Reformulation without f

Shakil Rafi

September 12, 2023

Contents

Part I

On Convergence of Brownian Motion Monte Carlo

Chapter 1

Introduction.

1.1 Notation, Definitions & Basic notions.

We introduce here basic notations that we will be using throughout this part. Large parts are taken from standard literature inspired by *Matrix Computations* by Golub and Van Loan (2013), and *Probability: Theory & Examples* by Rick Durrett (2019).

1.1.1 Norms and Inner Product

Definition 1.1.1 (Euclidean Norm). *Let $\|\cdot\|_E : \mathbb{R}^d \to [0,\infty)$ denote the Euclidean norm defined for all $d \in \mathbb{N}_0$ and $x = \{x_1, x_2, \cdots, x_d\} \in \mathbb{R}^d$ as:*

$$\|x\|_E = \left(\sum_{i=1}^{d} x_i^2 \right)^{\frac{1}{2}} \tag{1.1.1}$$

For the special case that $d = 1$, $d = 2$, and where it is clear from context we will denote $\|\cdot\|_E$ as $|\cdot|$.

Definition 1.1.2 (Max Norm). *Let $\|\cdot\|_\infty : \mathbb{R}^d \to [0,\infty)$ denote the max norm define for all $d \in \mathbb{N}_0$ and $x = \{x_1, x_2, \cdots, x_d\} \in \mathbb{R}^d$ as:*

$$\|x\|_\infty = \max_{i \in \{1,2,\cdots,d\}} \{|x_i|\} \tag{1.1.2}$$

Definition 1.1.3 (Frobenius Norm). *Let $\|\cdot\|_F : \mathbb{R}^{m \times n} \to [0,\infty)$ denote the Frobenius norm defined* *for all* $m, n \in \mathbb{N}$, $A \in \mathbb{R}^{m \times n}$ *as:*

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} [A]_{i,j}^2 \right)^{\frac{1}{2}} \tag{1.1.3}$$

Definition 1.1.4 (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 all* $d \in \mathbb{N}$, $\mathbb{R}^d \ni x = \{x_1, x_2, ..., x_d\}$, *and* $\mathbb{R}^d \ni y = \{y_1, y_2, ..., y_d\}$ *as:*

$$\langle x, y \rangle = \sum_{i=1}^{d} (x_i y_i) \tag{1.1.4}$$

1.1.2 Probability Space and Brownian Motion

Definition 1.1.5 (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 mapping: $\mathbb{P} : \mathcal{F} \to [0, 1]$ *assigning each event in the* ***event space*** *a probability between* 0 *and* 1*. 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 1.1.6 (Random Variable)**.** *Given a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *a random variable is a measurable function* $\mathcal{X} : \Omega \to \mathbb{R}^d$*.*

Definition 1.1.7 (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:*

$$\int_{\Omega} X d\mathbb{P} \tag{1.1.5}$$

Definition 1.1.8 (Stochastic Process)**.** *A stochastic process is a family of random variables over a fixed probability space* $(\Omega, \mathcal{F}, \mathbb{R})$*.*

Definition 1.1.9 (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 a total order, it is the case that* $\mathcal{F}_i$ *is a sub* σ*-algebra of* $\mathcal{F}$*.*

Definition 1.1.10 (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) $\mathcal{W}_t - \mathcal{W}_s$ *is a normal random variable with* $\mu = 0$ *and* $\sigma^2 = t - s$ *when* $s < t$*.*

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

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

Definition 1.1.11 ($(\mathbb{F}_t)_{t \in [0,T]}$-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 1.1.12 ($(\mathbb{F}_t)_{t \in [0,T]}$-adapted stopping time)**.** *Let* $T \in (0, \infty)$*,* $\tau \in [0,T]$*. Assume a filtered probability space* $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$*.* $\tau \in \mathbb{R}$ *is a stopping time if the stochastic process* $\mathcal{X} = (\mathcal{X}_t)_{t \in [0,T]}$ *define as:*

$$\mathcal{X}_t := \begin{cases} 1 \text{ if } t < \tau \\ 0 \text{ if } t \geqslant \tau \end{cases} \tag{1.1.6}$$

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

Definition 1.1.13 (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.* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t \in [0,T]})$ *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 every* $t \in [0,T]$*,* $x \in \mathbb{R}^d$*, let* $\mathcal{X}^{t,x} = (\mathcal{X}^{t,x}_s)_{s \in [t,T]} \times \Omega \to \mathbb{R}^d$ *be an* $(\mathbb{F}_s)_{s \in [t,T]}$*-adapted schochastic process with continuous sample* *paths satisfying that for all $t \in [0,T]$ we have $\mathbb{P}$-a.s. that:*

$$\mathcal{X}^{t,x} = \mathcal{X}_0 + \int_0^t \mu(r, \mathcal{X}_r^{t,x})dr + \int_0^t \sigma(r, \mathcal{X}_r^{t,x})d\mathcal{W}_r \tag{1.1.7}$$

A strong solution to the stochastic differential equation (1.1.7) on probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_t)_{t\in[0,T]})$, w.r.t Briwnian 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:

$$\mathcal{X}^{t,x} = \mathcal{X}_0 + \int_0^t \mu(r, \mathcal{X}_r^{t,x})dr + \int_0^t \sigma(r, \mathcal{X}_r^{t,x})d\mathcal{W}_r \tag{1.1.8}$$

Definition 1.1.14 (Strong Uniqueness Property for Solutions to Stochastic Differential Equations)**.** *Assume that whenever we have two strong solutions $\mathcal{X}, \widetilde{\mathcal{X}}$, w.r.t. process $\mathcal{W}$ and initial condition $\mathcal{X}_0 = 0$, as defined in Definition 1.1.13, it is the case that for all $t \in [0,T]$ we have $\mathbb{P}(\mathcal{X}_t = \widetilde{\mathcal{X}}_t) = 1$, we then say that the pair (μ, σ) exhibit a strong uniqueness property.*

1.1.3 Lipschitz and Related Notions

Definition 1.1.15 (Globally Lipschitz Function)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}^d$ is (globally) Lipschitz if there exists an $L \in (0,\infty)$ such that:*

$$\sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \left\| \frac{f(x) - f(y)}{x - y} \right\|_E \leqslant L \tag{1.1.9}$$

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

Corollary 1.1.15.1. *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 $x_i \in \mathcal{K}$:

$$\mathcal{K} \subseteq \bigcup_{i=1}^{N} B_\delta(x_i) \tag{1.1.10}$$

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

$$|f(x_i) - f(x_j)| < \epsilon \quad \forall i, j \in \{1, 2, ..., N\} \tag{1.1.11}$$

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

$$\mathfrak{L} = \max_{\substack{i,j \in \{1,2,...,N\} \\ i \neq j}} \left| \frac{f(x_i) - f(x_j)}{x_i - x_j} \right| \tag{1.1.12}$$

$\mathfrak{L}$ 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:

$$\begin{aligned}
|f(x_1) - f(x_2)| &\leqslant |f(x_1) - f(x_i)| + |f(x_i) - f(x_j)| + |f(x_j) - f(x_2)| \\
&\leqslant |f(x_i) - f(x_j)| + 2\epsilon \\
&\leqslant \mathfrak{L} \cdot |x_i - x_j| + 2\epsilon \\
&\leqslant \mathfrak{L} \cdot |x_1 - x_2| + 2\epsilon
\end{aligned}$$

for all $\epsilon \in (0, \infty)$. □

Definition 1.1.16 (Locally Lipschitz Function)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}^d$ is locally Lipschitz if for every $x_0 \in \mathbb{R}^d$ there exists a compact set $\mathcal{K} \subseteq$ Domain(f) containing x_0, and constant $L \in (0, \infty)$*

such that

$$\sup_{\substack{x,y \in \mathcal{K} \\ x \neq y}} \left\| \frac{f(x) - f(y)}{x - y} \right\|_E \leqslant L \tag{1.1.13}$$

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

Corollary 1.1.16.1. *A function $f : \mathbb{R}^d \to \mathbb{R}^d$ that is globally Lipschitz is also locally Lipschitz. More concisely* $\text{Lip}_G(X) \subsetneq \text{Lip}_L(X)$*.*

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

$$\frac{|f(x_1) - f(x_2)|}{x_1 - x_2} \geqslant \mathfrak{L} \tag{1.1.14}$$

This directly contradicts Definition 1.1.15. □

1.1.4 Kolmogorov Equations

Definition 1.1.17 (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 atleast Lipschitz a Kolmogorov Equation is an equation of the form:*

$$\begin{cases} \left(\frac{\partial}{\partial t} u\right)(t,x) = \frac{1}{2} \operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^* \left(\operatorname{Hess}_x u\right)(t,x)\right) + \left\langle \mu(t,x), \left(\nabla_x u\right)(t,x)\right\rangle \\ u(0,x) = g(x) \end{cases} \tag{1.1.15}$$

Definition 1.1.18 (Strict Solution to Kolmogorov Equation)**.** *A function $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ is a strict solution to (1.1.15) 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 any $x \in \text{Domain}(A)$, $u(\cdot,x)$ is continuously differentiable on $[0,\infty)$ and satisfies (1.1.15).

Definition 1.1.19 (Generalized Solution to Kolmogorov Equation)**.** *A generalized solution to (1.1.15) is defined as:*

$$u(t,x) = \mathbb{E}\left[g\left(\mathcal{X}^{t,x}\right)\right] \tag{1.1.16}$$

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

$$\mathcal{X}^{t,x} = \int_0^t \mu\left(\mathcal{X}_r^{t,x}\right) dr + \int_0^t \sigma\left(\mathcal{X}_r^{t,x}\right) dW_r \tag{1.1.17}$$

Definition 1.1.20 (Laplace Operator w.r.t. x)**.** *Given a function $f \in C^2(\mathbb{R}^d, \mathbb{R})$, the Laplace operator $\nabla_x^2 : C^2(\mathbb{R}^d, \mathbb{R}) \to \mathbb{R}$ is defined as:*

$$\Delta_x f = \nabla_x^2 f := \nabla \cdot \nabla f = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i} \tag{1.1.18}$$

1.1.5 Linear Algebra Notation and Definitions

Definition 1.1.21 (Identity, Zero Matrix, and the 1-matrix)**.** *We will define the identity matrix in dimension $d \in \mathbb{N}_0$ as the matrix $\mathbb{I}_d \in \mathbb{R}^{d \times d}$ where:*

$$\mathbb{I}_d \coloneqq [\mathbb{I}_d]_{i,j} = \begin{cases} 1 & i = j \\ 0 & else \end{cases} \tag{1.1.19}$$

Note that $\mathbb{I}_0 = 1$.

For $m, n \in \mathbb{N}$ the zero matrix $\mathbb{0}_{m,n} \in \mathbb{R}^{m \times n}$ as:

$$\mathbb{0}_{m,n} \coloneqq [\mathbb{0}_{m,n}]_{i,j} = 0 \quad \forall i, j \tag{1.1.20}$$

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

The matrix of ones $\mathbb{e}_{m,n} \in \mathbb{R}^{m \times n}$ *as:*

$$\mathbb{e}_{m,n} \coloneqq [\mathbb{e}]_{i,j} = 1 \quad \forall i, j \tag{1.1.21}$$

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 1.1.22 (Complex conjugate and transpose)**.** *Let* $m, n \in \mathbb{N}$*, and* $A \in \mathbb{C}^{m \times n}$*. We denote by* $A^* \in \mathbb{C}^{n \times m}$ *the matrix:*

$$[A^*]_{i,j} = \overline{[A]_{j,i}} \quad \forall i, j \tag{1.1.22}$$

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 1.1.23 (Column, Row, and General Vector Notation)**.** *Let* $m, n \in \mathbb{N}$ *and let* $A \in \mathbb{R}^{m \times n}$*. We denote the* i*-th row for* $1 \leqslant i \leqslant m$ *as:*

$$[A]_{i,*} = \begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,n} \end{bmatrix} \tag{1.1.23}$$

Similarly we done the j*-th row for* $1 \leqslant i \leqslant n$ *as:*

$$[A]_{*,j} = \begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix} \tag{1.1.24}$$

Definition 1.1.24 (Kronecker Product)**.** *Let* $m_1, n_1, m_2, n_2 \in \mathbb{N}$*. Given matrix* $A \in \mathbb{R}^{m_1, n_1}$ *and* $B \in \mathbb{R}^{m_2, n_2}$ *we define the Kronecker product* $A \otimes_K B \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$ *as the block matrix given by:*

$$A \otimes_K B \coloneqq [A \otimes_K B]_{i,j} = [A]_{i,j} \, B \tag{1.1.25}$$

1.1.6 O-type notation and function growth

Definition 1.1.25 (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 $c, x_0 \in \mathbb{R}$ such that:*

$$0 \leqslant f(x) \leqslant cg(x) \qquad \text{for all } x \geqslant x_0 \tag{1.1.26}$$

We say that $f \in \Omega(g(x))$ if there exists $c, x_0 \in \mathbb{R}$ such that:

$$0 \leqslant cg(x) \leqslant f(x) \qquad \text{for all } x \geqslant x_0 \tag{1.1.27}$$

We say that $f \in \Theta(g(x))$ if there exists $c_1, c_2, x_0 \in \mathbb{R}$ such that:

$$0 \leqslant c_1 g(x) \leqslant f \leqslant c_2 g(x) \qquad \text{for all } x \geqslant x_0 \tag{1.1.28}$$

Corollary 1.1.25.1 (Bounded functions and O-type notation)**.** *Let $f(x) \in C(\mathbb{R}, \mathbb{R})$, then:*

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

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

(iii) if f is bounded above and below 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 $x_0 \in R$ such that for all $x \geqslant x_0$ it is 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 $x_0 \in R$ such that for all $x \geqslant x_0$ it is 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).

□

Corollary 1.1.25.2. *Let $f \in O(x^n)$ for $n \in \mathbb{N}_0$. Then it is also the case that $f \in O\left(x^{n+1}\right)$.*

Proof. Let $f \in O(x^n)$. Then there exists $c, x_0 \in \mathbb{R}$, such that:

$$f(x) \leqslant cx^n \text{ for all } x \geqslant x_0 \tag{1.1.29}$$

Note however that for all $n \in \mathbb{N}$, there exists $x_1 \in \mathbb{R}$ such that:

$$x^n \leqslant x^{n+1} \text{ for all } x \geqslant x_1 \tag{1.1.30}$$

Thus:

$$f(x) \leqslant cx^n \leqslant cx^{n+1} \text{ for all } x \geqslant \max\{x_0, x_1\} \tag{1.1.31}$$

□

Chapter 2

Brownian Motion Monte Carlo

2.1 Brownian Motion Preliminaries

Lemma 2.1.1 (Time reversal 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. Let $\mathfrak{W} = W_{T-t} - W_t$. Then $\mathfrak{W}_s = \{\mathfrak{W}_s : s \in [t,T]\}$ is also a standard Brownian motion on $[0,T]$.*

Proof. $\mathfrak{W}$ is a Gaussian process, since a finite, linear combination of variables from theis process reduces to a finite, linear combination of variables from W. Next $\mathbb{E}[\mathfrak{W}_t] = \mathbb{E}[W_{T-t}] - \mathbb{E}(W_T) = 0$. Next if $s,t \in [0,T]$ with $s \leqslant t$ then

$$
\begin{aligned}
\mathrm{Cov}(\mathfrak{W}_\mathfrak{s}, \mathfrak{W}_\mathrm{t}) &= \mathrm{Cov}(W_{T-s} - W_T, W_{T-t} - W_t) \\
&= \mathrm{Cov}(W_{T-s}, W_{T-t}) - \mathrm{Cov}(W_{T-s}, W_T) - \mathrm{Cov}(W_T, W_{T-t}) + \mathrm{Cov}(W_T, W_t) \\
&= (T-t) - (T-s) - (T-t) + T = s \qquad (2.1.1)
\end{aligned}
$$

Finally $t \mapsto \mathfrak{W}_t$ is continuous of $[0,T]$ with probability 1, since $t \mapsto W_t$ is continuous on $[0,T]$ with probability 1. □

Lemma 2.1.2 (Shift 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 $\mathfrak{W}_t = W_{s+t} - W_s$. Then $\mathfrak{W} = \{\mathfrak{W}_t : t \in [0,\infty)\}$ is also a standard Brownian motion.*

Proof. Since W has stationary, independent increments, the process $\mathfrak{W}$ is equivalent in distribution to W. Clearly also, $\mathfrak{W}$ is continuous as W is. □

Lemma 2.1.3. *The product of a constant with a Brownian motion is a Brownian motion*

Lemma 2.1.4. *The sum of Brownian motions is a Brownian motion.*

Definition 2.1.5 (Of $\mathfrak{k}$)**.** *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 $\left(\mathbb{E}\left[|\mathcal{X} - \mathbb{E}\left[\mathcal{X}\right]\right)^p]\right)^{\frac{1}{p}} \leqslant c \left(\mathbb{E}\left[|\mathcal{X}|^p\right]\right)^{\frac{1}{p}}$.*

Definition 2.1.6 (Primary Setting)**.** *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:*

$$\max\{|g(x)|\} \leqslant \mathfrak{L}\left(1 + \|x\|_E^p\right) \tag{2.1.2}$$

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:

$$u(t,x) = \mathbb{E}\left[g\left(x + \mathcal{W}^0_{T-t}\right)\right] \tag{2.1.3}$$

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:

$$U^\theta_m(t,x) = \frac{1}{m}\left[\sum_{k=1}^{m} g\left(x + \mathcal{W}^{(\theta,0,-k)}_{T-t}\right)\right] \tag{2.1.4}$$

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 $\mathcal{W}^{(\theta,0,-k)}_{T-t}$ 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 $U^\theta : [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$.

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. □

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:*

$$\mathbb{E}\left[\left|U^\theta\left(t, x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \mathbb{E}\left[\left|g\left(x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \int_s^T \mathbb{E}\left[\left|U^\theta\left(r, x + \mathcal{W}^\theta_{r-s}\right)\right|\right] dr < \infty \tag{2.1.5}$$

Proof. Note that (2.1.2), 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:

$$\begin{aligned}\mathbb{E}\left[\left|g(x + \mathcal{W}^\theta_{t-s})\right|\right] &\leqslant \mathbb{E}\left[\mathfrak{L}\left(1 + \|x + \mathcal{W}^\theta_{t-s}\|^p_E\right)\right] \\ &\leqslant \mathfrak{L}\left[1 + 2^{\max\{p-1,0\}}\left(\|x\|^p_E + \mathbb{E}\left[\left\|\mathcal{W}^\theta_T\right\|^p_E\right]\right)\right] < \infty\end{aligned} \tag{2.1.6}$$

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

$$\mathbb{E}\left[\left|U^\theta\left(t, x + \mathcal{W}^\theta_{t-s}\right)\right|\right] + \int_s^T \mathbb{E}\left[\left[U^\theta\left(r, x + \mathcal{W}^\theta_{r-s}\right)\right|\right] dr < \infty \tag{2.1.7}$$

To prove this claim observe the triangle inequality and (2.1.4), demonstrate that for all $s \in [0,T]$,

$t \in [s,T]$, $\theta \in \Theta$, it holds that:

$$\mathbb{E}\left[\left|U^\theta\left(t, x+\mathcal{W}^\theta_{t-s}\right)\right|\right] \leqslant \frac{1}{m}\left[\sum_{i=1}^{m} \mathbb{E}\left[\left|g\left(x+\mathcal{W}^\theta_{t-s}+\mathcal{W}^{(\theta,0,-i)}_{T-t}\right)\right|\right]\right] \tag{2.1.8}$$

Now observe that (2.1.6) 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:

$$\mathbb{E}\left[\left|g\left(x+\mathcal{W}^\theta_{t-s}+\mathcal{W}^{(\theta,0,i)}_{T-t}\right)\right|\right] = \mathbb{E}\left[\left|g\left(x+\mathcal{W}^\theta_{(t-s)+(T-t)}\right)\right|\right] = \mathbb{E}\left[\left|g\left(x+\mathcal{W}^\theta_{T-s}\right)\right|\right] < \infty \tag{2.1.9}$$

Combining (2.1.8) and (2.1.9) demonstrate that for all $s \in [0,T]$, $t \in [s,T]$, $\theta \in \Theta$ it holds that:

$$\mathbb{E}\left[\left|U^\theta(t, x+\mathcal{W}^\theta_{t-s})\right|\right] < \infty \tag{2.1.10}$$

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

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

Combining (**??**), (2.1.10), and (2.1.11) completes the proof of Lemma 2.1.8.

□

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:*

$$\mathbb{E}\left[\left|U^0(t,x)\right|\right] + \mathbb{E}\left[\left|g\left(x+\mathcal{W}^{(0,0,-1)}_{T-t}\right)\right|\right] < \infty \tag{2.1.12}$$

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

$$\mathbb{E}\left[U^0(t,x)\right] = \mathbb{E}\left[g\left(x+\mathcal{W}^{(0,0,-1)}_{T-t}\right)\right] \tag{2.1.13}$$

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

$$\begin{aligned}
&\mathbb{E}\left[\left|U^0\left(t,x\right)\right|\right] + \mathbb{E}\left[\left|g\left(x+\mathcal{W}^{(0,0,-1)}_{T-t}\right)\right|\right] \\
&< \mathbb{E}\left[\left|U^\theta\left(t,x+\mathcal{W}^\theta_{t-s}\right)\right|\right] + \mathbb{E}\left[\left|g\left(x+\mathcal{W}^\theta_{t-s}\right)\right|\right] + \int_s^T \mathbb{E}\left[\left|U^\theta\left(r,x+\mathcal{W}^\theta_{r-s}\right)\right|\right] dr \\
&< \infty
\end{aligned} \tag{2.1.14}$$

Furthermore (ii) is a restatement of (4.0.7) with $\theta = 0$, $m = 1$, and $k = 1$. This completes the proof of Corollary 2.1.8.1. □

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:*

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

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

$$\mathbb{E}\left[\left|\mathbb{E}\left[\mathcal{X}_1\right] - \frac{1}{n}\left(\sum_{i=1}^n \mathcal{X}_i\right)\right|^p\right] = \mathbb{E}\left[\left|\frac{1}{n}\left(\sum_{i=1}^n \left(\mathbb{E}\left[\mathcal{X}_1\right] - \mathcal{X}_i\right)\right)\right|^p\right] = \frac{1}{n^p}\mathbb{E}\left[\left|\sum_{i=1}^n \left(\mathbb{E}\left[\mathcal{X}_i\right] - \mathcal{X}_i\right)\right|^p\right] \tag{2.2.2}$$

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:

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

This completes the proof of the lemma. □

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 $\mathbb{E}\left[|\mathcal{X}_1|\right] < \infty$. Then it holds that:*

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

Proof. Observe that e.g. (Grohs et al., 2018, Lemma 2.3) and Lemma 2.3.1 establish (2.2.5). □

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:*

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

Proof. This a direct consequence of Definition 2.1.5 and Corollary 2.2.1.1. □

2.3 Bounds and Covnvergence

Lemma 2.3.1. *Assume Setting 4.0.1. Then it holds for all $t \in [0,T]$, $x \in \mathbb{R}^d$*

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

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:

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

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 (1.16), (1.17), and item (ii) of Corollary 2.1.8.1 yields that:

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|U^0(t,x) - \mathbb{E}\left[U^0(t,x)\right]\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \\
&= \left(\mathbb{E}\left[\left|\frac{1}{m}\left[\sum_{k=1}^m G_k(t,x)\right] - \mathbb{E}\left[G_1(t,x)\right]\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} && (2.3.4)\\
&\leqslant \frac{\mathfrak{m}}{m^{\frac{1}{2}}}\left(\mathbb{E}\left[|G_1(t,x)|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} && (2.3.5)\\
&= \frac{\mathfrak{m}}{m^{\frac{1}{2}}}\left[\left(\mathbb{E}\left[\left|g\left(x+\mathcal{W}^1_{T-t}\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}\right] && (2.3.6)
\end{aligned}$$

This and the fact that $\mathcal{W}^0$ has independent increments ensure that for all $n \in$, $t \in [0,T]$, $x \in \mathbb{R}^d$ 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. □

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

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

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

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

This and (4.0.6) ensure that:

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

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 (4.0.6) 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:

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

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. □

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

$$\left(\mathbb{E}\left[\left|U^0\left(t,x+\mathcal{W}_t^0\right)-u\left(t,x+\mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}\leqslant\mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s\in[0,T]}\mathbb{E}\left[\left(1+\left\|x+\mathcal{W}_s^0\right\|_E^p\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.15}$$

Proof. Observe that Lemma 2.3.2 ensures that:

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

Observe next that (4.0.6) ensures that:

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

Which in turn yields that:

$$\mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\mathbb{E}\left[\left(1+\left\|x+\mathcal{W}_T^0\right\|_E^p\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}\leqslant\mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s\in[0,T]}\mathbb{E}\left[\left(1+\left\|x+\mathcal{W}_s^0\right\|_E^p\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.18}$$

Combining 2.3.16, 2.3.17, and 2.3.18 yields that:

$$\left(\mathbb{E}\left[\left|U^0\left(t,x+\mathcal{W}_t^0\right)-u\left(t,x+\mathcal{W}_t^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}\leqslant\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\mathbb{E}\left[\left|g\left(x+\mathcal{W}_T^0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}\leqslant$$
$$\mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s\in[0,T]}\mathbb{E}\left[\left(1+\left\|x+\mathcal{W}_s^0\right\|_E^p\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \tag{2.3.19}$$

This completes the proof of Lemma 2.3.3. □

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:*

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

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:

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

And it is also the case that:

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

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

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|U^0\left(t+\mathfrak{t},x+\mathcal{W}^0_t\right)-u\left(t+\mathfrak{t},x+\mathcal{W}^0_t\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \\
&\leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s\in[0,T-\mathfrak{t}]}\mathbb{E}\left[\left(1+\left\|x+\mathcal{W}^0_s\right\|^p_E\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \\
&\leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s\in[0,T]}\mathbb{E}\left[\left(1+\left\|x+\mathcal{W}^0_s\right\|^p_E\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}
\end{aligned} \tag{2.3.22}$$

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

$$\begin{aligned}
\left(\mathbb{E}\left[\left|U^0\left(\mathfrak{t},x\right)-u\left(\mathfrak{t},x\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} &= \left(\mathbb{E}\left[\left|U^0\left(\mathfrak{t},x+\mathcal{W}^0_0\right)-u\left(\mathfrak{t},x+\mathcal{W}^0_0\right)\right|^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \\
&\leqslant \mathfrak{L}\left(\frac{\mathfrak{m}}{m^{\frac{1}{2}}}\right)\left(\sup_{s\in[0,T]}\mathbb{E}\left[\left(1+\left\|x+\mathcal{W}^0_s\right\|^p_E\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}
\end{aligned} \tag{2.3.23}$$

This completes the proof of Corollary 2.3.3.1. □

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* $\max\{|g_d(x)|\} \leqslant$

$Ld^p\left(1+\Sigma_{k=1}^d |x_k|\right)$, 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:

$$u_d\left(t,x\right)=\mathbb{E}\left[g_d\left(x+\mathcal{W}^{d,0}_{T-t}\right)\right] \tag{2.3.24}$$

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:

$$U^{d,\theta}_m(t,x)=\frac{1}{m}\left[\sum_{k=1}^m g_d\left(x+\mathcal{W}^{d,(\theta,0,-k)}_{T-t}\right)\right] \tag{2.3.25}$$

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 comput 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:

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

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 $\mathfrak{m}_{\mathfrak{p}}=\sqrt{\mathfrak{p}-1}$, $\mathfrak{p}\in[2,\infty)$, 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:

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

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

(I) it holds for all $d\in\mathbb{N}$ that $\{A\in\mathcal{F}:\mathbb{P}(A)=0\}\subseteq\mathbb{F}^d_0$

(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 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:

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

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

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

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:

$$\begin{aligned}
&\sup_{s\in[0,T]} \left(\mathbb{E}\left[\left(1+\left\|x+W_s^{d,0}\right\|_E^q\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \leqslant 1 + \sup_{s\in[0,T]} \left(\mathbb{E}\left[\left\|x+W_s^{d,0}\right\|_E^{q\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}} \\
&\leqslant 1 + \sup_{s\in[0,T]} \left(\max\{T,1\}\left(\left(1+\|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{q\mathfrak{p}(q\mathfrak{p}+3)T}{2}\right)\right)^{\frac{1}{\mathfrak{p}}} \\
&\leqslant 1 + \max\{T^{\frac{1}{\mathfrak{p}}},1\}\left(\left(1+\|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{q(q\mathfrak{p}+3)T}{2}\right) \\
&\leqslant 2\left(\left(1+\|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{q(q\mathfrak{p}+3)T}{2}+\frac{T}{\mathfrak{p}}\right) \\
&\leqslant 2\left(\left(1+\|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{\left[q(q\mathfrak{p}+3)+1\right]T}{2}\right)
\end{aligned} \tag{2.3.31}$$

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:

$$\begin{aligned}
&L\left(\frac{\mathfrak{m}_{\mathfrak{p}}}{m^{\frac{1}{2}}}\right)\left(\sup_{x\in[-L,L]^d}\sup_{s\in[0,T]}\left(\mathbb{E}\left[\left(1+\left\|x+W_s^{d,0}\right\|_E^q\right)^{\mathfrak{p}}\right]\right)^{\frac{1}{\mathfrak{p}}}\right) \\
&\leqslant L\left(\frac{\mathfrak{m}_{\mathfrak{p}}}{m^{\frac{1}{2}}}\right)\left(\sup_{x\in[-L,L]^d}\left[\left(\left(1+\|x\|_E^2\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{\left[q(q\mathfrak{p}+3)+1\right]T}{2}\right)\right]\right) \\
&\leqslant L\left(\frac{\mathfrak{m}_{\mathfrak{p}}}{m^{\frac{1}{2}}}\right)\left(\left(1+L^2d\right)^{\frac{q\mathfrak{p}}{2}} + (q\mathfrak{p}+1)d^{\frac{q\mathfrak{p}}{2}}\right)\exp\left(\frac{\left[q(q\mathfrak{p}+3)+1\right]T}{2}\right)
\end{aligned} \tag{2.3.32}$$

Combining this with Corollary 2.3.3.1 tells us that:

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

This and the fact that for all $d \in \mathbb{N}$ and $\epsilon \in (0,\infty)$ 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:

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

This 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 $U^{d,0}_{\mathfrak{N}_{d,\epsilon}}(T,0) : \Omega \to \mathbb{R}$. 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:

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

Note that other than d and ε everything on the right hand side is a constant.

□

Chapter 3

That u is a viscosity solution

We can extend the work for the heat equation to generic parabolic partial differential equations. We do this by first introducing viscosity solutions to Kolmogorov PDEs as given in Crandall & Lions Crandall et al. (1992) and further extended, esp. in 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 first 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 $\mathfrak{S} = \operatorname{supp}(\mu) \cup \operatorname{supp}(\sigma) \subseteq [0,T] \times \mathbb{R}^d$. 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) $[(\mathbb{P}(\mathcal{X}_0 \notin \mathfrak{S}) = 1) \implies (\mathbb{P}(\forall t \in [0,T] : \mathcal{X}_t = \mathcal{X}_0) = 1)]$

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

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:

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

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 $t \in [0,T]$, $\omega \in \Omega$ 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). □

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 $\mathfrak{S} = \operatorname{supp}(\mu)\cup\operatorname{supp}(\sigma)$. 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 $\mathcal{X}^{t,x} = (\mathcal{X}^{t,x}_s)_{s\in[t,T]} : [t,T]\times\Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_s)_{s\in[t,T]}$-adapted stochastic process with continuous sample paths satisfying for all $s \in [t,T]$ with $\mathbb{P}$-almost surety that:*

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

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

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

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)

$$\begin{aligned} 0 &= \left(\frac{\partial}{\partial t}u\right)(t,x) \\ &= \left(\frac{\partial}{\partial t}u\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^{*}\left(\operatorname{Hess}_x u\right)(t,x)\right) + \left\langle \mu(t,x), \left(\nabla_x u\right)(t,x)\right\rangle \end{aligned} \tag{3.1.11}$$

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)

$$\left(\frac{\partial}{\partial t} u\right)(t,x) + \frac{1}{2} \operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^* \left(\operatorname{Hess}_x u\right)(t,x)\right) + \left\langle \mu(t,x), \left(\nabla_x u\right)(t,x)\right\rangle = 0 \tag{3.1.13}$$

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

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:

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

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

and hence:

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

□

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. □

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:*

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

where $(t,x,u(t,x),\mu(x),\sigma(x)\left[\sigma(x)\right]^*) \in (0,T) \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d$, 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 $H(t,x,r,p,A) = -G(t,x,-r,-p,-A) \geqslant -G(t,x,-r,-p,-B) = H(t,x,r,p,B)$.

□

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 $\left(\frac{\partial}{\partial t} u\right)(t,x) + G\left(t,x,u(t,x),\left(\nabla_x u\right)(t,x), \left(\operatorname{Hess}_x u\right)(t,x)\right) \geqslant 0$ 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:

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

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 $\left(\frac{\partial}{\partial t}u\right)(t,x) + G\left(t,x,u(t,x),\left(\nabla_x u\right)(t,x),\left(\operatorname{Hess}_x u\right)(t,x)\right) \leqslant 0$ 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 $t \in (0,T)$, $x \in \mathcal{O}$, $\phi \in C^{1,2}\left((0,T) \times \mathcal{O}, \mathbb{R}\right)$ with $\phi(t,x) = u(t,x)$ and $\phi \leqslant u$ that:

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

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 $\left(\frac{\partial}{\partial t}u_d\right)(t,x) + G(t,x,u(t,x),\nabla_x(x,t),(\operatorname{Hess}_x u_d)(t,x))$ if and only if:*

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

(ii) u is a viscosity supersolution of $\left(\frac{\partial}{\partial t}u_d\right)(t,x) + G(t,x,u(t,x),\nabla_x(x,t),(\operatorname{Hess}_x u_d)(t,x)) = 0$ 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 $\mathfrak{r} \in \mathcal{O}$, $\phi \in C^{1,2}\left((0,T) \times \mathcal{O}, \mathbb{R}\right)$, 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 $\left(\frac{\partial}{\partial t}u_d\right)(t,x) + G\left(t,x,u(t,x),\left(\nabla_x u_D\right)(t,x),\left(\operatorname{Hess}_x u_d\right)(t,x)\right) \geqslant 0$ 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:*

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

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:

$$\begin{aligned}
&\left(\frac{\partial}{\partial t}\phi\right)(\mathfrak{t},\mathfrak{r}) + G\left(\mathfrak{t},\mathfrak{r},u(\mathfrak{t},\mathfrak{r}),(\nabla_x)(\mathfrak{t},\mathfrak{r}),(\operatorname{Hess}_x \phi)(\mathfrak{t},\mathfrak{r})\right) \\
&= \left(\frac{\partial}{\partial t}\psi\right)(\mathfrak{t},\mathfrak{r}) + G\left(\mathfrak{t},\mathfrak{r},u(\mathfrak{t},\mathfrak{r}),(\nabla_x)(\mathfrak{t},\mathfrak{r}),(\operatorname{Hess}_x \psi)(\mathfrak{t},\mathfrak{r})\right) \geqslant 0 \qquad (3.2.12)
\end{aligned}$$

□

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:*

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

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

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

then u_0 is a viscosity solution of:

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

Proof. Let $(t_o, x_o) \in (0,T) \times \mathcal{O}$. 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 $\phi_0(t_0,x_0) = u_0(t_0,x_0)$, $\phi_0(t_0,x_0) \geqslant u_0(t_0,x_0)$, 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 $\{(s,y) \in \mathbb{R}^d \times \mathbb{R} : \max\left(|s-t_0|^2, \|y-x_0\|_E^2\right) \leqslant \delta\}$. 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 $n \geqslant \nu_\varepsilon$, and

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

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

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:

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

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 $n \in \mathbb{N} \cap [\nu_\varepsilon, \infty)$:

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

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

$$\begin{aligned}
&\limsup_{n\to\infty} [u_n(\tau_n^\varepsilon,\rho_n^\varepsilon) - \phi_\epsilon(\tau_n^\varepsilon,\rho_n^\varepsilon)] \\
&\geqslant \limsup_{n\to\infty} [u_n(\tau_n^\varepsilon,\rho_n^\varepsilon) - (\phi_0(\tau_n^\varepsilon,\rho_n^\varepsilon) + \varepsilon\left(|\tau_n^\varepsilon - t_0| + \|\rho_n^\varepsilon - x_0\|_E\right))] \\
&\geqslant \limsup_{n\to\infty} [u_n(\tau_n^\varepsilon,\rho_n^\varepsilon) - u_0(\tau_n^\varepsilon,\rho_n^\varepsilon) - \varepsilon\left(|\tau_n^\varepsilon - t_0| + \|\rho_n^\varepsilon - x_0\|_E\right)] \\
&= \limsup_{n\to\infty} [-\varepsilon\left(|\tau_n^\varepsilon - t_0| + \|\rho_n^\varepsilon - x_0\|_E\right)] \leqslant 0
\end{aligned} \tag{3.2.22}$$

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:

$$\begin{aligned}
&G_0\left(t_0,x_0,\phi_0(t_0,x_0),(\nabla_x\phi_0)(t_0,x_0),(\mathrm{Hess}_x\,\phi_0)(t_0,x_0) + \mathrm{Id}_{\mathbb{R}^d}\right) \\
&= G_0\left(t_0,x_0,u_0(t_0,x_0),(\nabla_x\phi_\varepsilon)(t_0,x_0),(\mathrm{Hess}_x\,\phi_\varepsilon)(t_0,x_0)\right) \\
&\geqslant \limsup_{n\to\infty}\left[G_0\left(\tau_n^\varepsilon,\rho_n^\varepsilon,u_n(\tau_n^\varepsilon,\rho_n^\varepsilon),(\nabla_x\phi_\varepsilon)(\tau_n^\varepsilon,\rho_n^\varepsilon),(\mathrm{Hess}_x\,\phi_\varepsilon)(\tau_n^\varepsilon,\rho_n^\varepsilon)\right)\right] && (3.2.23)\\
&\geqslant \limsup_{n\to\infty}\left[G_n\left(\tau_n^\varepsilon,\rho_n^\varepsilon,u_n(\tau_n^\varepsilon,\rho_n^\varepsilon),(\nabla_x\phi_\varepsilon)(\tau_n^\varepsilon,\rho_n^\varepsilon),(\mathrm{Hess}_x\,\phi_\varepsilon)(\tau_n^\varepsilon,\rho_n^\varepsilon)\right)\right] && (3.2.24)
\end{aligned}$$

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

$$\left(\frac{\partial}{\partial t}\phi_0\right)(t_0,x_0) + G_0\left(t_0,x_0,\phi_0(t_0,x_0),(\nabla_x\phi_0)(t_0,x_0),(\mathrm{Hess}_x\,\phi_0)(t_0,x_0) + \varepsilon\mathrm{Id}_{\mathbb{R}^d}\right) \geqslant 0 \tag{3.2.25}$$

That G_0 is upper semi-continuous then yields that:

$$\left(\frac{\partial}{\partial t}\phi_0\right)(t_0,x_0) + G_0\left(t_0,x_0,\phi_0(t_0,x_0),(\nabla_x\phi_0)(t_0,x_0),(\mathrm{Hess}_x\,\phi_0)(t_0,x_0) + \varepsilon\mathrm{Id}_{\mathbb{R}^d}\right) \geqslant 0 \tag{3.2.26}$$

This establishes (3.2.29) which establishes the lemma.

□

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:*

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

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

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

then u_0 is a viscosity solution of:

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

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 degenenerate 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:

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

This together with (3.2.27) establish that:

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

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

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

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

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

This completes the proof. □

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:*

$$\limsup_{n \to \infty} \left[\sup_{(t,x,r,p,A) \in \mathcal{K}} \left(\left| G_n(t,x,r,p,A) - G_0(t,x,r,p,A) \right| + \left| u_n(t,x) - u_0(t,x) \right| \right) \right] = 0 \tag{3.2.34}$$

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

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

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

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

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

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

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:

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

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. $\square$

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:*

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

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.

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

then it is the case that:

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

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:

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

Note now that (3.2.40) tells us that:

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

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:

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

Itô's isometry applied to the second summand yields:

$$\begin{aligned}\left(\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E\right]\right)^{\frac{1}{2}} \leqslant \int_t^s \left(\mathbb{E}\left[\left\|\mu_n(r, \mathcal{X}_r^{n,t,x}) - \mu_0(r, \mathcal{X}_r^{0,t,x})\right\|_E^2\right]\right)^{\frac{1}{2}} dr + \\ \left(\int_t^s \mathbb{E}\left[\left\|\sigma_n(r, \mathcal{X}_r^{n,t,x}) - \sigma_0(r, \mathcal{X}^{0,t,x})\right\|_F^2\right] dr\right)^{\frac{1}{2}} \end{aligned} \quad (3.2.46)$$

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:

$$\begin{aligned}\mathbb{E}\left[\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{n,t,x}\|_E^2\right] &\leqslant 2\left[\int_t^s \left(\mathbb{E}\left[\left\|\mu_n(r, \mathcal{X}_r^{n,t,x}) - \mu_0(r, \mathcal{X}_r^{0,t,x})\right\|_E^2\right]\right)^{\frac{1}{2}} dr\right]^2 \\ &+ 2\int_t^s \mathbb{E}\left[\left\|\sigma_n(r, \mathcal{X}_r^{nt,x}) - \sigma_0(r, \mathcal{X}_r^{0,t,x})\right\|_F^2\right] dr \\ &\leqslant 2T\int_t^s \mathbb{E}\left[\left\|\mu_n(r, \mathcal{X}_r^{n,t,x}) - \mu_0(r, \mathcal{X}_r^{0,t,x})\right\|_E^2\right] dr \\ &+ 2\int_t^s \mathbb{E}\left[\left\|\sigma_n(r, \mathcal{X}_r^{n,t,x}) - \sigma_0(r, \mathcal{X}_r^{0,t,x})\right\|_F^2\right] dr \end{aligned} \quad (3.2.47)$$

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:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|^2\right] \\
&\leqslant 2T\int_t^s \left(2\mathbb{E}\left[\left\|\mu_n(r,\mathcal{X}_r^{n,t,x}) - \mu_0(r,\mathcal{X}_r^{n,t,x})\right\|_E^2\right] + 2\mathbb{E}\left[\left\|\mu_0(r,\mathcal{X}_r^{n,t,x}) - \mu_0(r,\mathcal{X}_r^{0,t,x})\right\|_E^2\right]\right) dr \\
&+ 2\int_t^2 \left(2\mathbb{E}\left[\left\|\sigma_n(r,\mathcal{X}_r^{n,t,x}) - \sigma_0(r,\mathcal{X}_r^{n,t,x})\right\|_F^2\right] + 2\mathbb{E}\left[\left\|\sigma_0(r,\mathcal{X}_r^{n,t,x}) - \sigma_0(r,\mathcal{X}_r^{0,t,x})\right\|_F\right]\right) dr \quad (3.2.48)
\end{aligned}$$

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:

$$\begin{aligned}
\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E^2\right] &\leqslant 4L^2(T+1)\int_t^s \mathbb{E}\left[\left\|\mathcal{X}_r^{n,t,x} - \mathcal{X}_r^{0,t,x}\right\|_E^2\right] dr \\
&+ 4T(T+1)\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_0(r,y)\|_F^2\right)\right]
\end{aligned}$$

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:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E^2\right] \\
&\leqslant 4T(T+1)\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_)(r,y)\|_F^2\right)\right] e^{4L^2T(T+1)} \quad (3.2.49)
\end{aligned}$$

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:

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{E}\left[\left\|\mathcal{X}_s^{n,t,x} - \mathcal{X}_s^{0,t,x}\right\|_E^2\right] \\
&\leqslant \limsup_{n\to\infty}\left[4T(T+1)\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_0(r,y)\|_F^2\right)\right] e^{4L^2T(T+1)}\right] \\
&\leqslant 4T(T+1)\left[\limsup_{n\to\infty}\left[\sup_{r\in[0,T]}\sup_{y\in\mathbb{R}^d}\left(\|\mu_n(r,y) - \mu_0(r,y)\|_E^2 + \|\sigma_n(r,y) - \sigma_0(r,y)\|_F^2\right)\right]\right] e^{4L^2T(T+1)} \\
&\leqslant 0
\end{aligned}$$

This completes the proof. □

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* $\operatorname{supp}(\mu) \cup \operatorname{supp}(\sigma) \subseteq [0,T] \times \mathcal{O}$ *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:

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

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 are 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 $\bigcup_{n\in\mathbb{N}}\left[\operatorname{supp}(\mathfrak{s}_n) \cup \operatorname{supp}(\mathfrak{m}_n)\right] \subseteq [0,T] \times (-\rho,\rho)^d$ 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:

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

and

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

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

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

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

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

and:

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

Note that (Beck et al., 2021b, Lemma 2.2) with $g \curvearrowleft \mathfrak{g}_k$, $\mu \curvearrowleft \mathfrak{m}_n$, $\sigma \curvearrowleft \mathfrak{s}_n$, $\mathcal{X}^{t,x} \curvearrowleft \mathcal{X}^{n,t,x}$ 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:

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

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

$$\left(\frac{\partial}{\partial t}\mathfrak{u}^n\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\mathfrak{s}_n(t,x)\left[\mathfrak{s}_n(t,x)\right]^* \left(\operatorname{Hess}_x \mathfrak{u}^n\right)(t,x)\right) + \langle \mathfrak{m}_n(t,x), \left(\nabla_x \mathfrak{u}^n\right)(t,x)\rangle = 0 \tag{3.2.61}$$

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 $\mathbb{P}(\forall s \in [t,T] : \mathfrak{X}^{n,t,x}_s = x = \mathcal{X}^{t,x}_s) =$

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:

$$\begin{aligned}\sup_{t\in[0,T]} \sup_{x\in\mathbb{R}^d} \left[\left|\mathfrak{u}^n(t,x) - \mathfrak{u}^0(t,x)\right|\right] &= \sup_{t\in[0,T]} \sup_{x\in(-\rho,\rho)^d} \left[\left|\mathfrak{u}^n(t,x) - \mathfrak{u}^0(t,x)\right|\right] \\ &\leqslant \sup_{t\in[0,T]} \sup_{x\in(-\rho,\rho)^d} \left(\mathbb{E}\left[\left|\mathfrak{g}_k\left(\mathfrak{X}_T^{n,t,x}\right) - \mathfrak{g}\left(\mathcal{X}_T^{t,x}\right)\right|\right]\right) \end{aligned} \tag{3.2.62}$$

Note that Lemma 3.2.13 allows us to conclude that:

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

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

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

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:

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

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:

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

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:

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

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

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

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. □

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 $O_r = \{x \in \mathcal{O} : \left(\|x\|_E \leqslant r \text{ and } \{y \in \mathbb{R}^d : \|y - x\|_E < \frac{1}{r}\} \subseteq \mathcal{C}\right.$ 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:*

$$\sup\left(\left\{\frac{\|\mu(t,x) - \mu(t,y)\|_E + \|\sigma(t,x) - \sigma(t,y)\|_F}{\|x-y\|_E} : t \in [0,T], x, y \in O_r, x \neq y\right\} \cup \{0\}\right) < \infty \tag{3.2.69}$$

$$\left(\frac{\partial}{\partial t}V\right)(t,x) + \frac{1}{2}\text{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*\left(\text{Hess}_x V\right)(t,x)\right) + \langle \mu(t,x), \left(\nabla_x V\right)(t,x)\rangle \leqslant 0 \tag{3.2.70}$$

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 $\mathcal{X}^{t,x} = (\mathcal{X}^{t,x}_s)_{s\in[t,T]} : [t,T] \times \Omega \to \mathcal{O}$ 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:

$$\mathcal{X}^{t,x}_s = x + \int_t^s \mu(r, \mathcal{X}^{t,x}_r)dr + \int_t^s \sigma(r, \mathcal{X}^{t,x}_n)dW_r \tag{3.2.71}$$

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:*

$$u(t,x) = \mathbb{E}\left[u(T, \mathcal{X}_T^{t,x})\right] \tag{3.2.72}$$

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

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

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:

$$\left(\frac{\partial}{\partial t}\mathfrak{u}^{n,k}\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\mathfrak{s}_n(t,x)\left[\mathfrak{s}_n(t,x)\right]^*\left(\operatorname{Hess}_x \mathfrak{u}^{n,k}\right)(t,x)\right) + \langle \mathfrak{m}_n(t,x), \left(\nabla_x(\mathfrak{u}^{n,k}\right)(t,x)\rangle = 0 \tag{3.2.81}$$

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):

$$\mathbb{P}\left(\forall s \in [t,T] : \mathbb{1}_{\{s \leqslant \mathfrak{t}_n^{t,x}\}}\mathfrak{X}_s^{n,t,x} = \mathbb{1}_{\{s \leqslant \mathfrak{t}_n^{t,x}\}}\mathcal{X}_s^{t,x}\right) = 1 \tag{3.2.82}$$

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

$$\begin{aligned}\mathbb{E}\left[\left|\mathfrak{g}_k\left(\mathfrak{X}_T^{n,t,x}\right) - \mathfrak{g}_k(\mathcal{X}_T^{t,x}\right)\right|\right] &= \mathbb{E}\left[\mathbb{1}_{\{\mathfrak{t}_n^{t,x} < T\}}\left|\mathfrak{g}_k(\mathfrak{X}_T^{n,t,x}) - \mathfrak{g}_k(\mathcal{X}_T^{t,x})\right|\right] \\ &\leqslant 2\left[\sup_{y \in \mathcal{O}} |\mathfrak{g}_k(y)|\right]\mathbb{P}\left(\mathfrak{t}_n^{t,x} < T\right)\end{aligned}$$

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:

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

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

$$\limsup_{k \to \infty}\left[\sup_{(t,x) \in \mathcal{K}}\left(\left|\mathfrak{u}^{n,k}(t,x) - \mathfrak{u}^{0,k}(t,x)\right|\right)\right] = 0 \tag{3.2.83}$$

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:

$$\limsup_{n \to \infty}\left[\sup_{(t,x) \in \mathcal{K}}\left(\|\mathfrak{m}_n(t,x) - \mu(t,x)\|_E + \|\mathfrak{s}_n(t,x) - \sigma(t,x)\|_F\right)\right] = 0 \tag{3.2.84}$$

Note that (3.2.81), (3.2.83) 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:

$$\left(\frac{\partial}{\partial t}\mathfrak{u}^{0,k}\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^*\left(\operatorname{Hess}_x \mathfrak{u}^{0,k}\right)(t,x)\right) + \langle \mu(t,x), \left(\nabla_x \mathfrak{u}^{0,k}\right)(t,x)\rangle = 0 \tag{3.2.85}$$

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:

$$\limsup_{k \to \infty}\left[\sup_{(t,x) \in \mathcal{K}}\left|\mathfrak{u}^{0,k}(t,x) - u(t,x)\right|\right] = 0 \tag{3.2.86}$$

This together with (3.2.85), (3.2.74), Corollary 3.2.11.2 shows that u_0 is a viscosity solution to:

$$\left(\frac{\partial}{\partial t}u\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma(t,x)\left[\sigma(t,x)\right]^{*}\left(\operatorname{Hess}_x u\right)(t,x)\right) + \left\langle \mu(t,x), \left(\nabla_x u\right)\right\rangle = 0 \tag{3.2.87}$$

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.

□

3.3 Solutions, characterization, and computational bounds to the Kolmogorov backward equations

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:*

$$\left(\frac{\partial}{\partial t}u_d\right)(t,x) + \frac{1}{2}\operatorname{Trace}\left(\sigma_d(x)\left[\sigma_d(x)\right]^{*}\left(\operatorname{Hess}_x u_d\right)(t,x)\right) + \left\langle \mu_d(x), \left(\nabla_x u_d\right)(t,x)\right\rangle = 0 \tag{3.3.1}$$

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:

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

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

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

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. □

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 $\mu_d(x) = \mathbb{0}_{d,1}$ where $\mathbb{0}_d$ is the zero vector in dimension d for $d \in \mathbb{N}$. □

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:*

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

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:

$$\begin{aligned}
\cancel{-\alpha e^{-t\alpha}}v + e^{-t\alpha}v_t + \frac{1}{2}\left[e^{-t\alpha}\nabla_x^2 v + 2\langle \nabla_x v, \nabla_x e^{-t\alpha}\rangle + v\nabla_x^2 e^{-t\alpha}\right] + \cancel{\alpha e^{-t\alpha}}v &= 0 \\
e^{-t\alpha}v_t + \frac{1}{2}\left[e^{-t\alpha}\nabla_x^2 v - 2te^{-t\alpha}\langle \nabla_x v, \nabla_x \alpha\rangle + v\nabla_x^2 e^{-t\alpha}\right] &= 0 \\
e^{-t\alpha}v_t + \frac{1}{2}\left[e^{-t\alpha}\nabla_x^2 v - 2te^{-t\alpha}\langle \nabla_x v, \nabla_x \alpha\rangle - tve^{-t\alpha}\nabla_x^2 \alpha\right] &= 0 \\
v_t + \frac{1}{2}\left[\nabla_x^2 v - 2t\langle \nabla_x v, \nabla_x \alpha\rangle - tv\nabla_x^2 \alpha\right] &= 0 \\
v_t + \frac{1}{2}\left[\nabla_x^2 v - 2t\langle \nabla_x \alpha, \nabla_x v\rangle - tv\nabla_x^2 \alpha\right] &= 0 \\
v_t + \frac{1}{2}\nabla_x^2 v + \langle -t\nabla_x \alpha, \nabla_x v\rangle - \frac{1}{2}tv\nabla_x^2 \alpha &= 0 \qquad (3.3.12)
\end{aligned}$$

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 $f(t,x,v) = -\frac{1}{2}tv\nabla_x^2 \alpha$ 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:

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

Note that we then have by the Cauchy-Schwarz inequality:

$$
\begin{aligned}
\langle x, \mu(t,x)\rangle \leqslant \|\langle x, -t\nabla_x \alpha\rangle\|_E &\leqslant \|x\|_E \|t\nabla_x \alpha\|_E \\
&\leqslant T\left(\|x\|_E \mathfrak{B}\right) \\
&\leqslant T\left(\mathfrak{B}+d\right)\|x\|_E \\
&= L\|x\|_E \leqslant L\left(1+\|x\|_E^2\right)
\end{aligned}
\tag{3.3.14}
$$

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

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:

$$
\begin{aligned}
\|\mu(t,x)-\mu(t,y)\|_E + \|\sigma(t,x)-\sigma(t,y)\|_F &= \|\mu(t,x)-\mu(t,y)\|_E + 0 \\
&= \|t\nabla_x\alpha(x) - t\nabla_x\alpha(y)\|_E \\
&\leqslant T\|\nabla_x\alpha(x) - \nabla_x\alpha(y)\|_E \\
&\leqslant 2T\mathfrak{B}
\end{aligned}
\tag{3.3.15}
$$

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

$$
\begin{aligned}
\left|\frac{\mathfrak{f}(x+h)-\mathfrak{f}(x)}{h}\right| &\leqslant \mathscr{C}\left(\|x\|_E+\|x+h\|_E\right) \\
\lim_{h\to 0}\left|\frac{\mathfrak{f}(x+h)-\mathfrak{f}(x)}{h}\right| &\leqslant \lim_{h\to 0}\mathscr{C}\left(\|x\|_E+\|x+h\|_E\right) \\
\left|\nabla_x \mathfrak{f}\,(x)\right| &\leqslant 2\mathscr{C}\|x\|_E = \mathscr{K}\|x\|_E
\end{aligned}
\tag{3.3.16}
$$

This suggests that $\nabla_x\mathfrak{f} \in O\left(x\right)$ and in particular that $\mathfrak{f} \in O\left(x^2\right)$. However with $\mathfrak{f} \curvearrowleft \mu$ 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 1.1.25.1. However since $O(c) \subseteq O(x) \subseteq O\left(x^2\right)$ by Corollary 1.1.25.2 it is also the case that $\mu \in O\left(x^2\right)$, and hence there exists a $\mathfrak{C}$ satisfying the claim. This proves the claim. ☐

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:

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

☐

Note that we may rewrite (3.3.12) as:

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

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 $f\left(t,x,u(t,x)\right) = -\frac{1}{2}tu\nabla_x^2\alpha$ 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:

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

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:

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

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

$$\begin{aligned}
v(t,x) &= \mathbb{E}\left[v\left(T, \mathcal{X}_T^{t,x}\right) + \int_t^T f\left(s, \mathcal{X}_s^{t,x}, v\left(s, \mathcal{X}_s^{t,x}\right)\right) ds\right] \qquad (3.3.20) \\
v(t,x) &= \mathbb{E}\left[v\left(T, \mathcal{X}_T^{t,x}\right) - \frac{1}{2}\int_t^T tv\left(s, \mathcal{X}_s^{t,x}\right) \nabla_x^2\alpha\left(\mathcal{X}_s^{t,x}\right) ds\right] \\
e^{t\alpha(x)}u(t,x) &= \mathbb{E}\left[\exp\left[T\alpha\left(\mathcal{X}_T^{t,x}\right)\right] u\left(T, \mathcal{X}_T^{t,x}\right) - \frac{1}{2}\int_t^T t\exp\left[t\alpha\left(\mathcal{X}_s^{t,x}\right)\right] u\left(t, \mathcal{X}_s^{t,x}\right) \nabla_x^2\alpha\left(\mathcal{X}_s^{t,x}\right) ds\right] \\
u(t,x) &= \mathbb{E}\left[\exp\left[T\alpha\left(\mathcal{X}_T^{t,x}\right) - t\alpha(x)\right] u\left(T, \mathcal{X}_T^{t,x}\right)\right] \\
&\quad - \mathbb{E}\left[\frac{1}{2e^{t\alpha(x)}} \int_t^T t\exp\left[t\alpha\left(\mathcal{X}_s^{t,x}\right)\right] u\left(t, \mathcal{X}_s^{t,x}\right) \nabla_x^2\alpha\left(\mathcal{X}_s^{t,x}\right) ds\right]
\end{aligned}$$

□

Chapter 4

Brownian motion Monte Carlo of the non-linear case

We now seek to apply the techniques introduced in Chapter 2 on **??**. To do so we need a variation of Setting 4.0.1. To that end we define such a setting. Assume v, f, α from Lemma 3.3.2.

Definition 4.0.1 (Subsequent Setting)**.** *Let* $g \in C\left(\mathbb{R}^d, \mathbb{R}\right)$ *be the function defined by:*

$$g(x) = v(T, x) \tag{4.0.1}$$

Let $F : C\left([0, T] \times \mathbb{R}^d, \mathbb{R}\right) \to C\left([, T] \times \mathbb{R}^d, \mathbb{R}\right)$ *be the functional defined as:*

$$(F(v))(t, x) = f(t, x, v(t, x)) \tag{4.0.2}$$

Note also that by Claim 3.3.5 it is the case that:

$$|f(t, x, w) - f(t, x, \mathfrak{w})| \leqslant L\,|w - \mathfrak{w}| \tag{4.0.3}$$

Note also that since $f(t, x, 0) = 0$*, and since by (Beck et al., 2021a, Corollary 3.9),* v *is growing at most polynomially, it is then the case that:*

$$\max\left\{|f(t, x, 0)|, |g(x)|\right\} \leqslant \mathfrak{L}\left(1 + \|x\|^p\right) \tag{4.0.4}$$

Substituting (4.0.1) and (4.0.2) into (3.3.20) renders (3.3.20) as:

$$
\begin{aligned}
v(t,x) &= \mathbb{E}\left[v\left(T, \mathcal{X}_T^{t,x}\right) + \int_t^T f\left(s, \mathcal{X}_s^{t,x}, v\left(s, \mathcal{X}_s^{t,x}\right)\right) ds\right] \\
v(t,x) &= \mathbb{E}\left[v\left(T, \mathcal{X}_T^{t,x}\right)\right] + \mathbb{E}\left[\int_t^T f\left(s, \mathcal{X}_s^{t,x}, v\left(s, \mathcal{X}_s^{t,x}\right)\right) ds\right] \\
v(t,x) &= \mathbb{E}\left[v\left(T, \mathcal{X}_T^{t,x}\right)\right] + \int_t^T \mathbb{E}\left[f\left(s, \mathcal{X}_s^{t,x}, v\left(s, \mathcal{X}_s^{t,x}\right)\right) ds\right] \\
v(t,x) &= \mathbb{E}\left[g\left(\mathcal{X}_T^{t,x}\right)\right] + \int_t^T \mathbb{E}\left[\left(F(v)\right)\left(s, \mathcal{X}_s^{t,x}\right)\right] ds
\end{aligned}
$$

Let $d, m \in \mathbb{N}$, $T, \mathfrak{L}, p \in [0,\infty)$, $\mathfrak{p} \in [2,\infty)$ $\mathfrak{m} = \mathfrak{k}_{\mathfrak{p}}\sqrt{\mathfrak{p}-1}$, $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$, $f \in C\left([0,T]\times\mathbb{R}^d\times\mathbb{R}\right)$, $g \in C(\mathbb{R}^d,\mathbb{R})$, let $F : C\left([0,T]\times\mathbb{R}^d,\mathbb{R}\right) \to C\left([0,T]\times\mathbb{R}^d,\mathbb{R}\right)$ assume for all $t\in[0,T], x\in\mathbb{R}^d$ that:

$$
|f(t,x,w) - f(t,x,\mathfrak{w})| \leqslant L\,|w-\mathfrak{w}| \qquad \max\left\{|f(t,x,0)|, |g(x)|\right\} \leqslant \mathfrak{L}\left(1+\|x\|_E^p\right) \tag{4.0.5}
$$

and let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $\mathfrak{u}^\theta : \Omega \to [0,1]$, $\theta\in\Theta$ be i.i.d. random variables, and suume for all $\theta\in\Theta$, $r\in(0,1)$ that $\mathbb{P}\left(\mathfrak{u}^\theta \leqslant r\right) = r$, let $\mathcal{U}^\theta : [0,T]\times\Omega \to [0,T]$, $\theta\in\Theta$ satisty for all $t\in[0,T]$, $\theta\in\Theta$ that $\mathcal{U}_t^\theta = t + (T-t)\,\mathfrak{u}^\theta$, 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 $\mathbb{E}\left[\left|g\left(x+\mathcal{W}_{T-t}^0\right)\right|\right] + \int_t^T \mathbb{E}\left[\left(F(u)\right)\left(s, x+\mathcal{W}_{s-t}^0\right)\right] < \infty$ and:

$$
u(t,x) = \mathbb{E}\left[g\left(x+\mathcal{W}_{T-t}^0\right)\right] + \int_t^T \mathbb{E}\left[\left(F(u)\right)\left(s, x+\mathcal{W}_{s-t}^0\right)\right] ds \tag{4.0.6}
$$

and let let $U^\theta : [0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}$, $\theta\in\Theta$, $n\in\mathbb{Z}$ satisfy for all $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$, $n\in\mathbb{N}_0$ that:

$$
\begin{aligned}
U_n^\theta(t,x) = \frac{\mathbb{1}_{\mathbb{N}}(n)}{m^n}&\left[\sum_{k=1}^{m^n} g\left(x+\mathcal{W}_{T-t}^{(\theta,0,-k)}\right)\right] \\
&+ \sum_{i=1}^{n-1} \frac{T-t}{m^{n-i}}\left[\sum_{k=1}^{m^{n-i}} \left(F\left(U_i^{(\theta,i,k)}\right)\right)\left(\mathcal{U}^{(\theta,i,k)}, x+\mathcal{W}^{(\theta,i,k)}_{\mathcal{U}_t^{(\theta,i,k)}}\right)\right]
\end{aligned} \tag{4.0.7}
$$

Part II

A Structural Description of Artificial Neural Networks

Chapter 5

Introduction and Basic Notions

We seek here to introduce a unified framework for artificial neural networks. This framework borrows from the work presented in Grohs et al. (2018), and work done by Joshua Padgett, Benno Kuckuk, and Arnulf Jentzen (unpublished). With this framework in place we wish to study ANNs from a 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 required to model PDEs and their growth (exponential or otherwise) as dimensions d increase.

5.1 The Basic Definition of ANNs

Definition 5.1.1 (Hadamard Product)**.** *Let $m, n \in \mathbb{N}$. Let $A, B \in \mathbb{R}^{m \times n}$. We define the Hadamard product $\odot : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ as:*

$$A \odot B \coloneqq [A \odot B]_{i,j} = [A]_{i,j} \times [B]_{i,j} \quad \forall i, j \tag{5.1.1}$$

Definition 5.1.2 (Rectifier Function)**.** *Let $d \in \mathbb{N}$ and $x \in \mathbb{R}^d$. We denote by $\mathfrak{r}_d : \mathbb{R}^d \to \mathbb{R}^d$ the function given by:*

$$\mathfrak{r}_d(x) = (\max\{x_1, 0\}, \max\{x_2, 0\}, ..., \max\{x_d, 0\}) \tag{5.1.2}$$

Definition 5.1.3 (Multidimensionalization function)**.** *Let $d \in \mathbb{N}$, and let $f \in C(\mathbb{R}, \mathbb{R})$. We denote* *by* $\text{Mult}_f^d : \mathbb{R}^d \to \mathbb{R}^d$ *the function which, for all* $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ *is given by:*

$$\text{Mult}_f^d(x) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_d) \end{bmatrix} \tag{5.1.3}$$

Definition 5.1.4 (Artificial Neural Networks)**.** *Denote by* NN *the set given by:*

$$\text{NN} = \bigcup_{L\in\mathbb{N}} \bigcup_{l_0,l_1,...,l_L\in\mathbb{N}} \left(\bigtimes_{k=1}^{L} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right) \tag{5.1.4}$$

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

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

$$\mathcal{P}(\nu) = \sum_{k=1}^{L} l_k \left(l_{k-1} + 1 \right) \tag{5.1.5}$$

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

$$\mathcal{D}(\nu) = L \tag{5.1.6}$$

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

$$\mathcal{I}(\nu) = l_0 \tag{5.1.7}$$

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

$$\mathcal{O}(\nu) = l_L \tag{5.1.8}$$

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

output), given by:

$$\mathcal{H}(\nu) = L - 1 \tag{5.1.9}$$

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

$$\mathcal{L}(\nu) = (l_0, l_1, l_2, ..., l_L) \tag{5.1.10}$$

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

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

$$\mathcal{W}_i(\nu) = \begin{cases} l_i & i \leqslant L \\ 0 & i > L \end{cases} \tag{5.1.11}$$

Note that this implies that 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 also 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 often find it convenient to denote the neural network as ν^{l_0,l_L}, where special emphasis needs to be paid to the size of the input and output layer. Note that it is evident from (5.1.11) that $\mathcal{W}_0 \left(\nu^{i,j} \right) = i$ and $\mathcal{W}_L \left(\nu^{i,j} \right) = j$ for a neural network of depth L and $i, j \in \mathbb{N}$.

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

Definition 5.1.5 (Activation Functions)**.** *We will denote by* $\mathfrak{a} \in C(\mathbb{R}^d, \mathbb{R}^d)$ *the column matrix of* *functions given by:*

$$\mathfrak{a}(x) = \begin{bmatrix} \sigma_1(x_1) \\ \sigma_2(x_2) \\ \vdots \\ \sigma_d(x_d) \end{bmatrix} \tag{5.1.12}$$

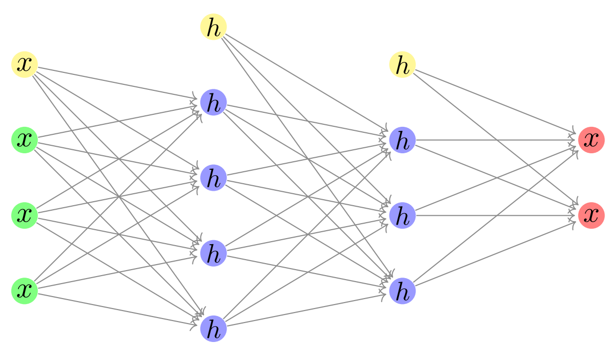
Where for each i, $\sigma_i \in C(\mathbb{R}, \mathbb{R})$. Each row represents a specific (not necessarily unique) activation function.

Definition 5.1.6 (Realizations of Artificial Neural Networks with Activation Functions)**.** *Let* $\mathsf{Act} \in C\left(\mathbb{R}^{L-1}, \mathbb{R}^{L-1}\right)$, *we denote by* $\mathfrak{R}_{\mathfrak{a}} : \mathrm{NN} \to \left(\bigcup_{k,l\in\mathbb{N}} C\left(\mathbb{R}^k, \mathbb{R}^l\right)\right)$ *the function satisfying for all* $L \in \mathbb{N}$, $l_0, l_1, ..., l_L \in \mathbb{N}$, $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L)) \in \left(\times_{k=1}^{L} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}\right]\right)$, $x_0 \in \mathbb{R}^{l_0}, x_1 \in \mathbb{R}^{l_1}, ..., x_{L-1} \in \mathbb{R}^{l_{L-1}}$ *with* $\forall k \in \{1, 2, ..., L\} : x_k = \mathrm{Mult}^{l_k}_{[\mathfrak{a}]_{k,1}}\left(W_k x_{k-1} + b_k\right)$ *such that:*

$$\mathfrak{R}_{\mathsf{Act}}(\nu) \in C\left(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}\right) \text{ and } \qquad \left(\mathfrak{R}_{\mathsf{Act}}(\nu)\right)(x_0) = W_L x_{L-1} + b_L \tag{5.1.13}$$

We will often denote the realized neural network ν^{l_0, l_L} taking $\mathbb{R}^{l_0}$ to $\mathbb{R}^{l_L}$ as $\nu^{l_0, l_L} : \mathbb{R}^{l_0} \rightarrowtail \mathbb{R}^{l_L}$ or simply as $\mathbb{R}^{l_0} \overset{\nu}{\rightarrowtail} \mathbb{R}^{l_L}$ where l_0 and l_L are obvious.

A neural network ν with $\mathcal{L}(\nu) = (4, 5, 4, 2)$



Lemma 5.1.7. *Let $\nu \in \mathrm{NN}$, it is then the case that:*

(i) $\mathcal{L}(\nu) \in \mathbb{N}^{\mathcal{D}(\nu)+1}$, *and*

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

Proof. By assumption:

$$\nu \in \text{NN} = \bigcup_{L \in \mathbb{N}} \bigcup_{(l_0, l_1, ..., l_L) \in \mathbb{N}^{L+1}} \left(\bigtimes_{k=1}^{L} \left[\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right] \right) \tag{5.1.14}$$

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

$$\nu \in \left(\bigtimes_{j=1}^{L} \left[\mathbb{R}^{l_j \times l_{j-1}} \times \mathbb{R}^{l_j} \right] \right) \tag{5.1.15}$$

This also ensures that $\mathcal{L}(\nu) = (l_0, l_1, ..., l_L) \in \mathbb{N}^{L+1} = \mathbb{N}^{\mathcal{D}(\nu)+1}$ and further that $\mathcal{I}(\nu) = l_0$, $\mathcal{O}(\nu) = l_L$, and that $\mathcal{D}(\nu) = L$. Together with (5.1.13) this proves the lemma. □

5.2 Composition and extensions of ANNs

The first operation we want to be able to do is to compose neural networks. This follows then naturally to the idea of neural network extensions.

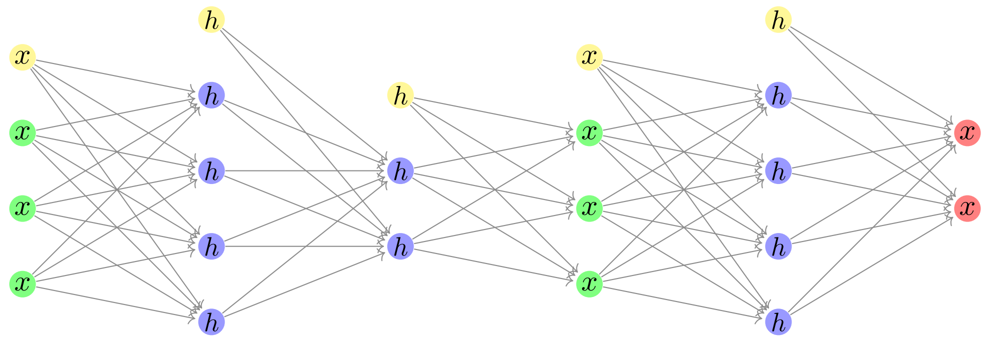
5.2.1 Composition

Definition 5.2.1 (Compositions of ANNs)**.** *We denote by* $(\cdot) \bullet (\cdot) : \{(\nu_1, \nu_2) \in \text{NN} \times \text{NN} : \mathcal{I}(\nu_1) = \mathcal{O}(\nu_1)\} \to \text{NN}$ *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* $\nu_2 = ((W'_1, b'_1), (W'_2, b'_2), ... (W'_M, b'_M)) \in \left(\bigtimes_{k=1}^{M} \left[\mathbb{R}^{m_k \times m_{k-1}} \times \mathbb{R}^{m_k} \right] \right)$ *with* $l_0 = \mathcal{I}(\nu_1) = \mathcal{O}(\nu_2) = m_M$ *and :*

$$\nu_1 \bullet \nu_2 = \begin{cases} ((W'_1, b'_1), (W'_2, b'_2), ...(W'_{M-1}, b'_{M-1}), (W_1 W'_M, W_1 b'_M + b_1), (W_2, b_2), & \\ ..., (W_L, b_L)) & : (L > 1) \wedge (M > 1) \\ ((W_1 W'_1, W_1 b'_1 + b_1), (W_2, b_2), (W_3, b_3), ..., (W_L b_L)) & : (L > 1) \wedge (M = 1) \\ ((W'_1, b'_1), (W'_2, b'_2), ..., (W'_{M-1}, b'_{M-1})(W_1, b'_M + b_1)) & : (L = 1) \wedge (M > 1) \\ ((W_1 W'_1, W_1 b'_1 + b_1)) & : (L = 1) \wedge (M = 1) \end{cases} \tag{5.2.1}$$

Diagrammatically this is represented as $\mathbb{R}^i \overset{\nu_2^{i,j}}{\rightarrowtail} \mathbb{R}^j \overset{\nu_1^{j,k}}{\rightarrowtail} \mathbb{R}^k$.

A neural network ν with $\mathcal{L}(\nu) = (4, 5, 4, 2)$



Lemma 5.2.2. *Let $\nu, \mu \in \mathrm{NN}$ be such that $\mathcal{O}(\mu) = \mathcal{I}(\nu)$. It is then the case that:*

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

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

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

Proof. This is a consequence of (5.2.1) which imply both (i) and (ii). □

Lemma 5.2.3. *Let $\nu_1, \nu_2, \nu_3 \in \mathrm{NN}$ satisfy that $\mathcal{I}(\nu_1) = \mathcal{O}(\nu_2)$ and $\mathcal{I}(\nu_2) = \mathcal{O}(\nu_3)$, it is then the case*

that:

$$(\nu_1 \bullet \nu_2) \bullet \nu_3 = \nu_1 \bullet (\nu_2 \bullet \nu_3) \tag{5.2.2}$$

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

Definition 5.2.4 (Powers of ANNs)**.** *We denote by $(\cdot)^{\bullet n} : \{\nu \in \mathrm{NN} : \mathcal{I}(\nu) = \mathcal{O}(\nu)\} \to \mathrm{NN}$, $n \in \mathbb{N}_0$,* *the function that satisfies for all $n \in \mathbb{N}_0$, $\nu \in \mathsf{NN}$, with $\mathcal{I}(\nu) = \mathcal{O}(\nu)$ that:*

$$\nu^{\bullet n} = \begin{cases} \left(\mathbb{I}_{\mathcal{O}(\nu)}, \mathbb{0}_{\mathcal{O}\,\nu,1}\right) \in \mathbb{R}^{\mathcal{O}(\nu)\times\mathcal{O}(\nu)} \times \mathbb{R}^{\mathcal{O}(\nu)} & : n = 0 \\ \nu \bullet \left(\nu^{\bullet(n-1)}\right) & : n \in \mathbb{N} \end{cases} \tag{5.2.3}$$

Diagrammatically this can be represented as $\overbrace{\mathbb{R}^i \overset{\nu^{i,i}}{\rightarrowtail} \mathbb{R}^i \overset{\nu^{i,i}}{\rightarrowtail} \ldots \overset{\nu^{i,i}}{\rightarrowtail} \mathbb{R}^i}^{n-times}$ *where $\mathcal{I}(\nu) = \mathcal{O}(\nu)$.*

5.2.2 Extensions

Often we need to be able to extend one neural network to be the same depth as another, hence the extension operation.

Definition 5.2.5 (Extensions of ANNs)**.** *Let $L \in \mathbb{N}$, $\mu \in \mathsf{NN}$ satisfy that $\mathcal{I}(\mu) = \mathcal{O}(\mu)$. We denote by $\mathfrak{E}_{L,\mu} : \{\mu \in \mathsf{NN} : (\mathcal{D}(\nu) \leqslant L \text{ and } \mathcal{O}(\nu) = \mathcal{I}(\mu))\} \to \mathsf{NN}$ the function satisfying for all $\nu \in \mathsf{NN}$ with $\mathcal{D}(\nu) \leqslant L$ and $\mathcal{O}(\nu) = \mathcal{I}(\mu)$ that:*

$$\mathfrak{E}_{L,\mu}(\nu) = \left(\mu^{\bullet(L-\mathcal{D}(\nu))}\right) \bullet \nu \tag{5.2.4}$$

Lemma 5.2.6. *Let $\mu, \nu \in \mathsf{NN}$ with $L \in \mathbb{N}$. It is then the case that:*

(i) $\mathcal{D}\left(\mathfrak{E}_{L,\mu(\nu)}\right) = \mathcal{D}\left(\nu\right) + \left[\left(L - \mathcal{D}\left(\nu\right)\right) \cdot \mathcal{D}\left(\mu\right) - \left(L - \mathcal{D}\left(\nu\right) - 1\right)\right] - 1$

5.3 Parallelization of ANNs

Definition 5.3.1 (Parallelization of ANNs of same length)**.** *Let $n \in \mathbb{N}$, we then denote by:*

$$\boxminus_{i=1}^{n} : \{(\nu_1, \nu_2, ..., \nu_n) \in \mathsf{NN}^n : \mathcal{D}(\nu_1) = \mathcal{D}(\nu_2) = ... = \mathcal{D}(\nu_n)\} \to \mathsf{NN} \tag{5.3.1}$$

the function satisfying for all $L \in \mathbb{N}$, $\nu_1, \nu_2, ..., \nu_n \in \mathrm{NN}$ and $L = \mathcal{D}(\nu_1) = \mathcal{D}(\nu_2) = ... = \mathcal{D}(\nu_n)$ that:

$$\boxminus_{i=1}^n \nu_i = \left(\left(\begin{bmatrix} \mathrm{Weight}_{1,\nu_1} & 0 & 0 & \cdots & 0 \\ 0 & \mathrm{Weight}_{1,\nu_2} & 0 & \cdots & 0 \\ 0 & 0 & \mathrm{Weight}_{1,\nu_3} & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \mathrm{Weight}_{1,\nu_n} \end{bmatrix}, \begin{bmatrix} \mathrm{Bias}_{1,\nu_1} \\ \mathrm{Bias}_{1,\nu_2} \\ \mathrm{Bias}_{1,\nu_3} \\ \vdots \\ \mathrm{Bias}_{1,\nu_n} \end{bmatrix} \right), \right.$$
$$\left(\begin{bmatrix} \mathrm{Weight}_{2,\nu_1} & 0 & 0 & \cdots & 0 \\ 0 & \mathrm{Weight}_{2,\nu_2} & 0 & \cdots & 0 \\ 0 & 0 & \mathrm{Weight}_{3,\nu_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathrm{Weight}_{2,\nu_n} \end{bmatrix}, \begin{bmatrix} \mathrm{Bias}_{2,\nu_1} \\ \mathrm{Bias}_{2,\nu_2} \\ \mathrm{Bias}_{2,\nu_3} \\ \vdots \\ \mathrm{Bias}_{2,\nu_n} \end{bmatrix} \right), ...,$$
$$\left. \left(\begin{bmatrix} \mathrm{Weight}_{L,\nu_1} & 0 & 0 & \cdots & 0 \\ 0 & \mathrm{Weight}_{L,\nu_2} & 0 & \cdots & 0 \\ 0 & 0 & \mathrm{Weight}_{L,\nu_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathrm{Weight}_{L,\nu_n} \end{bmatrix}, \begin{bmatrix} \mathrm{Bias}_{L,\nu_1} \\ \mathrm{Bias}_{L,\nu_2} \\ \mathrm{Bias}_{L,\nu_3} \\ \vdots \\ \mathrm{Bias}_{L,\nu_n} \end{bmatrix} \right) \right) \tag{5.3.2}$$

For the case where two neural networks ν_1, ν_2 are parallelized it is convenient to write $\nu_1 \boxminus \nu_2$. Diagrammatically this can be represented as: $\mathbb{R}^{\sum_{i=1}^n \mathcal{I}(\nu_i)} \overset{\boxminus_{i=1}^n \nu_i}{\rightarrowtail} \mathbb{R}^{\sum_{i=1}^n \mathcal{O}(\nu_i)}$. Or alternatively as $\mathbb{R}^{\sum_{i=1}^n \mathcal{I}(\nu_i)} \overset{\nu^{\sum_{i=1}^n \mathcal{I}(\nu_i), \sum_{i=1}^n \mathcal{O}(\nu_i)}}{\rightarrowtail} \mathbb{R}^{\sum_{i=1}^n \mathcal{O}(\nu_i)}$.

Remark 5.3.2. *Given $n, L \in \mathbb{N}$, $\nu_1, \nu_2, ..., \nu_n \in \mathrm{NN}$ such that $L = \mathcal{D}(\nu_1) = \mathcal{D}(\nu_2) = ... = \mathcal{D}(\nu_n)$ it is then the case, as seen from (5.3.2) that:*

$$\boxminus_{i=1}^n \nu_i \in \left(\bigtimes_{k=1}^{L} \left[\mathbb{R}^{\left(\sum_{j=1}^n \mathcal{W}_k(\nu_j)\right) \times \left(\sum_{j=1}^n \mathcal{W}_{k-1}(\nu_j)\right)} \times \mathbb{R}^{\left(\sum_{j=1}^n \mathcal{W}_k(\nu_j)\right)} \right] \right) \tag{5.3.3}$$

As a consequence:

$$\mathcal{P}(\boxminus_{i=1}^n \nu_i) = \sum_{i=1}^n \mathcal{P}\left(\nu_i\right) = \sum_{k=1}^L \left[\sum_{j=1}^n \mathcal{W}_k\left(\nu_j\right) \times \sum_{j=1}^n \mathcal{W}_{k-1}\left(\nu_j\right) + \sum_{j=1}^n \mathcal{W}_k\left(\nu_j\right) \right] \tag{5.3.4}$$

Lemma 5.3.3. *Given two neural networks $\nu_1, \nu_2 \in \text{NN}$. It is the case that $\mathfrak{R}_{\mathfrak{a}}(\nu_1 \boxminus \nu_2) = \mathfrak{R}_{\mathfrak{a}}(\nu_2 \boxminus \nu_1)$.*

Proof. Note that this is a consequence of the commutativity of summation in the exponents of (5.3.3). □

Lemma 5.3.4. *Let $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, and $\nu = \boxminus_{i=1}^{n} \nu_i$ satisfy the condition that $\mathcal{D}(\nu_1) = \mathcal{D}(\nu_2) = ... = \mathcal{D}(\nu_n)$. It is then the case that $\mathfrak{R}_{\mathfrak{a}}(\nu) \in C\left(\mathbb{R}^{\sum_{i=1}^{n} \mathcal{I}(\nu_i)}, \mathbb{R}^{\sum_{i=1}^{n} \mathcal{O}(\nu_i)}\right)$*

Proof. Let $L = \mathcal{D}(\nu_1)$, and let $l_{i,0}, l_{i,1}...l_{i,L} \in \mathbb{N}$ satisfy for all $i \in \{1, 2, ..., n\}$ that $\mathcal{L}(\nu_i) = (l_{i,0}, l_{i,1}, ..., l_{i,L})$. Furthermore let $((W_{i,1}, b_{i,1}), (W_{i,2}, b_{i,2}), ..., (W_{i,L}, b_{i,L})) \in \left(\bigtimes_{j=1}^{L} \left[\mathbb{R}^{l_{i,j} \times l_{i,j-1}} \times \mathbb{R}^{l_{i,j}}\right]\right)$ satisfy for all $i \in \{1, 2, ..., n\}$ that:

$$\nu_i = ((W_{i,1}, b_{i,1}), (W_{i,2}, b_{i,2}), ..., (W_{i,L}, b_{i,L})) \tag{5.3.5}$$

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 $((A_1, b_1), (A_2, b_2), ..., (A_L, b_L)) \in \left(\bigtimes_{j=1}^{L} \left[\mathbb{R}^{\alpha_j \times \alpha_{j-1}} \times \mathbb{R}^{\alpha_j}\right]\right)$ satisfy that:

$$\boxminus_{i=1}^{n} \nu_i = ((A_1, b_1), (A_2, b_2), ..., (A_L, b_L)) \tag{5.3.6}$$

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:

$$x_{i,j} = \text{Mult}_{\mathfrak{a}}^{l_{i,j}}\left(W_{i,j} x_{i,j-1} + b_{i,j}\right) \tag{5.3.7}$$

Note that (5.3.6) demonstrates that $\mathcal{I}\left(\boxminus_{i=1}^{n} \nu_i\right) = \alpha_0$ and $\mathcal{O}\left(\boxminus_{i=1}^{n} \nu_i\right) = \alpha_L$. This and Item(ii) of Lemma 5.1.7, and the fact that for all $i \in \{1, 2, ..., n\}$it is the case that $\mathcal{I}(\nu_i) = l_{i,0}$ and $\mathcal{O}(\nu_i) = l_{i,L}$ ensures that:

$$\begin{aligned}
\mathfrak{R}_{\mathfrak{a}}\left(\boxminus_{i=1}^{n}\right) \in C\left(\mathbb{R}^{\alpha_0}, \mathbb{R}^{\alpha_L}\right) &= C\left(\mathbb{R}^{\sum_{i=1}^{n} l_{i,0}}, \mathbb{R}^{\sum_{i=1}^{n} l_{i,L}}\right) \\
&= C\left(\mathbb{R}^{\sum_{i=1}^{n} \mathcal{I}(\nu_i)}, \mathbb{R}^{\sum_{i=1}^{n} \mathcal{O}(\nu_i)}\right)
\end{aligned}$$

This proves the lemma. □

5.4 Affine Linear Transformations as ANNs

Definition 5.4.1. *Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. We denote by $\mathsf{Aff}_{W,b} \in (\mathbb{R}^{m \times n} \times \mathbb{R}^m) \subseteq \mathrm{NN}$ the neural network given by $\mathsf{Aff}_{W,b} = (W, b)$.*

Lemma 5.4.2. *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) $\mathcal{L}(\mathsf{Aff}_{W,b}) = (n, m) \in \mathbb{N}^2$.

(ii) for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *it is the case that* $\mathfrak{R}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}) \in C(\mathbb{R}^m, \mathbb{R}^m)$

(iii) for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ *we have* $(\mathfrak{R}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}))(x) = Wx + b$

Proof. Note that (i) is a consequence of Definition 5.1.4 and 5.4.1. Note next that $\mathsf{Aff}_{W,b} = (W, b) \in (\mathbb{R}^{m \times n} \times \mathbb{R}^m) \subseteq \mathrm{NN}$. Note that (5.1.13) then tells us that $\mathfrak{R}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}) = Wx + b$ which in turn proves (ii) and (iii) □

Remark 5.4.3. *Given $W \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^{m \times 1}$, it is the case that according to Definition (5.1.5) we have: $\mathcal{P}(\mathsf{Aff}_{W,b}) = m \times n + m$*

Lemma 5.4.4. *Let $\nu \in \mathrm{NN}$. It is then the case that:*

(i) For all $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathcal{O}(\nu)}$

$$\mathcal{L}(\mathsf{Aff}_{W,B} \bullet \nu) = \left(\mathcal{W}_0(\nu), \mathcal{W}_1(\nu), ..., \mathcal{W}_{\mathcal{D}(\nu)-1}(\nu), m\right) \in \mathbb{N}^{\mathcal{D}(\nu)+1} \tag{5.4.1}$$

(ii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathcal{O}(\nu)}$, $B \in \mathbb{R}^m$, *we have that* $\mathfrak{R}_{\mathfrak{a}}(\mathsf{Aff}_{W,B} \bullet \nu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu)}, \mathbb{R}^m\right)$.

(iii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $m \in \mathbb{N}$, $W \in \mathbb{R}^{m \times \mathcal{O}(\nu)}$, $B \in \mathbb{R}^m$, $x \in \mathbb{R}^{\mathcal{I}(\nu)}$ *that:*

$$\left(\mathfrak{R}\left(\mathsf{Aff}_{W,b} \bullet \nu\right)\right)(x) = W\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu\right)\right)(x) + b \tag{5.4.2}$$

(iv) For all $n \in \mathbb{N}$, $W \in \mathbb{R}^{\mathcal{I}(\nu) \times n}$, $b \in \mathbb{R}^{\mathcal{I}(\nu)}$ *that:*

$$\mathcal{L}(\nu \bullet \mathsf{Aff}_{W,b}) = \left(n, \mathcal{W}_1(\nu), \mathcal{W}_2(\nu), ..., \mathcal{W}_{\mathcal{D}(\nu)}(\nu)\right) \in \mathbb{N}^{\mathcal{D}(\nu)+1} \tag{5.4.3}$$

(v) *For all* $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $W \in \mathbb{R}^{\mathcal{I}(\nu) \times n}$, $b \in \mathbb{R}^{\mathcal{I}(\nu)}$ *that* $\mathfrak{R}_{\mathfrak{a}} (\nu \bullet \mathsf{Aff}_{W,b}) \in C\left(\mathbb{R}^n, \mathbb{R}^{\mathcal{O}(\nu)}\right)$ *and,*

(vi) *For all* $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $n \in \mathbb{N}$, $W \in \mathbb{R}^{\mathcal{I}(\nu) \times n}$, $b \in \mathbb{R}^{\mathcal{I}(\nu)}$, $x \in \mathbb{R}^n$ *that:*

$$\left(\mathfrak{R}_{\mathfrak{a}} \left(\nu \bullet \mathsf{Aff}_{W,b}\right)\right)(x) = \left(\mathfrak{R}_{\mathfrak{a}}(\nu)\right)(Wx + b) \tag{5.4.4}$$

Proof. From Lemma 5.4.2 we see that $\mathfrak{R}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}) \in C(\mathbb{R}^n, \mathbb{R}^m)$ given by $\mathfrak{R}_{\mathfrak{a}}(\mathsf{Aff}_{W,b}) = Wx + b$. This and (Grohs et al., 2023, Proposition 2.6) prove $(i) - (vi)$. □

5.5 Sums of ANNs

Definition 5.5.1 (The Cpy Network). *We define the neural network,* $\mathsf{Cpy}_{n,k} \in \mathrm{NN}$ *for* $n, k \in \mathbb{N}$ *as the neural network given by:*

$$\mathsf{Cpy}_{n,k} = \mathsf{Aff}_{[\mathbb{I}_k \, \mathbb{I}_k \, \cdots \, \mathbb{I}_k]^T, \mathbb{0}_{nk}} \tag{5.5.1}$$

Lemma 5.5.2. *Let* $n, k \in \mathbb{N}$ *and let* $\mathsf{Cpy}_{n,k} \in \mathrm{NN}$, *it is then the case for all* $n, k \in \mathbb{N}$ *that:*

(i) $\mathcal{D}\left(\mathsf{Cpy}_{n,k}\right) = 1$

(ii) $\mathcal{P}\left(\mathsf{Cpy}_{n,k}\right) = nk^2 + nk$

Proof. Note that (i) is a consequence of Definition 5.4.1 and (ii) follows from the structure of $\mathsf{Cpy}_{n,k}$. □

Definition 5.5.3 (The Sm Network). *We define the neural network* $\mathsf{Sum}_{n,k}$ *for* $n, k \in \mathbb{N}$ *as the neural network given by:*

$$\mathsf{Sum}_{n,k} = \mathsf{Aff}_{[\mathbb{I}_k \, \mathbb{I}_k \, \cdots \, \mathbb{I}_k], \mathbb{0}_k} \tag{5.5.2}$$

Lemma 5.5.4. *Let* $n, k \in \mathbb{N}$ *and* $\mathsf{Sum}_{n,k} \in \mathrm{NN}$, *it is then the case for all* $n, k \in \mathbb{N}$ *that:*

(i) $\mathcal{D}\left(\mathsf{Sum}_{n,k}\right) = 1$

(ii) $\mathcal{P}\left(\mathsf{Cpy}_{n,k}\right) = nk^2 + k$

Proof. (i) is a consequence of Definition 5.4.1 and (ii) follows from the structure of $\mathsf{Sum}_{n,k}$. ☐

Definition 5.5.5 (Sum of ANNs of the same depth and same end widths)**.** *Let $u, v \in \mathbb{Z}$ with $u \leqslant v$. Let $\nu_u, \nu_{u+1}, ..., \nu_v \in$ NN satisfy for all $i \in \mathbb{N} \cap [u, v]$ that $\mathcal{D}(\nu_i) = \mathcal{D}(\nu_u)$, $\mathcal{I}(\nu_i) = \mathcal{I}(\nu_u)$, and $\mathcal{O}(\nu_i) = \mathcal{O}(\nu_u)$. 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:*

$$\oplus_{i=u}^{v} \nu_i \coloneqq \left(\mathsf{Sum}_{v-u+1,\mathcal{O}(\nu_2)} \bullet [\boxminus_{i=u}^{v} \nu_i] \bullet \mathsf{Cpy}_{(v-u+1),\mathcal{I}(\nu_1)} \right) \tag{5.5.3}$$

Or more concisely $\mathbb{R}^{\sum_{i=u}^{v} \mathcal{I}(\nu_i)} \overset{\oplus_{i=u}^{v} \nu_i}{\rightarrowtail} \mathbb{R}^{\sum_{i=u}^{v} \mathcal{O}(\nu_i)}$

5.5.1 Neural Network Sum Properties

Lemma 5.5.6. *Let $\nu_1, \nu_2 \in$ NN satisfy that $\mathcal{D}(\nu_1) = \mathcal{D}(\nu_2) = L$, $\mathcal{I}(\nu_1) = \mathcal{I}(\nu_2)$, and $\mathcal{O}(\nu_1) = \mathcal{O}(\nu_2)$, and $\mathcal{L}(\nu_1) = (l_{1,1}, l_{1,2}, ...l_{1,L})$ and $\mathcal{L}(\nu_2) = (l_{2,1}, l_{2,2}, ..., l_{2,L})$ it is then the case that:*

$$\begin{aligned} \mathcal{P}(\nu_1 \oplus \nu_2) &= \mathcal{P}\left(\mathsf{Aff}_{[\mathbb{I}_{\mathcal{O}(\nu_2)}\, \mathbb{I}_{\mathcal{O}(\nu_2)}], \mathbb{0}_{\mathcal{O}(\nu_2)}} \bullet [\nu_1 \boxminus \nu_2] \bullet \mathsf{Aff}_{[\mathbb{I}_{\mathcal{I}(\nu_1)}\, \mathbb{I}_{\mathcal{I}(\nu_1)}]^T, \mathbb{0}_{2 \cdot \mathcal{I}(\nu_1)}} \right) \\ &\leqslant 2l_{2,L}^2 + l_{2,L}(1 + l_{1,L-1} + l_{2,L-1}) + \mathcal{P}(\nu_1) + \mathcal{P}(\nu_2) + 2l_{1,0}^2 + l_{1,0}(2 + l_{1,1} + l_{2,1}) \end{aligned} \tag{5.5.4}$$

Proof. Observe, that by Definition 5.3.1 and Remark 5.4.2 we get that:

$$\begin{aligned} \mathcal{P}(\nu_1 \boxminus \nu_2) &= \mathcal{P}(\nu_1) + \mathcal{P}(\nu_2) \\ &= \sum_{k=1}^{L} \left[\left(\mathcal{W}_k(\nu_1) + \mathcal{W}_k(\nu_2) \right) \left(\mathcal{W}_{k-1}(\nu_1) + \mathcal{W}_{k-1}(\nu_2) \right) + \left(\mathcal{W}_k(\nu_1) + \mathcal{W}_k(\nu_2) \right) \right] \end{aligned} \tag{5.5.5}$$

Note also that by Remark 5.4.3 we have that:

$$\mathcal{P}\left(\mathsf{Aff}_{[\mathbb{I}_{\mathcal{O}(\nu_2)}\, \mathbb{I}_{\mathcal{O}(\nu_2)}], \mathbb{0}_{\mathcal{O}(\nu_2)}} \right) = 2\left(\mathcal{O}(\nu_2) \right)^2 + \mathcal{O}(\nu_2) \tag{5.5.6}$$

and:

$$\mathcal{P}\left(\mathsf{Aff}_{[\mathbb{I}_{\mathcal{I}(\nu_1)}\, \mathbb{I}_{\mathcal{I}(\nu_1)}]^T, \mathbb{0}_{2 \cdot \mathcal{I}(\nu_1)}} \right) = 2\left(\mathcal{I}(\nu_1) \right)^2 + 2\mathcal{I}(\nu_1) \tag{5.5.7}$$

Finally note that (Grohs et al., 2023, Proposition 2.6, Item (iv)) tells us that given neural networks $\nu_1, \nu_2 \in \text{NN}$, with $\mathcal{I}(\nu_1) = \mathcal{O}(\nu_2)$, and $\mathcal{L}(\nu_k) = \left(l_{k,1}, l_{k,2}, ..., l_{k,\mathcal{D}(\phi_k)}\right)$ it is then the case that:

$$\mathcal{P}\left(\nu_1 \bullet \nu_2\right) \leqslant \mathcal{P}\left(\nu_1\right) + \mathcal{P}\left(\nu_2\right) + l_{1,1}l_{2,L-1} \tag{5.5.8}$$

Combining (5.5.13),(**??**),(5.5.7), and (5.5.8) gives us that:

$$\begin{aligned}
&\mathcal{P}\left(\left(\nu_1 \boxminus \nu_2\right) \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{I}(\nu_1)}\ \mathbb{I}_{\mathcal{I}(\nu_1)}\right]^T, \mathbb{0}_{2\cdot\mathcal{I}(\nu_1)}}\right) \\
&\leqslant \mathcal{P}(\nu_1 \boxminus \nu_2) + \mathcal{P}\left(\mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{I}(\nu_1)}\ \mathbb{I}_{\mathcal{I}(\nu_1)}\right]^T, \mathbb{0}_{2\cdot\mathcal{I}(\nu_1)}}\right) + \left(l_{1,1} + l_{2,1}\right) \cdot \mathcal{I}(\nu_1) \\
&= \mathcal{P}\left(\nu_1\right) + \mathcal{P}\left(\nu_2\right) + 2l_{1,0}^2 + 2l_{1,0} + \left(l_{1,1} + l_{2,1}\right) l_{1,0} \\
&= \mathcal{P}\left(\nu_1\right) + \mathcal{P}\left(\nu_2\right) + 2l_{1,0}^2 + l_{1,0}\left(2 + l_{1,1} + l_{2,1}\right)
\end{aligned} \tag{5.5.9}$$

And again that:

$$\begin{aligned}
&\mathcal{P}\left(\mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{O}(\nu_2)}\ \mathbb{I}_{\mathcal{O}(\nu_2)}\right], \mathbb{0}_{\mathcal{O}(\nu_2)}} \bullet \left[\nu_1 \boxminus \nu_2\right] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{I}(\nu_1)}\ \mathbb{I}_{\mathcal{I}(\nu_1)}\right]^T, \mathbb{0}_{2\cdot\mathcal{I}(\nu_1)}}\right) \\
&\leqslant \mathcal{P}\left(\mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{O}(\nu_2)}\ \mathbb{I}_{\mathcal{O}(\nu_2)}\right], \mathbb{0}_{\mathcal{O}(\nu_2)}}\right) + \mathcal{P}\left(\left(\nu_1 \boxminus \nu_2\right) \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{I}(\nu_1)}\ \mathbb{I}_{\mathcal{I}(\nu_1)}\right]^T, \mathbb{0}_{2\cdot\mathcal{I}(\nu_1)}}\right) \\
&+ l_{2,L} \cdot \left(l_{1,L-1} + l_{2,L-1}\right) \\
&\leqslant 2l_{2,L}^2 + l_{2,L} \\
&+ \mathcal{P}\left(\nu_1\right) + \mathcal{P}\left(\nu_2\right) + 2l_{1,0}^2 + l_{1,0}\left(2 + l_{1,1} + l_{2,1}\right) \\
&+ l_{2,L} \cdot \left(l_{1,L-1} + l_{2,L-1}\right) \\
&= 2l_{2,L}^2 + l_{2,L}\left(1 + l_{1,L-1} + l_{2,L-1}\right) + \mathcal{P}\left(\nu_1\right) + \mathcal{P}\left(\nu_2\right) + 2l_{1,0}^2 + l_{1,0}\left(2 + l_{1,1} + l_{2,1}\right)
\end{aligned}$$

This completes the lemma. □

Lemma 5.5.7. *Let $\nu_1, \nu_2 \in \text{NN}$ satisfy that $\mathcal{D}(\nu_1) = \mathcal{D}(\nu_2) = L$, $\mathcal{I}(\nu_1) = \mathcal{I}(\nu_2)$, and $\mathcal{O}(\nu_1) = \mathcal{O}(\nu_2)$, and $\mathcal{L}(\nu_1) = (l_{1,1}, l_{1,2}, ...l_{1,L})$ and $\mathcal{L}\left(\nu_2\right) = (l_{2,1}, l_{2,2}, ..., l_{2,L})$ it is then the case that:*

$$\mathcal{D}\left(\nu_1 \boxminus \nu_2\right) = L \tag{5.5.10}$$

Proof. Note that $\mathcal{D}\left(\mathsf{Cpy}_{n,k}\right) = 1 = \mathcal{D}\left(\mathsf{Sum}_{n,k}\right)$ for all $n, k \in \mathbb{N}$. Note also that $\mathcal{D}\left(\nu_1 \boxminus \nu_2\right) =$

$\mathcal{D}(\nu_1) = \mathcal{D}(\nu_2)$ and that for $\nu, \mu \in \mathsf{NN}$ it is the case that $\mathcal{D}(\nu \bullet \mu) = \mathcal{D}(\nu) + \mathcal{D}(\mu) - 1$. Thus:

$$\begin{aligned}\mathcal{D}(\nu_1 \oplus \nu_1) = (\nu_1 \oplus \nu_2) &= \mathcal{D}\left(\mathsf{Aff}_{[\mathbb{I}_{\mathcal{O}(\nu_2)}\, \mathbb{I}_{\mathcal{O}(\nu_2)}],\mathbb{0}_{\mathcal{O}(\nu_2)}} \bullet [\nu_1 \boxminus \nu_2] \bullet \mathsf{Aff}_{[\mathbb{I}_{\mathcal{I}(\nu_1)}\, \mathbb{I}_{\mathcal{I}(\nu_1)}]^T,\mathbb{0}_{2\cdot\mathcal{I}(\nu_1)}}\right) \\ &= L\end{aligned}$$

□

Lemma 5.5.8. *Let $\nu_1, \nu_2 \in \mathsf{NN}$, such that $\mathcal{D}(\nu_1) = \mathcal{D}(\nu_2) = L$, $\mathcal{I}(\nu_1) = \mathcal{I}(\nu_2) = l_0$, and $\mathcal{O}(\nu_1) = \mathcal{O}(\nu_2) = l_L$. It is then the case that $\mathfrak{R}(\nu_1 \oplus \nu_2) = \mathfrak{R}(\nu_2 \oplus \nu_1)$, i.e. the realized 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 5.3.1 then tells us that:

$$\begin{aligned}\nu_1 \boxminus \nu_2 = \Bigg(&\left(\begin{bmatrix} W_1 & 0 \\ 0 & W_1' \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix}\right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix}\right), ..., \\ &\left(\begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \begin{bmatrix} b_L \\ b_L' \end{bmatrix}\right)\Bigg)\end{aligned}$$

Note also that by Claims **??** and **??** and Definition 5.4.1 we know that:

$$\mathsf{Aff}_{[\mathbb{I}_{\mathcal{I}(\nu_2)}\, \mathbb{I}_{\mathcal{I}(\nu_2)}]^T,\mathbb{0}_{2\,\mathcal{I}(\nu_2),1}} = \left(\begin{bmatrix} \mathbb{I}_{\mathcal{I}(\nu_2)} \\ \mathbb{I}_{\mathcal{I}(\nu_2)} \end{bmatrix}, \mathbb{0}_{2\,\mathcal{I}(\nu_2),1}\right) \tag{5.5.11}$$

and:

$$\mathsf{Aff}_{[\mathbb{I}_{\mathcal{O}(\nu_1)}\, \mathbb{I}_{\mathcal{O}(\nu_1)}],\mathbb{0}_{2\,\mathcal{O}(\nu_1),1}} = \left(\begin{bmatrix} \mathbb{I}_{\mathcal{O}(\nu_1)} & \mathbb{I}_{\mathcal{O}(\nu_1)} \end{bmatrix}, \mathbb{0}_{2\,\mathcal{O}(\nu_1),1}\right) \tag{5.5.12}$$

Applying Definition 5.2.1, specifically the second case, (5.5.3) and (**??**) yields that:

$$
\begin{aligned}
&[\nu_1 \boxminus \nu_2] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{I}(\nu_2)}\ \mathbb{I}_{\mathcal{I}(\nu_2)}\right]^T, \mathbb{0}_{2\mathcal{I}(\nu_2),1}} \\
&= \left(\left(\begin{bmatrix} W_1 & 0 \\ 0 & W_1' \end{bmatrix} \begin{bmatrix} \mathbb{I}_{\mathcal{I}(\nu_1)} \\ \mathbb{I}_{\mathcal{I}(\nu_1)} \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix} \right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix} \right), ..., \left(\begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \begin{bmatrix} b_L \\ b_L' \end{bmatrix} \right) \right) \\
&= \left(\left(\begin{bmatrix} W_1 \\ W_1' \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix} \right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix} \right), ..., \left(\begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \begin{bmatrix} b_L \\ b_L' \end{bmatrix} \right) \right)
\end{aligned}
$$

Applying Claim **??** and especially the third case of Definition 5.2.1 to to the above then gives us:

$$
\begin{aligned}
&\mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{O}(\nu_1)}\ \mathbb{I}_{\mathcal{O}(\nu_1)}\right],0} \bullet [\nu_1 \boxminus \nu_2] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{I}(\nu_2)}\ \mathbb{I}_{\mathcal{I}(\nu_2)}\right]^T,0} \\
&= \left(\left(\begin{bmatrix} W_1 \\ W_1' \end{bmatrix}, \begin{bmatrix} B_1 \\ B_1' \end{bmatrix} \right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix} \begin{bmatrix} b_2 \\ b_2' \end{bmatrix} \right), ..., \left(\left[\mathbb{I}_{\mathcal{O}(\nu_2)}\ \mathbb{I}_{\mathcal{O}(\nu_2)} \right] \begin{bmatrix} W_L & 0 \\ 0 & W_L' \end{bmatrix}, \left[\mathbb{I}_{\mathcal{O}(\nu_2)}\ \mathbb{I}_{\mathcal{O}(\nu_2)} \right] \begin{bmatrix} b_L \\ b_L' \end{bmatrix} \right) \right) \\
&= \left(\left(\begin{bmatrix} W_1 \\ W_1' \end{bmatrix}, \begin{bmatrix} b_1 \\ b_1' \end{bmatrix} \right), \left(\begin{bmatrix} W_2 & 0 \\ 0 & W_2' \end{bmatrix}, \begin{bmatrix} b_2 \\ b_2' \end{bmatrix} \right), ..., \left(\begin{bmatrix} W_L & W_L' \end{bmatrix}, b_L + b_L' \right) \right) \qquad (5.5.13)
\end{aligned}
$$

Now note that:

$$
\begin{aligned}
\nu_2 \boxminus \nu_1 = \Bigg(&\left(\begin{bmatrix} W_1' & 0 \\ 0 & W_1 \end{bmatrix}, \begin{bmatrix} b_1' \\ b_1 \end{bmatrix} \right), \left(\begin{bmatrix} W_2' & 0 \\ 0 & W_2 \end{bmatrix}, \begin{bmatrix} b_2' \\ b_2 \end{bmatrix} \right), ..., \\
&\left(\begin{bmatrix} W_L' & 0 \\ 0 & W_L \end{bmatrix}, \begin{bmatrix} b_L' \\ b_L \end{bmatrix} \right) \Bigg)
\end{aligned}
$$

And thus:

$$
\begin{aligned}
&\mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{O}(\nu_2)}\ \mathbb{I}_{\mathcal{O}(\nu_2)}\right],0} \bullet [\nu_2 \boxminus \nu_1] \bullet \mathsf{Aff}_{\left[\mathbb{I}_{\mathcal{I}(\nu_1)}\ \mathbb{I}_{\mathcal{I}(\nu_1)}\right]^T,0} \\
&= \left(\left(\begin{bmatrix} W_1' \\ W_1 \end{bmatrix}, \begin{bmatrix} b_1' \\ b_1 \end{bmatrix} \right), \left(\begin{bmatrix} W_2' & 0 \\ 0 & W_2 \end{bmatrix}, \begin{bmatrix} b_2' \\ b_2 \end{bmatrix} \right), ..., \left(\begin{bmatrix} W_L' & W_L \end{bmatrix}, \begin{bmatrix} b_L' + b_L \end{bmatrix} \right) \right) \qquad (5.5.14)
\end{aligned}
$$

Let $x \in \mathbb{R}^{\mathcal{I}(\nu_1)}$, 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 realization of (5.5.13) is then given by:

$$\mathfrak{R}\left(\begin{bmatrix} W_L & W_L' \end{bmatrix} \begin{bmatrix} W_{L-1}(...(W_2\left(W_1 x + b_1\right) + b_2) + ...) + b_{L-1} \\ W_{L-1}'(...(W_2'\left(W_1' x + b_1'\right) + b_2') + ...) + b_{L-1}' \end{bmatrix} + b_L + b_L' \right) \tag{5.5.15}$$

The full realization of (5.5.14) is then given by:

$$\mathfrak{R}\left(\begin{bmatrix} W_L' & W_L \end{bmatrix} \begin{bmatrix} W_{L-1}'(...(W_2'\left(W_1' x + b_1'\right) + b_2') + ...) + b_{L-1}' \\ W_{L-1}(...(W_2\left(W_1 x + b_1\right) + b_2) + ...) + b_{L-1} \end{bmatrix} + b_L + b_L' \right) \tag{5.5.16}$$

Since (5.5.25) and (5.5.16) are the same this proves that $\nu_1 \oplus \nu_2 = \nu_2 \oplus \nu_1$. □

Note that this is a special case of (Grohs et al., 2022, Lemma 3.28).

Lemma 5.5.9. *Let $l_0, l_1, ..., l_L \in \mathbb{N}$. Let $\nu \in \mathsf{NN}$ with $\mathcal{L}(\nu) = (l_0, l_1, ..., l_L)$. There then exists a neural network $\mathsf{Zr}_{l_0,l_1,...,l_L} \in \mathsf{NN}$ such that $\mathfrak{R}(\nu \oplus \mathsf{Zr}_{l_0,l_1,...,l_L}) = \mathfrak{R}(\mathsf{Zr}_{l_0,l_1,...,l_L} \oplus \nu) = \nu$.*

Proof. Let $\nu = \left(\left(W_1, b_1\right), \left(W_2, b_2\right), ..., \left(W_L, b_L\right)\right)$, 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 $\mathsf{Zr}_{l_0,l_1,...,l_L}$ the neural network which for all $l_0, l_1, ..., l_L \in \mathbb{N}$ is given by:

$$\mathsf{Zr}_{l_0,l_1,...,l_L} = \left(\left(\mathbb{0}_{l_1,l_0}, \mathbb{0}_{l_1}\right), \left(\mathbb{0}_{l_2,l_1}, \mathbb{0}_{l_2}\right), ..., \left(\mathbb{0}_{l_L,l_{L-1}}, \mathbb{0}_{l_L}\right)\right) \tag{5.5.17}$$

Thus, by (5.5.25), we have that:

$$\begin{aligned} \mathfrak{R}(\mathsf{Zr}_{l_0,l_1,...,l_L} \oplus \nu) &= \begin{bmatrix} 0 & W_L \end{bmatrix} \begin{bmatrix} 0 \\ W_{L-1}(...(W_2\left(W_1 x + b_1\right) + b_2) + ...) + b_{L-1} \end{bmatrix} + b_L \\ &= W_L(W_{L-1}(...W_2\left(W_1 x + b_1\right) + b_2) + ...) + b_{L-1}) + b_L \end{aligned} \tag{5.5.18}$$

$$\mathfrak{R}(\nu \oplus \mathrm{Zr}_{l_0,l_1,...,l_L}) = \begin{bmatrix} W_L & 0 \end{bmatrix} \begin{bmatrix} W_{L-1}(...(W_2\left(W_1x+b_1\right)+b_2)+...)+b_{L-1} \\ 0 \end{bmatrix} + b_L$$

$$= W_L(W_{L-1}(...W_2\left(W_1x+b_1\right)+b_2)+...)+b_{L-1})+b_L \tag{5.5.19}$$

And finally:

$$\mathfrak{R}(\nu) = W_L(W_{L-1}(...W_2\left(W_1x+b_1\right)+b_2)+...)+b_{L-1})+b_L \tag{5.5.20}$$

This completes the proof. □

Lemma 5.5.10. *Given neural networks $\nu_1, \nu_2, \nu_3 \in \mathrm{NN}$ with fixed depth L, fixed starting width of l_0 and fixed finishing width of l_L, it is then the case that $\mathfrak{R}\left((\nu_1 \oplus \nu_2) \oplus \nu_3\right) = \mathfrak{R}\left(\nu_1 \oplus (\nu_2 \oplus \nu_3)\right)$, i.e. the realization with a continuous activation function of $\oplus$ is associative.*

Proof. Let $\nu_1 = \left(\left(W_1^1, b_1^1\right), \left(W_2^1, b_2^1\right), ..., \left(W_L^1, b_L^1\right)\right)$, $\nu_2 = \left(\left(W_1^2, b_1^2\right), \left(W_2^2, b_2^2\right), ..., \left(W_L^2, b_L^2\right)\right)$, and $\nu_3 = \left(\left(W_1^3, b_1^3\right), \left(W_2^3, b_2^3\right), ..., \left(W_L^3, b_L^3\right)\right)$. Then (5.5.25) tells us that:

$$\mathfrak{R}(\nu_1 \oplus \nu_2) = \begin{bmatrix} W_L^1 & W_L^2 \end{bmatrix} \begin{bmatrix} W_{L-1}^1\left(...\left(W_2^1\left(W_1^1x+b_1^1\right)+b_2^1\right)+...\right)+b_{L-1}^1 \\ W_{L-1}^2\left(...\left(W_2^2\left(W_1^2x+b_1^2\right)+b_2^2\right)+...\right)+b_{L-1}^2 \end{bmatrix} + b_L^1 + b_L^2$$

And thus:

$$\mathfrak{R}\left((\nu_1 \oplus \nu_2) \oplus \nu_3\right)(x) =$$

$$\mathfrak{R}\left(\begin{bmatrix} \mathbb{I} & W_L^3 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} W_L^1 & W_L^2 \end{bmatrix} \begin{bmatrix} W_{L-1}^1\left(...\left(W_2^1\left(W_1^1x+b_1^1\right)+b_2^1\right)+...\right)+b_{L-1}^1 \\ W_{L-1}^2\left(...\left(W_2^2\left(W_1^2x+b_1^2\right)+b_2^2\right)+...\right)+b_{L-1}^2 \end{bmatrix} + b_L^1 + b_L^2 \\ W_{L-1}^3\left(...\left(W_2^3\left(W_1^3x+b_1^3\right)+b_2^3\right)+...\right)+b_{L-1}^3 \end{bmatrix} + b_L^3\right) \tag{5.5.21}$$

Similarly we have that:

$$\mathfrak{R}_{\mathfrak{a}}\left(\nu_1 \oplus \left(\nu_2 \oplus \nu_3\right)\right)(x) =$$
$$\mathfrak{R}\left(\begin{bmatrix} W_L^1 & \mathbb{I}\end{bmatrix}\begin{bmatrix} W_{L-1}^1\left(...\left(W_2^1\left(W_1^1 x + b_1^1\right) + b_2^1\right) + ...\right) + b_{L-1}^1 \\ \begin{bmatrix} W_L^2 & W_L^3\end{bmatrix}\begin{bmatrix} W_{L-1}^2\left(...\left(W_2^2\left(W_1^2 x + b_1^2\right) + b_2^2\right) + ...\right) + b_{L-1}^2 \\ W_{L-1}^3\left(...\left(W_2^3\left(W_1^3 x + b_1^3\right) + b_2^3\right) + ...\right) + b_{L-1}^3 \end{bmatrix} + b_L^2 + b_L^3 \end{bmatrix} + b_L^1\right) \tag{5.5.22}$$

Note that the associativity of matrix-vector multiplication, ensures that (5.5.21) and (5.5.22) are the same. □

Definition 5.5.11 (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 5.5.12. *For fixed depth, and layer widths the set of realized neural networks $\nu \in \mathrm{NN}$ form a commutative semi-group under the operation of $\oplus$.*

Proof. This is a consequence of Lemmas 5.5.8, 5.5.9, and 5.5.10. □

Lemma 5.5.13. *Let $\nu, \mu \in \mathrm{NN}$, with same length and end-widths. It is then the case that $\mathfrak{R}_{\mathfrak{a}}(\nu \oplus \mu) = \mathfrak{R}_{\mathfrak{a}}(\nu) + \mathfrak{R}_{\mathfrak{a}}(\mu)$.*

Proof. Let $\nu = \left(\left(W_1, b_1\right), \left(W_2, b_2\right), ..., \left(W_L, b_L\right)\right)$ and $\mu = \left(\left(W_1', b_1'\right), \left(W_2', b_2'\right), ..., \left(W_L', b_L'\right)\right)$. Note now that by (5.5.25) we have that:

$$\mathfrak{R}_{\mathfrak{a}}(\nu) = W_L\left(W_{L-1}(...(W_2\left(W_1 x + b_1\right) + b_2) + ...\right) + b_{L-1}) + b_L \tag{5.5.23}$$

and:

$$\mathfrak{R}_{\mathfrak{a}}(\mu) = W_L'\left(W_{L-1}'(...(W_2'\left(W_1' x + b_1'\right) + b_2') + ...) + b_{L-1}'\right) + b_L' \tag{5.5.24}$$

and in addition:

$$\mathfrak{R}_{\mathfrak{a}}(\nu \oplus \mu) = \begin{bmatrix} W_L & W'_L \end{bmatrix} \begin{bmatrix} W_{L-1}(...(W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1} \\ W'_{L-1}(...(W'_2(W'_1 x + b'_1) + b'_2) + ...) + b'_{L-1} \end{bmatrix} + b_L + b'_L \tag{5.5.25}$$

This proves the lemma. □

Definition 5.5.14 (Sum of ANNs of different lengths but same end widths)**.** *Let $u, v \in \mathbb{N}$ with $u \leqslant v$. Let $\nu_u, \nu_{u+1}, ..., \nu_v, \mu$ be neural networks such that it is the case for all $i \in \mathbb{N} \cap [u,v]$ that $\mathcal{I}(\nu_i) = \mathcal{I}(\nu_u)$, $\mathcal{O}(\nu_i) = \mathcal{I}(\mu) = \mathcal{O}(\mu)$ and $\mathcal{H}(\mu) = 1$. We then denote by $\boxplus^v_{i=u,\mu} \nu_i$, denoted $(\nu_u \,_u\boxplus_\mu\, \nu_{u+1} \,_u\boxplus_\mu \cdots \,_u\boxplus_\mu\, \nu_v$ the neural network given by:*

$$\boxplus^v_{i=u,\mu} \nu_i = \left[\oplus^v_{i=u} \mathfrak{E}_{\max_{j \in \{u,u+1,...,v\}} \mathcal{D}(\nu_j), \mu}(\nu_i) \right] \in \mathrm{NN} \tag{5.5.26}$$

5.6 Linear Combinations of ANNs

Definition 5.6.1 (Scalar left-multiplication with an ANN)**.** *Let $\lambda \in \mathbb{R}$. We will denote by $(\cdot) \circledast (\cdot) : \mathbb{R} \times \mathrm{NN} \to \mathrm{NN}$ the function that satisfy for all $\lambda \in \mathbb{R}$ and $\nu \in \mathrm{NN}$ that $\lambda \circledast \nu = \mathsf{Aff}_{\lambda \mathbb{I}_{\mathcal{O}(\nu)},0} \bullet \nu$. Diagrammatically this can be represented as:*

Definition 5.6.2 (Scalar right-multiplication with an ANN)**.** *Let $\lambda \in \mathbb{R}$. We will denote by $(\cdot) \circledast (\cdot) : \mathrm{NN} \times \mathbb{R} \to \mathrm{NN}$ the function satisfying for all $\nu \in \mathrm{NN}$ and $\lambda \in \mathbb{R}$ that $\nu \circledast \lambda = \nu \bullet \mathsf{Aff}_{\lambda \mathbb{I}_{\mathcal{I}(\nu)},0}$.*

Lemma 5.6.3. *Let $\lambda \in \mathbb{R}$ and $\nu \in \mathrm{NN}$. it is then the case that:*

(i) $\mathcal{L}(\lambda \circledast \nu) = \mathcal{L}(\nu)$

(ii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ that $\mathfrak{R}_{\mathfrak{a}}(\lambda \circledast \nu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu)}, \mathbb{R}^{\mathcal{O}(\nu)}\right)$

(iii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, and $x \in \mathbb{R}^{\mathcal{I}(\nu)}$ that:

$$\mathfrak{R}_{\mathfrak{a}}(\lambda \circledast \nu) = \lambda \mathfrak{R}_{\mathfrak{a}}(\nu) \tag{5.6.1}$$

Proof. Let $\nu \in \mathrm{NN}$ such that $\mathcal{L}(\nu) = (l_1, l_2, ..., l_L)$ and $\mathcal{D}(\nu) = L$ where $l_1, l_2, ..., l_L, L \in \mathbb{N}$. Then

Item (i) of Lemma 5.4.2 tells us that:

$$\mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{O}(\nu)},0}\right) = (\mathcal{O}(\nu), \mathcal{O}(\nu)) \tag{5.6.2}$$

This and Item (i) from Lemma 5.4.4 gives us that:

$$\mathcal{L}(\lambda \circledast \nu) = \mathcal{L}\left(\mathsf{Aff}_{\lambda\mathbb{I}_{\mathcal{O}(\nu)},0} \bullet \nu\right) = (l_0, l_1, ..., l_{L-1}, \mathcal{O}(\nu)) = \mathcal{L}(\nu) \tag{5.6.3}$$

Which proves (i). Item $(ii) - (iii)$ of Lemma 5.4.2 then prove that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\nu)}$, that $\mathfrak{R}_{\mathfrak{a}}(\lambda \circledast \nu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu), \mathcal{O}(\nu)}\right)$ given by:

$$\begin{aligned}(\mathfrak{R}_{\mathfrak{a}}(\lambda \circledast \nu))(x) &= \left(\mathfrak{R}_{\mathfrak{a}}\left(\mathsf{Aff}_{\lambda\mathbb{I}_{\mathcal{O}(\nu),0}} \bullet \nu\right)\right)(x) \\ &= \lambda\mathbb{I}_{\mathcal{O}(\nu)}((\mathfrak{R}_{\mathfrak{a}}(\nu))(x)) = \lambda((\mathfrak{R}_{\mathfrak{a}}(\nu))(x))\end{aligned} \tag{5.6.4}$$

This then establishes Items $(ii) - (iii)$, completing the proof. □

Lemma 5.6.4. *Let* $\lambda \in \mathbb{R}$ *and* $\nu \in \mathrm{NN}$. *It is then the case that:*

(i) $\mathcal{L}(\nu \circledast \lambda) = \mathcal{L}(\nu)$

(ii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathfrak{R}_{\mathfrak{a}}(\nu \circledast \lambda) \in C\left(\mathbb{R}^{\mathcal{I}(\nu)}, \mathbb{R}^{\mathcal{O}(\nu)}\right)$

(iii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, *and* $x \in \mathbb{R}^{\mathcal{I}(\nu)}$ *that:*

$$\mathfrak{R}_{\mathfrak{a}}(\nu \circledast \lambda) = \mathfrak{R}_{\mathfrak{a}}(\nu)(\lambda x) \tag{5.6.5}$$

Proof. Let $\nu \in \mathrm{NN}$ such that $\mathcal{L}(\nu) = (l_1, l_2, ..., l_L)$ and $\mathcal{D}(\nu) = L$ where $l_1, l_2, ..., l_L, L \in \mathbb{N}$. Then Item (i) of Lemma 5.4.2 tells us that:

$$\mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu)},0}\right) = (\mathcal{I}(\nu), \mathcal{I}(\nu)) \tag{5.6.6}$$

This and Item (iv) of Lemma 5.4.4 tells us that:

$$\mathcal{L}(\nu \circledast \lambda) = \mathcal{L}\left(\nu \bullet \mathsf{Aff}_{\lambda\mathbb{I}_{\mathcal{I}(\nu)}}\right) = (\mathcal{I}(\nu), l_1, l_2, ..., l_L) = \mathcal{L}(\nu) \tag{5.6.7}$$

Which proves (i). Item $(v)-(vi)$ of Lemma 5.4.4 then prove that for all $\mathfrak{a} \in C(\mathbb{R},\mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\nu)}$ that $\mathfrak{R}_{\mathfrak{a}}(\nu \circledast \lambda) \in C\left(\mathbb{R}^{\mathcal{I}(\nu),\mathcal{O}(\nu)}\right)$ given by:

$$\begin{aligned}\left(\mathfrak{R}_{\mathfrak{a}}(\nu \circledast \lambda)\right)(x) &= \left(\mathfrak{R}_{\mathfrak{a}}\left(\nu \bullet \mathsf{Aff}_{\lambda \mathbb{I}_{\mathcal{I}(\nu)},0}\right)\right)(x) \\ &= \left(\mathfrak{R}_{\mathfrak{a}}(\nu)\right)\left(\mathsf{Aff}_{\lambda \mathbb{I}_{\mathcal{I}(\nu)}}\right)(x) \\ &= \left(\mathfrak{R}_{\mathfrak{a}}(\nu)\right)(\lambda x) \end{aligned} \tag{5.6.8}$$

This completes the proof. □

Lemma 5.6.5. *Let $\nu, \mu \in \mathrm{NN}$ with the same length and the same end-widths, and $\lambda \in \mathbb{R}$. It is then the case, for all $\mathfrak{a} \in C(\mathbb{R},\mathbb{R})$ that:*

$$\mathfrak{R}_{\mathfrak{a}}(\lambda \circledast (\nu \oplus \mu))(x) = \mathfrak{R}_{\mathfrak{a}}((\lambda \circledast \nu) \oplus (\lambda \circledast \mu))(x) \tag{5.6.9}$$

$$= (\lambda\, \mathfrak{R}_{\mathfrak{a}}(\nu))(x) + (\lambda\, \mathfrak{R}_{\mathfrak{a}}(\mu))(x) \tag{5.6.10}$$

Proof. Let $\nu = ((W_1,B_1),(W_2,B_2),...,(W_L,B_L))$ and $\mu = ((W_1',B_1'),(W_2',B_2'),...,(W_L',B_L'))$. From Lemma 5.6.3 and (5.5.25) we have that:

$$\begin{aligned}&\mathfrak{R}_{\mathfrak{a}}(\lambda \circledast (\nu \oplus \mu))(x) = \lambda\, \mathfrak{R}_{\mathfrak{a}}(\nu \oplus \mu)(x) \\ &= \lambda \left(\begin{bmatrix} W_L & W_L' \end{bmatrix} \begin{bmatrix} W_{L-1}(...(W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1} \\ W_{L-1}'(...(W_2'(W_1' x + b_1') + b_2') + ...) + b_{L-1}' \end{bmatrix} + b_L + b_L' \right)\end{aligned}$$

Note that:

$$(\lambda\, \mathfrak{R}_{\mathfrak{a}}(\nu))(x) = \lambda \Big[W_L\left(W_{L-1}(...(W_2(W_1 x + b_1) + b_2) + ...) + b_{L-1}\right) + b_L \Big] \tag{5.6.11}$$

and that:

$$(\lambda\, \mathfrak{R}_{\mathfrak{a}}(\mu))(x) = \lambda \Big[W_L'\left(W_{L-1}'(...(W_2'\left(W_1' x + b_1'\right) + b_2') + ...) + b_{L-1}'\right) + b_L' \Big] \tag{5.6.12}$$

This combined with Lemma 5.5.13 completes the proof. □

Lemma 5.6.6. *Let $\nu, \mu \in \mathsf{NN}$ with the same length and the same end-widths, and $\lambda \in \mathbb{R}$. It is then the case, for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ that:*

$$\mathfrak{R}_{\mathfrak{a}}\left(\left(\nu \oplus \mu\right) \circledast \lambda\right)(x) = \mathfrak{R}_{\mathfrak{a}}\left(\left(\nu \circledast \lambda\right) \oplus \left(\mu \circledast \lambda\right)\right)(x) \tag{5.6.13}$$

$$= \left(\mathfrak{R}_{\mathfrak{a}}(\nu)\right)(\lambda x) + \left(\mathfrak{R}_{\mathfrak{a}}(\mu)\right)(\lambda x) \tag{5.6.14}$$

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 5.6.4 and (5.5.25) we have that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu \oplus \mu\right) \circledast \lambda\right)(x) &= \left(\mathfrak{R}_{\mathfrak{a}}\left(\nu \oplus \mu\right)\right)(\lambda x) \\
&= \begin{bmatrix} W_L & W_L' \end{bmatrix} \begin{bmatrix} W_{L-1}(...(W_2\left(W_1 \lambda x + b_1\right) + b_2) + ...) + b_{L-1} \\ W_{L-1}'(...(W_2'\left(W_1' \lambda x + b_1'\right) + b_2') + ...) + b_{L-1}' \end{bmatrix} + b_L + b_L'
\end{aligned}$$

Note that:

$$\left(\mathfrak{R}_{\mathfrak{a}}(\nu)\right)(\lambda x) = W_L\left(W_{L-1}(...(W_2\left(W_1 \lambda x + b_1\right) + b_2) + ...) + b_{L-1}\right) + b_L \tag{5.6.15}$$

and that:

$$\left(\mathfrak{R}_{\mathfrak{a}}(\mu)\right)(\lambda x) = W_L'\left(W_{L-1}'(...(W_2'\left(W_1' \lambda x + b_1'\right) + b_2') + ...) + b_{L-1}'\right) + b_L' \tag{5.6.16}$$

This together with Lemma 5.5.13 completes the proof. □

Lemma 5.6.7. *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 $\nu_u, \nu_{u+1}, ..., \nu_v, \mu \in \mathsf{NN}$, $B_u, B_{u+1}, ..., B_v \in \mathbb{R}^{\mathcal{I}(\mu)}$ satisfy that $\mathcal{L}(\nu_u) = \mathcal{L}(\nu_{u+1}) = ... = \mathcal{L}(\nu_v)$ and further that:*

$$\mu = \left[\oplus_{i=u}^{v}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_1)}, B_i}\right)\right)\right] \tag{5.6.17}$$

It then holds:

(i) That:

$$\mathcal{L}(\mu) = \left(\mathcal{I}(\nu_u), \sum_{i=u}^{v} \mathcal{W}_1\left(\nu_u\right), \sum_{i=u}^{v} \mathcal{W}_2\left(\nu_u\right), ..., \sum_{i=u}^{v} \mathcal{W}_{\mathcal{D}(\nu_u)-1}\left(\nu_u\right), \mathcal{O}(\nu_u)\right)$$
$$= \left(\mathcal{I}(\nu_u), n\,\mathcal{W}_1(\nu_u), n\,\mathcal{W}_2(\nu_u), ..., n\,\mathcal{W}_{\mathcal{D}(\nu_u-1)}, \mathcal{O}(\nu_u)\right)$$

(ii) that for all $\mathfrak{a} \in C\left(\mathbb{R}, \mathbb{R}\right)$*, that* $\mathfrak{R}_{\mathfrak{a}}(\mu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)$*, and*

(iii) for all $\mathfrak{a} \in C\left(\mathbb{R}, \mathbb{R}\right)$ *and* $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$ *that:*

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(\mu\right)\right)(x) = \sum_{i=u}^{v} c_i \left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(x + B_i\right) \tag{5.6.18}$$

Proof. Assume hypothesis that $\mathcal{L}(\nu_u) = \mathcal{L}(\nu_{u+1}) = ... = \mathcal{L}(\nu_v)$. Note that Item (i) of Lemma 5.4.2 gives us that for all $i \in \{u, u+1, ..., v\}$ that:

$$\mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i}\right) = \mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_u)}}\right) = \left(\mathcal{I}\left(\nu_u\right), \mathcal{I}\left(\nu_u\right)\right) \in \mathbb{N}^2 \tag{5.6.19}$$

This together with (Grohs et al., 2023, Proposition 2.6, Item (i)) assures us that for all $i \in \{u, u+1, ..., v\}$ it is the case that:

$$\mathcal{L}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i}\right) = \left(\mathcal{I}(\nu_u), \mathcal{W}_1\left(\nu_u\right), \mathcal{W}_2\left(\nu_u\right), ..., \mathcal{W}_{\mathcal{D}(\nu_u)}\left(\nu_u\right)\right) \tag{5.6.20}$$

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:

$$\mathcal{L}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i}\right)\right) = \mathcal{L}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i}\right) \tag{5.6.21}$$

This, (5.6.20), and (Grohs et al., 2022, Lemma 3.28, Item (ii)) then yield that:

$$\begin{aligned}
\mathcal{L}(\mu) &= \mathcal{L}\left(\oplus_{i=u}^{v}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},B_i}\right)\right)\right) \\
&= \left(\mathcal{I}(\nu_u), \sum_{i=u}^{v} \mathcal{W}_1\left(\nu_u\right), \sum_{i=u}^{v} \mathcal{W}_2\left(\nu_u\right), ..., \sum_{i=u}^{v} \mathcal{W}_{\mathcal{D}(\nu_u)-1}\left(\nu_u\right), \mathcal{O}\left(\nu_u\right)\right) \\
&= \left(\mathcal{I}(\nu_u), n\,\mathcal{W}_1(\nu_u), n\,\mathcal{W}_2(\nu_u), ..., n\,\mathcal{W}_{\mathcal{D}(\nu_u)-1}(\nu_u), \mathcal{O}(\nu_u)\right)
\end{aligned} \tag{5.6.22}$$

This establishes item (i). Items (v) and (vi) from Lemma 5.4.4 tells us that for all $i \in \{u, u{+}1, ..., v\}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$, it is the case that $\mathfrak{R}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},B_i}\right) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)$ and further that:

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)(x) = \left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)(x + b_i) \tag{5.6.23}$$

This along with (Grohs et al., 2022, Lemma 3.14) ensures that for all $i \in \{u, u+1, ..., v\}$, $\mathfrak{a} \in C\left(\mathbb{R}, \mathbb{R}\right)$, $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$, it is the case that:

$$\mathfrak{R}_{\mathfrak{a}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},B_i}\right)\right) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right) \tag{5.6.24}$$

and:

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)\right)(x) = c_i\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)(x + b_i) \tag{5.6.25}$$

Now observe that (Grohs et al., 2022, Lemma 3.28) and (5.6.21) ensure that for all $\mathfrak{a} \in C\left(\mathbb{R}, \mathbb{R}\right)$, $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$, it is the case that $\mathfrak{R}_{\mathfrak{a}}\left(\mu\right) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)$ and that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{a}}\left(\mu\right)\right)(x) &= \left(\mathfrak{R}_{\mathfrak{a}}\left(\oplus_{i=u}^{v}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)\right)\right)(x) \\
&= \sum_{i=u}^{v}\left(\mathfrak{R}_{\mathfrak{a}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)\right)(x) \\
&= \sum_{i=u}^{v} c_i\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)(x + b_i)
\end{aligned}$$

This establishes items (ii)–(iii) and thus the proof is complete. □

Lemma 5.6.8. *Let $u, v \in \mathbb{Z}$ with $u \leqslant v$. Let $\lambda_u, \lambda_{u+1}, ..., \lambda_v \in \mathbb{R}$. Let $\nu_u, \nu_{u+1}, ..., \nu_v, \mu \in \text{NN}$, $B_u, B_{u+1}, ..., B_v \in \mathbb{R}^{\mathcal{I}(\mu)}$ satisfy that $\mathcal{L}(\nu_u) = \mathcal{L}(\nu_{u+1}) = ... = \mathcal{L}(\nu_v)$ and further that:*

$$\mu = \left[\oplus_{i=u}^{v}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_1)}, b_i} \bullet \nu\right) \circledast c_i\right)\right] \tag{5.6.26}$$

It then holds:

(i) That:

$$\begin{aligned}\mathcal{L}(\mu) &= \left(\mathcal{I}(\nu_u), \sum_{i=u}^{v} \mathcal{W}_1(\nu_u), \sum_{i=u}^{v} \mathcal{W}_2(\nu_u), ..., \sum_{i=u}^{v} \mathcal{W}_{\mathcal{D}(\nu_u)-1}(\nu_u), \mathcal{O}(\nu_u)\right) \\ &= \left(\mathcal{I}(\nu_u), n\,\mathcal{W}_1(\nu_u), n\,\mathcal{W}_2(\nu_u), ..., n\,\mathcal{W}_{\mathcal{D}(\nu_u - 1)}, \mathcal{O}(\nu_u)\right)\end{aligned} \tag{5.6.27}$$

(ii) that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, that $\mathfrak{R}_{\mathfrak{a}}(\mu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)$, and

(iii) for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ and $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$ that:

$$\left(\mathfrak{R}_{\mathfrak{a}}(\mu)\right)(x) = \sum_{i=u}^{v} \left(\mathfrak{R}_{\mathfrak{a}}(\nu_i)\right)(c_i x + b_i) \tag{5.6.28}$$

Proof. Assume hypothesis that $\mathcal{L}(\nu_u) = \mathcal{L}(\nu_{u+1}) = ... = \mathcal{L}(\nu_v)$. Note that Item (i) of Lemma 5.4.2 gives us that for all $i \in \{u, u+1, ..., v\}$ that:

$$\mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i}\right) = \mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_u)}}\right) = \left(\mathcal{I}(\nu_u), \mathcal{I}(\nu_u)\right) \in \mathbb{N}^2 \tag{5.6.29}$$

Note then that (Grohs et al., 2023, Proposition 2.6, Item (ii)) tells us that for all $i \in \{u, u+1, ..., v\}$ it is the case that:

$$\mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i} \bullet \nu\right) = \left(\mathcal{I}(\nu_u), \mathcal{W}_1(\nu_u), \mathcal{W}_2(\nu_u), ..., \mathcal{W}_{\mathcal{D}(\nu_u)}(\nu_u)\right) \tag{5.6.30}$$

This and Item (i) of Lemma 5.6.4 tells us that for all $i \in \{u, u+1, ..., v\}$ it is the case that:

$$\mathcal{L}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, b_i} \bullet \nu\right) \circledast c_i\right) = \mathcal{L}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, b_i} \bullet \nu\right) \tag{5.6.31}$$

This, (5.6.30), and (Grohs et al., 2022, Lemma 3.28, Item (ii)) tell us that:

$$\begin{aligned}
\mathcal{L}(\mu) &= \mathcal{L}\left(\oplus_{i=u}^{v}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\bullet\nu_i\right)\circledast c_i\right)\right) \\
&= \left(\mathcal{I}(\nu_u),\sum_{i=u}^{v}\mathcal{W}_1\left(\nu_u\right),\sum_{i=u}^{v}\mathcal{W}_2\left(\nu_u\right),...,\sum_{i=u}^{v}\mathcal{W}_{\mathcal{D}(\nu_u)-1}\left(\nu_u\right),\mathcal{O}\left(\nu_u\right)\right) \\
&= \left(\mathcal{I}(\nu_u),n\,\mathcal{W}_1(\nu_u),n\,\mathcal{W}_2(\nu_u),...,n\,\mathcal{W}_{\mathcal{D}(\nu_u)-1}(\nu_u),\mathcal{O}(\nu_u)\right)
\end{aligned} \tag{5.6.32}$$

This establishes Item (i). Items (i) and (ii) from Lemma 5.4.4 tells us that for all $i\in\{u,u+1,...,v\}$, $\mathfrak{a}\in C(\mathbb{R},\mathbb{R})$, $x\in\mathbb{R}^{\mathcal{I}(\nu_u)}$, it is the case that $\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\bullet\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},B_i}\right)\in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)},\mathbb{R}^{\mathcal{O}(\nu_u)}\right)$ and further that:

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\bullet\nu_i\right)\right)(x)=\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)(x)+b_i \tag{5.6.33}$$

This along with Lemma 5.6.4 ensures that for all $i\in\{u,u+1,...,v\}$, $\mathfrak{a}\in C\left(\mathbb{R},\mathbb{R}\right)$, $x\in\mathbb{R}^{\mathcal{I}(\nu_u)}$, it is the case that:

$$\mathfrak{R}_{\mathfrak{a}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\bullet\nu_i\right)\circledast c_i\right)\in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)},\mathbb{R}^{\mathcal{O}(\nu_u)}\right) \tag{5.6.34}$$

and:

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\bullet\nu_i\right)\circledast c_i\right)\right)(x)=\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(c_ix+b_i\right) \tag{5.6.35}$$

Now observe that (Grohs et al., 2022, Lemma 3.28) and (5.5.8) ensure that for all $\mathfrak{a}\in C\left(\mathbb{R},\mathbb{R}\right)$, $x\in\mathbb{R}^{\mathcal{I}(\nu_u)}$, it is the case that $\mathfrak{R}_{\mathfrak{a}}\left(\mu\right)\in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)},\mathbb{R}^{\mathcal{O}(\nu_u)}\right)$ and that:

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(\mu\right)\right)(x)=\left(\mathfrak{R}_{\mathfrak{a}}\left(\oplus_{i=u}^{v}\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\bullet\nu_i\right)\right)\circledast c_i\right)(x) \tag{5.6.36}$$

$$=\sum_{i=u}^{v}\left(\mathfrak{R}_{\mathfrak{a}}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\bullet\nu_i\right)\circledast c_i\right)\right)(x) \tag{5.6.37}$$

$$=\sum_{i=u}^{v}\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)\left(c_ix+b_i\right)$$

This establishes items (ii)–(iii) and thus the proof is complete. □

Lemma 5.6.9. *Let $L \in \mathbb{N}$, $u, v \in \mathbb{Z}$ with $u \leqslant v$. Let $c_u, c_{u+1}, ..., c_v \in \mathbb{R}$. $\nu_u, \nu_{u+1}, ..., \nu_v, \mu, \mathfrak{I} \in \mathsf{NN}$, $B_u, B_{u+1}, ..., B_v \in \mathbb{R}^{\mathcal{I}(\nu_u)}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, satisfy for all $j \in \mathbb{N} \cap [u, v]$ that $L = \max_{i \in \mathbb{N} \cap [u,v]} \mathcal{D}(\nu_i)$, $\mathcal{I}(\nu_j) = \mathcal{I}(\nu_u)$, $\mathcal{O}(\nu_j) = \mathcal{I}(\mathfrak{I}) = \mathcal{O}(\mathfrak{I})$, $\mathcal{H}(\mathfrak{I}) = 1$, $\mathfrak{R}_{\mathfrak{a}}(\mathfrak{I}) = \mathbb{I}_{\mathbb{R}}$, and that:*

$$\mu = \boxplus_{i=u,\mathfrak{I}}^{v} \left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},,b_i}\right)\right) \tag{5.6.38}$$

We then have:

(i) it holds that:

$$\mathcal{L}(\mu) = \left(\mathcal{I}(\nu_u), \sum_{i=u}^{v} \mathcal{W}_1\left(\mathfrak{E}_{L,\mathfrak{I}}(\nu_i)\right), \sum_{i=u}^{v} \mathcal{W}_2\left(\mathfrak{E}_{L,\mathfrak{I}}(\nu_i)\right), ..., \sum_{i=u}^{v} \mathcal{W}_{L-1}\left(\mathfrak{E}_{I,\mathfrak{I}}(\nu_i), \mathcal{O}(\nu_u)\right)\right) \tag{5.6.39}$$

(ii) it holds that $\mathfrak{R}_{\mathfrak{a}}(\mu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)$, and that,

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

$$\left(\mathfrak{R}_{\mathfrak{a}}(\mu)\right)(x) = \sum_{i=u}^{v} c_i \left(\mathfrak{R}_{\mathfrak{a}}(\nu_i)\right)(x + b_i) \tag{5.6.40}$$

Proof. Note that Item(i) from Lemma 5.6.7 establish Item(i) and (5.5.26), in addition, items (v) and (vi) from Lemma 5.4.4 tell us that for all for all $i \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\nu_u}$, it holds that $\mathfrak{R}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},B_i} \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)\right)$ and further that:

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},B_i}\right)\right)(x) = \left(\mathfrak{R}_{\mathfrak{a}}(\nu_i)\right)(x + b_k) \tag{5.6.41}$$

This, Lemma 5.6.3 and (Grohs et al., 2023, Lemma 2.14, Item (ii)) show that for all $i \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$, it holds that:

$$\mathfrak{R}_{\mathfrak{a}}\left(\mathfrak{E}_{L,\mathfrak{I}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)\right) = \mathfrak{R}_{\mathfrak{a}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right) \tag{5.6.42}$$

and:

$$\left(\mathfrak{R}_{\mathfrak{a}}\left(\mathfrak{E}_{L,\mathfrak{I}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)\right)\right)(x) = \left(\mathfrak{R}_{\mathfrak{a}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)\right)(x)$$
$$= c_i\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu_i\right)\right)(x+b_i) \tag{5.6.43}$$

This combined with (Grohs et al., 2022, Lemma 3.28) and (5.6.21) demonstrate that for all $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$ it holds that $\mathfrak{R}_{\mathfrak{a}}(\mu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)$ and that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{a}}(\mu)\right)(x) &= \left(\mathfrak{R}_{\mathfrak{a}}\left(\boxplus_{i=u,\mathfrak{I}}^{v}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}}\right)\right)\right)\right)(x) \\
&= \left(\mathfrak{R}_{\mathfrak{a}}\left(\oplus_{i=u}^{v} \mathfrak{E}_{L,\mathfrak{I}}\left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i}\right)\right)\right)\right)(x) \\
&= \sum_{i=u}^{v} c_i\left(\mathfrak{R}_{\mathfrak{a}}(\nu_i)\right)(x+b_i)
\end{aligned} \tag{5.6.44}$$

This establishes Items(ii)–(iii) thus proving the lemma. □

Lemma 5.6.10. *Let $L \in \mathbb{N}$, $u, v \in \mathbb{Z}$ with $u \leqslant v$. Let $c_u, c_{u+1}, ..., c_v \in \mathbb{R}$. $\nu_u, \nu_{u+1}, ..., \nu_v, \mu, \mathfrak{I} \in \mathrm{NN}$, $B_u, B_{u+1}, ..., B_v \in \mathbb{R}^{\mathcal{I}(\nu_u)}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, satisfy for all $j \in \mathbb{N} \cap [u, v]$ that $L = \max_{i \in \mathbb{N} \cap [u,v]} \mathcal{D}(\nu_i)$, $\mathcal{I}(\nu_j) = \mathcal{I}(\nu_u)$, $\mathcal{O}(\nu_j) = \mathcal{I}(\mathfrak{I}) = \mathcal{O}(\mathfrak{I})$, $\mathcal{H}(\mathfrak{I}) = 1$, $\mathfrak{R}_{\mathfrak{a}}(\mathfrak{I}) = \mathbb{I}_{\mathbb{R}}$, and that:*

$$\mu = \boxplus_{i=u,\mathfrak{I}}^{v}\left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)},b_i} \bullet \nu_i\right) \circledast c_i\right) \tag{5.6.45}$$

We then have:

(i) it holds that:

$$\mathcal{L}(\mu) = \left(\mathcal{I}(\nu_u), \sum_{i=u}^{v} \mathcal{W}_1\left(\mathfrak{E}_{L,\mathfrak{I}}(\nu_i)\right), \sum_{i=u}^{v} \mathcal{W}_2\left(\mathfrak{E}_{L,\mathfrak{I}}(\nu_i)\right), ..., \sum_{i=u}^{v} \mathcal{W}_{L-1}\left(\mathfrak{E}_{L,\mathfrak{I}}(\nu_i), \mathcal{O}(\nu_u)\right)\right) \tag{5.6.46}$$

(ii) it holds that $\mathfrak{R}_{\mathfrak{a}}(\mu) \in C\left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)}\right)$, and that,

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

$$\left(\mathfrak{R}_{\mathfrak{a}}(\mu)\right)(x) = \sum_{i=u}^{v}\left(\mathfrak{R}_{\mathfrak{a}}(\nu_i)\right)(c_i x + b_i) \tag{5.6.47}$$

Proof. Note that Item(i) from Lemma 5.6.8 establish Item(i) and (5.5.26), in addition, items (ii) and (iii) from Lemma 5.4.4 tell us that for all for all $i \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\nu_u}$, it holds that $\mathfrak{R}_{\mathfrak{a}} \left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i} \bullet \nu_i \in C \left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)} \right) \right)$ and further that:

$$\left(\mathfrak{R}_{\mathfrak{a}} \left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, B_i} \bullet \nu_i \right) \right) (x) = \left(\mathfrak{R}_{\mathfrak{a}} \left(\nu_i \right) \right) (x) + b_k \tag{5.6.48}$$

This, Lemma 5.6.4 and (Grohs et al., 2023, Lemma 2.14, Item (ii)) show that for all $i \in \mathbb{N} \cap [u, v]$, $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$, it holds that:

$$\mathfrak{R}_{\mathfrak{a}} \left(\mathfrak{E}_{L, \mathfrak{I}} \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, b_i} \bullet \nu_i \right) \circledast c_i \right) \right) = \mathfrak{R}_{\mathfrak{a}} \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, b_i} \bullet \nu_i \right) \circledast c_i \right) \in C \left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)} \right) \tag{5.6.49}$$

and:

$$\begin{aligned} \left(\mathfrak{R}_{\mathfrak{a}} \left(\mathfrak{E}_{L, \mathfrak{I}} \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, b_i} \bullet \nu_i \right) \circledast c_i \right) \right) \right) (x) &= \left(\mathfrak{R}_{\mathfrak{a}} \left(c_i \circledast \left(\nu_i \bullet \mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, b_i} \right) \right) \right) (x) \\ &= \left(\mathfrak{R}_{\mathfrak{a}} \left(\nu_i \right) \right) \left(c_i x + b_i \right) \end{aligned} \tag{5.6.50}$$

This and (Grohs et al., 2022, Lemma 3.28) and (5.6.31) demonstrate that for all $x \in \mathbb{R}^{\mathcal{I}(\nu_u)}$ it holds that $\mathfrak{R}_{\mathfrak{a}} (\mu) \in C \left(\mathbb{R}^{\mathcal{I}(\nu_u)}, \mathbb{R}^{\mathcal{O}(\nu_u)} \right)$ and that:

$$\begin{aligned} \left(\mathfrak{R}_{\mathfrak{a}} (\mu) \right) (x) &= \left(\mathfrak{R}_{\mathfrak{a}} \left(\boxplus_{i=u, \mathfrak{I}}^{v} \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}} \bullet \nu_i \right) \circledast c_i \right) \right) \right) (x) \\ &= \left(\mathfrak{R}_{\mathfrak{a}} \left(\oplus_{i=u}^{v} \mathfrak{E}_{L, \mathfrak{I}} \left(\left(\mathsf{Aff}_{\mathbb{I}_{\mathcal{I}(\nu_i)}, b_i} \bullet \nu_i \right) \circledast c_i \right) \right) \right) (x) \\ &= \sum_{i=u}^{v} \left(\mathfrak{R}_{\mathfrak{a}} \left(\nu_i \right) \right) \left(c_i x + b_i \right) \end{aligned} \tag{5.6.51}$$

This completes the proof. □

Chapter 6

ANN Product Approximations

6.1 Approximation for simple products

Lemma 6.1.1. *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:*

$$A_k = \begin{bmatrix} 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ 2 & -4 & 2 & 0 \\ -c_k & 2c_k & -c_k & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ -\frac{1}{2} \\ -1 \\ 0 \end{bmatrix} \quad C_k = \begin{bmatrix} -c_k & 2c_k & -c_k & 1 \end{bmatrix} \tag{6.1.1}$$

and that:

$$c_k = 2^{1-2k} \tag{6.1.2}$$

It is then the case that

(i) There exists unique $\xi_k \in \text{NN}$, $k \in \mathbb{N}$ *which satisfies for all* $k \in [2,\infty) \cap \mathbb{N}$ *that* $\xi_1 = (\mathsf{Aff}_{C_1,0} \bullet \mathfrak{i}_4) \bullet \mathsf{Aff}_{\mathbb{e}_4,B}$*. Note that for all* $d \in \mathbb{N}$, $\mathfrak{i}_d = ((\mathbb{I}_d, \mathbb{0}_d), (\mathbb{I}_d, \mathbb{0}_d))$ *(explained in detail in Definition 9.1.1), and that:*

$$\xi_k = (\mathsf{Aff}_{C_k,0} \bullet \mathfrak{i}_4) \bullet \left(\mathsf{Aff}_{A_{k-1},B} \bullet \mathfrak{i}_4\right) \bullet \cdots \bullet (\mathsf{Aff}_{A_1,B} \bullet \mathfrak{i}_4) \bullet \mathsf{Aff}_{\mathbb{e}_4,B} \tag{6.1.3}$$

(ii) for all $k \in \mathbb{N}$ *we have* $\mathfrak{R}_{\mathfrak{r}} \in C(\mathbb{R},\mathbb{R})$

(iii) for all $k \in \mathbb{N}$ *we have* $\mathcal{L}(\xi_k) = (1, 4, 4, ..., 4, 1) \in \mathbb{N}^{k+2}$

(iv) for all $k \in \mathbb{N}$, $x \in \mathbb{R} \setminus [0,1]$ *that* $(\mathfrak{R}_{\mathfrak{r}}(\xi_k))(x) = \mathfrak{r}(x)$

(v) for all $k \in \mathbb{N}$, $x \in [0,1]$, *we have* $\left|x^2 - (\mathfrak{R}_{\mathfrak{r}}(\xi_k))(x)\right| \leqslant 2^{-2k-2}$, *and*

(vi) for al $k \in \mathbb{N}$, *we have that* $\mathcal{P}(\xi_k) = 20k - 7$

Proof. 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:

$$g_1(x) = \begin{cases} 2x & : x \in \left[0, \frac{1}{2}\right) \\ 2 - 2x & : x \in \left[\frac{1}{2}, 1\right] \\ 0 & : x \in \mathbb{R} \setminus [0,1] \end{cases} \tag{6.1.4}$$

$$g_{k+1} = g_1(g_k)$$

and let $f_k : [0,1] \to [0,1]$, $k \in \mathbb{N}_0$ be the functions satisfying for all $k \in \mathbb{N}_0$, $n \in \{0, 1, ..., 2^k - 1\}$, $x \in \left[\frac{n}{2^k}, \frac{n+1}{2^k}\right)$ that $f_k(1) = 1$ and:

$$f_k(x) = \left[\frac{2n+1}{2^k}\right] x - \frac{n^2+n}{2^{2k}} \tag{6.1.5}$$

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:

$$r_1(x) = \begin{bmatrix} r_{1,1}(x) \\ r_{2,1}(x) \\ r_{3,1}(x) \\ r_{4,1}(x) \end{bmatrix} = \mathfrak{r}\left(\begin{bmatrix} x \\ x - \frac{1}{2} \\ x - 1 \\ x \end{bmatrix}\right) \tag{6.1.6}$$

$$r_{k+1} = A_{k+1} r_k(x)$$

Note that since it is the case that for all $x \in \mathbb{R}$ that $\mathfrak{r}(x) = \max\{x, 0\}$, (6.1.4) and (6.1.6) shows that it holds for all $x \in \mathbb{R}$ that:

$$\begin{aligned}2r_{1,1}(x) - 4r_{2,1}(x) + 2r_{3,1}(x) &= 2\,\mathfrak{r}(x) - 4\,\mathfrak{r}\left(x - \frac{1}{2}\right) + 2\,\mathfrak{r}\,(x-1) \\ &= 2\max\{x,0\} - 4\max\left\{x - \frac{1}{2}, 0\right\} + 2\max\{x-1,0\} \\ &= g_1(x) \end{aligned} \tag{6.1.7}$$

Note also that combined with (6.1.5), the fact that for all $x \in [0,1]$ it holds that $f_0(x) = x = \max\{x,0\}$ tells us that for all $x \in \mathbb{R}$:

$$r_{4,1}(x) = \max\{x,0\} = \begin{cases} f_0(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R} \setminus [0,1] \end{cases} \tag{6.1.8}$$

We next claim that for all $k \in \mathbb{N}$ it is the case that:

$$(\forall x \in \mathbb{R} : 2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x) = g(x)) \tag{6.1.9}$$

and that:

$$\left(\forall x \in \mathbb{R} : r_{4,k}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R} \setminus [0,1] \end{cases}\right) \tag{6.1.10}$$

We prove (6.1.9) and (6.1.10) by induction. The base base of $k = 1$ is proved by (6.1.7) and (6.1.8). 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:

$$2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x) = g_k(x) \tag{6.1.11}$$

and:

$$r_{4,k}(x) = \begin{cases} f_{k-1}(x) & : x \in [0,1] \\ \max\{x,0\} & : x \in \mathbb{R} \setminus [0,1] \end{cases} \tag{6.1.12}$$

Note that then (6.1.4),(6.1.6), and (6.1.7) then tells us that for all $x \in \mathbb{R}$ it is the case that:

$$\begin{aligned}
g_{k+1}(x) = g_1(g_k(x)) &= g_1(2r_{1,k}(x) + 4r_{2,k}(x) + 2r_{3,k}(x)) \\
&= 2\,\mathfrak{r}\,(2r_{1,k}(x)) + 4r_{2,k} + 2r_{3,k}(x)) \\
&- 4\,\mathfrak{r}\left(2r_{1,k}(x) - 4r_{2,k} + 2r_{3,k}(x) - \frac{1}{2}\right) \\
&+ 2\,\mathfrak{r}\,(2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x) - 1) \\
&= 2r_{1,k+1}(x) - 4r_{2,k+1}(x) + 2r_{3,k+1}(x)
\end{aligned} \tag{6.1.13}$$

In addition note that (6.1.5), (6.1.6), and (6.1.8) tells us that for all $x \in \mathbb{R}$:

$$\begin{aligned}
r_{4,k+1}(x) &= \mathfrak{r}\left((-2)^{3-2(k+1)}\,r_{1,k}(x) + 2^{4-2(k+1)}r_{2,k}(x) + (-2)^{3-2(k+1)}\,r_{3,k}(x) + r_{4,k}(x)\right) \\
&= \mathfrak{r}\left((-2)^{1-2k}\,r_{1,k}(x) + 2^{2-2k}r_{k,2}(x) + (-2)^{1-2k}\,r_{3,k}(x) + r_{4,k}(x)\right) \\
&= \mathfrak{r}\left(2^{-2k}\left[-2r_{1,k}(x) + 2^2r_{2,k}(x) - 2r_{3,k}(x)\right] + r_{4,k}(x)\right) \\
&= \mathfrak{r}\left(-\left[2^{-2k}\right]\left[2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x)\right] + r_{4,k}(x)\right) \\
&= \mathfrak{r}\left(-\left[2^{-2k}\right]g_k(x) + r_{4,k}(x)\right)
\end{aligned} \tag{6.1.14}$$

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 $f_k(x) \geqslant 0$, (6.1.12), shows that for all $x \in [0,1]$ it holds that:

$$\begin{aligned}
r_{4,k+1}(x) &= \mathfrak{r}\left(-2\left[2^{-2k}g_k\right] + f_{k-1}(x)\right) = \mathfrak{r}\left(-2\left(2^{-2k}g_k(x)\right) + x - \left[\sum_{j=1}^{k-1}\left(2^{-2j}g_j(x)\right)\right]\right) \\
&= \mathfrak{r}\left(x - \left[\sum_{j=1}^{k}2^{-2j}g_j(x)\right]\right) = \mathfrak{r}\left(f_k(x)\right) = f_k(x)
\end{aligned} \tag{6.1.15}$$

Note next that (6.1.12) and (6.1.14) then tells us that for all $x \in \mathbb{R} \setminus [0,1]$:

$$r_{4,k+1}(x) = \max\left\{-\left(2^{-2k}g_x(x)\right) + r_{4,k}(x)\right\} = \max\{\max\{x,0\},0\} = \max\{x,0\} \tag{6.1.16}$$

Combining (6.1.13) and (6.1.15) proves (6.1.9) and (6.1.10). Note that then (6.1.1) and (6.1.9) assure that for all $k \in \mathbb{N}$, $x \in \mathbb{R}$ it holds that $\mathfrak{R}_{\mathfrak{r}}(\xi_k) \in C(\mathbb{R},\mathbb{R})$ and that:

$$\begin{aligned}
&(\mathfrak{R}_{\mathfrak{r}}(\xi_k))(x) \\
&= \left(\mathfrak{R}_{\mathfrak{r}}\left((\mathsf{Aff}_{C_k,0} \bullet \mathfrak{i}_4) \bullet \left(\mathsf{Aff}_{A_{k-1},B} \bullet \mathfrak{i}_4\right) \bullet \cdots \bullet (\mathsf{Aff}_{A_1,B} \bullet \mathfrak{i}_4) \bullet \mathsf{Aff}_{\mathbb{e}_4,B}\right)\right)(x) \\
&= (-2)^{1-2k} r_{1,k}(x) + 2^{2-2k} r_{2,k}(x) + (-2)^{1-2k} r_{3,k}(x) + r_{4,k}(x) \\
&= (-2)^{2-2k}\left(\left[\frac{r_{1,k}(x) + r_{3,k}(x)}{-2}\right] + r_{2,k}(x)\right) + r_{4,k}(x) \\
&= 2^{2-2k}\left(\left[\frac{r_{1,k}(x) + r_{3,k}(x)}{-2}\right] + r_{2,k}(x)\right) + r_{4,k}(x) \\
&= 2^{-2k}\left(4r_{2,k}(x) - 2r_{1,k}(x) - 2r_{3,k}(x)\right) + r_{4,k}(x) \\
&= -\left[2^{-2k}\right]\left[2r_{1,k}(x) - 4r_{2,k}(x) + 2r_{3,k}(x)\right] + r_{4,k}(x) = -\left[2^{-2k}\right] g_k(x) + r_{4,k}(x) \qquad (6.1.17)
\end{aligned}$$

This and (6.1.10) tell us that:

$$\begin{aligned}
(\mathfrak{R}_{\mathfrak{r}}(\xi_k))(x) = -\left(2^{-2k} g_k(x)\right) + f_{k-1}(x) &= -\left(2^{-2k} g_k(x)\right) + x - \left[\sum_{j=1}^{k-1} 2^{-2j} g_j(x)\right] \\
&= x - \left[\sum_{j=1}^{k} 2^{-2j} g_j(x)\right] = f_k(x)
\end{aligned}$$

Which then implies for all $k \in \mathbb{N}$, $x \in [0,1]$ that it holds that:

$$\left\| x^2 - (\mathfrak{R}_{\mathfrak{r}}(\xi_k))(x) \right\| \leqslant 2^{-2k-2} \qquad (6.1.18)$$

This in turn establish Item (i).

Finally observe that (6.1.17) then tells us that for all $k \in \mathbb{N}$, $x \in \mathbb{R} \setminus [0,1]$ it holds that:

$$(\mathfrak{R}_{\mathfrak{r}}(\xi_k))(x) = -2^{-2k} g_k(x) + r_{4,k}(x) = r_{4,k}(x) = \max\{x, 0\} = \mathfrak{r}(x) \qquad (6.1.19)$$

This establishes Item(iv). Note next that Item(iii) ensures for all $k \in \mathbb{N}$ that $\mathcal{D}(\xi_k) = k+1$, and:

$$\mathcal{P}(\xi_k) = 4(1+1) + \left[\sum_{j=2}^{k} 4(4+1)\right] + (4+1) = 8 + 20(k-1) + 5 = 20k - 7 \qquad (6.1.20)$$

This in turn proves Item(vi). The proof of the lemma is thus complete. □

Lemma 6.1.2. *Let* $\delta, \epsilon \in (0,\infty)$, $\alpha \in (0,\infty)$, $q \in (2,\infty)$, $\Phi \in \mathrm{NN}$ *satisfy that* $\delta = 2^{\frac{-2}{q-2}}\varepsilon^{\frac{q}{q-2}}$, $\alpha = \left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}}$, $\Phi \in C(\mathbb{R},\mathbb{R})$, $\mathcal{D}(\Phi) \leqslant \max\left\{\frac{1}{2}\log_2(\delta^{-1})+1, 2\right\}$, $\mathcal{P}(\Phi) \leqslant \max\left\{10\log_2\left(\delta^{-1}\right)-7, 13\right\}$, $\sup_{x\in\mathbb{R}\setminus[0,1]} \left|\left(\mathfrak{R}_{\mathfrak{r}}(\Phi) - \mathfrak{r}(x)\right|\right. = 0$, *and* $\sup_{x\in[0,1]}\left|x^2 - \left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(x)\right| \leqslant \delta$, *then:*

(i) there exists a unique $\Psi \in \mathrm{NN}$, *satisfying:* $\Psi = \left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}\right) \bigoplus \left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}\right)$

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

(iii) it holds that $\left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(0) = 0$

(iv) it holds for all $x \in \mathbb{R}$ *that* $0 \leqslant \left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(x) \leqslant \varepsilon + |x|^2$

(v) it holds for all $x \in \mathbb{R}$ *that* $\left|x^2 - \left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(x)\right| \leqslant \varepsilon \max\{1, |x|^q\}$

(vi) it holds that $\mathcal{D}(\Psi) \leqslant \max\left\{1 + \frac{1}{q-2} + \frac{q}{2(q-2)}\log_2\left(\varepsilon^{-1}\right), 2\right\}$, *and*

(vii) it holds that $\mathcal{P}(\Psi) \leqslant \max\left\{\left[\frac{40q}{q-2}\right]\log_2\left(\varepsilon^{-1}\right) + \frac{80}{q-2} - 28, 52\right\}$

Proof. Note that for all $x \in \mathbb{R}$ it is the case that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(x) &= \left(\mathfrak{R}_{\mathfrak{r}}\left(\left(\mathsf{Aff}_{\alpha^{-2}} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}\right) \oplus \left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}\right)\right)\right)(x) \\
&= \left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{\alpha,0}\right)\right)(x) + \left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Aff}_{\alpha^{-2},0} \bullet \Phi \bullet \mathsf{Aff}_{-\alpha,0}\right)\right)(x) \\
&= \frac{1}{\alpha^2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(\alpha x) + \frac{1}{\alpha^2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(-\alpha x) \\
&= \frac{1}{\left(\frac{\varepsilon}{2}\right)^{\frac{2}{q-2}}}\left[\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)\left(\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x\right) + \left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)\left(-\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x\right)\right] \qquad (6.1.21)
\end{aligned}$$

This and the assumption that $\Phi \in C(\mathbb{R},\mathbb{R})$ along with the assumption that $\sup_{x\in\mathbb{R}\setminus[0,1]}\left|\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(x) - \mathfrak{r}(x)\right| = 0$ tells us that for all $x \in \mathbb{R}$ it holds that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(0) &= \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}}\left[\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(0) + \left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(0)\right] \\
&= \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}}\left[\mathfrak{r}(0) + \mathfrak{r}(0)\right] \\
&= 0 \qquad (6.1.22)
\end{aligned}$$

This in turn establishes Item (iii). Observe next that from the assumption that $\mathfrak{R}_{\mathfrak{r}}(\Phi) \in C(\mathbb{R},\mathbb{R})$ and the assumption that $\sup_{x\in\mathbb{R}\setminus[0,1]} |(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ ensure that for all $x \in \mathbb{R}\setminus[-1,1]$ it holds that:

$$\begin{aligned}[\mathfrak{R}_{\mathfrak{r}}(\Phi)](x) + [\mathfrak{R}_{\mathfrak{r}}(\Phi)(-x)] = \mathfrak{r}(x) + \mathfrak{r}(-x) &= \max\{x,0\} + \max\{-x,0\} \\ &= |x| \end{aligned} \tag{6.1.23}$$

The assumption that for all $\sup_{x\in\mathbb{R}\setminus[0,1]} |(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - \mathfrak{r}(x)| = 0$ and the assumption that $\sup_{x\in[0,1]} |x^2 - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x)| \leqslant \delta$ show that:

$$\begin{aligned}&\sup_{x\in[-1,1]} \left|x^2 - \left([\mathfrak{R}_{\mathfrak{r}}(\Phi)](x) + [\mathfrak{R}_{\mathfrak{r}}(\Phi)(x)]\right)\right| \\ &= \max\left\{\sup_{x\in[-1,0]} \left|x^2 - \left(\mathfrak{r}(x) + [\mathfrak{R}_{\mathfrak{r}}(\Phi)](-x)\right)\right|, \sup_{x\in[0,1]} \left|x^2 - \left([\mathfrak{R}_{\mathfrak{r}}(\Phi)](x) + \mathfrak{r}(-x)\right)\right|\right\} \\ &= \max\left\{\sup_{x\in[-1,0]} \left|(-x)^2 - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(-x)\right|, \sup_{x\in[0,1]} \left|x^2 - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x)\right|\right\} \\ &= \sup_{x\in[0,1]} \left|x^2 - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x)\right| \leqslant \delta \end{aligned} \tag{6.1.24}$$

Next observe that (6.1.21) and (6.1.23) 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:

$$\begin{aligned}0 \leqslant [\mathfrak{R}_{\mathfrak{r}}(\Psi)](x) &= \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} \left([\mathfrak{R}_{\mathfrak{r}}(\Phi)]\left(\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x\right) + [\mathfrak{R}_{\mathfrak{r}}(\Phi)]\left(-\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x\right)\right) \\ &= \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} \left|\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x\right| = \left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}|x|} \leqslant |x|^2 \end{aligned} \tag{6.1.25}$$

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:

$$\begin{aligned}\left|x^2 - (\mathfrak{R}_{\mathfrak{r}}(\Psi))(x)\right| &= \left|x^2 - \left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}} |x|\right| \leqslant \left(|x|^2 + \left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}} |x|\right) \\ &= \left(|x|^q\, |x|^{-(q-2)} + \left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}} |x|^q\, |x|^{-(q-1)}\right) \\ &\leqslant \left(|x|^q \left(\frac{\varepsilon}{2}\right)^{\frac{q-2}{q-2}} + \left(\frac{\varepsilon}{2}\right)^{\frac{-1}{q-2}} |x|^q \left(\frac{\varepsilon}{2}\right)^{\frac{q-1}{q-2}}\right) \\ &= \left(\frac{\varepsilon}{2} + \frac{\varepsilon}{2}\right) |x|^q = \varepsilon\, |x|^q \leqslant \varepsilon \max\{1, |x|^q\} \end{aligned} \tag{6.1.26}$$

Note that (6.1.24), (6.1.21) 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:

$$\begin{aligned} &\left| x^2 - \left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(x) \right| \\ &= \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} \left| \left(\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x \right)^2 - \left(\left[\mathfrak{R}_{\mathfrak{r}}(\Phi)\right] \left(\left(\frac{\varepsilon}{2}\right)^{\frac{1}{q-2}} x \right) + \left[\mathfrak{R}_{\mathfrak{r}}(\Phi)\right](-y) \right) \right| \\ &\leqslant \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} \left[\sup_{y \in [-1,1]} \left| y^2 - \left[\mathfrak{R}_{\mathfrak{r}}(\Phi)\right](y) + \left[\mathfrak{R}_{\mathfrak{r}}(\Phi)\right](-y) \right| \right] \\ &\leqslant \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} \delta = \left(\frac{\varepsilon}{2}\right)^{\frac{-2}{q-2}} 2^{\frac{-2}{q-2}} \varepsilon^{\frac{q}{q-2}} = \varepsilon \leqslant \varepsilon \max\{1, |x|^q\} \end{aligned} \tag{6.1.27}$$

Now note that this and (6.1.26) tells us that for all $x \in \mathbb{R}$ it is the case that:

$$\left| x^2 - \left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(x) \right| \leqslant \varepsilon \max\{1, |x|^q\} \tag{6.1.28}$$

This establishes Item (v). Note that, (6.1.27) 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:

$$\left| \left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(x) \right| \leqslant \left| x^2 - \left(\mathfrak{R}_{\mathfrak{r}}(\Psi)\right)(x) \right| + |x|^2 \leqslant \varepsilon + |x|^2 \tag{6.1.29}$$

This and (6.1.26) tells us that for all $x \in \mathbb{R}$:

$$\left| \left(\mathfrak{R}_{\mathfrak{r}}\right)(x) \right| \leqslant \varepsilon + |x|^2 \tag{6.1.30}$$

This establishes Item (iv). Notice next that

□

Remark 6.1.3. *Note that from here onward we will refer to the neural network network Ψ defined in Lemma 9.1.3 Item(i) as the* Sqr *neural network.*

Lemma 6.1.4. *Let $\delta, \varepsilon \in (0,\infty)$, $q \in (2,\infty)$, $A_1, A_2, A_3 \in \mathbb{R}^{1\times 2}$, $\Phi \in \mathbb{N}$ satisfy for all $x \in \mathbb{R}$ that $\delta = \varepsilon \left(2^{q-1}+1\right)^{-1}$, $A_1 = [1 \quad 1]$, $A_2 = [1 \quad 0]$, $A_3 = [0 \quad 1]$, $\mathfrak{R}_{\mathfrak{r}} \in C(\mathbb{R},\mathbb{R})$, $\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(0) = 0$, $0 \leqslant \left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(x) \leqslant \delta + |x|^2$, $|x^2 - \left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(x)| \leqslant \delta \max\{1, |x|^q\}$, $\mathcal{D}(\Phi) \leqslant \max\{1 + \frac{1}{q-2} + \frac{q}{2(q-2)} \log_2\left(\delta^{-1}\right), 2\}$, and $\mathcal{P}(\Phi) \leqslant \max\left\{ \left[\frac{40q}{q-2}\right] \log_2\left(\delta^{-1}\right) + \frac{80}{q-2} - 28, 52 \right\}$, then:*

(i) there exists a unique $\Gamma \in \mathrm{NN}$ *satisfying:*

$$\Gamma = \left(\frac{1}{2} \circledast \left(\Phi \bullet \mathsf{Aff}_{A_1,0}\right)\right) \bigoplus \left(\left(-\frac{1}{2}\right) \circledast \left(\Phi \bullet \mathsf{Aff}_{A_2,0}\right)\right) \bigoplus \left(\left(-\frac{1}{2}\right) \circledast \left(\Phi \bullet \mathsf{Aff}_{A_3,0}\right)\right) \tag{6.1.31}$$

(ii) it holds that $\mathfrak{R}_{\mathfrak{r}}(\Gamma) \in C\left(\mathbb{R}^2, \mathbb{R}\right)$

(iii) it holds for all $x \in \mathbb{R}$ *that* $\left(\mathfrak{R}_{\mathfrak{r}}(\Gamma)\right)(x,0) = \left(\mathfrak{R}_{\mathfrak{r}}(\Gamma)\right)(0,y) = 0$

(iv) it holds for any $x, y \in \mathbb{R}$ *that* $\left| xy - \left(\mathfrak{R}_{\mathfrak{r}}(\Gamma)\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)\right| \leqslant \varepsilon \max\{1, |x|^q, |y|^q\}$

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

(vi) it holds that $\mathcal{D}(\Gamma) \leqslant \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right]$

Proof. Note that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{r}}(\Gamma)\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= \mathfrak{R}_{\mathfrak{r}}\left(\left(\frac{1}{2} \circledast \left(\Phi \bullet \mathsf{Aff}_{A_1,0}\right)\right) \bigoplus \left(\left(-\frac{1}{2}\right) \circledast \left(\Phi \bullet \mathsf{Aff}_{A_2,0}\right)\right) \bigoplus\right. \qquad (6.1.32) \\
&\left.\left(\left(-\frac{1}{2}\right) \circledast \left(\Phi \bullet \mathsf{Aff}_{A_3,0}\right)\right)\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \\
&= \mathfrak{R}_{\mathfrak{r}}\left(\frac{1}{2} \circledast \left(\Phi \bullet \mathsf{Aff}_{A_1,0}\right)\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) + \mathfrak{R}_{\mathfrak{r}}\left(\left(-\frac{1}{2}\right) \circledast \left(\Phi \bullet \mathsf{Aff}_{A_2,0}\right)\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \\
&+ \mathfrak{R}_{\mathfrak{r}}\left(\left(-\frac{1}{2}\right) \circledast \left(\Phi \bullet \mathsf{Aff}_{A_3,0}\right)\right)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \\
&= \frac{1}{2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)\left(\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right) - \frac{1}{2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)\left(\begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right) \\
&- \frac{1}{2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)\left(\begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right) \\
&= \frac{1}{2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(x+y) - \frac{1}{2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(x) - \frac{1}{2}\left(\mathfrak{R}_{\mathfrak{r}}(\Phi)\right)(y) \qquad (6.1.33)
\end{aligned}$$

Note that this, and the assumption that $(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) \in C(\mathbb{R},\mathbb{R})$ and that $(\mathfrak{R}_{\mathfrak{r}}(\Phi))(0) = 0$ ensures:

$$\begin{aligned}
(\mathfrak{R}_{\mathfrak{r}}(\Gamma))\left(\begin{bmatrix} x \\ 0 \end{bmatrix}\right) &= \frac{1}{2}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x+0) - \frac{1}{2}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - \frac{1}{2}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(0) \\
&= 0 \\
&= \frac{1}{2}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(0+y) - \frac{1}{2}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(0) - \frac{1}{2}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(y) \\
&= (\mathfrak{R}_{\mathfrak{r}}(\Gamma))\left(\begin{bmatrix} 0 \\ y \end{bmatrix}\right)
\end{aligned} \tag{6.1.34}$$

Next, observe that since by assumption it is the case for all $x, y \in \mathbb{R}$ that $|x^2 - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x)| \leqslant \delta \max\{1, |x|^q\}$, $xy = \frac{1}{2}|x+y|^2 - \frac{1}{2}|x|^2 - \frac{1}{2}|y|^2$, triangle Inequality and from (6.1.33) we have that:

$$\begin{aligned}
&|(\mathfrak{R}_{\mathfrak{r}}(\Gamma)(x,y)) - xy| \\
&= \left|\frac{1}{2}\left[(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x+y) - |x+y|^2\right] - \frac{1}{2}\left[(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - |x|^2\right] - \frac{1}{2}\left[(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - |y|^2\right]\right| \\
&\leqslant \left|\frac{1}{2}\left[(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x+y) - |x+y|^2\right] + \frac{1}{2}\left[(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - |x|^2\right] + \frac{1}{2}\left[(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - |y|^2\right]\right| \\
&\leqslant \frac{\delta}{2}\left[\max\{1, |x+y|^q\} + \max\{1, |x|^q\} + \max\{1, |y|^q\}\right]
\end{aligned}$$

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:

$$\begin{aligned}
&|(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - xy| \\
&\leqslant \frac{\delta}{2}\left[\max\left\{1, 2^{q-1}|x|^q + 2^{q-1}|y|^q\right\} + \max\{1, |x|^q\} + \max\{1, |y|^q\}\right] \\
&\leqslant \frac{\delta}{2}\left[\max\left\{1, 2^{q-1}|x|^q\right\} + 2^{q-1}|y|^q + \max\{1, |x|^q\} + \max\{1, |y|^q\}\right] \\
&\leqslant \frac{\delta}{2}\left[2^q + 2\right]\max\{1, |x|^q, |y|^q\} = \varepsilon \max\{1, |x|^q, |x|^q\}
\end{aligned}$$

This proves Item(iv). Note that $\mathcal{P}(\mathsf{Aff}_{A_i}) = 2$ for $i = \{1, 2, 3\}$. This, combined with (Grohs et al., 2023, Lemma 2.6, Item(iv)), and the fact that $\mathcal{P}(\Phi) \leqslant \max\left\{\left[\frac{40q}{q-2}\right]\log_2(\delta^{-1}) + \frac{80}{q-2} - 28, 52\right\}$, and

Lemma 5.4.2 Item (i) tells us that for $i = 1, 2, 3$:

$$\begin{aligned}
&\mathcal{P}\left(\frac{1}{2} \circledast \left(\Phi \bullet \mathsf{Aff}_{A_i,0}\right)\right) \\
&\leqslant \mathcal{O}(\Phi)^2 + \mathcal{P}(\Phi) + \mathcal{P}\left(\mathsf{Aff}_{A_i,0}\right) + \mathcal{W}_1(\Phi) \cdot \mathcal{W}_1\left(\mathsf{Aff}_{A_i,0}\right) \\
&= 1^2 + \mathcal{P}(\Phi) + (1 \cdot 2 + 1) + (l_{\phi,1} \cdot 1) \\
&= 6 + \mathcal{P}\left(\Phi\right) \\
&\leqslant 6 + \max\left\{\left[\frac{40q}{q-2}\right] \log_2\left(\delta^{-1}\right) + \frac{80}{q-2} - 28, 52\right\}
\end{aligned}$$

Notice now that by Lemma 5.5.6 we have that

$$\begin{aligned}
&\mathcal{P}\left(\left(\frac{1}{2} \circledast \left(\Phi \bullet \mathsf{Aff}_{A_1,0}\right)\right) \oplus \left(\left(-\frac{1}{2}\right) \circledast \left(\Phi \bullet \mathsf{Aff}_{A_2,0}\right)\right)\right) \\
&\leqslant 6 + \max\left\{\left[\frac{40q}{q-2}\right] \log_2\left(\delta^{-1}\right) + \frac{80}{q-2} - 28, 52\right\} \\
&+ \max\left\{\left[\frac{40q}{q-2}\right] \log_2\left(\delta^{-1}\right) + \frac{80}{q-2} - 28, 52\right\}
\end{aligned}$$

□

Remark 6.1.5. *We shall refer to this neural network for a given q and given ε from now on as* $\mathsf{Prd}^{q,\varepsilon}$.

6.2 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* so that we are able to effectively parallelize two neural networks not of the same length.

6.2.1 The Tun Neural Network

Definition 6.2.1 (The Tunneling Neural Network)**.** *We define the tunneling neural network, denoted as* Tun_n *for* $n \in \mathbb{N}$ *by:*

$$\mathsf{Tun}_n = \begin{cases} \mathsf{Aff}_{1,0} & : n = 1 \\ \mathsf{Id}_1 & : n = 2 \\ \bullet^{n-2} \mathsf{Id}_1 & n \in \mathbb{N} \cap [3, \infty) \end{cases} \tag{6.2.1}$$

Where Id_1 *is as in Definition 9.2.1.*

Lemma 6.2.2. *Let* $n \in \mathbb{N}$*,* $x \in \mathbb{R}$ *and* $\mathsf{Tun}_n \in \mathrm{NN}$*. For all* $n \in \mathbb{N}$ *and* $x \in \mathbb{R}$*, it is then the case that:*

(i) $\mathfrak{R}_{\mathfrak{r}} (\mathsf{Tun}_n) \in C (\mathbb{R}, \mathbb{R})$

(ii) $\mathcal{D} (\mathsf{Tun}_n) = n$

(iii) $(\mathfrak{R}_{\mathfrak{r}} (\mathsf{Tun}_n)) (x) = x$

(iv) $\mathcal{P} (\mathsf{Tun}_n) = \begin{cases} 2 & : n = 1 \\ 7 + 6(n-2) & n \in \mathbb{N} \cap [2, \infty) \end{cases}$

(v) $\mathcal{L} (\mathsf{Tun}_n) = (l_0, l_1, ..., l_{L-1}, l_L) = (1, 2, ..., 2, 1)$

Proof. Note that by Lemma 5.4.2 it is the case that $\mathcal{D} (\mathsf{Aff}_{1,0}) = 1$ and by Lemma 9.2.1 it is the case that $\mathcal{D} (\mathsf{Id}_1) = 2$. Assume now that for all $n \leqslant N$ that $\mathcal{D} (\mathsf{Tun}_n) = n$, then for the inductive step, by Lemma 5.2.2 we have that:

$$\begin{aligned} \mathcal{D} (\mathsf{Tun}_{n+1}) &= \mathcal{D} \left(\bullet^{n-1} \mathsf{Id}_1\right) \\ &= \mathcal{D} \left(\left(\bullet^{n-2} \mathsf{Id}_1\right) \bullet \mathsf{Id}_1\right) \\ &= n + 2 - 1 = n + 1 \end{aligned} \tag{6.2.2}$$

This completes the induction and hence Item (i) and Item (iii). Note next that by (5.1.13) we have that:

$$\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Aff}_{1,0}\right)\right)(x) = x \tag{6.2.3}$$

Lemma 9.2.2, Item (iii) also tells us that:

$$\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Id}_1\right)\right)(x) = \mathfrak{r}(x) - \mathfrak{r}(-x) = x \tag{6.2.4}$$

Assume now that for all $n \leqslant N$ that $\mathsf{Tun}_n(x) = x$. For the inductive step, by Lemma 9.2.2, Item (iii), and we then have that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tun}_{n+1}\right)\right)(x) &= \left(\mathfrak{R}_{\mathfrak{r}}\left(\bullet^{n-1}\mathsf{Id}_1\right)\right)(x)(x) \\
&= \left(\mathfrak{R}_{\mathfrak{r}}\left(\left(\bullet^{n-2}\mathsf{Id}_1\right)\bullet \mathsf{Id}_1\right)\right) \\
&= \left(\left(\mathfrak{R}_{\mathfrak{r}}\left(\bullet^{n-2}\mathsf{Id}_1\right)\right)\circ\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Id}_1\right)\right)\right)(x) \\
&= \left(\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tun}_n\right)\right)\circ\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Id}_1\right)\right)\right)(x) \\
&= x
\end{aligned} \tag{6.2.5}$$

This proves Item (ii). Next note that $\mathcal{P}\left(\mathsf{Aff}_{1,0}\right) = 2$. Note also that:

$$\begin{aligned}
\mathcal{P}\left(\mathsf{Id}_1\right) &= \mathcal{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\right] \\
&= 7
\end{aligned}$$

And that by definition of composition:

$$\begin{aligned}
\mathcal{P}\left(\mathsf{Tun}_2\right) &= \mathcal{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\bullet\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\right] \\
&= \mathcal{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\right] \\
&= 13
\end{aligned}$$

Now for the inductive step assume that for all $n \leqslant N \in \mathbb{N}$, it is the case that $\mathcal{P}(\mathsf{Tun}_n) = 7+6(n-2)$. For the inductive step we then have:

$$\begin{aligned}
&\mathcal{P}(\mathsf{Tun}_{n+1}) = \mathcal{P}(\mathsf{Tun}_n \bullet \mathsf{Id}_1) \\
&= \mathcal{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\cdots,\left(\begin{bmatrix}1 & -1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\bullet \mathsf{Id}_1\right] \\
&= \mathcal{P}\left[\left(\left(\begin{bmatrix}1\\-1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\cdots,\left(\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix},\begin{bmatrix}0\\0\end{bmatrix}\right),\left(\begin{bmatrix}1 & -1\end{bmatrix},\begin{bmatrix}0\end{bmatrix}\right)\right)\right] \\
&= 7+6(n-2)+6 = 7+6\left((n+1)-2\right) \qquad (6.2.6)
\end{aligned}$$

This proves Item (iv).

Note finally that Item (v) is a consequence of Lemma 9.2.2, Item (i) and (Grohs et al., 2023, Proposition 2.6) ☐

6.2.2 The Pwr and Tay Neural Networks

Definition 6.2.3 (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 $\mathsf{Pwr}_n^{q,\varepsilon} \in \mathrm{NN}$, denoted for $n \in \mathbb{N}_0$ as:*

$$\mathsf{Pwr}_n^{q,\varepsilon} = \begin{cases} \mathsf{Aff}_{0,1} & n = 0 \\ \mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathcal{D}(\mathsf{Pwr}_{n-1}^{q,\varepsilon})} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right] \bullet \mathsf{Cpy}_{2,1} & \mathbb{N} \cap [1,\infty) \end{cases}$$

Lemma 6.2.4. *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 $\mathsf{Pwr}_n \in \mathrm{NN}$. It is then the case for all $n \in \mathbb{N}_0$, and $x \in \mathbb{R}$ that:*

(i) $\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)\right)(x) \in C(\mathbb{R},\mathbb{R})$

(ii) $\mathcal{D}(\mathsf{Pwr}_n^{q,\varepsilon}) \leqslant \begin{cases} 1 & n = 0 \\ \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q\right] + \mathcal{D}\left(\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right) - 1 & n \geqslant 1 \end{cases}$

(iii) $\mathcal{P}(\mathsf{Pwr}_n^{q,\varepsilon}) \leqslant \begin{cases} 2 & n=0 \\ \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q+1\right]+4\left(\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right)+\mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) & \\ +\mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)+2\,\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-2}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-235 & n \geqslant 1 \end{cases}$

(iv) $\left|x^n - \left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)\right)(x)\right| \leqslant \begin{cases} 0 & n=0,1 \\ \varepsilon\max\left\{1,|x|^q,\left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right|^q\right\} & n \geqslant 2 \end{cases}$

Proof. Note that Lemma 9.2.2 ensures that $\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_0\right) \in C\left(\mathbb{R},\mathbb{R}\right)$. Note next that by (Grohs et al., 2023, Proposition 2.6), with $\Phi_1 \curvearrowleft \nu_1, \Phi_2 \curvearrowleft \nu_2, a \curvearrowleft \mathfrak{r}$, we have that:

$$\left(\mathfrak{R}_{\mathfrak{r}}\left(\nu_1 \bullet \nu_2\right)\right)(x) = \left(\left(\mathfrak{R}_{\mathfrak{r}}\left(\nu_1\right)\right) \circ \left(\mathfrak{R}_{\mathfrak{r}}\left(\nu_2\right)\right)\right)(x) \tag{6.2.7}$$

This, with the fact that, the composition of continuous functions is continuous, the fact the parallelization of continuous realized neural networks is continuous tells us that $\left(\mathfrak{R}_{\mathfrak{r}}\,\mathsf{Pwr}_n\right) \in C\left(\mathbb{R},\mathbb{R}\right)$ for $n \in \mathbb{N} \cap [2,\infty)$.

Note next that by Lemma 9.2.2, it is the case that $\mathcal{D}\left(\mathsf{Id}_1\right) = 1$. By Lemmas 5.5.2 and 5.2.2 it is also the case that: $\mathcal{D}\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathcal{D}(\mathsf{Pwr}_{n-1}^{q,\varepsilon})} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right] \bullet \mathsf{Cpy}\right) = \mathcal{D}\left(\mathsf{Prd}^{q,\varepsilon} \bullet \left[\mathsf{Tun}_{\mathcal{D}(\mathsf{Pwr}_{n-1}^{q,\varepsilon})} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right]\right)$. Note also that by parallelization properties we have that $\mathcal{D}\left(\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) = \mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)$. This with (Grohs et al., 2023, Proposition 2.6) and the fact that $\bullet$ is associative then yields that for $n \geqslant 2$ that:

$$\begin{aligned} \mathcal{D}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right) &= \mathcal{D}\left(\mathsf{Prd} \bullet \left[\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right] \bullet \mathsf{Cpy}_{2,1}\right) \\ &= \mathcal{D}\left(\mathsf{Prd} \bullet \left[\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right]\right) \\ &= \mathcal{D}\left(\mathsf{Prd}\right) + \mathcal{D}\left(\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right) - 1 \\ &\leqslant \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q\right] + \mathcal{D}\left(\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right) - 1 \end{aligned} \tag{6.2.8}$$

Next note that by Lemma 5.4.2 we have that:

$$\mathcal{P}\left(\mathsf{Pwr}_0^{q,\varepsilon}\right) = \mathcal{P}\left(\mathsf{Aff}_{0,1}\right) = 2 \tag{6.2.9}$$

Next note that by (Grohs et al., 2023, Proposition 2.6) we then have that for $n \geqslant 2$:

$$
\begin{aligned}
&\mathcal{P}\left(\left[\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right] \bullet \mathsf{Cpy}_{2,1}\right) \\
&\leqslant \mathcal{P}\left(\left[\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right]\right) + \mathcal{P}\left(\mathsf{Cpy}_{2,1}\right) + \left(2 + \mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right) \cdot 1 \\
&= \mathcal{P}\left(\left[\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)} \boxminus \mathsf{Pwr}_{n-1}^{q,\varepsilon}\right]\right) + \mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 6 \\
&= \mathcal{P}\left(\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)}\right) + \mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + \mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 6 \\
&= 4\left(\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right) + \mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + \mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 13
\end{aligned} \tag{6.2.10}
$$

and that:

$$
\begin{aligned}
&\mathcal{P}\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) \\
&= \mathcal{P}\left(\mathsf{Prd}\right) + \mathcal{P}\left(\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) + \mathcal{W}_1\left(\mathsf{Prd}\right) \cdot \left[\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-2}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 2\right] \\
&\leqslant \mathcal{P}\left(\mathsf{Prd}\right) + 4\left(\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right) + \mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + \mathcal{W}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 13 \\
&+ \mathcal{W}_1\left(\mathsf{Prd}\right) \cdot \left[\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-2}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 2\right] \\
&\leqslant \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] - 252 + 4\left(\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right) + \mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + \mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 13 \\
&+ \mathcal{W}_1\left(\mathsf{Prd}\right) \cdot \left[\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-2}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 2\right] \\
&= \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] - 252 + 4\left(\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right) + \mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + \mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 13 \\
&+ 2\,\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-2}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + 4 \\
&= \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q + 1\right] + 4\left(\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right) + \mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) + \mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) \\
&+ 2\,\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-2}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) - 235
\end{aligned}
$$

Next note that $\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{0,1}\right)$ and $\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{1,0}\right)$ are exactly 1 and x respectively. Note also that the realizations of Tun_n and $\mathsf{Cpy}_{2,1}$ are exact. Thus it is the case that for all $n \in \mathbb{N}$, we have that $\mathfrak{R}(\mathsf{Pwr}_n^{q,\varepsilon})(x) = \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(\mathfrak{R}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x), x\right)\right)$ Note then that Lemma 6.1.4 then gives us that for all $n \in \mathbb{N}$ it is the case that:

$$\begin{aligned}\left|x^n - \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_n^{q,\varepsilon}\right)\right| &= \left|x^{n-1} \cdot x - \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Prd}^{q,\varepsilon}\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x), x\right)\right)\right| \\ &\leqslant \varepsilon \max\left\{1, |x|^q, \left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)(x)\right|^q\right\} \end{aligned} \tag{6.2.11}$$

This completes the lemma. □

Remark 6.2.5. *Note we may now define what we will call neural network polynomials as objects of the form:*

$$\bigcup_{n \in \mathbb{N}_0} \left[\bigoplus_{i=0}^{n} \left(c_i \circledast \mathfrak{E}_{\max_{i \in \{0,1,\dots,n\}}\left\{\mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\} - \mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right), \mathsf{Id}_1}\left[\mathsf{Pwr}_i^{q,\varepsilon}\right]\right)\right] \subseteq \mathrm{NN} \tag{6.2.12}$$

Where $c_i \in \mathbb{R}$, for all $i \in \{0, 1, ..., n\}$.

Definition 6.2.6 (Taylor Approximations for e^x around $x = 0$)**.** *Let $\delta, \varepsilon \in (0, \infty)$, $q \in (2, \infty)$ and $\delta = \varepsilon\left(2^{q-1} + 1\right)^{-1}$. We define, for all $n \in \mathbb{N}_0$, the family of neural networks $\mathsf{Tay}_{n,q,\varepsilon}^{\exp}$ as:*

$$\mathsf{Tay}_{n,q,\varepsilon}^{\exp} = \bigoplus_{i=0}^{n} \left[\frac{1}{i!} \circledast \mathfrak{E}_{\max_{i \in \{0,1,\dots,n\}}\left\{\mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\} - \mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right), \mathsf{Id}_1}\left[\mathsf{Pwr}_i^{q,\varepsilon}\right]\right] \tag{6.2.13}$$

Lemma 6.2.7. *Let $\nu_1, \nu_2 \in \mathrm{NN}$, $f, g \in C(\mathbb{R}, \mathbb{R})$, and $\varepsilon_1, \varepsilon_2 \in (0, \infty)$ such that for all $x \in \mathbb{R}$ it holds that $|f(x) - \mathfrak{R}_{\mathfrak{r}}(\nu_1)| \leqslant \varepsilon_1$ and $|g(x) - \mathfrak{R}_{\mathfrak{r}}(\nu_2)| \leqslant \varepsilon_2$. It is then the case for all $x \in \mathbb{R}$ that:*

$$\left|[f + g](x) - \mathfrak{R}_{\mathfrak{r}}\left([\nu_1 \oplus \nu_2]\right)(x)\right| \leqslant \varepsilon_1 + \varepsilon_2 \tag{6.2.14}$$

Proof. Note that the triangle inequality then tells us:

$$\begin{aligned}\left|[f + g](x) - \mathfrak{R}_{\mathfrak{r}}[\nu_1 \oplus \nu_2](x)\right| &= \left|f(x) + g(x) - \mathfrak{R}_{\mathfrak{r}}(\nu_1)(x) - \mathfrak{R}_{\mathfrak{r}}(\nu_2)(x)\right| \\ &\leqslant \left|f(x) - \mathfrak{R}_{\mathfrak{r}}(\nu_1)(x)\right| + \left|g(x) - \mathfrak{R}_{\mathfrak{r}}(\nu_2)(x)\right| \\ &\leqslant \varepsilon_1 + \varepsilon_2\end{aligned}$$

□

Lemma 6.2.8. *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, $\left| f_i(x) - \mathfrak{R}_{\mathfrak{r}}(\nu_i)(x)\right| \leqslant \varepsilon_i$. It is then the case for all $x \in \mathbb{R}$, that:*

$$\left|\sum_{i=1}^{n} f_i(x) - \bigoplus_{i=1}^{n} \left(\mathfrak{R}_{\mathfrak{r}}(\nu_i)\right)(x)\right| \leqslant \sum_{i=1}^{n} \varepsilon_i \tag{6.2.15}$$

Proof. This is a consequence of a finite number of applications of (6.2.14). □

Lemma 6.2.9. *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) $\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right) \in C(\mathbb{R},\mathbb{R})$

(ii) $\mathcal{D}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right) \leqslant \begin{cases} 2 & n=0 \\ \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q\right] + \mathcal{D}\left(\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right)}\right) - 1 & n \geqslant 1 \end{cases}$

(iii) $\mathcal{P}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right) \leqslant \begin{cases} 2 & n=0 \\ \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q+1\right] + 4\left(\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right)\right) + \mathcal{P}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right) & \\ +\mathcal{W}_1\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right) + 2\,\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right)-2}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right) - 235 & n \geqslant 1 \end{cases}$

(iv) $\left|\sum_{i=0}^{n}\left[\frac{x^i}{i!}\right] - \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right)\right| \leqslant \sum_{i=1}^{n}\frac{\varepsilon}{i!}\max\left\{1, |x|^q, \left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}^{q,\varepsilon}_{i-1}\right)(x)\right|^q\right\}$

Proof. Note that by Lemma 5.6.3, Lemma 6.2.4, and (Grohs et al., 2023, Proposition 2.6) for all $n \in \mathbb{N}_0$ it is the case that:

$$\begin{aligned}\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right) &= \mathfrak{R}_{\mathfrak{r}}\left[\bigoplus_{i=0}^{n}\left[\frac{1}{i!} \circledast \mathfrak{E}_{\max_{i\in\{0,1,\dots,n\}}\left\{\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_i\right)\right\}-\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_i\right),\mathsf{Id}_1}\left[\mathsf{Pwr}^{q,\varepsilon}_i\right]\right]\right] \\ &= \sum_{i=0}^{n}\frac{1}{i!}\,\mathfrak{R}_{\mathfrak{r}}\left[\mathfrak{E}_{\max_{i\in\{0,1,\dots,n\}}\left\{\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_i\right)\right\}-\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_i\right),\mathsf{Id}_1}\left[\mathsf{Pwr}^{q,\varepsilon}_i\right]\right] \\ &= \sum_{i=0}^{n}\frac{1}{i!}\,\mathfrak{R}_{\mathfrak{r}}\left[\mathsf{Pwr}^{q,\varepsilon}_i\right]\end{aligned} \tag{6.2.16}$$

Since $\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}^{q,\varepsilon}_n\right)\right)(x) \in C(\mathbb{R},\mathbb{R})$, for all $n \in \mathbb{N}_0$ and since the finite sum of continuous functions is continuous, this proves Item (i).

Note that $\mathsf{Tay}^{\exp}_{n,q,\varepsilon}$ is only as deep as the deepest of the $\mathsf{Pwr}^{q,\varepsilon}_i$ networks, which from definition is $\mathsf{Pwr}^{q,\varepsilon}_n$, which in turn has the largest bound. Therefore, by Lemma 5.5.7 and Lemma 5.4.4 we have that:

$$\begin{aligned}\mathcal{D}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right) &= \mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_n\right) \\ &\leqslant \begin{cases} 2 & n = 0 \\ \frac{q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right) + q\right] + \mathcal{D}\left(\mathsf{Tun}_{\mathcal{D}\left(\mathsf{Pwr}^{q,\varepsilon}_{n-1}\right)}\right) - 1 & n \geqslant 1 \end{cases}\end{aligned}$$

Note that $\mathcal{P}\left(\mathsf{Id}_1\right) = 7$ and further Definition 5.2.1 and (Grohs et al., 2023, Proposition 2.6) tells us that for $\nu = \left(\left(W_1, b_1\right), \left(W_2, b_2\right), ..., \left(W_L, b_L\right)\right) \in \mathrm{NN}$ it is the case that:

$$\mathcal{P}\left(\mathsf{Id}_1 \bullet \nu\right) \leqslant 7 + \mathcal{P}\left(\nu\right) + 2 \cdot \mathcal{W}_{L-1}\left(\nu\right) \tag{6.2.17}$$

Which then in turn implies that for $L \in \mathbb{N}$ and $\nu \in \mathrm{NN}$, it is the case that:

$$\begin{aligned}\mathcal{P}\left(\mathfrak{E}_{L,\mathsf{Id}_1}\left(\nu\right)\right) &= \mathcal{P}\left(\mathsf{Id}_1^{\bullet(L-\mathcal{D}(\nu))} \bullet \nu\right) \\ &\leqslant \mathcal{P}\left(\mathsf{Id}_1^{\bullet(L-\mathcal{D}(\nu))}\right) + \mathcal{P}\left(\nu\right) + 2 \cdot \mathcal{W}_{\mathcal{D}(\nu)-1}\left(\nu\right) \\ &= 7 + 6\left(L - \mathcal{D}\left(\nu\right) - 1\right) + \mathcal{P}\left(\nu\right) + 2 \cdot \mathcal{W}_{\mathcal{D}(\nu)-1}\left(\nu\right) \\ &= 6L - 6\,\mathcal{D}\left(\nu\right) + 1 + \mathcal{P}\left(\nu\right) + 2 \cdot \mathcal{W}_{\mathcal{D}(\nu)-1}\left(\nu\right)\end{aligned}$$

Note that each neural network summand in $\mathsf{Tay}^{\exp}_{n,q,\varepsilon}$ 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, i.e. Tun_k. This, finally, with Lemma 6.2.4 then implies that:

$$\begin{aligned}
\mathcal{P}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right) &= \mathcal{P}\left[\bigoplus_{i=0}^{n}\left[\frac{1}{i!}\circledast \mathfrak{E}_{\max_{i\in\{0,1,\dots,n\}}\left\{\mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\}-\mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right),\mathsf{Id}_1}\left[\mathsf{Pwr}_i^{q,\varepsilon}\right]\right]\right] \\
&\leqslant n\cdot\mathcal{P}\left(\frac{1}{n!}\circledast \mathsf{Pwr}_n\right) \\
&\leqslant n\cdot\mathcal{P}\left(\mathsf{Pwr}_n\right)+2 \\
&\leqslant \begin{cases} 2 & n=0 \\ \frac{360q}{q-2}\left[\log_2\left(\varepsilon^{-1}\right)+q+1\right]+4\left(\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)\right)+\mathcal{P}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right) & \\ +\mathcal{W}_1\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)+2\,\mathcal{W}_{\mathcal{D}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-2}\left(\mathsf{Pwr}_{n-1}^{q,\varepsilon}\right)-235 & n\geqslant 1 \end{cases}
\end{aligned}$$

Finally, note that for all $i\in\mathbb{N}$, Lemma 6.2.4, and by absolute homogenity of norms, the fact that $\frac{1}{i!}\geqslant 0$ for all $i\in\mathbb{N}$, and (Grohs et al., 2023, Proposition 2.6) then tells us that it is the case that:

$$\begin{aligned}
\left|x^i-\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_i\right)\right)(x)\right| &\leqslant \varepsilon\max\left\{1,|x|^q,\left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}\right)(x)\right|^q\right\} \\
\frac{1}{i!}\left|x^i-\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_i\right)\right)(x)\right| &\leqslant \frac{\varepsilon}{i!}\max\left\{1,|x|^q,\left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}\right)(x)\right|^q\right\} \\
\left|\frac{x^i}{i!}-\left(\mathfrak{R}_{\mathfrak{r}}\left(\frac{1}{i!}\circledast\mathsf{Pwr}_i\right)\right)(x)\right| &\leqslant \frac{\varepsilon}{i!}\max\left\{1,|x|^q,\left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}\right)(x)\right|^q\right\}
\end{aligned} \tag{6.2.18}$$

This, Lemma 6.2.9, and the fact that realization of the tunneling neural network leads to the identity function (Lemma 6.2.2 and (Grohs et al., 2023, Proposition 2.6)) then tells us that:

$$\begin{aligned}
&\left|\sum_{i=0}^{n}\left[\frac{x^i}{i!}\right]-\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right)\right| \\
&=\left|\sum_{i=0}^{n}\left[\frac{x^i}{i!}\right]-\mathfrak{R}_{\mathfrak{r}}\left[\bigoplus_{i=0}^{n}\left[\frac{1}{i!}\circledast \mathfrak{E}_{\max_{i\in\{0,1,\dots,n\}}\left\{\mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right)\right\}-\mathcal{D}\left(\mathsf{Pwr}_i^{q,\varepsilon}\right),\mathsf{Id}_1}\left[\mathsf{Pwr}_i^{q,\varepsilon}\right]\right]\right]\right| \\
&\leqslant \sum_{i=1}^{n}\frac{\varepsilon}{i!}\max\left\{1,|x|^q,\left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}_{i-1}^{q,\varepsilon}\right)(x)\right|^q\right\}
\end{aligned}$$

□

Lemma 6.2.10. *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* $x \in [a,b] \subseteq \mathbb{R}_{\geqslant 0}$ *that:*

$$\left|e^x - \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right)(x)\right| \leqslant \sum_{i=1}^{n} \frac{\varepsilon}{i!} \max\left\{1, |x|^q, \left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}^{q,\varepsilon}_{i-1}\right)(x)\right|^q\right\} + \left|\frac{e^b \cdot b^{n+1}}{(n+1)!}\right| \tag{6.2.19}$$

Proof. Note that Taylor's theorem states that for $x \in [a,b] \subseteq \mathbb{R}_{\geqslant 0}$ it is the case that:

$$e^x = \sum_{i=0}^{n}\left[\frac{x^i}{i!}\right] + \frac{e^{\xi} \cdot x^{n+1}}{(n+1)!} \tag{6.2.20}$$

Where $\xi \in [0,x]$ in the Lagrange form of the remainder. Note then, for all $n \in \mathbb{N}_0$, $x \in [a,b] \subseteq \mathbb{R}$, and $\xi \in [0,x]$ it is the case that the second summand is bounded by:

$$\frac{e^{\xi} \cdot x^{n+1}}{(n+1)!} \leqslant \frac{e^b \cdot b^{n+1}}{(n+1)!} \tag{6.2.21}$$

This, and the triangle inequality then indicates that for all $x \in [a,b] \subseteq \mathbb{R}_{\geqslant 0}$:

$$\begin{aligned}
\left|e^x - \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right)(x)\right| &= \left|\sum_{i=0}^{n}\left[\frac{x^i}{i!}\right] + \frac{e^{\xi} \cdot x^{n+1}}{(n+1)!} - \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right)(x)\right| \\
&\leqslant \left|\sum_{i=0}^{n}\left[\frac{x^i}{i!}\right] - \mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Tay}^{\exp}_{n,q,\varepsilon}\right)(x)\right| + \left|\frac{e^b \cdot b^{n+1}}{(n+1)!}\right| \\
&\leqslant \sum_{i=1}^{n} \frac{\varepsilon}{i!} \max\left\{1, |x|^q, \left|\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Pwr}^{q,\varepsilon}_{i-1}\right)(x)\right|^q\right\} + \left|\frac{e^b \cdot b^{n+1}}{(n+1)!}\right|
\end{aligned}$$

□

Chapter 7

A modified Multi-Level Picard and associated neural network

We now look at neural networks in the context of multi-level Picard iterations.

Lemma 7.0.1. *Let $\alpha, \beta, M \in [0,\infty)$, $U_n \in [0,\infty)$, for $n \in \mathbb{N}_0$ satisfy for all $n \in \mathbb{N}$ that:*

$$U_n \leqslant \alpha M^n + \sum_{i=0}^{n-1} M^{n-i} \left(\max\left\{\beta, U_i\right\} + \mathbb{1}_{\mathbb{N}}\left(i\right) \max\left\{\beta, U_{\max\{i-1,0\}}\right\}\right) \tag{7.0.1}$$

It is then also the case that for all $n \in \mathbb{N}$ that $U_n \leqslant (2M+1)^n \max\{\alpha, \beta\}$.

Proof. Let:

$$S_n = M^n + \sum_{i=0}^{n-1} M^{n-i} \left[(2M+1)^i + \mathbb{1}_{\mathbb{N}}\left(i\right)\left(2M+1\right)^{\max\{i-1,0\}}\right] \tag{7.0.2}$$

We prove this by induction. The base case of $n=0$ already implies that $U_0 \leqslant \alpha \leqslant \max\{\alpha,\beta\}$. Next assume that $U_n \leqslant (2M+1)^n \max\{\alpha,\beta\}$ holds for all integers upto and including n, it is then the case that:

$$
\begin{aligned}
U_{n+1} &\leqslant \alpha M^{n+1} + \sum_{i=0}^{n} M^{n+1-i} \left(\max\{\beta, U_i\} + \mathbb{1}_{\mathbb{N}}(i) \max\left\{\beta, U_{\max\{i-1,0\}}\right\}\right) \\
&\leqslant \alpha M^{n+1} + \sum_{i=0}^{n} M^{n+1-i} \left[\max\left\{\beta, (2M+1)^k \max\{\alpha,\beta\}\right\} \right. \\
&\left. + \mathbb{1}_{\mathbb{N}}(i) \max\left\{\beta, (2M+1)^{\max\{k-1,0\}} \max\{\alpha,\beta\}\right\}\right] \\
&\leqslant \alpha M^{n+1} + \max\{\alpha,\beta\} \sum_{i=0}^{n} M^{n+1-i} \left[(2M+1)^i + \mathbb{1}_{\mathbb{N}}(i)(2M+1)^{\max\{i-1,0\}}\right] \\
&\leqslant \max\{\alpha,\beta\} S_{n+1}
\end{aligned} \tag{7.0.3}
$$

Then (7.0.2) and the assumption that $M \in [0, \infty)$ tells us that:

$$
\begin{aligned}
S_{n+1} &= M^{n+1} + \sum_{i=0}^{n} M^{n+1-i} \left[(2M+1)^i + \mathbb{1}_{\mathbb{N}}(i)(2M+1)^{\max\{i-1,0\}}\right] \\
&= M^{n+1} \sum_{i=0}^{n} M^{n+1-i} (2M+1)^k + \sum_{i=1}^{n} M^{n+1-i} (2M+1)^{i-1} \\
&= M^{n+1} + M \left[\frac{(2M+1)^{n+1} - M^{n+1}}{M+1} \right] + M \left[\frac{(2M+1)^n - M^n}{M+1} \right] \\
&= M^{n+1} + \frac{M(2M+1)^{n+1}}{M+1} + \frac{(2M+1)^n}{M+1} - M \left[\frac{M^{n+1} + M^n}{M+1} \right] \\
&\leqslant M^{n+1} + \frac{M(2M+1)^{n+1}}{M+1} + \frac{(2M+1)^{n+1}}{M+1} - M^{n+1} \left[\frac{\cancel{M+1}}{\cancel{M+1}} \right] \\
&= (2M+1)^{n+1}
\end{aligned} \tag{7.0.4}
$$

This completes the induction step proving (7.0.1). □

Lemma 7.0.2. *Let* $\Theta = \left(\bigcup^{n \in \mathbb{N}} \mathbb{Z}^n\right)$, $d, M \in \mathbb{N}$, $T \in (0, \infty)$, $f \in C(\mathbb{R}, \mathbb{R})$, $g, \in C\left(\mathbb{R}^d, \mathbb{R}\right)$, $\mathsf{F}, \mathsf{G} \in$ NN *satisfy that* $\mathfrak{R}_{\mathfrak{r}}(\mathsf{F}) = f$ *and* $\mathfrak{R}_{\mathfrak{r}}(\mathsf{G}) = g$, *let* $\mathfrak{u}^\theta \in [0,1]$, $\theta \in \Theta$, *and* $\mathcal{U}^\theta : [0,T] \to [0,T]$, $\theta \in \Theta$, *satisfy for all* $t \in [0,T]$, *theta* $\in \Theta$ *that* $\mathcal{U}^\theta_t = t + (T-t)\mathfrak{u}^\theta$, *let* $\mathcal{W}^\theta : [0,T] \to \mathbb{R}^d$, $\theta \in \Theta$, *for every* $\theta \in \Theta$, $t \in [0,T]$, $s \in [t,T]$, *let* $\mathcal{Y}^\theta_{t,s} \in \mathbb{R}$ *satisfy* $\mathcal{Y}^\theta_{t,s} = \mathcal{W}^\theta_s - \mathcal{W}^\theta_t$ *and let* $\mathcal{U}^\theta_n : [0,T] \times \mathbb{R}^d \to \mathbb{R}$,

$n \in \mathbb{N}_0$, $\theta \in \Theta$, *satisfy for all* $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$, $x \in \mathbb{R}^d$ *that:*

$$U_n^\theta(t,x) = \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n}\left[\sum_{k=1}^{M^n} g\left(x + \mathcal{Y}_{t,T}^{(\theta,0,-k)}\right)\right] + \sum_{i=0}^{n-1}\frac{T-t}{M^{n-i}}\left[\sum_{k=1}^{M^{n-i}}\left(\left(f \circ U_i^{(\theta,i,k)}\right) - \mathbb{1}_{\mathbb{N}}(i)\left(f \circ U_{\max\{i-1,0\}}^{(\theta,-i,k)}\right)\right)\left(\mathcal{U}_t^{(\theta,i,k)}, x + \mathcal{Y}_{t,\mathcal{U}_t^{(\theta,i,k)}}^{(\theta,i,k)}\right)\right] \tag{7.0.5}$$

it is then the case that:

(i) *there exists unique* $\mathsf{U}_{n,t}^\theta \in \mathrm{NN}$, $t \in [0,T]$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, *which satisfy for all* $\theta_1, \theta_2 \in \Theta$, $n \in \mathbb{N}_0$, $t_1, t_2 \in [0,T]$ *that* $\mathcal{L}\left(\mathsf{U}_{n,t_1}^{\theta_1}\right) = \mathcal{L}\left(\mathsf{U}_{n,t_2}^{\theta_2}\right)$.

(ii) *for all* $\theta \in \Theta$, $t \in [0,T]$ *that* $\mathsf{U}_{0,t}^\theta = \left(\begin{bmatrix}0 & 0 & \cdots & 0\end{bmatrix}, [0]\right) \in \mathbb{R}^{1\times d} \times \mathbb{R}^1$

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

$$\begin{aligned}\mathsf{U}_{n,t}^\theta = &\left[\bigoplus_{k=1}^{M^n}\left(\frac{1}{M^n} \circledast \left(\mathsf{G} \bullet \mathsf{Aff}_{\mathbb{I}_d, \mathcal{Y}_{t,T}^{(\theta,0,-k)}}\right)\right)\right] \\ &\boxplus_{\mathbb{I}}\left[\boxplus_{i=0,\mathbb{I}}^{n-1}\left[\left(\frac{T-t}{M^{n-i}}\right) \circledast \left(\boxplus_{k=1,\mathbb{I}}^{M^{n-i}}\left(\left(\mathsf{F} \bullet \mathsf{U}_{i,\mathcal{U}_t^{(\theta,i,k)}}^{(\theta,i,k)}\right) \bullet \mathsf{Aff}_{\mathbb{I}_d}, \mathcal{Y}_{t,\mathcal{U}_t^{(\theta,i,k)}}^{(\theta,i,k)}\right)\right)\right]\right] \\ &\boxplus_{\mathbb{I}}\left[\boxplus_{i=0,\mathbb{I}}^{n-1}\left[\left(\frac{(t-T)\mathbb{1}_{\mathbb{N}}}{M^{n-i}}\right) \circledast \left(\boxplus_{k=1,\mathbb{I}}^{M^{n-i}}\left(\left(\mathsf{F} \bullet \mathsf{U}_{\max\{i-1,0\},\mathcal{U}_t^{(\theta,i,k)}}^{(\theta,-i,k)}\right) \bullet \mathsf{Aff}_{\mathbb{I}_d,\mathcal{Y}_{t,\mathcal{U}_t^{(\theta,i,k)}}^{(\theta,i,k)}}\right)\right)\right]\right]\end{aligned} \tag{7.0.6}$$

(iv) *that for all* $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$, *that* $\mathcal{D}\left(\mathsf{U}_{n,t}^\theta\right) = n \cdot \mathcal{H}(\mathsf{F}) + \max\{1, \mathbb{1}_{\mathbb{N}}(n)\,\mathcal{D}(\mathsf{G})\}$

(v) *that for all* $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$, *that* $\left\|\mathcal{L}\left(\mathsf{U}_{\mathsf{n,t}}^\theta\right)\right\|_{\max} \leqslant (2M+1)^n \max\left\{2, \|\mathcal{L}(\mathsf{F})\|_{\max}, \|\mathcal{L}(\mathsf{G})\|_F\right\}$

(vi) *it holds for all* $\theta \in \Theta$, $n \in \mathbb{N}_0$, $t \in [0,T]$, $x \in \mathbb{R}^d$ *that* $U_n^\theta(t,x) = \left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{U}_{n,t}^\theta\right)\right)(x)$, *and*

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

$$\mathcal{P}\left(\mathsf{U}_{n,t}^\theta\right) \leqslant 2n\,\mathcal{H}(\mathsf{F}) + \max\{1, \mathbb{1}_{\mathbb{N}}(n)\,\mathcal{D}(\mathsf{G})\}\left[(2M+1)^n \max\left\{2, \|\mathcal{L}(\mathsf{F})\|_{\max}, \|\mathcal{L}(\mathsf{G})\|_{\max}\right\}\right]^2 \tag{7.0.7}$$

Chapter 8

Some categorical ideas about neural networks

Chapter 9

ANN first approximations

9.1 Activation Function as Neural Networks

Definition 9.1.1 (Activation ANN). *Let* $n \in \mathbb{N}$. *We denote by* $\mathfrak{i}_n \in \left(\left(\mathbb{R}^{n \times n} \times \mathbb{R}^n\right) \times \left(\mathbb{R}^{n \times n} \times \mathbb{R}^n\right)\right) \subseteq \mathsf{NN}$ *the neural network given by* $\mathfrak{i}_n = \left(\left(\mathbb{I}_n, \mathbb{0}_n\right), \left(\mathbb{I}_n, \mathbb{0}_n\right)\right)$

Lemma 9.1.2. *Let* $n \in \mathbb{N}$, *it then holds that:*

(i) $\mathcal{L}(\mathfrak{i}_n) = (n, n, n) \in \mathbb{N}^3$.

(ii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathfrak{R}_{\mathfrak{a}}(\mathfrak{i}_n) \in C(\mathbb{R}^n, \mathbb{R}^n)$ *and,*

(iii) For all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathfrak{R}_{\mathfrak{a}}(\mathfrak{i}_n) = \mathrm{Mult}^n_{\mathfrak{a}}$

Proof. The fact that $\mathfrak{i}_n \in \left(\left(\mathbb{R}^{n \times n} \times \mathbb{R}^n\right) \times \left(\mathbb{R}^{n \times n} \times \mathbb{R}^n\right)\right) \subseteq \mathsf{NN}$ tells us that $\mathcal{L}(\mathfrak{i}_n) = (n, n, n) \in \mathbb{N}^3$. This establishes Item (i). Note next that 5.1.13 establishes that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$ it holds that $\mathfrak{R}_{\mathfrak{a}}(\mathfrak{i}_n) \in C(\mathbb{R}^n, \mathbb{R}^n)$ and that:

$$\left(\mathfrak{R}_{\mathfrak{a}}(\mathfrak{i}_n)\right)(x) = \mathbb{I}_n\left(\mathrm{Mult}^n_{\mathfrak{a}}\left(\mathbb{I}_n x + \mathbb{0}_{n,1}\right)\right) + \mathbb{0}_{n,1} = \mathrm{Mult}^n_{\mathfrak{a}}(x) \tag{9.1.1}$$

□

Lemma 9.1.3. *Let* $\nu \in \mathsf{NN}$. *Then:*

(i) It holds that:

$$\mathcal{L}\left(\mathsf{i}_{\mathcal{O}(\nu)} \bullet \nu\right) = \left(\mathcal{I}(\nu), \mathcal{W}_1(\nu), \mathcal{W}_2(\nu), ..., \mathcal{W}_{\mathcal{D}(\nu)-1}, \mathcal{O}(\nu), \mathcal{O}(\nu)\right) \in \mathbb{N}^{\mathcal{D}(\nu)+2} \tag{9.1.2}$$

(ii) It holds that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathfrak{R}_{\mathfrak{a}}\left(\mathsf{i}_{\mathcal{O}(\nu)} \bullet \nu\right) \in C\left(\mathbb{R}^{\mathcal{I}(\nu)}, \mathbb{R}^{\mathcal{O}(\nu)}\right)$

(iii) It holds that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^{\mathcal{I}(\nu)}$ *that* $\left(\mathfrak{R}_{\mathfrak{a}}\left(\mathsf{i}_{\mathcal{O}(\nu)} \bullet \nu\right)\right)(x) = \mathrm{Mult}^n_{\mathfrak{a}}\left(\left(\mathfrak{R}_{\mathfrak{a}}(\nu)\right)(x)\right)$

(iv) It holds that:

$$\mathcal{L}\left(\nu \bullet \mathsf{i}_{\mathcal{I}(\nu)}\right) = \left(\mathcal{I}(\nu), \mathcal{I}(\nu), \mathcal{W}_1(\nu), \mathcal{W}_2(\nu), ..., \mathcal{W}_{\mathcal{D}(\nu)-1}(\nu), \mathcal{O}(\nu)\right) \in \mathbb{N}^{\mathcal{D}(\nu)+2} \tag{9.1.3}$$

(v) It holds that for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *that* $\mathfrak{R}_{\mathfrak{a}}\left(\nu \bullet \mathsf{i}_{\mathcal{I}(\nu)}\right) \in C\left(\mathbb{R}^{\mathcal{I}(\nu)}, \mathbb{R}^{\mathcal{O}(\nu)}\right)$*, and*

(vi) It holds for all $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$ *that* $\left(\mathfrak{R}_{\mathfrak{a}}\left(\nu \bullet \mathsf{i}_{\mathcal{I}(\nu)}\right)\right)(x) = \left(\mathfrak{R}_{\mathfrak{a}}(\nu)\right)\left(\mathrm{Mult}^{\mathcal{I}(\nu)}_{\mathfrak{a}}(x)\right)$

Proof. Note that Lemma 9.1.3 implies that for all $n \in \mathbb{N}$, $\mathfrak{a} \in C(\mathbb{R}, \mathbb{R})$, $x \in \mathbb{R}^n$, it holds that $\mathfrak{R}_{\mathfrak{a}}(\mathsf{i}_n) \in C(\mathbb{R}^n, \mathbb{R}^n)$ and:

$$\left(\mathfrak{R}_{\mathfrak{a}}(\mathsf{i}_n)\right)(x) = \mathrm{Mult}^n_{\mathfrak{a}}(x) \tag{9.1.4}$$

This and (Grohs et al., 2023, Proposition 2.6) establishes Items (i)–(vi). This completes the proof of the lemma. □

9.2 ANN Representations for One-Dimensional Identity

Definition 9.2.1 (Identity Neural Network)**.** *We will denote by* $\mathsf{Id}_d \in \mathrm{NN}$ *the neural network satisfying for all* $d \in \mathbb{N}$ *that:*

(i)

$$\mathsf{Id}_1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)\left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix}\right)\right) \in \left(\left(\mathbb{R}^{2\times 1} \times \mathbb{R}^2\right) \times \left(\mathbb{R}^{1\times 2} \times \mathbb{R}^1\right)\right) \tag{9.2.1}$$

(ii)

$$\mathsf{Id}_d = \boxminus_{i=1}^{d} \mathsf{Id}_1 \tag{9.2.2}$$

For $d > 1$.

Lemma 9.2.2. *Let $d \in \mathbb{N}$, it is then the case that:*

(i) $\mathcal{L}(\mathsf{Id}_d) = (d, 2d, d) \in \mathbb{N}^3$.

(ii) $\mathfrak{R}_{\mathfrak{r}}(\mathsf{Id}_d) \in C\left(\mathbb{R}^d, \mathbb{R}^d\right)$.

(iii) For all $x \in \mathbb{R}^d$ that:

$$\left(\mathfrak{R}_{\mathfrak{r}}(\mathsf{Id}_d)\right)(x) = x$$

Proof. Note that (9.2.1) ensure that $\mathcal{L}(\mathsf{Id}_d) = (1, 2, 1)$. Furthermore, (9.2.2) and Remark 5.3.6 prove that $\mathcal{L}(\mathsf{Id}_d) = (d, 2d, d)$ which in turn proves Item (i). Note now that Remark 5.3.6 tells us that:

$$\mathsf{Id}_d = \boxminus_{i=1}^{d}(\mathsf{Id}_1) \in \left(\bigtimes_{i=1}^{L}\left[\mathbb{R}^{dl_i \times dl_{i-1}} \times \mathbb{R}^{dl_i}\right]\right) = \left(\left(\mathbb{R}^{2d \times d} \times \mathbb{R}^{2d}\right) \times \left(\mathbb{R}^{d \times 2d} \times \mathbb{R}^{d}\right)\right) \tag{9.2.3}$$

Note that 9.2.1 ensures that for all $x \in \mathbb{R}$ it is the case that:

$$\left(\mathfrak{R}_{\mathfrak{r}}(\mathsf{Id}_1)\right)(x) = \mathfrak{r}(x) - \mathfrak{r}(-x) = \max\{x, 0\} - \max\{-x, 0\} = x \tag{9.2.4}$$

And Lemma 5.3.4 shows us that for all $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ it is the case that $\mathfrak{R}_{\mathfrak{r}}(\mathsf{Id}_d) \in C\left(\mathbb{R}^d, \mathbb{R}^d\right)$ and that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{a}}(\mathsf{Id}_d)\right)(x) &= \left(\mathfrak{R}_{\mathfrak{a}}\left(\boxminus_{i=1}^{d}(\mathsf{Id}_1)\right)\right)(x_1, x_2, ..., x_d) \\
&= \left(\left(\mathfrak{R}_{\mathfrak{a}}(\mathsf{Id}_1)\right)(x_1), \left(\mathfrak{R}_{\mathfrak{a}}(\mathsf{Id}_1)\right)(x_1), ..., \left(\mathfrak{R}_{\mathfrak{a}}(\mathsf{Id}_1)\right)(x_d)\right) \\
&= (x_1, x_2, ..., x_d) = x
\end{aligned} \tag{9.2.5}$$

This proves Item (ii)–(iii), thus establishing the lemma. □

Note here the difference between Definition 9.1.1 and Definition 9.2.1.

Lemma 9.2.3. *Let* $\nu \in \mathrm{NN}$ *with end-widths* d*. It is then the case that* $\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Id}_d \bullet \nu\right)(x) = \mathfrak{R}_{\mathfrak{r}}\left(\nu \bullet \mathsf{Id}_d\right) = \mathfrak{R}_{\mathfrak{r}}(\nu)$*, i.e.* Id_d *acts as a compositional identity.*

Proof. From (5.2.1) and Definition 9.2.1 we have eight cases.

Case 1 where $d = 1$ and subcases:

(1.i) $\mathsf{Id}_d \bullet \nu$ where $\mathcal{D}(\nu) = 1$

(1.ii) $\mathsf{Id}_d \bullet \nu$ where $\mathcal{D}(\nu) > 1$

(1.iii) $\nu \bullet \mathsf{Id}_d$ where $\mathcal{D}(\nu) = 1$

(1.iv) $\nu \bullet \mathsf{Id}_d$ where $\mathcal{D}(\nu) > 1$

Case 2 where $d > 1$ and subcases:

(2.i) $\mathsf{Id}_d \bullet \nu$ where $\mathcal{D}(\nu) = 1$

(2.ii) $\mathsf{Id}_d \bullet \nu$ where $\mathcal{D}(\nu) > 1$

(2.iii) $\nu \bullet \mathsf{Id}_d$ where $\mathcal{D}(\nu) = 1$

(2.iv) $\nu \bullet \mathsf{Id}_d$ where $\mathcal{D}(\nu) > 1$

Case 1.i: Let $\nu = ((W_1, b_1))$. Deriving from Definitions 9.2.1 and 5.2.1 we have that:

$$\mathsf{Id}_1 \bullet \nu = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} W_1, \begin{bmatrix} 1 \\ -1 \end{bmatrix} b_1 + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1, \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \tag{9.2.6}$$

$$= \left(\left(\begin{bmatrix} W_1 \\ -W_1 \end{bmatrix}, \begin{bmatrix} b_1 \\ -b_1 \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \tag{9.2.7}$$

Let $x \in \mathbb{R}$. Upon realization with $\mathfrak{r}$ and $d = 1$ we have:

$$\begin{aligned} \left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{Id}_1 \bullet \nu\right)\right)(x) &= \mathfrak{r}(W_1 x + b_1) - \mathfrak{r}(-W_1 x - b_1) \\ &= \max\{W_1 x + b_1, 0\} - \max\{-W_1 x - b_1, 0\} \\ &= W_1 x + b_1 \\ &= \mathfrak{R}_{\mathfrak{r}}(\nu) \end{aligned}$$

Case 1.ii: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definition 9.2.1 and 5.2.1 we have that:

$$\begin{aligned}
\mathsf{Id}_1 \bullet \nu &= \left((W_1, b_1), (W_2, b_2), ..., (W_{L-1}, b_{L-1}), \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} W_L, \begin{bmatrix} 1 \\ -1 \end{bmatrix} b_L + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right) \\
&= \left((W_1, b_1), (W_2, b_2), ..., (W_{L-1}, b_{L-1}), \left(\begin{bmatrix} W_L \\ -W_L \end{bmatrix}, \begin{bmatrix} b_L \\ -b_L \end{bmatrix} \right), \left(\begin{bmatrix} 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 \end{bmatrix} \right) \right)
\end{aligned}$$

Let $x \in \mathbb{R}$. Note that upon realization with $\mathfrak{r}$ the last two layers are:

$$\begin{aligned}
&\mathfrak{r}(W_L x + b_L) - \mathfrak{r}(-W_L x - b_L, 0) \\
&= \max\{W_L x + b_L, 0\} - \max\{-W_L x - b_L, 0\} \\
&= W_L x + b_L
\end{aligned} \tag{9.2.8}$$

This, along with Case 1.i implies that the unrealized 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 9.2.1 and 5.2.1 we have:

$$\begin{aligned}
\nu \bullet \mathsf{Id}_1 &= \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(W_1 \begin{bmatrix} 1 & -1 \end{bmatrix}, W_1 \begin{bmatrix} 0 \end{bmatrix} + b_1 \right) \right) \\
&= \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} W_1 & -W_1 \end{bmatrix}, b_1 \right) \right)
\end{aligned}$$

Let $x \in \mathbb{R}$. Upon realization with $\mathfrak{r}$ we have that:

$$\begin{aligned}
\left(\mathfrak{R}_{\mathfrak{r}} \left(\nu \bullet \mathsf{Id}_1 \right) \right)(x) &= \begin{bmatrix} W_1 & -W_1 \end{bmatrix} \mathfrak{r} \left(\begin{bmatrix} x \\ -x \end{bmatrix} \right) + b_1 \\
&= W_1 \, \mathfrak{r}(x) - W_1 \, \mathfrak{r}(-x) + b_1 \\
&= W_1 \left(\mathfrak{r}(x) - \mathfrak{r}(-x) \right) + b_1 \\
&= W_x + b_1 = \mathfrak{R}_{\mathfrak{r}}(\nu)
\end{aligned} \tag{9.2.9}$$

Case 1.iv: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definitions 9.2.1 and 5.2.1 we have that:

$$\nu \bullet \mathsf{Id}_1 = \left(\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} W_1 & -W_1 \end{bmatrix}, b_1 \right), (W_2, b_2), ..., (W_L, b_L) \right) \tag{9.2.10}$$

Let $x \in \mathbb{R}$. Upon realization with $\mathfrak{r}$ we have that the first two layers are:

$$\begin{aligned} &\begin{bmatrix} W_1 & -W_1 \end{bmatrix} \mathfrak{r} \left(\begin{bmatrix} x \\ -x \end{bmatrix} \right) + b_1 \\ &= W_1 \, \mathfrak{r}(x) - W_1 \, \mathfrak{r}(-x) + b_1 \\ &= W_1 \left(\mathfrak{r}(x) - \mathfrak{r}(-x) \right) + b_1 \\ &= W_1 x + b_1 = \mathfrak{R}_{\mathfrak{r}} (\nu) \end{aligned} \tag{9.2.11}$$

This along with Case 1.iii implies that the unrealized first layer is equivalent (W_1, b_1) whence we have that $\nu \bullet \mathsf{Id}_1 = \nu$.

Observe that Definitions 5.3.1 and 9.2.1 tells us that:

$$\boxminus_{i=1}^{d} \mathsf{Id}_i = \left(\left(\overbrace{\begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{bmatrix}}^{d-many}, \mathbb{0}_{2d} \right), \left(\overbrace{\begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,2} \end{bmatrix}}^{d-many}, \mathbb{0}_{d} \right) \right)$$

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 5.2.1 and 9.2.1 we have:

$$\mathsf{Id}_d \bullet \nu = \left(\left(\begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{bmatrix} W_1, \begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{bmatrix} b_1 \right), \left(\begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,2} \end{bmatrix}, \mathbb{0}_d \right) \right)$$

$$= \left(\left(\begin{bmatrix} [W_1]_{1,*} \\ -[W_1]_{1,*} \\ \vdots \\ [W_1]_{d,*} \\ -[W_1]_{d,*} \end{bmatrix}, \begin{bmatrix} [b_1]_1 \\ -[b_1]_1 \\ \vdots \\ [b_1]_d \\ -[b_1]_d \end{bmatrix} \right), \left(\begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,2} \end{bmatrix}, \mathbb{0}_d \right) \right)$$

Let $x \in \mathbb{R}^d$. Upon realization with $\mathfrak{r}$ we have that:

$$\begin{aligned} &\left(\mathfrak{R}_{\mathfrak{r}} \left(\mathsf{Id}_d \bullet \nu \right) \right) (x) \\ &= \mathfrak{r}([W_1]_{1,*} \cdot x + [b_1]_1) - \mathfrak{r}(-[W_1]_{1,*} \cdot x - [b_1]_1) + \cdots \\ &+ \mathfrak{r}([W_1]_{d,*} \cdot x + [b_1]_d) - \mathfrak{r}(-[W_1]_{d,*} \cdot x - [b_1]_d) \\ &= [W_1]_{1,*} \cdot x + [b_1]_1 + \cdots + [W_1]_{d,*} \cdot x + [b_1]_d \\ &= W_1 x + b_1 = \mathfrak{R}_{\mathfrak{r}} (\nu) \end{aligned}$$

Case 2.ii: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definition 9.2.1 and 5.2.1 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 realization with $\mathfrak{r}$ the last two layers become:

$$\begin{aligned}
&\mathfrak{r}([W_L]_{1,*} \cdot x + [b_L]_1) - \mathfrak{r}(-[W_L]_{1,*} \cdot x - [b_L]_1) + \cdots \\
&+ \mathfrak{r}([W_L]_{d,*} \cdot x + [b_L]_d) - \mathfrak{r}(-[W_L]_{d,*} \cdot x - [b_L]_d) \\
&= [W_L]_{1,*} \cdot x + [b_L]_1 + \cdots + [W_L]_{d,*} \cdot x + [b_L]_d \\
&= W_L x + b_L \qquad (9.2.12)
\end{aligned}$$

This, along with Case 2.i implies that the unrealized 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 9.2.1 and 5.2.1 we have:

$$\begin{aligned}
&\nu \bullet \mathsf{Id}_d \\
&= \left(\left(\begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{bmatrix}, \mathbb{0}_{2d} \right), \left(W_1 \begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,2} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,2} \end{bmatrix}, b_1 \right) \right)
\end{aligned}$$

Upon realization with $\mathfrak{r}$ we have that:

$$\begin{aligned}
&\left(\mathfrak{R}_{\mathfrak{r}}\left(\nu\right)\right)(x) \qquad (9.2.13) \\
&= \begin{bmatrix} [W_1]_{*,1} & -[W_1]_{*,1} & \cdots & [W_1]_{*,d} & -[W_1]_{*,d} \end{bmatrix} \mathfrak{r} \left(\begin{bmatrix} [x]_1 \\ -[x]_1 \\ \vdots \\ [x]_d \\ -[x]_d \end{bmatrix} \right) + b_1 \\
&= [W_1]_{*,1}\, \mathfrak{r}([x]_1) - [W_1]_{*,1}\, \mathfrak{r}(-[x]_1) + \cdots + [W_1]_{*,d}\, \mathfrak{r}([x]_d) - [W_1]_{*,d}\, \mathfrak{r}(-[x]_d) + b_1 \\
&= [W_1]_{*,1} \cdot [x]_1 + \cdots + [W_1]_{*,d} \cdot [x]_d \\
&= W_1 x + b_1 = \mathfrak{R}_{\mathfrak{r}}(\nu) \qquad (9.2.14)
\end{aligned}$$

Case 2.iv: Let $\nu = ((W_1, b_1), (W_2, b_2), ..., (W_L, b_L))$. Deriving from Definitions 9.2.1 and 5.2.1 we have:

$$\nu \bullet \mathsf{Id}_d$$
$$= \left(\left(\begin{bmatrix} \text{Weight}_{\mathsf{Id}_1,1} & & \\ & \ddots & \\ & & \text{Weight}_{\mathsf{Id}_1,1} \end{bmatrix}, \mathbb{0}_{2d} \right), \left(\begin{bmatrix} [W_1]_{*,1} & -[W_1]_{*,1} & \cdots & [W_1]_{*,d} & -[W_1]_{*,d} \end{bmatrix}, b_1 \right), ... \right.$$
$$\left. (W_2, b_2), ..., (W_L, b_L) \right)$$

Upon realization with $\mathfrak{r}$ we have that the first two layers are:

$$\begin{aligned} &\left(\mathfrak{R}_{\mathfrak{r}}(\nu)\right)(x) && (9.2.15) \\ &= \begin{bmatrix} [W_1]_{*,1} & -[W_1]_{*,1} & \cdots & [W_1]_{*,d} & -[W_1]_{*,d} \end{bmatrix} \mathfrak{r} \left(\begin{bmatrix} [x]_1 \\ -[x]_1 \\ \vdots \\ [x]_d \\ -[x]_d \end{bmatrix} \right) + b_1 \\ &= [W_1]_{*,1}\,\mathfrak{r}([x]_1) - [W_1]_{*,1}\,\mathfrak{r}(-[x]_1) + \cdots + [W_1]_{*,d}\,\mathfrak{r}([x]_d) - [W_1]_{*,d}\,\mathfrak{r}(-[x]_d) + b_1 \\ &= [W_1]_{*,1} \cdot [x]_1 + \cdots + [W_1]_{*,d} \cdot [x]_d \\ &= W_1 x + b_1 && (9.2.16) \end{aligned}$$

This, along with Case 2.iii implies that the unrealized first layer is equivalent to (W_L, b_L) whence $\mathsf{Id}_d \bullet \nu = \nu$.

This completes the proof. □

Definition 9.2.4 (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 9.2.5. *Let $d \in \mathbb{N}$. For a fixed d, the set of all neural networks $\nu \in \mathsf{NN}$ with realizations in $\mathfrak{r}$ and end-widths d form a monoid under the operation of $\bullet$.*

Proof. This is a consequence of Lemma 9.2.3 and Lemma 5.2.3. ☐

Remark 9.2.6. *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".*

9.3 Modulus of Continuity

Definition 9.3.1. *Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}$. We denote the modulus of continuity $\omega_f : [0, \infty] \to [0, \infty]$ as the function given for all $h \in [0, \infty]$ as:*

$$\omega_f(h) = \sup\left(\{|f(x) - f(y)| \in [0, \infty) : (x, y \in A, |x - y| \leqslant h)\} \cup \{0\}\right) \tag{9.3.1}$$

Lemma 9.3.2. *Let $\alpha \in [-\infty, \infty]$, $b \in [a, \infty]$, and let $f : [a, b] \cap \mathbb{R} \to \mathbb{R}$ be a function. It is then the case that for all all $x, y \in [a, b] \cap \mathbb{R}$ that $|f(x) - f(y)| \leqslant \omega_f(|x - y|)$.*

Proof. Note that (9.3.1) implies the lemma. ☐

Lemma 9.3.3. *Let $A \subseteq \mathbb{R}$, $L \in [0, \infty)$, and let $f : A \to \mathbb{R}$ satisfy for all $x, y \in A$ that $|f(x) - f(y)| \leqslant L|x - y|$. It is then the case that for all $h \in [0, \infty)$ that $\omega_f(h) \leqslant Lh$.*

Proof. Since it holds for all $x, y \in \mathbb{R}$ that $|f(x) - f(y)| \leqslant L|x - y|$, it then, with (9.3.1) imply for all $h \in [0, \infty)$ that:

$$\begin{aligned}
\omega_f(h) &= \sup\left(\{|f(x) - f(y)| \in [0, \infty) : (x, y \in A, |x - y| \leqslant h)\} \cup \{0\}\right) \\
&\leqslant \sup\left(\{L|x - y| \in [0, \infty) : (x, y \in A, |x - y| \leqslant h)\} \cup \{0\}\right) \\
&\leqslant \sup\left(\{Lh, 0\}\right) = Lh
\end{aligned} \tag{9.3.2}$$

This completes the proof of the lemma. ☐

9.4 Linear Interpolation

Note that in order to approximate more complex function we need to have a framework for approximating genertic 1-dimensional continuous functions. We introduce the linear interpolation operator, and later on see how neural networks can approximate 1-dimensional conitinuous functions to arbitrary precision.

9.4.1 The Linear Interpolation Operator

Definition 9.4.1 (Linear Interpolation Operator)**.** *Let $n \in \mathbb{N}$, $x_0, x_1, ..., x_n, y_0, y_1, ..., y_n \in \mathbb{R}$. Let it also be the case that $x_0 \leqslant x_1 \leqslant \cdots \leqslant x_n$. We denote by $\mathrm{Lin}^{y_0,y_1,...,y_n}_{x_0,x_1,...,x_n} : \mathbb{R} \to \mathbb{R}$, the function that satisfies for $i \in \{1, 2, ..., n\}$, and for all $w \in (-\infty, x_0)$, $x \in [x_{i-1}, x_i)$, $z \in [x_n, \infty)$ that:*

(i) $\mathrm{Lin}^{y_0,y_1,...,y_n}_{x_0,x_1,...,x_n}(w) = y_0$

(ii) $\mathrm{Lin}^{y_0,y_1,...,y_n}_{x_0,x_1,...,x_n}(x) = y_{i-1} + \frac{y_i - y_{i-1}}{x_i - x_{i-1}}(x - x_{i-1})$

(iii) $\mathrm{Lin}^{y_0,y_1,...,y_n}_{x_0,x_1,...,x_n}(z) = y_n$

Lemma 9.4.2. *Let $n \in \mathbb{N}$, $x_0, x_1, ..., x_n, y_0, y_1, ..., y_n \in \mathbb{R}$ with $x_0 \leqslant x_1 \leqslant \cdots \leqslant x_n$, it is then the case that:*

(i) for all $i \in \{0, 1, ..., n\}$ that:

$$\left(\mathrm{Lin}^{y_0,y_1,...,y_n}_{x_0,x_1,...,x_n}\right)(x_i) = y_i \tag{9.4.1}$$

(ii) for all $i \in \{0, 1, ..., n\}$ and $x \in [x_{i-1}, x_i]$ that:

$$\left(\mathrm{Lin}^{y_0,y_1,...,y_n}_{x_0,x_1,...,x_n}\right)(x) = \left(\frac{x_i - x}{x_i - x_{i-1}}\right) y_{i-1} + \left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) y_i \tag{9.4.2}$$

Proof. Note that (9.4.1) is a direct consequence of Definition 9.4.1. Item (i) then implies for all $i \in \{1, 2, ..., n\}$ $x \in [x_{i-1}, x_i]$ that:

$$\begin{aligned}
\left(\mathrm{Lin}^{y_0,y_1,...,y_n}_{x_0,x_1,...,x_n}\right)(x) &= \left[\left(\frac{x_i - x_{i-1}}{x_i - x_{i-1}}\right) - \left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right)\right] y_{i-1} + \left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) y_i \\
&= \left(\frac{x_i - x}{x_i - x_{i-1}}\right) y_{i-1} + \left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) y_i
\end{aligned}$$

□

Lemma 9.4.3. *Let $N \in \mathbb{N}$, $L, x_0, x_1, ..., x_N \in \mathbb{R}$ satisfy $x_0 < x_1 < \cdots < x_N$, and set let $f : [x_0, x_N] \to \mathbb{R}$ satisfy for all $x, y \in [x_0, x_N]$ that $|f(x) - f(y)| \leqslant L|x - y|$, it is then the case that:*

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

$$\left| \left(\text{Lin}_{x_0, x_1, \ldots, x_N}^{f(x_0), f(x_1), \ldots, f(x_N)} \right)(x) - \left(\text{Lin}_{x_0, x_1, \ldots, x_N}^{f(x_0), f(x_1), \ldots, f(x_N)} \right)(y) \right| \leqslant L|x - y| \tag{9.4.3}$$

, and

(ii) that:

$$\sup_{x \in [x_0, x_N]} \left| \left(\text{Lin}_{x_1, x_2, \ldots, x_N}^{f(x_0), f(x_1), \ldots, f(x_N)} \right)(x) - f(x) \right| \leqslant L \left(\max_{i \in \{1, 2, \ldots N\}} |x_i - x_{i-1}| \right) \tag{9.4.4}$$

Proof. The assumption that for all $x, y \in [x_0, x_k]$ it is the case that $|f(x) - f(y)| \leqslant L|x - y|$ and Lemma 9.3.3 prove Item (i) and Item (ii). □

9.4.2 Neural Networks to approximate the Lin operator

Lemma 9.4.4. *Let $\alpha, \beta, h \in \mathbb{R}$. Denote by $\mathsf{ReLU} \in \text{NN}$ the neural network given by $\mathsf{ReLU} = h \circledast (\mathsf{i}_1 \bullet \mathsf{Aff}_{\alpha,\beta})$. It is then the case that:*

(i) $\mathsf{ReLU} = ((\alpha, \beta), (h, 0))$

(ii) $\mathcal{L}(\mathsf{ReLU}) = (1, 1, 1) \in \mathbb{N}^3$.

(iii) $\mathfrak{R}_{\mathfrak{r}}(\mathsf{ReLU}) \in C(\mathbb{R}, \mathbb{R})$

(iv) for all $x \in \mathbb{R}$ that $(\mathfrak{R}_{\mathfrak{r}}(\mathsf{ReLU}))(x) = h \max\{\alpha x + \beta, 0\}$

Proof. Note that by Definition 5.4.1 we know that $\mathsf{Aff}_{\alpha,\beta} = ((\alpha, \beta))$, this with Definition 9.1.1, and Definition 5.2.1 together tell us that $\mathsf{i}_1 \bullet \mathsf{Aff}_{\alpha,\beta} = (\alpha, \beta)$. A further application of Definition 5.2.1, and an application of Definition 5.6.1 yields that $h \circledast (\mathsf{i}_1 \bullet \mathsf{Aff}_{\alpha,\beta}) = ((\alpha, \beta), (h, 0))$. This proves Item (i).

Note that $\mathcal{L}(\mathsf{Aff}_{\alpha,\beta}) = (1, 1)$, $\mathcal{L}(\mathsf{i}_1) = (1, 1, 1)$, and $\mathcal{L}(h) = 1$. Item (i) of Lemma 9.1.3 therefore tells us that $\mathcal{L}(\mathsf{ReLU}) = \mathcal{L}(h \circledast (\mathsf{i}_1 \bullet \mathsf{Aff}_{\alpha,\beta}))$. This proves Item (ii).

Note that Lemmas 9.1.2 and 9.1.3 tell us that:

$$\forall x \in \mathbb{R} : \left(\mathfrak{R}_{\mathfrak{r}}\left(\mathfrak{i}_1 \bullet \mathsf{Aff}_{\alpha,\beta}\right)\right)(x) = \mathfrak{r}\left(\mathfrak{R}_{\mathfrak{r}}\right)(x) = \max\{\alpha x + \beta\} \tag{9.4.5}$$

This and Lemma 5.6.1 ensures that $\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{ReLU}\right) \in C\left(\mathbb{R},\mathbb{R}\right)$ and further that:

$$\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{ReLU}\right)\right)(x) = h\left(\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathfrak{i}_1 \bullet \mathsf{Aff}_{\alpha,\beta}\right)\right)(x)\right) = h \max\{\alpha x + \beta, 0\} \tag{9.4.6}$$

This proves Item (iii)-(iv). This completes the proof of the lemma. □

Lemma 9.4.5. *Let $N \in \mathbb{N}$, $x_0, x_1, ..., x_N, y_0, y_1, ..., y_N \in \mathbb{R}$ and further that $x_0 \leqslant x_2 \leqslant \cdots \leqslant x_N$. Let $\Phi \in \mathrm{NN}$ satisfy that:*

$$\Phi = \mathsf{Aff}_{1,y_0} \bullet \left(\bigoplus_{i=0}^{N} \left[\left(\frac{y_{\min\{i+1,N\}} - y_i}{x_{\min\{i+1,N\}} - x_{\min\{i,N-1\}}} - \frac{y_i - y_{\max\{i-1,0\}}}{x_{\max\{i,1\}} - x_{\max\{i-1,0\}}} \right) \circledast \left(\mathfrak{i}_1 \bullet \mathsf{Aff}_{1,-x_i}\right) \right] \right) \tag{9.4.7}$$

It is then the case that:

(i) $\mathcal{L}(\Phi) = (1, N+1, 1) \in \mathbb{N}^3$

(ii) $\mathfrak{R}_{\mathfrak{r}}\left(\Phi\right) \in C\left(\mathbb{R},\mathbb{R}\right)$

(iii) $\left(\mathfrak{R}_{\mathfrak{r}}\left(\Phi\right)\right)(x) = \mathrm{Lin}_{x_0,x_1,...,x_N}^{y_0,y_1,...,y_N}(x)$

(iv) $\mathcal{P}(\Phi) = 3N + 4$

Proof. For notational convenience, let it be the case that for all $i \in \{0, 1, 2, ..., N\}$:

$$h_i = \frac{y_{\min\{i+1,N\}} - y_i}{x_{\min\{i+1,N\}} - x_{\min\{i,N-1\}}} - \frac{y_i - y_{\max\{i-1,0\}}}{x_{\max\{i,1\}} - x_{\max\{i-1,0\}}} \tag{9.4.8}$$

Note that $\mathcal{L}\left(\mathfrak{i}_1 \bullet \mathsf{Aff}_{1,-x_0}\right) = (1,1,1)$, and further that for all $i \in \{0,1,2,...,N\}$, $h_i \in \mathbb{R}$. Lemma 9.4.4 then tells us that for all $i \in \{0,1,2,...,N\}$, $\mathcal{L}\left(h_i \circledast \left(\mathfrak{i}_1 \bullet \mathsf{Aff}_{1,-x_i}\right)\right) = (1,1,1)$, $\mathfrak{R}_{\mathfrak{r}}\left(h_i \circledast \left(\mathfrak{i}_1 \bullet \mathsf{Aff}_{1,-x_i}\right)\right) \in C\left(\mathbb{R},\mathbb{R}\right)$, and that $\left(\mathfrak{R}_{\mathfrak{a}}\left(h_i \circledast \left(\mathfrak{i}_1 \bullet \mathsf{Aff}_{1,-x_i}\right)\right)\right)(x) = h_i \max\{x - x_k, 0\}$. This, (9.4.7), Lemma 5.4.4, and (Grohs et al., 2022, Lemma 3.28) ensure that $\mathcal{L}(\Phi) = (1, N+1, 1) \in \mathbb{N}^3$ and that $\mathfrak{R}_{\mathfrak{r}}\left(\Phi\right) \in C\left(\mathbb{R},\mathbb{R}\right)$ establishing Items (i)–(ii).

In addition, note that Item (i) and (5.1.11), tell us that:

$$\mathcal{P}(\Phi) = \overbrace{(N+1)}^{W_1} + \underbrace{(N+1)}_{b_1} + \overbrace{(N+1)}^{W_2} + \underbrace{1}_{b_2} = 3N+4 \tag{9.4.9}$$

Which proves Item (iv). For all $i \in \{0,1,2,...,N\}$, let ϕ_i be $\phi_i = h_i \circledast (\mathfrak{i} \bullet \mathsf{Aff}_{1,-x_i})$. Next note that 9.4.8, Lemma 5.4.4, and (Grohs et al., 2022, Lemma 3.28) then tell us that:

$$\left(\mathfrak{R}_{\mathfrak{r}}\left(\Phi\right)\right)(x) = y_0 + \sum_{i=1}^{n} \left(\mathfrak{R}_{\mathfrak{a}}\left(\phi_i\right)\right)(x) = y_0 + \sum_{i=1}^{n} h_i \max\{x - x_i, 0\} \tag{9.4.10}$$

Since $x_0 \leqslant x_i$ for all $i \in \{1,2,...,n\}$, it then is the case for all $x \in (\infty, x_0]$ that:

$$\left(\mathfrak{R}_{\mathfrak{r}}\left(\Phi\right)\right)(x) = y_0 + 0 = y_0 \tag{9.4.11}$$

Claim 9.4.6. *For all $i \in \{1,2,...,N\}$ it is the case that :*

$$\sum_{j=0}^{i-1} h_j = \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \tag{9.4.12}$$

We prove this claim by induction. For the base case of $i=1$ we have:

$$\sum_{j=0}^{0} h_0 = h_0 = \frac{y_1 - y_0}{x_1 - x_0} - \frac{y_0 - y_0}{x_1 - x_0} = \frac{y_1 - y_0}{x_1 - x_0} \tag{9.4.13}$$

This proves the base base for (9.4.12). Assume next that this holds for k, for the $(k+1)$-th induction step we have:

$$\begin{aligned}\sum_{j=0}^{k+1} h_j = \sum_{j=0}^{k} h_j + h_{k+1} &= \frac{y_k - y_{k-1}}{x_k - x_{k-1}} + h_{k+1} \\ &= \frac{y_k - y_{k-1}}{x_k - x_{k-1}} + \frac{y_{k+2} - y_{k-1}}{x_{k+2} - x_{k+1}} - \frac{y_{k+1} - y_k}{x_{k+1} - x_k} \\ &= \frac{y_{k+1} - y_k}{x_{k+1} - x_k}\end{aligned} \tag{9.4.14}$$

This proves (9.4.12). In addition, note that (9.4.10), (9.4.12), and the fact that for all $i \in \{1,2,...,n\}$

it is the case that $x_{i-1} \leqslant x_i$ tells us that for all $i \in \{1, 2, ..., n\}$ and $x \in [x_{i-1}, x_i]$ it is the case that:

$$\begin{aligned}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - (\mathfrak{R}_{\mathfrak{a}}(\Phi))(x_{i-1}) &= \sum_{j=0}^{n} h_j \left(\max\{x - x_j, 0\} - \max\{x_{i-1} - x_j, 0\}\right) \\ = \sum_{j=0}^{i-1} c_j \left[(x - x_j) - (x_{i-1} - x_j)\right] &= \sum_{j=0}^{i-1} c_j (x - x_{i-1}) = \left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right)(x - x_{i-1}) \end{aligned} \tag{9.4.15}$$

Claim 9.4.7. *For all $i \in \{1, 2, ..., N\}$, $x \in [x_{i-1}, x_i]$ it is the case that:*

$$(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) = y_{i-1} + \left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right)(x - x_{i-1}) \tag{9.4.16}$$

We will prove this claim by induction. For the base case of $i = 1$, (9.4.15) and (9.4.12) tell us that:

$$\begin{aligned}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) &= (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x_{i-1}) + (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x_{i-1}) \\ &= y_0 + \left(\frac{y_1 - y_0}{x_i - x_{i-1}}\right)(x - x_{i-1}) \end{aligned} \tag{9.4.17}$$

For the induction step notice that (9.4.15) implies that for all $i \in \{2, 3, ..., N\}$, $x \in [x_{i-1}, x_i]$, with the realization that $\forall x \in [x_{i-2}, x_{i-1}] : (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) = y_{i-2} + \left(\frac{y_{i-1} - y_{i-2}}{x_{i-1} - x_{i-2}}\right)(x - x_{i-2})$, it is then the case that:

$$\begin{aligned}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) &= (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x_{i-1}) + (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x_{i-1}) \\ &= y_{i-2} + \left(\frac{y_{i-1} - y_{i-2}}{x_{i-1} - x_{i-2}}\right)(x_{i-1} + x_{i-2}) + \left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right)(x - x_{i-1}) \\ &= y_{i-1} + \left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right)(x - x_{i-1}) \end{aligned} \tag{9.4.18}$$

Thus induction proves (9.4.16). Furthermore note that (9.4.12) and (9.4.8) tell us that:

$$\sum_{i=0}^{N} h_i = c_N + \sum_{i=0}^{N-1} h_i = -\frac{y_N - y_{N-1}}{x_N - x_{N-1}} + \frac{y_N - y_{N-1}}{x_N - x_{N-1}} = 0 \tag{9.4.19}$$

The fact that $\forall i \in \{0,1,...,N\} : x_i \leqslant x_N$, together with (9.4.10) imply for all $x \in [x_N, \infty)$ that:

$$\begin{aligned}(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) - (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x_N) &= \left[\sum_{i=0}^{N} h_i \left(\max\{x - x_i, 0\} - \max\{x_N - x_i, 0\}\right)\right] \\ &= \sum_{i=0}^{N} h_i \left[(x - x_i) - (x_N - r_i)\right] = \sum_{i=0}^{N} h_i (x - x_N) = 0\end{aligned}$$

This and (9.4.16) tells us that for all $x \in [x_N, \infty)$ we have:

$$(\mathfrak{R}_{\mathfrak{r}}(\Phi))(x) = (\mathfrak{R}_{\mathfrak{r}}(\Phi))(x_N) = y_{N-1} + \left(\frac{y_N - y_{N-1}}{x_N - x_{N-1}}\right)(x_N - x_{N-1}) = x_N \tag{9.4.20}$$

Together with (9.4.11), (9.4.16), and Definition 9.4.1 establishes Item (iii) thus proving the lemma.

□

9.5 Neural network approximation of 1-dimensional functions.

Lemma 9.5.1. *Let $N \in \mathbb{N}$, $L.a.x_0, x_1, ..., x_N \in \mathbb{R}$, $b \in (a, \infty)$, satisfy for all $i \in \{0,1,...,N\}$ that $x_i = a + \frac{i(b-a)}{N}$. Let $f : [a,b] \to \mathbb{R}$ satisfy for all $x, y \in [a,b]$ that $|f(x) - f(y)| \leqslant L|x - y|$ and let $\mathsf{F} \in \mathrm{NN}$ satisfy:*

$$\mathsf{F} = \mathsf{Aff}_{1,f(x_0)} \bullet \left[\bigoplus_{i=0}^{N} \left(\left(\frac{N\left(f\left(x_{\min\{i+1,N\}}\right) - 2f(x_i) + f\left(x_{\max\{i-1,0\}}\right)\right)}{b-a}\right) \circledast \left(\mathsf{Id}_1 \bullet \mathsf{Aff}_{1,-x_k}\right)\right)\right] \tag{9.5.1}$$

It is then the case that:

(i) $\mathcal{L}(\mathsf{F}) = (1, N+1, 1)$

(ii) $\mathfrak{R}_{\mathfrak{r}}(\mathsf{F}) \in C(\mathbb{R}, \mathbb{R})$

(iii) $\mathfrak{R}_{\mathfrak{r}}(\mathsf{F}) = \mathrm{Lin}^{f(x_0), f(x_1), ..., f(x_N)}_{x_1, x_2, ..., x_N}$

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

(v) it holds that $\sup_{x \in [a,b]} |(\mathfrak{R}_{\mathfrak{r}}(\mathsf{F}))(x) - f(x)| \leqslant \frac{L(b-a)}{N}$*, and*

(vi) $\mathcal{P}(\mathsf{F}) = 3N + 4$.

Proof. Note that since it is the case that for all $i \in \{0, 1, ..., N\}$: $x_{\min\{i+1,N\}} - x_{\min\{i,N-1\}} = x_{\max\{i,1\}} - x_{\max\{i-1,0\}} = \frac{b-a}{N}$, we have that:

$$\frac{f\left(x_{\min\{i+1,N\}}\right) - f\left(x_i\right)}{x_{\min\{i+1,N\}} - x_{\min\{i,N-1\}}} - \frac{f(x_i) - f\left(x_{\max\{i-1,0\}}\right)}{x_{\max\{i,1\}} - x_{\max\{i-1,0\}}} = \frac{N\left(f\left(x_{\min\{i+1,N\}}\right) - 2f\left(x_i\right) + f\left(x_{\max\{i-1,0\}}\right)\right)}{b-a} \tag{9.5.2}$$

Thus Items (i)-(iv) of Lemma 9.4.5 prove Items (i)-(iii), and (vi) of this lemma. Item (iii) combined with the assumption that for all $x, y \in [a, b] : |f(x) - f(y)| \leqslant |x - y|$ and Item (i) in Lemma 9.4.3 establish Item (iv). Furthermore, note that Item (iii), the assumption that for all $x, y \in [a, b]$: $|f(x) - f(y)| \leqslant L|x - y|$, Item (ii) in Lemma 9.4.3 and the fact that for all $i \in \{1, 2, ..., N\}$: $x_i - x_{i-1} = \frac{b-a}{N}$ demonstrate for all $x \in [a, b]$ it holds that:

$$\left|\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{F}\right)\right)(x) - f\left(x\right)\right| \leqslant L\left(\max_{i\in\{1,2,...,N\}}|x_i - x_{i-1}|\right) = \frac{L(b-a)}{N} \tag{9.5.3}$$

□

Lemma 9.5.2. *Let $L, a \in \mathbb{R}$, $b \in [a, \infty)$, $\xi \in [a, b]$, let $f : [a, b] \to \mathbb{R}$ satisfy for all $x, y \in [a, b]$ that $|f(x) - f(y)| \leqslant L|x - y|$, and let $\mathsf{F} \in \mathrm{NN}$ satisfy $\mathsf{F} = \mathsf{Aff}_{1,f(\xi)} \bullet (0 \circledast (\mathsf{i}_1 \bullet \mathsf{Aff}_{1,-\xi}))$, it is then the case that:*

(i) $\mathcal{L}(\mathsf{F}) = (1, 1, 1)$

(ii) $\mathfrak{R}_{\mathfrak{r}}(\mathsf{F}) \in C(\mathbb{R}, \mathbb{R})$

(iii) for all $x \in \mathbb{R}$, we have $(\mathfrak{R}_{\mathfrak{r}}(\mathsf{F}))(x) = f(\xi)$

(iv) $\sup_{x\in[a,b]} |(\mathfrak{R}_{\mathfrak{r}}(\mathsf{F}))(x) - f(x)| \leqslant L \max\{\xi - a, b - \xi\}$

(v) $\mathcal{P}(\mathsf{F}) = 4$

Proof. Note that Item (i) is a consequence of the fact that $\mathsf{Aff}_{1,-\xi}$ is a neural network with a real number as weight and a real number as a bias, and the fact that $\mathcal{L}(\mathsf{i}_1) = (1, 1, 1)$. Note also that Item (iii) of Lemma 9.4.4 prove Item (iii).

Note that from the construction of Aff we have that:

$$\begin{aligned} \left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{F}\right)\right)(x) &= \left(\mathfrak{R}_{\mathfrak{r}}\left(0 \circledast \left(\mathsf{i}_1 \bullet \mathsf{Aff}_{1,-\xi}\right)\right)\right)(x) + f\left(\xi\right) \\ &= 0\left(\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{i}_1 \bullet \mathsf{Aff}_{1,-\xi}\right)\right)(x)\right) + f\left(\xi\right) = f\left(\xi\right) \end{aligned} \tag{9.5.4}$$

Which establishes Item (iii). Note that (9.5.4), the fact that $\xi \in [a,b]$ and the fact that for all $x, y \in [a,b]$ it is the case that $|f(x) - f(y)| \leqslant |x-y|$ give us that for all $x \in [a,b]$ it holds that:

$$\left|\left(\mathfrak{R}_{\mathfrak{r}}\left(\mathsf{F}\right)\right)(x) - f\left(x\right)\right| = \left|f\left(\xi\right) - f\left(x\right)\right| \leqslant L\left|x - \xi\right| \leqslant L \max\left\{\xi - a, b - \xi\right\} \tag{9.5.5}$$

This establishes Item (iv). Note a simple parameter count yields that:

$$\mathcal{P}\left(\mathsf{F}\right) = 1(1+1) + 1(1+1) = 4 \tag{9.5.6}$$

Establishing Item (v) and hence the lemma. □

9.6 p-norm Approximations

Definition 9.6.1 (p-norm)**.** *Let $d \in \mathbb{N}$, and $p \in \mathbb{N} \cap [1,\infty]$. We denote by $\|\cdot\|_p : \mathbb{R}^d \to [0,\infty)$ the p-norm given for all $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ by $\|x\|_p = \left[\sum_{i=1}^d x_i\right]^{\frac{1}{d}}$ and by $\|\cdot\|_\infty : \mathbb{R}^d \to [0,\infty)$ the max norm, given for all $x = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$ by $\|x\|_\infty = \max_{i \in \{1,2,...,d\}} |x_i|$.*

Lemma 9.6.2. *Let $\varepsilon \in (0,\infty)$, $L \in [0,\infty)$, $a, b \in \mathbb{R}$ with $a \leqslant b$. and let $f : [a,b] \to \mathbb{R}$ satisfy for all $x, y \in [a,b]$ that $|f(x) - f(y)| \leqslant L|x-y|$. It is then the case that there exists a neural network $\phi \in \mathrm{NN}$ such that:*

(i) $\mathfrak{R}_{\mathfrak{r}}\left(\phi\right) \in C\left(\mathbb{R}, \mathbb{R}\right)$

(ii) $\mathcal{H}(\phi) = 1$

(iii) $\mathcal{W}_1(\phi) \leqslant L(b-a)\varepsilon^{-1} + 2$

(iv) for all $x, y \in \mathbb{R}$, $\left|\left(\mathfrak{R}_{\mathfrak{r}}\left(\phi\right)\right)(x) - \left(\mathfrak{R}_{\mathfrak{r}}\left(\phi\right)\right)(x)\right| \leqslant L|x-y|$

(v) it holds that $\sup_{x \in [a,b]} \left|\left(\mathfrak{R}_{\mathfrak{a}}\left(\phi\right)\right)(x) - f(x)\right| \leqslant \varepsilon$

(vi) it holds that $\mathcal{P}(\phi) = 3(\mathcal{W}_1(\phi)) + 1 \leqslant 3L(b-a)\varepsilon^{-1} + 7$ *, and*

(vii) $\|\mathcal{T}(\phi)\|_\infty \leqslant \max\{1, |a|, |b|, 2L, |f(x)|\}$

Proof. Note □

Part III

A deep-learning solution for u and Brownian motions

Chapter 10

ANN representations of Brownian Motion Monte Carlo

This is tentative without any reference to f.

Lemma 10.0.1. *Let $d, M \in \mathbb{N}$, $T \in (0,\infty)$, $\mathfrak{a} \in C(\mathbb{R},\mathbb{R})$, $\Gamma \in \mathrm{NN}$, satisfy that $\mathfrak{R}_{\mathfrak{a}}(\mathsf{G}) \in C\left(\mathbb{R}^d,\mathbb{R}\right)$, 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:*

$$U^\theta(t,x) = \frac{1}{M}\left[\sum_{k=1}^{M}\left(\mathfrak{R}_{\mathfrak{a}}\left(\Gamma\right)\right)\left(x + \mathcal{W}^{(\theta,0,-k)}\right)\right] \tag{10.0.1}$$

Let $\mathsf{U}^\theta_t \in \mathrm{NN}$, $\theta \in \Theta$ satisfy for all $\theta \in \Theta$, $t \in [0,T]$ that:

$$\mathsf{U}^\theta_t = \left[\bigoplus_{k=1}^{M}\left(\frac{1}{M} \circledast \left(\mathsf{G} \bullet \mathsf{Aff}_{\mathbb{I}_d,\mathcal{W}^{(\theta,0,-k)}_{T-t}}\right)\right)\right] \tag{10.0.2}$$

It is then the case that:

(i) *for all $\theta_1,\theta_2 \in \Theta$, $t_1,t_2 \in [0,T]$ that $\mathcal{L}\left(\mathsf{U}^{\theta_1}_{t_1}\right) = \mathcal{L}\left(\mathsf{U}^{\theta_2}_{t_2}\right)$.*

(ii) *for all $\theta \in \Theta$, $t \in [0,T]$, that $\mathcal{D}\left(\mathsf{U}^\theta_t\right) \leqslant \mathcal{D}(\mathsf{G})$*

(iii) *for all $\theta \in \Theta$, $t \in [0,T]$ that:*

$$\left\|\mathcal{L}\left(\mathsf{U}^\theta_t\right)\right\|_{\max} \leqslant \|\,\mathcal{L}\left(\mathsf{G}\right)\|_{\max}\left(1+\sqrt{2}\right)M \tag{10.0.3}$$

(iv) for all $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ that $U^\theta(t,x) = \left(\mathfrak{R}_{\mathfrak{a}}\left(\mathbf{U}_t^\theta\right)\right)(x)$ and

(v) for all $\theta \in \Theta$, $t \in [0,T]$ that:

$$\mathcal{P}\left(\mathsf{U}_t^\theta\right) \leqslant 2\,\mathcal{D}\left(\mathsf{G}\right)\left[\left(1+\sqrt{2}\right)M\left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_{\max}\right]^2 \tag{10.0.4}$$

Proof. Throughout the proof let $\mathsf{P}_t^\theta \in \mathrm{NN}$, $\theta \in \Theta$, $t \in [0,T]$ satisfy for all $\theta \in \Theta$, $t \in [0,T]$ that:

$$\mathsf{P}_t^\theta = \bigoplus_{k=1}^{M}\left[\frac{1}{M} \circledast \left(\mathsf{G} \bullet \mathsf{Aff}_{\mathbb{I}_d, \mathcal{W}_{T-t}^{\theta,0,-k}}\right)\right] \tag{10.0.5}$$

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 5.6.7 applied for every $\theta \in \Theta$ $t \in [0,T]$ with $v \curvearrowleft M$, $c_{i \in \{u,u+1,\dots,v\}} \curvearrowleft \left(\frac{1}{M}\right)_{i \in \{u,u+1,\dots,v\}}$, $(B_i)_{i \in \{u,u+1,\dots,v\}} \curvearrowleft \left(\mathcal{W}_{T-t}^{(\theta,0,-k)}\right)_{k \in \{1,2,\dots,M\}}$, $(\nu_i)_{i \in \{u,u+1,\dots,v\}} \curvearrowleft (\mathsf{G})_{i \in \{u,u+1,\dots,v\}}$, $\mu \curvearrowleft \Phi_t^\theta$ and with the notation of Lemma 5.6.7 tells us that for all $\theta \in \Theta$, $t \in [0,T]$, and $x \in \mathbb{R}^d$ it holds that:

$$\mathcal{L}\left(\mathsf{P}_t^\theta\right) = \left(d, M\,\mathcal{W}_1\left(\mathsf{G}\right), M\,\mathcal{W}_2\left(\mathsf{G}\right), \dots, M\,\mathcal{W}_{\mathcal{D}(\mathsf{G})-1}\left(\mathsf{G}\right), 1\right) = \mathcal{L}\left(\mathsf{P}_0^0\right) \in \mathbb{N}^{\mathcal{D}(\mathsf{G})+1} \tag{10.0.6}$$

and that:

$$\begin{aligned}\left(\mathfrak{R}_{\mathfrak{a}}\left(\mathsf{P}_t^\theta\right)\right)(x) &= \frac{1}{M}\left[\sum_{k=1}^{M}\left(\mathfrak{R}_{\mathfrak{a}}\left(\mathsf{G}\right)\right)\left(x + \mathcal{W}_{T-t}^{(\theta,0,-k)}\right)\right] \\ &= \mathsf{U}^\theta\left(t,x\right) \end{aligned} \tag{10.0.7}$$

This proves Item (i).

Note that (10.0.6), and (10.0.7) also implies that:

$$\begin{aligned}\mathcal{L}\left(\mathsf{U}_t^\theta\right) &= \mathcal{L}\left(\mathsf{P}_t^\theta\right) \\ &= \left(d, \mathcal{W}_1\left(\mathsf{P}_t^\theta\right), \mathcal{W}_2\left(\mathsf{P}_t^\theta\right), \dots, \mathcal{W}_{\mathcal{D}(\mathsf{G})}\left(\mathsf{P}_t^\theta\right), t\right) \\ &= \mathcal{L}\left(\mathsf{U}_0^0\right) \in \mathbb{N}^{\mathcal{D}(\mathsf{G})+1} \end{aligned} \tag{10.0.8}$$

This indicates that for all $\theta \in \Theta$, $t \in [0,T]$ it is the case that:

$$\begin{aligned}\left\| \mathcal{L}\left(\mathsf{U}_t^\theta\right)\right\|_\infty &= \left\|\mathcal{L}\left(\mathsf{U}_0^0\right)\right\|_\infty \\ &= \max_{k \in \{1,2,\dots,\mathcal{D}(\mathsf{G})\}} \left(\mathcal{W}_k\left(\mathsf{P}_0^0\right)\right)\end{aligned}$$

This, (10.0.6), and (Grohs et al., 2023, Proposition 2.6) ensure that for all $\theta \in \Theta$, $t \in [0,T]$ it is the case that:

$$\begin{aligned}\left\| \mathcal{L}\left(\mathsf{U}_t^\theta\right)\right\|_\infty &= \left\|\mathcal{L}\left(\mathsf{U}_0^0\right)\right\|_\infty \leqslant \left\|\mathcal{L}\left(\mathsf{P}_0^0\right)\right\|_\infty \leqslant M \left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_\infty \\ &\leqslant M \left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_\infty + M\left[\left\|\mathcal{L}\left(\mathsf{U}_0^0\right)\right\|_\infty\right] \end{aligned} \tag{10.0.9}$$

Then (Hutzenthaler et al., 2021, Corollary 4.3), with $\gamma \curvearrowleft 0$, $\beta \curvearrowleft M$, $k \curvearrowleft 1$, $\alpha_0 \curvearrowleft \left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_\infty$, $\alpha_1 \curvearrowleft 0$, $(x_i)_{i \in \{0,1,\dots,k\}} \curvearrowleft \left(\left\|\mathcal{L}\left(\mathsf{U}_0^0\right)\right\|_\infty\right)_{i \in \{0,1,\dots,n\}}$ in the notation of (Hutzenthaler et al., 2021, Corollary 4.3) yields for all $\theta \in \Theta$, $t \in [0,T]$ that:

$$\begin{aligned}\left\| \mathcal{L}\left(\mathsf{U}_t^\theta\right)\right\|_\infty &\leqslant \frac{1}{2}\left(\left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_\infty\right)\left(1+\sqrt{2}\right)M \\ &\leqslant \left(\left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_\infty\right)\left(1+\sqrt{2}\right)M\end{aligned}$$

Note that (Grohs et al., 2023, Proposition 2.6, Item (ii)) proves that for all $\theta \in \Theta$, $t \in [0,T]$ it is the case that:

$$\mathcal{D}\left(\mathsf{U}_t^\theta\right) = \mathcal{D}\left(\mathsf{U}_0^0\right) = \mathcal{D}\left(\mathsf{G}\right) \tag{10.0.10}$$

This proves Items (ii)–(iii) and (10.0.7) proves Item (iv).

Items (ii)–(iii) together shows that for all $\theta \in \Theta$, $t \in [0,T]$ it is the case that:

$$\begin{aligned}
\mathcal{P}\left(\mathsf{U}_t^\theta\right) &\leqslant \sum_{k=1}^{\mathcal{D}\left(\mathsf{U}_t^\theta\right)} \left\|\mathcal{L}\left(\mathsf{U}_t^\theta\right)\right\|_{\max} \\
&= \mathcal{D}\left(\mathsf{U}_t^\theta\right)\left\|\mathcal{L}\left(\mathsf{U}_t^\theta\right)\right\|_\infty \\
&\leqslant \mathcal{D}\left(\mathsf{U}_t^\theta\right)\left(\left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_\infty\right)\left(1+\sqrt{2}\right)M \\
&= \mathcal{D}\left(\mathsf{G}\right)\left(\left\|\mathcal{L}\left(\mathsf{G}\right)\right\|_\infty\right)\left(1+\sqrt{2}\right)M
\end{aligned}$$

This proves Item (v) and hence the whole lemma. □

Bibliography

Bass, R. F. (2011). *Brownian Motion*, page 612. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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 feynmankac formulas for viscosity solutions of semilinear parabolic partial differential equations. *Stochastics and Dynamics*, 21(08):2150048.

Crandall, M. G., Ishii, H., and Lions, P.-L. (1992). Users 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.

Durrett, R. (2019). *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Golub, G. and Van Loan, C. (2013). *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press.

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).

Gyöngy, I. and Krylov, N. V. (1996). Existence of strong solutions for Itô's stochastic equations via approximations. *Probability Theory and Related Fields*, 105:143–158.

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 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.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus.* Graduate Texts in Mathematics (113) (Book 113). Springer New York.

Rio, E. (2009). Moment Inequalities for Sums of Dependent Random Variables under Projective Conditions. *J Theor Probab*, 22(1):146–163.

Appendices

```python
import numpy as np
import matplotlib.pyplot as plt

# Set the number of steps and the step size
num_steps = 5000
step_size = 0.1

# Generate the random steps
steps = np.random.normal(0, 1, (2, num_steps)) * step_size ** 0.5

# Calculate the Brownian motion
brownian_motion = np.cumsum(steps, axis=1)

# Plot the Brownian motion
plt.plot(brownian_motion[0], brownian_motion[1])
plt.title('Brownian␣Motion')
plt.xlabel('X')
plt.ylabel('Y')
plt.show()
```

Listing 1: Python