_must_be_>=_2")
  } else if (eps <= 0) {
    stop("Too_small_eps,_eps_must_be_>=_0")
  } else if (exponent %% 1 != 0 || exponent < 0) {
    stop("Exponent_must_be_a_non-negative_integer")
  } else {
    if (exponent == 0) {
      Aff(0, 1) -> return_network
      return(return_network)
    } else if (exponent >= 1) {
      Cpy(2, 1) -> first_third
      Pwr(q, eps, exponent - 1) |> stk(Pwr(q, eps, exponent - 1) |> dep() |>
          Tun()) -> mid_third
      Prd(q, eps) -> last_third
      last_third |>
        comp(mid_third) |>
        comp(first_third) -> return_network
    } else {
      return("Invalid_exponent,_must_be_non-negative_integer")
    }
    return(return_network)
  }
}
```

Listing 9.21: R code simulations involving $\mathsf{Pwr}_3^{q,\varepsilon}$

```r
source("Pwr.R")
library(tidyverse)

#' Pwr_3_diff function
#'
#' @param q parameter for Pwr_3
#' @param eps parameter for Pwr_3
#' @param x the number to be cubed
#' @param exponent = 3, i.e. cubing a number

Pwr_3_diff <- function(q, eps, x, exponent = 3) {
  return <- (Pwr(q, eps, exponent = 3) |> rlz(ReLU, x) - x^3) |> abs()
  return(return)
}

Pwr_3_diff_v <- Vectorize(Pwr_3_diff)

Pwr_3_data <- expand.grid(
  q = seq(2.1, 4, length.out = 50),
  eps = seq(0.01, 2, length.out = 50),
  x = seq(-5, 5, length.out = 50)
)

Pwr_3_data$diff <- Pwr_3_diff_v(Pwr_3_data$q, Pwr_3_data$eps, Pwr_3_data$x)

library(ggplot2)

ggplot(Pwr_3_data, aes(diff)) +
  scale_x_log10() +
  geom_density() +
  theme_minimal()

library(plotly)

fig <- plot_ly(
  type = "isosurface",
  x = Pwr_3_data$x,
  y = Pwr_3_data$q,
  z = Pwr_3_data$eps,
  value = Pwr_3_data$diff,
  isomin = 0.0001,
  isomax = 5,
  colorscale = "RdBu"
) |>
  layout(scene = list(
    xaxis = list(title = "x"),
    yaxis = list(title = "q"),
    zaxis = list(title = "eps")
  )) |>
  layout(scene = list(legend = list(title = "Diff_from_x^2")))

fig
```

```r
Pwr_3_data_aux <- expand.grid(
  q = seq(2.1, 10, length.out = 100),
  eps = seq(0.01, 4, length.out = 100)
)

Pwr_3_data_aux$param <- 0

for (k in 1:10000) {
  Pwr_3_data_aux$param[k] <- Pwr(Pwr_3_data_aux$q[k], Pwr_3_data_aux$eps[k],
      exponent = 3) |> param()
}

experimental_params <- ggplot(Pwr_3_data_aux, aes(x = q, y = eps, z = log10(
    param))) +
  geom_contour_filled() +
  theme_minimal() +
  labs(fill = "log_10_#_of_parameters")

Pwr_3_data_aux$dep <- 0

for (k in 1:10000) {
  Pwr_3_data_aux$dep[k] <- Pwr(Pwr_3_data_aux[k, ]$q, Pwr_3_data_aux[k, ]$eps,
      exponent = 3) |> dep()
}

experimental_deps <- ggplot(Pwr_3_data_aux, aes(x = q, y = eps, z = log10(dep)
    )) +
  geom_contour_filled(alpha = 0.8, breaks = seq(0, 10, 1)) +
  # scale_fill_continuous(breaks = seq(0, max(Pwr_3_data_aux$dep), by = 1)) +
  theme_minimal() +
  labs(fill = "log10_depths")

#' The param_upper_limit funnction
#'
#' @param q parameter for the Pwr network
#' @param eps parameter for the Pwr network
#'
#' @return the theoretical upper limit for the number of parameters

param_upper_limit <- function(q, eps) {
  4^(4.5) -> first_summand
  (4^4-1)/3 -> second_summand_a
  ((360*q)/(q-2))*(log2(1/eps)+q+1)+372 -> second_summand_b
  first_summand + (second_summand_a * second_summand_b) -> result
  return(result)
}


#' The dep_upper_limit function
#'
#' @param q parameter for the Pwr_3 network
#' @param eps parameter for the Pwr_3 network
#'
```

```
103  #' @return the theoretical upper limit for the depth
104
105  dep_upper_limit <- function(q, eps) {
106    ((q / (q - 2)) * (log2(1 / eps) + q) - 1) * 3 + 1
107  }
108
109  Pwr_3_data_aux$param_upper_limit <- 0
110
111  for (k in 1:10000) {
112    Pwr_3_data_aux$param_upper_limit[k] <- param_upper_limit(Pwr_3_data_aux[k, ]
           $q, Pwr_3_data_aux[k, ]$eps) |>
113      ceiling()
114  }
115
116  param_theoretical_upper_limits <- ggplot(Pwr_3_data_aux, aes(x = q, y = eps, z
         = log10(param_upper_limit))) +
117    geom_contour_filled() +
118    theme_minimal() +
119    labs(fill = "Log10_upper_limits_of_parameters")
120
121  Pwr_3_data_aux$dep_upper_limit <- 0
122
123  for (k in 1:10000) {
124    Pwr_3_data_aux$dep_upper_limit[k] <- dep_upper_limit(Pwr_3_data_aux[k, ]$q,
           Pwr_3_data_aux[k, ]$eps)
125  }
126
127  dep_theoretical_upper_limits <- ggplot(Pwr_3_data_aux, aes(x = q, y = eps, z =
         log10(dep_upper_limit))) +
128    geom_contour_filled() +
129    theme_minimal() +
130    labs(fill = "Log10_upper_limits_of_depth")
131
132  ggsave("Pwr_3_properties/param_theoretical_upper_limits.png", plot = param_
         theoretical_upper_limits, width = 6, height = 5, units = "in")
133  ggsave("Pwr_3_properties/dep_theoretical_upper_limits.png", plot = dep_
         theoretical_upper_limits, width = 6, height = 5, units = "in")
134  ggsave("Pwr_3_properties/experimental_deps.png", plot = experimental_deps,
         width = 6, height = 5, units = "in")
135  ggsave("Pwr_3_properties/experimental_params.png", plot = experimental_params,
         width = 6, height = 5, units = "in")
```

Listing 9.22: R code simulations involving Nrm_1^d

```
1  source("R/Aff.R")
2  source("R/stacking.R")
3  source("R/comp.R")
4  source("R/nn_sum.R")
5
6  #' @title Nrm
7  #'
8  #' @description
9  #' A function that creates the \eqn{\mathsf{Nrm}} neural networks.that take
```

```r
#' the 1- norm of a \eqn{d}-dimensional vector when instantiated with ReLU
#' activation.
#'
#'
#' @param d the dimensions of the vector being normed.
#'
#' @return a neural network that takes the 1-norm of a vector of
#' size d.under ReLU activation. This is the neural network that is:
#' \deqn{
#' \mathsf{Nrm}^1_1 = \left( \left( \begin{bmatrix} 1 \\ -1\end{bmatrix},
#' \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right), \left( \begin{bmatrix}1 && 1\
    end{bmatrix},
#' \begin{bmatrix}0\end{bmatrix}\right) \right) \in \left( \mathbb{R}^{2 \
    times 1} \times
#' \mathbb{R}^2 \right) \times \left( \mathbb{R}^{1 \times 2} \times \mathbb{R
    }^1 \right) \quad d=1 \\
#' \mathsf{Nrm}_1^d = \mathsf{Sum}_{d,1} \bullet \left[ \boxminus_{i=1}^d \
    mathsf{Nrm}_1^1 \right] \quad d>1
#' }
#'
#'
#'
#' \emph{Note:} This function is split into two cases
#' much like the definition itself.
#'
#' @references Lemma 4.2.1. Jentzen, A., Kuckuck, B., and von Wurstemberger, P
    . (2023).
#' Mathematical introduction to deep learning: Methods, implementations,
#' and theory. \url{https://arxiv.org/abs/2310.20360}

#' @export
#'
Nrm <- function(d) {
  if (d %% 1 != 0 || d < 1) {
    stop("d_must_be_a_natural_number")
  } else {
    if (d == 1) {
      c(1, -1) |> matrix() -> W_1
      c(0, 0) |> matrix() -> b_1
      c(1, 1) |> matrix(1, 2) -> W_2
      0 |> matrix() -> b_2

      list(W = W_1, b = b_1) -> layer_1
      list(W = W_2, b = b_2) -> layer_2

      list(layer_1, layer_2) -> return_network

      return(return_network)
    } else if (d > 1) {
      1 |> Nrm() -> first_compose
      for (i in 1:(d - 1)) {
        first_compose |> stk(Nrm(1)) -> first_compose
```

```
58          }
59          Sum(d, 1) |> comp(first_compose) -> return_network
60          return(return_network)
61      } else {
62          stop("Possibly_taking_the_norm_of_an_invalid_sized_array")
63      }
64    }
65 }
```

Listing 9.23: R code simulations involving Mxm_d

```
1 source("R/Aff.R")
2 source("R/stacking.R")
3 source("R/comp.R")
4 source("R/nn_sum.R")
5 source("R/Id.R")
6
7 #' @title Mxm
8 #' @description The function that returns the \eqn{\mathsf{Mxm}} neural
     networks.
9 #' These are neural networks of the type:
10 #' \deqn{
11 #'\mathsf{Mxm}^1 = \mathsf{Aff}_{1,0} \quad d = 1 \\
12 #'\\
13 #'\mathsf{Mxm}^2 = \left( \left( \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1\end
     {bmatrix},
14 #'\begin{bmatrix} 0 \\ 0 \\0\end{bmatrix}\right), \left( \begin{bmatrix}1&1&
     -1\end{bmatrix},
15 #'\begin{bmatrix}0\end{bmatrix}\right)\right) \quad d = 2 \\
16 #'\\
17 #'\mathsf{Mxm}^{2d} = \mathsf{Mxm}^d \bullet \left[ \boxminus_{i=1}^d \mathsf{
     Mxm}^2\right] \quad d > 2\\
18 #'\mathsf{Mxm}^{2d-1} = \mathsf{Mxm}^d \bullet \left[ \left( \boxminus^d_{i=1}
      \mathsf{Mxm}^2 \right)
19 #'\boxminus \mathsf{Id}_1\right] \quad d > 2
20 #'
21 #'}
22 #'
23
24 #' \emph{Note:} Because of certain quirks of R we will have split
25 #' into five cases. We add an extra case for \eqn{d = 3}. Unlike the paper
26 #' we will simply reverse engineer the appropriate \emph{d}.
27 #'
28 #' @param d The dimension of the input vector on instantiation.
29 #'
30 #' @return The neural network that will ouput the maximum of a vector of
31 #' size \eqn{d} when activated with the ReLU function.
32 #'
33 #' @references Lemma 4.2.4. Jentzen, A., Kuckuck, B., and von Wurstemberger, P
     . (2023).
34 #' Mathematical introduction to deep learning: Methods, implementations,
35 #' and theory. \url{https://arxiv.org/abs/2310.20360}
36
37
```

```r
#' @export

Mxm <- function(d) {
  if (d %% 1 != 0 || d < 1) {
    stop("d_must_be_a_natural_number")
  }
  if (d == 1) {
    return(Aff(1, 0))
  } else if (d == 2) {
    c(1, 0, 0, -1, 1, -1) |> matrix(3, 2) -> W_1
    c(0, 0, 0) |> matrix() -> b_1
    c(1, 1, -1) |> matrix(1, 3) -> W_2
    0 |> matrix() -> b_2
    list(W = W_1, b = b_1) -> layer_1
    list(W = W_2, b = b_2) -> layer_2
    list(layer_1, layer_2) -> return_network
    return(return_network)
  } else if (d == 3) {
    Mxm(2) |> stk(Id()) -> first_compose
    Mxm(2) |> comp(first_compose) -> return_network
    return(return_network)
  } else if ((d %% 2 == 0) & (d > 3)) {
    d / 2 -> d
    Mxm(2) -> first_compose
    for (i in 1:(d - 1)) {
      first_compose |> stk(Mxm(2)) -> first_compose
    }
    Mxm(d) |> comp(first_compose) -> return_network
    return(return_network)
  } else if ((d %% 2 != 0) & (d > 3)) {
    (d - 1) / 2 -> d

    Mxm(2) -> first_compose
    for (i in 1:(d - 1)) {
      first_compose |> stk(Mxm(2)) -> first_compose
    }
    first_compose |> stk(Id()) -> first_compose
    Mxm(d + 1) |> comp(first_compose) -> return_network
    return(return_network)
  } else {
    stop("Possibly_taking_max_of_vector_of_length_0")
  }
}
```

Listing 9.24: R code simulations involving Tay, note that this implementation is different from how it is presented in the exposition. We chose to explicitly define the Tay network, and let neural network exponentials, cosines, and sines be instantiations of this network with various different coefficients.

```r
source("R/Pwr.R")
source("R/nn_sum.R")
source("R/scalar_mult.R")
source("R/Aff.R")
```

```r
#' The Tay function
#'
#' @param f the function to be Taylor approximated, for now "exp", "sin"
#' and "cos". NOTE use the quotation marks when using this arguments
#' @param n the extent of Taylor approximations, a natural number
#' @param q argument for the Pwr networks \eqn{q \in (2,\infty)}
#' @param eps argument for the Pwr networks \eqn{eps \in (0,\infty)}
#'
#' @return a neural network that approximates the function f

Tay <- function(f, n, q, eps) {
  if (n %% 1 != 0 || n < 0) {
    stop("Number_of_Taylor_iteration_must_be_a_non_negative_integer")
  } else if (q < 2 || eps < 0) {
    stop("q_must_be_>_2_and_eps_must_be_>_0")
  } else if (f != "exp" && f != "sin" && f != "cos") {
    stop("For_now,_only_Taylor_approximations_for_exp,_sin,_and_cos_is_
        available")
  } else {
    if (f == "exp") {
      (1 / factorial(0)) |> slm(Pwr(q, eps, 0)) -> return_network
      if (n == 0) {
        return(return_network)
      }
      for (i in 1:n) {
        return_network |> nn_sum((1 / factorial(i)) |> slm(Pwr(q, eps, i))) ->
              return_network
      }
      return(return_network)
    }

    if (f == "cos") {
      1 |> slm(Pwr(q, eps, 0)) -> return_network
      if (n == 0) {
        return(return_network)
      }

      for (i in 1:n) {
        ((-1)^i) / factorial(2 * i) -> coeff
        return_network |> nn_sum(coeff |> slm(Pwr(q, eps, 2 * i))) -> return_
            network
      }
      return(return_network)
    }

    if (f == "sin") {
      Tay("cos", n, q, eps) -> return_network
      return_network |> comp(Aff(1, -pi / 2)) -> return_network
      return(return_network)
    }
  }
}
```

Listing 9.25: R code simulations for $\mathsf{Csn}_n^{q,\varepsilon}$

```r
source("R/Tay.R")

#' @title Csn
#' @description The function that returns \eqn{\mathsf{Csn}}.
#'
#' @param n The number of Taylor iterations
#' @param q a real number in \eqn{(2,\infty)}
#' @param eps a real number in \eqn{(0,\infty)}
#'
#' \emph{Note: } In practice for most desktop uses
#' \eqn{q < 2.05} and \eqn{\varepsilon< 0.05} tends to cause problems in
#' "too long a vector", atleaast as tested on my computer.
#'
#' @return A neural network that approximates \eqn{\cos} under instantiation
#' with ReLU activation. See also \code{\link{Sne}}.
#' @export

Csn <- function(n, q, eps) {
  if (q <= 2 || eps <= 0) {
    stop("q_must_be_>_2_and_eps_must_be_>_0")
  } else if (n %% 1 != 0 || n < 0) {
    stop("The_number_of_Taylor_iterations_must_be_non_negative_integer")
  } else {
    Tay("cos", n, q, eps) -> return_network
    return(return_network)
  }
}
```

Listing 9.26: R code simulations for $\mathsf{Sne}_n^{q,\varepsilon}$

```r
source("R/Tay.R")

#' @title Sne
#' @description Returns the \eqn{\mathsf{Sne}} neural networks
#'
#'
#' @param n the number of Taylor iterations
#' @param q a real number in \eqn{(2,\infty)}
#' @param eps a real number in \eqn{(0,\infty)}
#'
#' \emph{Note: } In practice for most desktop uses
#' \eqn{q < 2.05} and \eqn{\varepsilon< 0.05} tends to cause problems in
#' "too long a vector", atleaast as tested on my computer.
#'
#' @return a neural network that approximates \eqn{\sin} when given
#' an appropriate \eqn{n,q,\varepsilon} and instantiated with ReLU
#' activation and given value \eqn{x}.
#' @export

Sne <- function(n, q, eps) {
  if (q <= 2 || eps <= 0) {
    stop("q_must_be_>_2_and_eps_must_be_>_0")
```

```
23    } else if (n %% 1 != 0 || n < 0) {
24       stop("The_number_of_Taylor_iterations_must_be_non_negative_integer")
25    } else {
26       return(Tay("sin", n, q, eps))
27    }
28 }
```

Listing 9.27: R code simulations involving Etr

```
1  #' @title Trp
2  #' @description The function that returns the \eqn{\mathsf{Trp}} networks.
3  #'
4  #' @param h the horizontal distance between two mesh points
5  #'
6  #' @return The \eqn{\mathsf{Trp}} network that gives the area
7  #' when activated with ReLU and two meshpoint values x_1 and x_2.
8  #' @export
9
10 Trp <- function(h) {
11    if (h |> is.numeric() &&
12       h |> length() == 1 &&
13       h |> is.finite() &&
14       h > 0) {
15       c(h / 2, h / 2) |> matrix(1, 2) -> W
16       0 |> matrix() -> b
17       list(list(W = W, b = b)) -> return_network
18       return(return_network)
19    } else {
20       stop("h_must_be_a_positive_real_number")
21    }
22 }
23
24 #' @title Etr
25 #' @description The function that returns the \eqn{\mathsf{Etr}} networks.
26 #'
27 #' @param n number of trapezoids to make. Note this will result in a set of
28 #' trapezoids.
29 #' Note that this will result in n+1 meshpoints including the starting a and
30 #' ending b
31 #'
32 #' \emph{Note: } Upon instantiation with any continuous function this neural
33 #' network must be fed with \eqn{n+1} real numbers representing the values
34 #' of the function being approximated at the \eqn{n+1} meshpoints which are
35 #' the legs of the \eqn{n} triangles as stipulated in the input parameters.
36 #'
37 #' @param h width of trapezoids
38 #'
39 #' @return an approximation for area of the integral
40 #' @export
41
42 Etr <- function(n, h) {
43    if (h |> is.numeric() &&
44       h |> length() == 1 &&
45       h |> is.finite() &&
```

```
46      h > 0 &&
47      n %% 1 == 0 &&
48      n > 1) {
49      c(h / 2, rep(h, n - 1), h / 2) |>
50        matrix() |>
51        t() -> W
52      0 |> matrix() -> b
53      list(list(W = W, b = b)) -> return_network
54      return(return_network)
55    } else {
56      stop("n_must_be_a_natural_number_and_h_must_be_a_positive_real_number.")
57    }
58 }
```

Listing 9.28: R code simulations involving Etr

```
1 source("R/stacking.R")
2 source("R/comp.R")
3 source("R/Nrm.R")
4 source("R/Mxm.R")
5
6 #' The P neural network
7 #'
8 #' @param X a matrix. More precisely a collection of n row vectors of size d.
9 #' That is to say an d by n matrix
10 #' @param y a row vector of size m that is the ouput of the function at each
11 #' of the n sample points
12 #' @param L the Lipschitz constant for the function.
13 #'
14 #' @references Lemma 4.2.9. Jentzen, A., Kuckuck, B., and von Wurstemberger, P
       . (2023).
15 #' Mathematical introduction to deep learning: Methods, implementations,
16 #' and theory. \url{https://arxiv.org/abs/2310.20360}.
17 #'
18 #' @return A neural network that gives the maximum convolution approximation
19 #' of a function whose outputs is \eqn{y} at \eqn{n} sample points given by
20 #' each row of \eqn{X}, when instantiated with ReLU.
21 #' @export
22
23 P <- function(X, y, L) {
24   if (X |> is.matrix() == FALSE) (X |> matrix() -> X)
25   if (y |> is.matrix() == FALSE) (y |> matrix() -> y)
26   X |> nrow() -> d # the dimensionality of our x samples
27   X |> ncol() -> n # the number of samples to be taken
28
29   if (n == 1) {
30     return("Enter_atleast_2_interpolating_points")
31   }
32   Cpy(n, d) -> first_compose # the first neural network to be hit with x
33   Nrm(d) |> comp(Aff(diag(d), -X[, 1])) -> second_compose # the second neural
        network to be hit
34   for (j in 2:n) {
35     second_compose |> stk(Nrm(d) |> comp(Aff(diag(d), -X[, j]))) -> second_
          compose
```

```
 36    }
 37   Aff(-L * diag(n), y) -> third_compose # the third neural network to be hit
 38   n |> Mxm() -> fourth_compose # the fourth neural network to be hit
 39   fourth_compose |>
 40     comp(third_compose) |>
 41     comp(second_compose) |>
 42     comp(first_compose) -> return_network
 43   # the final neural network
 44   return(return_network)
 45 }
```

Vita

The author was born in November 1st, 1992 in the city of Dhaka in the heart of Bangladesh. He grew up in the large city with a childhood that included setting things on fire, and very occasionally focusing on mathematics. He failed to achieve his childhood goal of becoming an astronomer however when he entered college at Troy University in 2011 and realized it would involve cold nights outside, and so chose mathematics instead. He has continued his pursuits in mathematics and is now a graduate student at the University of Arkansas trying to graduate.