UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

UNA NUEVA VISIÓN PARA EL LAPLACIANO FRACCIONARIO VIA REDES NEURONALES PROFUNDAS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MATEMÁTICA APLICADA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

NICOLÁS ESTEBAN VALENZUELA FIGUEROA

PROFESOR GUÍA:
CLAUDIO MUÑOZ CERÓN

MIEMBROS DE LA COMISIÓN:
DANIEL REMENIK ZISIS
ERWIN TOPP PAREDES
JUAN POZO VERA

Este trabajo ha sido parcialmente financiado por los proyectos Fondecyt 1191412,
CMM ANID PIA AFB170001, CMM ANID BASAL ACE210010
y CMM ANID BASAL FB210005

SANTIAGO DE CHILE
2022

RESUMEN TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MATEMÁTICAS APLICADAS
Y MEMORIA PARA OPTAR AL TITULO DE
INGENIERO CIVIL MATEMÁTICO
POR: NICOLÁS ESTEBAN VALENZUELA FIGUEROA
FECHA: 2022
PROF. GUÍA: CLAUDIO MUÑOZ CERÓN

UNA NUEVA VISIÓN PARA EL LAPLACIANO FRACCIONARIO VIA REDES NEURONALES PROFUNDAS

Este trabajo de tesis está basado en el estudio, via redes neuronales profundas, del problema de Dirichlet fraccionario con condiciones de borde sobre un dominio acotado d-dimensional. El problema de estudio se introduce en el Capítulo 1, presentando motivaciones, además del resultado principal de esta tesis. Este consiste en demostrar que la solución del problema de Dirichlet fraccionario puede ser aproximado con redes neuronales profundas a precisión arbitraria, superando la maldición de la dimensionalidad.

Para demostrar el teorema principal es necesario ciertas herramientas estocásticas y de aprendizaje profundo: en el Capítulo 2 se definen los procesos de Lévy, y los procesos isotrópicos α-estables que estarán relacionados con el Laplaciano fraccionario. En el Capítulo 3 se definen las redes neuronales profundas y las operaciones clásicas entre estos objetos. El Capítulo 4 define los procesos llamados Walk-on-Spheres, que se relacionan con los procesos α-estables de manera natural.

El Capítulo 5 muestra que la solución del Problema de Dirichlet fraccionario se puede representar de manera estocástica, a partir de los procesos α-estables y a partir de los procesos Walk-on-Spheres.

En los Capítulos 6, 7 y 8 se demuestra que la solución del Problema de Dirichlet fraccionario se puede aproximar mediante redes neuronales profundas que superan la maldición de la dimensionalidad a una precisión arbitraria. El Capítulo 6 involucra el caso con término de fuente nula, y el Capítulo 7 utiliza la parte asociada al término de fuente en la solución del Problema de Dirichlet fraccionario. En Capítulo 8 se unen los resultados de los Capítulos 6 y 7 para concluir el Teorema principal de esta Tesis.

Finalmente, en el Capítulo 9 se discute sobre los resultados obtenidos, en las diferencias entre este trabajo y los resultados para el Problema de Dirichlet clásico. Se discute además sobre el trabajo a futuro.

RESUMEN TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MATEMÁTICAS APLICADAS
Y MEMORIA PARA OPTAR AL TITULO DE
INGENIERO CIVIL MATEMÁTICO
POR: NICOLÁS ESTEBAN VALENZUELA FIGUEROA
FECHA: 2022
PROF. GUÍA: CLAUDIO MUÑOZ CERÓN

UNA NUEVA VISIÓN PARA EL LAPLACIANO FRACCIONARIO VIA REDES NEURONALES PROFUNDAS

This work is devoted on the study, via deep neural networks, of the fractional Dirichlet boundary value Problem in d dimensions over a bounded domain. The setting of the Problem is introduced in Chapter 1, with some motivation and the principal result of this Thesis: we prove that the solution of the fractional Dirichlet Problem can be approximated with deep neural networks overcoming the curse of dimensionality and having arbitrary precision.

In order to prove the main theorem, we need certain stochastic and deep learning tools: in Chapter 2 we define the Lévy processes, and the isotropic α-stable processes, which are related with the fractional Laplacian. In Chapter 3 we define deep neural networks and their classical operations. In Chapter 4 we define the Walk-on-Spheres processes, which are related with the α-stable processes in a natural fashion.

In Chapter 5 we show that the solution of the fractional Dirichlet Problem can be represented in a stochastic form, using the α-stable and the Walk-on-Spheres processes.

In Chapters 6, 7 and 8 we prove that the solution of the fractional Dirichlet Problem can be approximated by deep neural networks overcoming the curse of dimensionality with arbitrary accuracy. Chapter 6 involves the case without source term, and Chapter 7 deal with the associated source term. In Chapter 8 we join the results of Chapters 6 and 7 to conclude the main theorem of this thesis.

Finally, in Chapter 9 we discuss the results of this thesis and the differences between this work and recent results for the classical Dirichlet Problem. We also discuss some future work.

Lo más terrible se aprende enseguida,
y lo hermoso nos cuesta la vida.

Silvio Rodríguez

Agradecimientos

En primer lugar quiero agracadecer a mi familia por todo el apoyo incondicional y el cariño que me han otorgado a lo largo de mi vida. En particular a madre María Virginia y a mi padre Mario, que me enseñaron los valores para ser la persona que soy hoy en día. También a mis hermanos Gabriel y Mario, por tantos momentos especiales.

Muchas gracias a Claudio Muñoz, por su excelente guía y gran apoyo que me brindó en todo este trabajo, por todas las reuniones que hicimos y por abrirme espacios para exponer mi trabajo. Gracias a los miembros de la comisión por ser parte de esta y a todos los académicos, académicas, funcionarios y funcionarias que fueron parte de mi proceso educativo.

Quiero agradecer a mis amigas Andrea y Catalina, que a pesar de la distancia, son las mejores amigas que podría tener. También agradecer a cada amigo y amiga que hice en plan común, que al estar en distintas carreras no nos vemos tanto, pero el cariño sigue estando presente. A mis amigos del DIM, los Chacras: Azócar, Barri, Benja, Bruno, Choco, Javo, Magaly, Majo, Pablo, Pelela, Vicho y Vivi, por acompañarme en mi paso por especialidad. Gracias por tantas matracas, salidas y juegos.

Finalmente agradezco especialmente a Bárbara, por todo el amor y apoyo que me ha entregado durante tantos años. Gracias por todo lo que hemos vivido juntos, por acompañarme en los peores momentos, y por motivarme a mejorar cada día.

Tabla de Contenido

Chapter 1

Introduction

1.1. Motivation: using DNN to describe PDEs

Some studies of partial differential equations (PDEs) are interested in the approximation of their solutions using numerical methods, such as finite differences and finite elements [2]. In low dimensions, these methods are efficient aproximations of the solution of PDEs with arbitrary precision, but they fail when we increase the dimension. This is because the computational cost of these methods depends exponentially on the dimension of the domain with the reciprocal of the accuracy as base.

In order to avoid this problem, deep neural networks (DNNs) have become key actors in the approximation of solutions of PDEs [23, 32, 33]. Among them, deep learning based algorithms have provided numerical simulations that approximates efficiently certain PDEs in high dimensions, see, e.g. [5, 6, 18, 32, 33]. The corresponding simulations suggest that the approximation by DNNs overcome the so-called *curse of dimensionality*, in the sense that the number of real parameters that describe the DNN is bounded by a polynomial on the dimension d, and on the reciprocal of the accuracy of the approximation.

Even better, recent works have theoretically proved that certain PDEs can be approximated by DNNs, overcoming the curse of dimensionality, see e.g. [7, 16, 17, 19, 23, 24]. The work of Hutzenthaler et al. [23] proved that Parabolic PDEs in the whole space $\mathbb{R}^d$ can be approximated by DNNs overcoming the curse of dimensionality. Similar results have been proved with other PDEs, such as nonlinear PDEs [24], partial integrodifferential equations [16] and elliptic PDEs with boundary conditions [17].

1.1.1. The case of the Laplacian

In order to describe the previous results in more detail, we start by considering the classical Dirichlet boundary value Problem in d-dimensions over a bounded, convex domain $D \subset \mathbb{R}^d$:

$$\begin{cases} -\Delta u(x) = f(x) & x \in D, \\ u(x) = g(x) & x \in \partial D, \end{cases}$$

where f, g are suitable continuous functions. In a recent work, Grohs and Herrmann [17] proved that DNNs overcome the curse of dimensionality in the approximation of solution of the above problem. More precisely, they used stochastic techniques such as the Feynman-Kac formula, the so-called *Walk-on-Spheres* (WoS) processes (defined in Chapter 4) and Monte Carlo simulations in order to show that DNNs approximate the exact solution, with arbitrary precision.

The Feynam-Kac representation and Monte Carlo simulations are usual tools in similar literature. Indeed, after Feynman and Kac proved that the solution of a Parabolic PDE is related with a stochastic process over a probability space, which involves the boundary conditions of the Problem and Brownian motions, several PDEs have been treated of this way. Moreover, if the stochastic process has finite expectation, then the expected value of that process can be approximated by the mean of M independent copies of the stochastic process. The corresponding mean is called a Monte Carlo simulation, and satisfy by strong law of large numbers that his limit, when $M \to \infty$, is equal to the expected value of the process.

1.2. Main Objective

The main purpose of this work is to extend the nice results obtained by Grohs and Herrmann in the case of the fractional Laplacian $(-\Delta)^{\alpha/2}$, with $\alpha \in (0,2)$, formally defined in $\mathbb{R}^d$ as

$$-(-\Delta)^{\alpha/2}u(x) = c_{d,\alpha} \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}^d \setminus B(0,\varepsilon)} \frac{u(y)-u(x)}{|y-x|^{d+\alpha}} dy, \qquad x \in \mathbb{R}^d, \tag{1.1}$$

where

$$c_{d,\alpha} = -\frac{2^\alpha \Gamma((d+\alpha)/2)}{\pi^{d/2}\Gamma(-\alpha/2)},$$

and $\Gamma(\cdot)$ is the classical Gamma function.

We prove that in this general case, there exist DNNs that approximate the solution of the problem with arbitrary precision. We also prove that the DNNs overcome the curse of dimensionality, a hard problem, specially because of the nonlocal character of the problem. However, some recent findings are key to fully describe the problem here. Indeed, Kyprianou et al. [25] showed that the Feynman-Kac formula and the **WoS processes** (to be described below) are also valid in the nonlocal case. We will deeply rely on these results to reproduce the Grohs and Herrmann program.

1.3. Setting: the Fractional Laplacian

Let $\alpha \in (0,2)$, $d \in \mathbb{N}$ and $D \subset \mathbb{R}^d$ a bounded domain. Consider the following Dirichlet boundary value problem

$$\begin{cases} (-\Delta)^{\alpha/2}u(x) = f(x) & x \in D, \\ u(x) = g(x) & x \in D^c. \end{cases} \tag{1.2}$$

Here, $(-\Delta)^{\alpha/2}$ is the fractional Laplacian defined in (1.1), and f, g are functions that satisfy suitable assumptions. Problem (1.2) has attracted considerable interest in past decades. Starting from the foundational work by Caffarelli and Silvestre [11], the study of fractional problems has always required a great amount of detail and very technical mathematics. The reader can consult the monographs by [4, 10, 20, 25, 28]. In addition, Machine Learning techniques proved numerically that there are efficient approximations for the solution of the fractional Problem, see, e.g. [21].

1.3.1. First hypotheses

In order to work with Problem (1.2), we will use the results of Kyprianou et al.. More precisely, we ask for the following conditions for the functions f, g:

- $g : D^c \to \mathbb{R}$ is a L_g-Lipschitz continuous function in $L^1_\alpha(D^c)$, $L_g > 0$, that is to say

$$\int_{D^c} \frac{|g(x)|}{1 + |x|^{d+\alpha}} dx < \infty. \tag{Hg-0}$$

- $f : D \to \mathbb{R}$ is a L_f-Lipschitz continuous function, $L_f > 0$, such that

$$f \in C^{\alpha+\varepsilon_0}(\overline{D}) \qquad \text{for some fixed} \qquad \varepsilon_0 > 0. \tag{Hf-0}$$

1.4. Stochastic representation and Montecarlo approximation

The previous assumptions are required to give a rigorous sense to the **continuous solution** in $L^1_\alpha(\mathbb{R}^d)$ of (1.2) in terms of the stochastic representation

$$u(x) = \mathbb{E}_x \left[g(X_{\sigma_D}) \right] + \mathbb{E}_x \left[\int_0^{\sigma_D} f(X_s) ds \right], \tag{1.3}$$

where $(X_t)_{t \geq 0}$ is an α-stable isotropic Lévy process and σ_D is the exit time of D for this process. See Theorem 5.1 in Chapter 5 below for full details. Representation (1.3) provides an approximation via Monte Carlo simulations. In fact, let $(X^i_t)_{t \geq 0}$ i.i.d. copies of $(X_t)_{t \geq 0}$ with their respective exit times σ^i_D. By strong law of large numbers one has

$$u(x) = \lim_{M \to \infty} \frac{1}{M} \sum_{i=1}^{M} g(X^i_{\sigma^i_D}) + \int_0^{\sigma^i_D} f(X^i_s) ds,$$

almost surely. In practice we will choose a M large enough to have an approximation of $u(x)$ with an error depending on the reciprocal of M. Although Monte Carlo simulations are effective approximations of the solution $u(x)$, these simulations can be numerically inefficient, because the approximation of $u(x)$ depends of each $x \in \mathbb{R}^d$. In addition, D is an arbitrary bounded domain, therefore it is not possible to determinate the cost of trajectories of the process X_t in the exit time σ_D, nor the quantity of trajectories M needed to have a approximation with small error.

1.5. Use of WoS processes in the fractional case

For the simulation of trajectories, Kyprianou et al. [25] propose the WoS processes. These discrete processes allow us to obtain the expected value of X_t in the exit time σ_D, without the necessity of simulate the entire trajectory. From $\rho_0 = x$, the WoS process $(\rho_n)_{n \in \mathbb{N}}$ is defined using specific points of the process X_t, for some t. The number of points that define the WoS process is described as a random variable N, which is finite almost surely. Even better, the distributions of ρ_N and X_{σ_D} are the same. For more detail, see Chapter 4.

The work of Kyprianou et al. proved that the solution of Problem (1.2) can be represented with the WoS processes, namely

$$u(x) = \mathbb{E}_x\left[g(\rho_N)\right] + \mathbb{E}_x\left[\sum_{n=1}^{N} r_n^{\alpha} V_1(0, f(\rho_{n-1} + r_n \cdot))\right], \tag{1.4}$$

where $(\rho_n)_{n=0}^{N}$ is the WoS process starting at $\rho_0 = x \in D$, $N = \min\{n \in \mathbb{N} : \rho_n \notin D\}$ and for any $n = 1, ..., N$, $r_n = \text{dist}(\rho_{n-1}, \partial D)$. $V_1(0, f(\cdot))$ represents the integral of $f(y)$ over the expected occupation measure of the stable process exiting the unitary ball centered at the origin $V_1(0, dy)$, for $|y| < 1$. For more details, see Chapter 5 and Lemma 5.1. For convenience we normalize the measure $V_1(0, dy)$ to obtain a probability measure over the ball $B(0, 1)$. This allow us to write equation 1.4 in function of expectations, and to approximate the solution using Monte Carlo simulations.

1.6. Main tools: Deep Neural Networks (DNN)

As said before, is a hard problem the approximation, via Monte Carlo simulations, of the solution over a domain D, because we need to do an independent simulation for each $x \in D$. This problem will be solved by the use of deep neural networks for the approximation. DNNs are sets of parameters:

$$\Phi = ((W_i, B_i)) \in \prod_{i=1}^{H+1} (\mathbb{R}^{k_i \times k_{i-1}} \times \mathbb{R}^{k_i}),$$

where $H \in \mathbb{N}$, $k_0, ..., k_{H+1} \in \mathbb{N}$. These parameters define a continuous function, called the realization of the DNN $\mathcal{R}(\Phi) : \mathbb{R}^{k_0} \to \mathbb{R}^{k_{H+1}}$, that satisfies

$$\left(\mathcal{R}(\Phi)\right)(x_0) = W_{H+1} x_H + B_{H+1},$$

where $x_i \in \mathbb{R}^{k_i}$, $i = 1, ..., H$ are defined as

$$x_i = A_{k_i}(W_i x_{i-1} + B_i).$$

Here the function $A_{k_i} : \mathbb{R}^{k_i} \to \mathbb{R}^{k_i}$ is the activation function of the DNN. Along this Thesis we will work with ReLu activation functions:

$$A_{k_i}(z) = (\max\{z_1, 0\}, ..., \max\{z_{k_i}, 0\}).$$

For the DNN Φ we define H as the number of hidden layers, $\mathbb{R}^{k_0}$ is the input layer, and $\mathbb{R}^{k_{H+1}}$ is the output layer. Figure 1.1 shows a scheme of the realization of Φ.

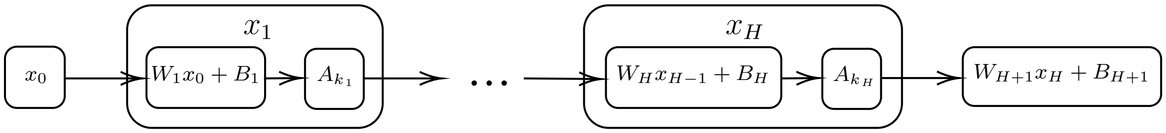

Figure 1.1: Scheme of the realization of a deep neural network Φ for $x_0 \in \mathbb{R}^{k_0}$. Each x_i lives in $\mathbb{R}^{k_i}$, and $(\mathcal{R}(\Phi))(x_0)$ lives in $\mathbb{R}^{k_{H+1}}$. The boxes that define x_i, $i = 1, ..., H$ are called hidden layers.

1.6.1. The curse of dimensionality

The number of parameters required to describe Φ is the sum of the entries of the matrices W_i and the vectors B_i. Now, if the realization of Φ approximates a function f with accuracy $\varepsilon > 0$, we say that the approximation overcomes the curse of dimensionality if the number of parameters that describe Φ are bounded by a polinomyal on the input dimension k_0 and on the reciprocal of the accuracy ε.

1.7. Results of this thesis

The results of this thesis are focused in the approximation via deep neural networks of the fractional Dirichlet Problem (1.2) with boundary conditions.

These results are part of the following work:

- N. V. *A new approach for the fractional Laplacian via deep neural networks*, preprint submitted 2022, available at https://arxiv.org/abs/2205.05229.

In order to enunciate the main theorem, we need to assume some hypothesis on the involved functions in equation (1.4). In particular, we assume that for all δ_g, δ_{dist}, δ_α, $\delta_f \in (0,1)$ there exists ReLu DNNs Φ_g, Φ_{dist}, Φ_α, Φ_f such that the functions $g, \text{dist}(\cdot, \partial D)$, $(\cdot)^\alpha$ and f are approximated properly with their respective accuracy, overcoming the curse of dimensionality.

The full details of these DNNs are in Assumptions 1, 2 in Chapter 6 and Assumption 3 in Chapter 7.

With Assumptions 1, 2 and 3 we enunciate the main theorem of this thesis.

Theorem 1.7.1 Let $\alpha \in (1,2)$, $p \in (1,\alpha)$ as in Assumption 1, $s \in (1,\alpha)$ such that $s < \frac{\alpha}{p}$ and $q \in \left[s, \frac{\alpha}{p}\right)$. Assume that (Hg-0) and (Hf-0) are satisfied. Suppose that for every $\delta_g, \delta_{\text{dist}}, \delta_\alpha, \delta_f \in (0,1)$ there exist ReLu DNNs Φ_g, $\Phi_{\text{dist}}, \Phi_\alpha$ and Φ_f satisfying Assumptions 1, 2 and 3, respectively.

Then for every $\epsilon \in (0,1)$, there exists a ReLu DNN Ψ_ϵ with continuous realization $\mathcal{R}(\Psi_\epsilon)$: $D \to \mathbb{R}$ such that:

1. Proximity in $L^q(D)$: If u is the solution of (1.2)

$$\left(\int_D |u(x) - (\mathcal{R}(\Psi_\epsilon))(x)|^q \, dx\right)^{\frac{1}{q}} \leq \epsilon. \tag{1.5}$$

2. Bounds: There exists $\widehat{B}, \eta > 0$ such that

$$\mathcal{P}(\Psi_\epsilon) \leq \widehat{B}|D|^\eta d^\eta \epsilon^{-\eta}. \tag{1.6}$$

The constant $\widehat{B}$ depends on $\|f\|_{L^\infty(D)}$, the Lipschitz constants of g, $\mathcal{R}(\Phi_f)$ and $\mathcal{R}(\Phi_\alpha)$, and on $\operatorname{diam}(D)$.

1.7.1. Idea of proof

Theorem 1.7.1 is consequence of Propositions 6.1 and 7.1, in Chapters 6 and 7, respectively, and his proof is available in Chapter 8. In this Section we sketch the proof of Proposition 6.1. The proof of Proposition 7.1 is pretty similar, with a few changes on the Monte Carlo simulations. Additionally, the function f is defined over a bounded domain, and then Assumption 3 is simpler than Assumption 1.

The idea of Proposition 6.1 is to approximate via deep neural networks the solution of Problem (1.2), with source term $f \equiv 0$, namely, when the solution takes the form.

$$u(x) = \mathbb{E}_x[g(\rho_N)].$$

The proof will be divided in several steps:

1.7.2. Step 1

First of all we define the following operator

$$E_M(x) = \frac{1}{M} \sum_{i=1}^{M} (\mathcal{R}(\Phi_g)) \left(\rho^i_{N_i}\right),$$

where $(\rho^i_{N_i}, N_i)$ are i.i.d. copies of (ρ_N, N). We prove that the norm of the difference between $u(x)$ and $E_M(x)$, in the space $L^q(\Omega, \mathbb{P}_x)$, is bounded. From Fubini we can integrate over the domain D, in order to prove that the quantity

$$\mathbb{E}_x \left[\int_D |u(x) - E_M(x)|^q dx\right],$$

is bounded.

1.7.3. Step 2

Note that the WoS process $\rho^i_{N_i}$ depends on the random variable N_i and on the sum of the absolute value of N_i copies of the isotropic α-stable process exiting the ball $B(0,1)$, denoted as $\sum_{n=1}^{N_i} |Y_{i,n}|$. Then we bound the norm of differences of $N_i, \sum_{n=1}^{N_i} |Y_{i,n}|$ and its respective Monte Carlo simulations, in the space $L^q(\Omega, \mathbb{P}_x)$. We then find a bound for the following quantities

$$\mathbb{E}_x \left[\left|N - \frac{1}{M}\sum_{i=1}^{M} N_i\right|^q\right], \qquad \text{and} \qquad \mathbb{E}_x \left[\left|\sum_{n=1}^{N} |Y_n| - \frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N_i} |Y_{i,n}|\right|^q\right].$$

Using these bounds, we prove that the sum of the three expectations are bounded by a positive constant, denoted by error_g^q. This implies that there exists independent copies of the random variable and the processes involved, such that the bound is also valid without the expectation. Notice that the copies found satisfy that their means:

$$\frac{1}{M}\sum_{i=1}^{M}\overline{N}_i, \qquad \frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|,$$

are bounded in function of error_g.

1.7.4. Step 3

Next Step is to prove that $E_M(x)$ can be approximated by a deep neural network. For this we prove that each copy $\rho^i_{\overline{N}_i}$ has an approach by DNN, namely $\Phi_{i,\overline{N}_i}$, such that the difference

$$|\rho^i_{\overline{N}_i} - (\mathcal{R}(\Phi_{i,\overline{N}_i}))(x)|,$$

is bounded properly in function of $\overline{N}_i$, $\sum_{n=1}^{\overline{N}_i}$ and δ_{dist}. This implies that $E_M(x)$ has an approach by DNN, namely $\Psi_{1,\varepsilon}$, that satisfies

$$\left(\mathcal{R}(\Psi_{1,\varepsilon})\right)(x) = \frac{1}{M}\sum_{i=1}^{M}\Big(\mathcal{R}(\Phi_g)\circ\mathcal{R}(\Phi_{i,\overline{N}})\Big)(x),$$

and the difference

$$|E_M(x) - \left(\mathcal{R}(\Psi_{1,\varepsilon})\right)(x)|,$$

is bounded if function of δ_g and the means of $\overline{N}_i$ and $\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|$. This bound implies that the norm of the difference between u and $\mathcal{R}(\Psi_{1,\varepsilon})$ in the space $L^q(D)$, is bounded depending on δ_g, error_g and on δ_{dist}. We prove additionally that error_g is bounded in function of M and δ_g.

1.7.5. Step 4

Finally, from the bounds obtained and the correct choice of δ_g, δ_{dist} and M, we prove that

$$\left(\int_D |u(x) - \left(\mathcal{R}(\Psi_{1,\varepsilon})\right)(x)|^q\,dx\right)^{\frac{1}{q}} \leq \varepsilon.$$

For the second point of Proposition 6.1, we study the DNN $\Psi_{1,\varepsilon}$. We show that de number of parameters that describe the DNN are given by operations between DNNs, such as compositions, sums and the fact that the identity function can be expressed as a DNN with arbitrary number of hidden layers. The choice of δ_g, δ_{dist} and M allow us to prove that the number of parameters that describes $\Psi_{1,\varepsilon}$ is at most a polynomial on the dimension d, on the measure of the space $|D|$ and on the reciprocal of the accuracy ε.

Chapter 2

Preliminaries

2.1. Notation

Along this paper, we shall use the following conventions:

- $\mathbb{N} = \{1, 2, 3, ...\}$ will be the set of Natural numbers.
- For any $q \geq 1$, $(\Omega, \mathcal{F}, \mu)$ measure space, $L^q(\Omega, \mu)$ denotes the Lebesgue space of order q with the measure μ. If μ is the Lebesgue measure, then the Lebesgue space will be denoted as $L^q(\Omega)$.

2.2. A quick review on Lévy processes

Let us introduce a brief review on the Lévy processes needed for the proof of the main results. For a detailed account on these processes, see e.g. [3, 8, 26, 29].

Definition 2.1 *$L := (L_t)_{t\geq 0}$ is a Lévy process in $\mathbb{R}^d$ if it satisfies $L_0 = 0$ and*

i) L has independent increments, namely, for all $n \in \mathbb{N}$ and for each $0 \leq t_1 < ... < t_n < \infty$, the random variables $(L_{t_2} - L_{t_1}, ..., L_{t_n} - L_{t_{n-1}})$ are independent.

ii) L has stationary increments, namely, for all $s \geq 0$, $L_{t+s} - L_s$ and L_t have the same law.

*iii) L_t is continuous on the right and has limit on the left for all $t > 0$ (i.e., $(L_t)_{t\geq 0}$ is **càdlàg**).*

Examples of Lévy processes are the Brownian motion, but also processes with jumps such as the *Poisson process* and the *compound Poisson process* [3].

Definition 2.2 *The Poisson process of intensity $\lambda > 0$ is a Lévy process N taking values in $\mathbb{N} \cup \{0\}$ wherein each $N(t) \sim Poisson(\lambda t)$, then we have*

$$\mathbb{P}(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Definition 2.3 *Let $(Z(n))_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables taking values in $\mathbb{R}^d$ with law μ and let N be a Poisson process with intensity λ that is independent of all the*

$Z(n)$. The compound Poisson process Y is defined as follows:

$$Y(t) = Z(1) + ... + Z(N(t)),$$

for each $t \geq 0$.

Another important element is the so-called Lévy's characteristic exponent.

Definition 2.4 *Let $(L_t)_{t\geq 0}$ be a Lévy process in $\mathbb{R}^d$.*

- *Its characteristic exponent $\Psi : \mathbb{R}^d \to \mathbb{C}$ is the continuous function that satisfies $\Psi(0) = 0$ and for all $t \geq 0$,*
$$\mathbb{E}\left[e^{i\xi\cdot L_t}\right] = e^{-t\Psi(\xi)}, \qquad \xi \in \mathbb{R}^d \setminus \{0\}. \tag{2.1}$$

- *A Lévy triple is (b, A, Π), where $b \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ is a positive semi-definite matrix, and Π is a Lévy measure in $\mathbb{R}^d$, i.e.*
$$\Pi(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} (1 \wedge |z|^2)\Pi(dz) < \infty. \tag{2.2}$$

A Lévy process is uniquely determined via its Lévy triple and its characteristic exponent.

Theorem 2.1 (Lévy-Khintchine, [26]) *Let (b, A, Π) be a Lévy triple. Define for each $\xi \in \mathbb{R}^d$*

$$\Psi(\xi) = ib \cdot \xi + \frac{1}{2}\xi \cdot A\xi + \int_{\mathbb{R}^d} \left(1 - e^{i\xi\cdot z} + i\xi \cdot z\mathbf{1}_{\{|z|<1\}}\right) \Pi(dz). \tag{2.3}$$

If Ψ is the characteristic exponent of a Lévy process with triple (b, A, Π) in the sense of (2.1)*, then it necessarily satisfies* (2.3)*. Conversely, given* (2.3) *there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, on which a Lévy process is defined having characteristic exponent Ψ in the sense of* (2.1).

For the next Theorem we define the *Poisson random measure.*

Definition 2.5 *Let $(S, \mathcal{S}, \eta)$ be an arbitrary σ-finite measure space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Let $N : \Omega \times \mathcal{S} \to \mathbb{N} \cup \{0, \infty\}$ such that $(N(\cdot, A))_{A\in\mathcal{S}}$ is a family of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. For convenience we supress the dependency of N on ω. N is called a Poisson random measure on S with intensity η if*

i) For mutually disjoint $A_1, ..., A_n$ in $\mathcal{S}$, the variables $N(A_1), ..., N(A_n)$ are independent.

ii) For each $A \in \mathcal{S}$, $N(A) \sim$ Poisson$(\eta(A))$,

iii) $N(\cdot)$ is a measure $\mathbb{P}$-almost surely.

Remark 2.2.1 In the next theorem we use $S \subset [0, \infty) \times \mathbb{R}^d$, and the intensity η will be defined on the product space.

From the Lévy-Khintchine formula, every Lévy process can be decomposed in three components: a Brownian part with drift, the large jumps and the compensated small jumps of the process L.

Theorem 2.2 (Lévy-Itô Decomposition) *Let L be a Lévy process with triple (b, A, Π). Then there exists process $L^{(1)}$, $L^{(2)}$ y $L^{(3)}$ such that for all $t \geq 0$*

$$L_t = L_t^{(1)} + L_t^{(2)} + L_t^{(3)},$$

where

1. $L_t^{(1)} = bt + AB_t$ *and* B_t *is a standard* d*-dimensional Brownian motion.*

2. $L_t^{(2)}$ *satisfies*

$$L_t^{(2)} = \int_0^t \int_{|z|\geq 1} zN(ds, dz),$$

where $N(ds, dz)$ *is a Poisson random measure on* $[0, \infty) \times \{z \in \mathbb{R}^d : |z| \geq 1\}$ *with intensity*

$$\Pi(\{z \in \mathbb{R}^d : |z| \geq 1\})dt \times \frac{\Pi(dz)}{\Pi(\{z \in \mathbb{R}^d : |z| \geq 1\})}.$$

If $\Pi(\{z \in \mathbb{R}^d : |z| \geq 1\}) = 0$*, then* $L^{(2)}$ *is the process identically equal to 0. In other words* $L^{(2)}$ *is a compound Poisson process.*

3. *The process* $L_t^{(3)}$ *satisfies*

$$L_t^{(3)} = \int_0^t \int_{|z|<1} z\widetilde{N}(ds, dz),$$

where $\widetilde{N}(ds, dz)$ *is the compensated Poisson random measure, defined by*

$$\widetilde{N}(ds, dz) = N(ds, dz) - ds\Pi(dz),$$

with $N(ds, dz)$ *the Poisson random measure on* $[0, \infty) \times \{z \in \mathbb{R}^d : |z| < 1\}$ *with intensity*

$$ds \times \Pi(dz)|_{\{z\in\mathbb{R}^d:|z|<1\}}.$$

2.3. Lévy processes and the Fractional Laplacian

A particular set of Lévy processes are the so-called isotropic α-stable processes, for $\alpha \in (0, 2)$. The following definitions can be found in [26] in full detail.

Definition 2.6 *Let* $\alpha \in (0, 2)$*.* $X := (X_t)_{t\geq 0}$ *is an isotropic* α*-stable process if* X *has a Lévy triple* $(0, 0, \Pi)$*, with*

$$\Pi(dz) = 2^{-\alpha}\pi^{-d/2}\frac{\Gamma((d+\alpha)/2)}{|\Gamma(-\alpha/2)|}\frac{1}{|z|^{\alpha+d}}dz, \quad z \in \mathbb{R}^d. \tag{2.4}$$

Recall that Γ *here is the Gamma function.*

Definition 2.7 (Equivalent definitions of an isotropic α-stable process)

1. X *is an isotropic* α*-stable process iff*

$$\text{for all} \quad c > 0, \quad (cX_{c^{-\alpha}t})_{t\geq 0} \quad \text{and} \quad X \quad \text{have the same law,} \tag{2.5}$$

and for all the orthogonal transformations U on $\mathbb{R}^d$,

$$(UX_t)_{t\geq 0} \quad \text{and} \quad X \quad \text{have the same law.}$$

2. *An isotropic α-stable process is an stable process whose characteristic exponent is given by*

$$\Psi(\xi) = |\xi|^\alpha, \qquad \xi \in \mathbb{R}^d. \tag{2.6}$$

Remark 2.3.1 For the first definition, we say that X satisfies the *scaling property* and is *rotationally invariant*, respectively.

Note by Definition 2.6 and by Lévy-Itô decomposition (Theorem 2.2),an isotropic α-stable process can be decomposed as

$$X_t = \int_0^t \int_{|z|\geq 1} zN(ds,dz) + \int_0^t \int_{|z|<1} z\widetilde{N}(ds,dz). \tag{2.7}$$

From this equation, we can conclude that every isotropic α-stable process is a pure jump process, whose jumps are determined by the Lévy measure defined in (2.4). We enunciate further properties about these processes:

Theorem 2.3 *Let g be a locally bounded, submultiplicative function and let L a Lévy process, then the following are equivalent:*

1. $\mathbb{E}[g(L_t)] < \infty$ *for some* $t > 0$.
2. $\mathbb{E}[g(L_t)] < \infty$ *for all* $t > 0$.
3. $\int_{|z|>1} g(z)\Pi(dz) < \infty$.

An important result from the previous theorem gives necessary and sufficient conditions for the existence of the p moment of an isotropic α-stable process.

Corollary 2.1 *Let X be an α-stable process and $p > 0$, then the following are equivalent*

1. $p < \alpha$.
2. $\mathbb{E}[|X_t|^p] < \infty$ *for some* $t > 0$.
3. $\mathbb{E}[|X_t|^p] < \infty$ *for all* $t > 0$.
4. $\int_{|z|>1} |z|^p\Pi(dz) < \infty$.

Remark 2.3.2 If $\alpha \in (0,1)$, by this corollary we have that X has no first moment. Otherwise, if $\alpha \in (1,2)$ then it has finite first moment, but no second moment.

2.4. Type s spaces and Monte Carlo Methods.

We now introduce some results that controls the difference between the expectation of a random variable and a Monte Carlo operator associated to his expectation in L^p norm, $p > 1$. For more details see [12]. In the following results are simplified the results of [12]. Along this

Section we work with real valued Banach spaces.

We start with some concepts related to Banach spaces. The reader can consult [12, 27] for more details in this topic.

Definition 2.8 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let J be a set, and let $r_j : \Omega \to \{-1, 1\}$, $j \in J$, be a family of independent random variables with for all $j \in J$,*

$$\mathbb{P}(r_j = 1) = \mathbb{P}(r_j = -1) = \frac{1}{2}.$$

Then we say that $(r_j)_{j \in J}$ is a $\mathbb{P}$-Rademacher family.

Definition 2.9 *Let $(r_j)_{j \in \mathbb{N}}$ a $\mathbb{P}$-Rademacher family. Let $s \in (0, \infty)$. A Banach space $(E, \|\cdot\|_E)$ is said to be of type s if there is a constant C such that for all finite sequences (x_j) in E,*

$$\mathbb{E}\left[\left\|\sum_j r_j x_j\right\|_E^s\right]^{\frac{1}{s}} \leq C \left(\sum_j \|x_j\|_E^s\right)^{\frac{1}{s}}.$$

The supremum of the constants C is called the type s-constant of E and it is denoted as $\mathscr{T}_s(E)$.

Remark 2.4.1 The existence of a finite constant C in Definition 2.9 is valid for $s \leq 2$ only (see, e.g. [27] Section 9 for more details).

Remark 2.4.2 Any Banach space $(E, \|\cdot\|_E)$ is of type 1. Moreover, triangle inequality ensures that $\mathscr{T}_1(E) = 1$.

Remark 2.4.3 Notice that for all Banach spaces $(E, \|\cdot\|_E)$, the function $(0, \infty) \ni s \to \mathscr{T}_s(E) \in [0, \infty]$ is non-decreasing. This implies for all $s \in (0, 1]$ and all Banach spaces $(E, \|\cdot\|_E)$ with $E \neq \{0\}$ that $\mathscr{T}_s(E) = 1$.

Remark 2.4.4 For all $s \in (0, 2]$ and all Hilbert spaces $(H, \langle\cdot,\cdot\rangle_H, \|\cdot\|_H)$ with $H \neq \{0\}$ it holds that $\mathscr{T}_s(H) = 1$.

Definition 2.10 *Let $(r_j)_{j \in \mathbb{N}}$ a $\mathbb{P}$-Rademacher family. Let $q, s \in (0, \infty)$ and $(E, \|\cdot\|_E)$ be a Banach space. The $\mathscr{K}_{q,s}$ (q, s)-Kahane-Khintchine constant of the space E is the extended real number given by the supremum of a constant C such that for all finite sequences (x_j) in E,*

$$\mathbb{E}\left[\left\|\sum_j r_j x_j\right\|_E^q\right]^{\frac{1}{q}} \leq C\, \mathbb{E}\left[\left\|\sum_j r_j x_j\right\|_E^s\right]^{\frac{1}{s}}$$

Remark 2.4.5 For all $q, s \in (0, \infty)$ it holds that $\mathscr{K}_{q,s} < \infty$. Moreover, if $q \leq s$ by Jensen's inequality implies that $\mathscr{K}_{q,s} = 1$.

Definition 2.11 *Let $q, s \in (0, \infty)$ and let $(E, \|\cdot\|_E)$ be a Banach space. Then we denote by*

$\Theta_{q,s}(E) \in [0,\infty]$ *the extended real number given by*

$$\Theta_{q,s}(E) = 2\mathscr{T}_s(E)\mathscr{K}_{q,s}.$$

Consider the case $(\mathbb{R}, |\cdot|)$. Notice that $(\mathbb{R}, \langle \cdot,\cdot \rangle_{\mathbb{R}}, |\cdot|)$ is a Hilbert space with the inner product $\langle x, y\rangle_{\mathbb{R}} = xy$. Then it holds for all $s \in (0,2]$ that

$$\mathscr{T}_s := \mathscr{T}_s(\mathbb{R}) = 1,$$

in other words, $(\mathbb{R}, |\cdot|)$ has type s for all $s \in (0,2]$. Moreover, it holds for all $q \in (0,\infty)$, $s \in (0,2]$ that

$$\Theta_{q,s} := \Theta_{q,s}(\mathbb{R}) = 2\mathscr{K}_{q,s} < \infty.$$

With this in mind, we enunciate a particular version of the Corollary 5.12 found in [12], replacing the Banach space $(E, \|\cdot\|_E)$ by $(\mathbb{R}, |\cdot|)$.

Corollary 2.2 *Let $M \in \mathbb{N}$, $s \in [1,2]$, $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\xi_j \in L^1(\mathbb{P}, |\cdot|)$, $j \in \{1, ..., M\}$, be independent and identically distributed. Then, for all $q \in [s,\infty]$,*

$$\begin{aligned} \left\| \mathbb{E}[\xi_1] - \frac{1}{M}\sum_{j=1}^{M}\xi_j \right\|_{L^q(\Omega,\mathbb{P})} &= \frac{1}{M}\mathbb{E}\left[\left|\sum_{j=1}^{M}\xi_j - \mathbb{E}\left[\sum_{j=1}^{M}\xi_j\right]\right|^q\right]^{\frac{1}{q}} \\ &\leq \frac{\Theta_{q,s}}{M^{1-\frac{1}{s}}}\mathbb{E}\left[|\xi_1 - \mathbb{E}[\xi_1]|^q\right]^{\frac{1}{q}}. \end{aligned} \tag{2.8}$$

Remark 2.4.6 The choice of the Banach space as $(\mathbb{R}, |\cdot|)$ ensures that $\Theta_{q,s}$ is finite and the bound above converges for suitable M large.

Chapter 3

Deep Neural Networks

In this Chapter we review recent results on the mathematical analysis of neural networks needed for the proof of the main theorem. For a detailed description, see e.g. [18, 23, 24].

3.1. Setting

For $d \in \mathbb{N}$ define

$$A_d : \mathbb{R}^d \to \mathbb{R}^d$$

the ReLU activation function such that for all $z \in \mathbb{R}^d$, $z = (z_1, ..., z_d)$, with

$$A_d(z) = (\max\{z_1, 0\}, ..., \max\{z_d, 0\}).$$

Let also

(NN1) $H \in \mathbb{N}$ be the number of hidden layers;

(NN2) $(k_i)_{i=0}^{H+1}$ be a positive integer sequence;

(NN3) $W_i \in \mathbb{R}^{k_i \times k_{i-1}}$, $B_i \in \mathbb{R}^{k_i}$, for any $i = 1, ..., H+1$ be the weights and biases, respectively;

(NN4) $x_0 \in \mathbb{R}^{k_0}$, and for $i = 1, ..., H$ let

$$x_i = A_{k_i}(W_i x_{i-1} + B_i). \tag{3.1}$$

We call

$$\Phi := (W_i, B_i)_{i=1}^{H+1} \in \prod_{i=1}^{H+1} \left(\mathbb{R}^{k_i \times k_{i-1}} \times \mathbb{R}^{k_i}\right) \tag{3.2}$$

the DNN associated to the parameters in (NN1)-(NN4). The space of all DNNs in the sense of (3.2) is going to be denoted by $\mathbf{N}$, namely

$$\mathbf{N} = \bigcup_{H \in \mathbb{N}} \bigcup_{(k_0, ..., k_{H+1}) \in \mathbb{N}^{H+2}} \left[\prod_{i=1}^{H+1} \left(\mathbb{R}^{k_i \times k_{i-1}} \times \mathbb{R}^{k_i}\right)\right].$$

Define the realization of the DNN $\Phi \in \mathbf{N}$ as

$$\mathcal{R}(\Phi)(x_0) = W_{H+1} x_H + B_{H+1}. \tag{3.3}$$

Notice that $\mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}})$. For any $\Phi \in \mathbf{N}$ define

$$\mathcal{P}(\Phi) = \sum_{n=1}^{H+1} k_n(k_{n-1}+1), \qquad \mathcal{D}(\Phi) = (k_0, k_1, ..., k_{H+1}), \tag{3.4}$$

and

$$|||\mathcal{D}(\Phi)||| = \max\{k_0, k_1, ..., k_{H+1}\}. \tag{3.5}$$

The entries of $(W_i, B_i)_{i=1}^{H+1}$ will be the weights of the DNN, $\mathcal{P}(\Phi)$ represents the number of parameters used to describe the DNN, working always with fully connected DNNs, and $\mathcal{D}(\Phi)$ representes the dimension of each layer of the DNN. Notice that $\Phi \in \mathbf{N}$ has $H+2$ layers: H of them hidden, one input and one output layer.

Remark 3.1.1 For $\Phi \in \mathbf{N}$ one has

$$|||\mathcal{D}(\Phi)||| \leq \mathcal{P}(\Phi) \leq (H+1)|||\mathcal{D}(\Phi)|||(|||\mathcal{D}(\Phi)|||+1).$$

Indeed, from the definition of $|||\cdot|||$,

$$|||\mathcal{D}(\Phi)||| \leq \sum_{n=1}^{H+1} k_n \leq \mathcal{P}(\Phi).$$

In addition, the definition of $\mathcal{P}(\Phi)$ implies that

$$\mathcal{P}(\Phi) \leq \sum_{n=1}^{H+1} |||\mathcal{D}(\Phi)|||(|||\mathcal{D}(\Phi)|||+1) = (H+1)|||\mathcal{D}(\Phi)|||(|||\mathcal{D}(\Phi)|||+1).$$

Remark 3.1.2 From the previous remark one has

$$\mathcal{P}(\Phi) \leq 2(H+1)|||\mathcal{D}(\Phi)|||.$$

If $|||\mathcal{D}(\Phi)|||$ grows at most polynomially in both the dimension of the input layer and the reciprocal of the accuracy ε of the DNN, then $\mathcal{P}(\Phi)$ satisfies that bound too. This means that, with the right bound on $|||\mathcal{D}(\Phi)|||$, the DNN Φ do not suffer the curse of dimensionality, in the sense established in Section 1.

3.2. Operations

In this section we summarize that some operations between DNNs are also DNNs. We start with the definition of two vector operators

Definition 3.1 *Let* $\boldsymbol{D} = \bigcup_{H\in\mathbb{N}} \mathbb{N}^{H+2}$.

1. *Define* $\odot : \boldsymbol{D} \times \boldsymbol{D} \to \boldsymbol{D}$ *such that for all* $H_1, H_2 \in \mathbb{N}$, $\alpha = (\alpha_0, ..., \alpha_{H_1+1}) \in \mathbb{N}^{H_1+2}$, $\beta = (\beta_0, ..., \beta_{H_2+1}) \in \mathbb{N}^{H_2+2}$ *it satisfied*

$$\alpha \odot \beta = (\beta_0, \beta_1, ..., \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, ..., \alpha_{H_1+1}) \in \mathbb{N}^{H_1+H_2+3}. \tag{3.6}$$

2. *Define* $\boxplus : \boldsymbol{D} \times \boldsymbol{D} \to \boldsymbol{D}$ *such that for all* $H \in \mathbb{N}$, $\alpha = (\alpha_0, ..., \alpha_{H+1}) \in \mathbb{N}^{H+2}$, $\beta = (\beta_0, ..., \beta_{H+1}) \in \mathbb{N}^{H+2}$ *it satisfied*

$$\alpha \boxplus \beta = (\alpha_0, \alpha_1 + \beta_1, ..., \alpha_H + \beta_H, \beta_{H+1}) \in \mathbb{N}^{H+2}. \tag{3.7}$$

3. *Define* $\mathfrak{n}_n \in \boldsymbol{D}$, $n \in \mathbb{N}$, $n \geq 3$ *as*

$$\mathfrak{n}_n = (1, \underbrace{2, \ldots, 2}_{(n-2)\text{-}times}, 1) \in \mathbb{N}^n. \tag{3.8}$$

Remark 3.2.1 From these definitions and the norm $|||\cdot|||$ defined in (3.5), we the following bounds are clear

1. For $H_1, H_2 \in \mathbb{N}$, $\alpha \in \mathbb{N}^{H_1+2}$ and $\beta \in \mathbb{N}^{H_2+2}$,

$$|||\alpha \odot \beta||| \leq \max\{|||\alpha|||, |||\beta|||, \alpha_0 + \beta_{H_2+1}\}.$$

2. For $H \in \mathbb{N}$ and $\alpha, \beta \in \mathbb{N}^{H+2}$,

$$|||\alpha \boxplus \beta||| \leq |||\alpha||| + |||\beta|||.$$

3. For $n \in \mathbb{N}$, $n \geq 3$, $|||\mathfrak{n}_n||| = 2$.

Now we state classical operations between DNNs. For a full details of the next Lemmas, the reader can consult e.g. [23, 24].

Lemma 3.1 *Let* $Id_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ *be the identity function on* $\mathbb{R}$ *and let* $H \in \mathbb{N}$. *Then* $Id_{\mathbb{R}} \in \mathcal{R}\left(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \mathfrak{n}_{H+2}\}\right)$.

Remark 3.2.2 A similar consequence is valid in $\mathbb{R}^d$. Let $Id_{\mathbb{R}^d} : \mathbb{R}^d \to \mathbb{R}^d$ be the identity function on $\mathbb{R}^d$ and let $H \in \mathbb{N}$. Therefore $Id_{\mathbb{R}^d} \in \mathcal{R}\left(\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = d\mathfrak{n}_{H+2}\}\right)$. The case with $H = 3$ is proved on [24].

Remark 3.2.3 Let $H \in \mathbb{N}$ and $\Phi \in \mathbf{N}$ such that $\mathcal{R}(\Phi) = Id_{\mathbb{R}^d}$. Then by Remark 3.3 we have that $|||\mathcal{D}(\Phi)||| = 2d$.

Lemma 3.2 *Let* $d_1, d_2, d_3 \in \mathbb{N}$, $f \in C(\mathbb{R}^{d_2}, \mathbb{R}^{d_3})$, $g \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$, $\alpha, \beta \in \boldsymbol{D}$ *such that* $f \in \mathcal{R}(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \alpha\})$ *and* $g \in \mathcal{R}(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \beta\})$. *Therefore* $(f \circ g) \in \mathcal{R}(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \alpha \odot \beta\})$.

Remark 3.2.4 Let $\Phi_f, \Phi_g, \Phi \in \mathbb{N}$ such that $\mathcal{R}(\Phi_f) = f$, $\mathcal{R}(\Phi_g) = g$ and $\mathcal{R}(\Phi) = f \circ g$. Then by Remark 3.3 it follows that

$$|||\mathcal{D}(\Phi)||| \leq \max\{|||\mathcal{D}(\Phi_f)|||, |||\mathcal{D}(\Phi_g)|||, 2d_2\}.$$

Lemma 3.3 *Let $M, H, p, q \in \mathbb{N}$, $h_i \in \mathbb{R}$, $\beta_i \in \boldsymbol{D}$, $f_i \in C(\mathbb{R}^p, \mathbb{R}^q)$, $i = 1, ..., M$ such that for all $i = 1, ..., M$ $\dim(\beta_i) = H + 2$ and $f_i \in \mathcal{R}(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \beta_i\})$. Then*

$$\sum_{i=1}^{M} h_i f_i \in \mathcal{R}\left(\left\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \boxplus_{i=1}^{M} \beta_i\right\}\right). \tag{3.9}$$

Remark 3.2.5 For $i = 1, ..., M$ let $\Phi_i \in \mathbf{N}$ such that $\mathcal{R}(\Phi_i) = f_i$ and let $\Phi \in \mathbf{N}$ such that

$$\mathcal{R}(\Phi) = \sum_{i=1}^{M} h_i f_i.$$

It follows from Remark 3.3 that

$$|||\mathcal{D}(\Phi)||| \leq \sum_{i=1}^{M} |||\mathcal{D}(\Phi_i)|||.$$

The following Lemma comes from [14] and is adapted to our notation.

Lemma 3.4 *Let $H, d, d_i \in \mathbb{N}$, $\beta_i \in \boldsymbol{D}$, $f_i \in C(\mathbb{R}^d, \mathbb{R}^{d_i})$, $i = 1, 2$ such that for $i = 1, 2$ $\dim(\beta_i) = H + 2$ and $f_i \in \mathcal{R}(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \beta_i\})$. Then*

$$(f_1, f_2) \in \mathcal{R}\left(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = (d, \beta_{1,1} + \beta_{2,1}, ..., \beta_{1,H+1} + \beta_{2,H+1})\}\right). \tag{3.10}$$

Remark 3.2.6 Let $\Phi_1, \Phi_2, \Phi \in \mathbf{N}$ such that $\mathcal{R}(\Phi_i) = f_i$, $i = 1, 2$ and $\mathcal{R}(\Phi) = (f_1, f_2)$. Notice by Lemma 3.4 and definition of the norm $|||\cdot|||$ in (3.5) that

$$|||\mathcal{D}(\Phi)||| \leq |||\mathcal{D}(\Phi_1)||| + |||\mathcal{D}(\Phi_2)|||.$$

For sake of completeness, we state the following lemma with his proof. We continue with the notation from [23]:

Lemma 3.5 *Let $H, p, q, r \in \mathbb{N}$, $M \in \mathbb{R}^{r \times q}$, $\alpha \in \boldsymbol{D}$, $f \in C(\mathbb{R}^p, \mathbb{R}^q)$, such that $dim(\alpha) = H + 2$ and $f \in \mathcal{R}(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = \alpha\})$. Then*

$$Mf \in \mathcal{R}\left(\{\Phi \in \boldsymbol{N} : \mathcal{D}(\Phi) = (\alpha_0, ..., \alpha_H, r)\}\right). \tag{3.11}$$

PROOF. Let $H, \alpha_0, ..., \alpha_{H+1} \in \mathbb{N}$, $\Phi_f \in \mathbf{N}$ satisfying that

$$\alpha = (\alpha_0, ..., \alpha_{H+1}), \qquad \mathcal{R}(\Phi_f) = f, \qquad \text{and} \qquad \mathcal{D}(\Phi_f) = \alpha.$$

Note that $p = \alpha_0$ and $q = \alpha_{H+1}$. Let $((W_1, B_1), ..., (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} (\mathbb{R}^{\alpha_n \times \alpha_{n-1}} \times \mathbb{R}^{\alpha_n})$ satisfty that

$$\Phi_f = ((W_1, B_1), ..., (W_{H+1}, B_{H+1})).$$

Let $M \in \mathbb{R}^{r \times \alpha_{H+1}}$ and define

$$\Phi = ((W_1, B_1), ..., (W_H, B_H), (MW_{H+1}, MB_{H+1})).$$

Notice that $(MW_{H+1}, MB_{H+1}) \in \mathbb{R}^{r \times \alpha_H} \times \mathbb{R}^r$, then $\Phi \in \mathbb{N}$. For $y_0 \in \mathbb{R}^{\alpha_0}$, and y_i, $i = 1, ..., H$

defined as in (NN4) we have

$$(\mathcal{R}(\Phi))(y_0) = MW_{H+1}y_H + MB_{H+1} = M(W_{H+1}y_H + B_{H+1}) = M(\mathcal{R}(\Phi_f))(y_0).$$

Therefore

$$\mathcal{R}(\Phi) = Mf, \qquad \text{and} \qquad \mathcal{D}(\Phi) = (\alpha_0, ..., \alpha_H, r),$$

and the Lemma is proved. $\square$

Remark 3.2.7 Let $\Phi_f, \Phi \in \mathbf{N}$ such that $\mathcal{R}(\Phi_f) = f$ and $\mathcal{R}(\Phi) = Mf$. From previous Lemma and the definition of $|||\cdot|||$ it follows that

$$|||\mathcal{D}(\Phi)||| \leq \max\{|||\mathcal{D}(\Phi_f)|||, r\}.$$

The following Lemma is from [17].

Lemma 3.6 *There exists constants $C_1, C_2, C_3, C_4 > 0$ such that for all $\kappa > 0$ and for all $\delta \in \left(0, \frac{1}{2}\right)$ there exists a ReLu DNN $\Upsilon \in \boldsymbol{N}$, with $\mathcal{R}(\Upsilon) \in C(\mathbb{R}^2, \mathbb{R})$ such that*

$$\sup_{a,b \in [-\kappa,\kappa]} |ab - (\mathcal{R}(\Upsilon))(a,b)| \leq \delta. \tag{3.12}$$

Moreover, for all $\delta \in \left(0, \frac{1}{2}\right)$,

$$\mathcal{P}(\Upsilon) \leq C_1 \left(\log_2\left(\frac{\max\{\kappa, 1\}}{\delta}\right)\right) + C_2, \tag{3.13}$$

$$\dim(\mathcal{D}(\Upsilon)) \leq C_3 \left(\log_2\left(\frac{\max\{\kappa, 1\}}{\delta}\right)\right) + C_4. \tag{3.14}$$

Chapter 4

Walk-on-spheres Processes

We start with some key notation that will be extensively used along this paper.

Let $(\Omega, \mathbb{P}, \mathcal{F})$ be a filtered probability space with $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$. Let $(X_t)_{t\geq 0}$ be an isotropic α-stable process starting at X_0. For $x \in D$ denote $\mathbb{P}_x$ the probability measure conditional to $X_0 = x$ and $\mathbb{E}_x$ the respective expectation. Finally, define for any $B \subset \mathbb{R}^d$ the exit time for the set B as

$$\sigma_B = \inf\{t \geq 0 : X_t \notin B\}.$$

Now we introduce the classical WoS process.

Definition 4.1 ([25]) *The Walk-on-Spheres (WoS) process $\rho := (\rho_n)_{n\in\mathbb{N}}$ is defined as follows:*

- $\rho_0 = x$, $x \in D$;
- *given ρ_{n-1}, $n \geq 1$, the distribution of ρ_n is chosen according to an independent sample of $X_{\sigma_{B_n}}$ under $\mathbb{P}_{\rho_{n-1}}$, where B_n is the ball centered on ρ_{n-1} and radius $r_n = \mathrm{dist}(\rho_{n-1}, \partial D)$.*

Remark 4.0.1 Notice by the Markov property that the process ρ can be written as the recurrence

$$\rho_n = \rho_{n-1} + Z_n, \quad n \in \mathbb{N},$$

where Z_n is an independent sample of $X_{\sigma_{B(0,r_n)}}$ under $\mathbb{P}_0$.

From the previous Remark, it is possible to rewrite ρ_n for $n \in \mathbb{N}$ depending on $x \in D$ and on n independent processes distributing accord $X_{\sigma_{B(0,1)}}$, as indicates the following Lemma.

Lemma 4.0.1 The WoS process $\rho := (\rho_n)_{n\in\mathbb{N}}$ can be defined as follows

- $\rho_0 = x$, $x \in D$;
- for $n \geq 1$,

$$\rho_n = \rho_{n-1} + r_n Y_n, \tag{4.1}$$

  where Y_n is an independent sample of $X_{\sigma_{B(0,1)}}$ and $r_n = \mathrm{dist}(\rho_{n-1}, \partial D)$.

PROOF. Note by the scaling property (2.5) that

$$X_t \qquad \text{and} \qquad r_n X_{r_n^{-\alpha} t}$$

have the same distribution for all $n \in \mathbb{N}$. Therefore

$$\begin{aligned}\sigma_{B(0,r_n)} &= \inf\{t \geq 0 : X_t \notin B(0,r_n)\} \\ &= r_n^{\alpha} \inf\{r_n^{-\alpha} t \geq 0 : r_n X_{r_n^{-\alpha} t} \notin B(0,r_n)\} \\ &= r_n^{\alpha} \inf\{s \geq 0 : r_n X_s \notin B(0,r_n)\} \\ &= r_n^{\alpha} \inf\{s \geq 0 : X_s \notin B(0,1)\} = r_n^{\alpha} \sigma_{B(0,1)}.\end{aligned} \tag{4.2}$$

This equality and the scaling property implies that

$$X_{\sigma_{B(0,1)}} \qquad \text{and} \qquad r_n^{-1} X_{r_n^{\alpha} \sigma_{B(0,1)}} \tag{4.3}$$

are equal in law under $\mathbb{P}_0$, and then from Remark 4.1 Z_n and $r_n X_{\sigma_{B(0,1)}}$ have the same distribution under $\mathbb{P}_0$. We can conclude that for $n \geq 1$, ρ_n can be written as the recurrence

$$\rho_n = \rho_{n-1} + r_n Y_n,$$

where Y_n is an independent sample of $X_{\sigma_{B(0,1)}}$. □

To study the WoS processes, we need to know about the processes $X_{\sigma_{B(0,1)}}$. The following result gives the distribution density of $X_{\sigma_{B(0,1)}}$.

Theorem 4.1 (Blumenthal, Getoor, Ray, 1961. [9]) *Suppose that $B(0,1)$ is a unit ball centered at the origin and write $\sigma_{B(0,1)} = \inf\{t > 0 : X_t \notin B(0,1)\}$. Then,*

$$\mathbb{P}_0\left(X_{\sigma_{B(0,1)}} \in dy\right) = \pi^{-(d/2+1)} \Gamma\left(\frac{d}{2}\right) \sin(\pi\alpha/2) \left|1 - |y|^2\right|^{-\alpha/2} |y|^{-d} dy, \qquad |y| > 1. \tag{4.4}$$

Using this result, one can prove a key result for the expectation of $X_{\sigma_{B(0,1)}}$ moments.

Corollary 4.1 *For all $\alpha \in (0,2)$, $\beta \in [0,\alpha)$ we have*

$$\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^{\beta}\right] = \frac{\sin(\pi\alpha/2)}{\pi} \frac{\Gamma\left(1 - \frac{\alpha}{2}\right)\Gamma\left(\frac{\alpha-\beta}{2}\right)}{\Gamma\left(1 - \frac{\beta}{2}\right)} =: K(\alpha,\beta). \tag{4.5}$$

Remark 4.0.2 Notice that the value of $K(\alpha,\beta)$ does not depend of the dimension d.

Remark 4.0.3 the condition $\beta < \alpha$ is necessary due to Corollary 2.1. If $\beta \geq \alpha$ then $\mathbb{E}[|X_t|^{\beta}] = \infty$ for all $t > 0$. Moreover, the integral

$$\int_1^{\infty} \frac{r^{\beta-1}}{(r^2-1)^{\alpha/2}} dr,$$

obtained in the proof of the Corollary 4.1 does not converges if $\beta \geq \alpha$.

Proof of Corollary 4.1. Let $\alpha \in (0,2)$, $\beta \in [0,\alpha)$. Notice by Theorem 4.1 and definition of the expectation that

$$\begin{aligned}\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^{\beta}\right] &= \int_{|y|>1}|y|^{\beta}\mathbb{P}_0\left(X_{\sigma_{B(0,1)}}\in dy\right)\\ &= \pi^{-(d/2+1)}\Gamma\left(\frac{d}{2}\right)\sin(\pi\alpha/2)\int_{|y|>1}\left|1-|y|^2\right|^{-\alpha/2}|y|^{\beta-d}dy.\end{aligned} \tag{4.6}$$

Using spherical coordinates one has

$$\int_{|y|>1}\left|1-|y|^2\right|^{-\alpha/2}|y|^{\beta-d}dy = \int_{\mathbb{S}^{d-1}}\int_1^{\infty}\left|1-r^2\right|^{-\alpha/2}r^{\beta-d}r^{d-1}drdS,$$

where $\mathbb{S}^{d-1}$ is the surface area of the unit $(d-1)$-sphere embedded in dimension d. One has that [31]

$$\left|\mathbb{S}^{d-1}\right| = \frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)},$$

and then

$$\int_{|y|>1}\left|1-|y|^2\right|^{-\alpha/2}|y|^{\beta-d}dy = \frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)}\int_1^{\infty}\frac{r^{\beta-1}}{(r^2-1)^{\alpha/2}}dr.$$

Replacing this result into (4.6) give us that

$$\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^{\beta}\right] = \frac{2}{\pi}\sin(\pi\alpha/2)\int_1^{\infty}\frac{r^{\beta-1}}{(r^2-1)^{\alpha/2}}dr.$$

Now we are able to use a change of variables $u=1/r$, then

$$\int_1^{\infty}\frac{r^{\beta-1}}{(r^2-1)^{\alpha/2}}dr = \int_0^1\frac{1}{u^2}u^{1-\beta}\frac{u^{\alpha}}{(1-u^2)^{-\alpha/2}}du = \int_0^1 u^{\alpha-\beta-1}(1-u^2)^{-\alpha/2}du.$$

Using another change of variable, $t=u^2$ we have

$$\int_0^1\frac{1}{2t^{\frac{1}{2}}}t^{\frac{\alpha-\beta-1}{2}}(1-t)^{-\frac{\alpha}{2}}dt = \frac{1}{2}\int_0^1 t^{\frac{\alpha-\beta}{2}-1}(1-t)^{1-\frac{\alpha}{2}-1}dt.$$

This result implies that

$$\mathbb{E}_x\left[\left|X_{\sigma_{B(0,1)}}\right|^{\beta}\right] = \frac{\sin(\pi\alpha/2)}{\pi}\int_0^1 t^{\frac{\alpha-\beta}{2}-1}(1-t)^{1-\frac{\alpha}{2}-1}dt.$$

The integral has the form of the *Beta function*, formally defined as:

$$B(z,w) := \int_0^1 u^{z-1}(1-u)^{w-1}du,$$

For full details of the Beta function see, e.g. [13]. In particular, the Beta function satisfies

$$B(z,w) = \frac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)}.$$

Finally

$$\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^{\beta}\right] = \frac{\sin(\pi\alpha/2)}{\pi} B\left(\frac{\alpha-\beta}{2}, 1-\frac{\alpha}{2}\right) = \frac{\sin(\pi\alpha/2)}{\pi} \frac{\Gamma\left(1-\frac{\alpha}{2}\right)\Gamma\left(\frac{\alpha-\beta}{2}\right)}{\Gamma\left(1-\frac{\beta}{2}\right)}.$$

□

4.1. Relation between WoS and isotropic α-stable process

Recall from Chapter 4 that the process $(\rho_n)_{n\geq 0}$ is related to a family of processes distributing accord $X_{\sigma_{B(0,1)}}$. We want now to have a relation between the processes $(X_t)_{t\geq 0}$ and $(\rho_n)_{n\geq 0}$. For this define $\widetilde{r}_1 := \text{dist}(x, \partial D)$, $\widetilde{B}_1 := B(x, r_1)$, $\tau_1 := \sigma_{\widetilde{B}_1}$ and for all $n \geq 1$ define:

$$\widetilde{r}_{n+1} := \text{dist}(X_{\mathcal{I}(n)}, \partial D), \tag{4.7}$$

$$\widetilde{B}_{n+1} := B\left(X_{\mathcal{I}(n)}, \widetilde{r}_{n+1}\right), \tag{4.8}$$

$$\tau_{n+1} := \inf\{t \geq 0 : X_{t+\mathcal{I}(n)} \notin \widetilde{B}_{n+1}\}, \tag{4.9}$$

where

$$\mathcal{I}(n) := \sum_{i=1}^{n} \tau_i, \qquad \mathcal{I}(0) = 0. \tag{4.10}$$

$\mathcal{I}(n)$ represents the total time of the process X_t takes to exit the n balls $\widetilde{B}_1, ..., \widetilde{B}_n$. The following Lemma establishes that for all $n \in \mathbb{N}$, ρ_n is equally distributed to the process $(X_t)_{t\geq 0}$ exiting the n balls $\widetilde{B}_1, ..., \widetilde{B}_n$, that is, $X_{\mathcal{I}(n)}$.

Lemma 4.1.1 For all $n \geq 0$ and $x \in D$, ρ_n and $X_{\mathcal{I}(n)}$ have the same distribution starting at x.

PROOF. Note that under $\mathbb{P}_{X_{\mathcal{I}(n-1)}}$, $X_{\mathcal{I}(n)}$ has the same distribution as $X_{\sigma_{\widetilde{B}_n}}$. Thus, by the Markov property and the scaling property one has

$$X_{\mathcal{I}(n)} = X_{\mathcal{I}(n-1)} + \widetilde{r}_n Y'_n,$$

where Y'_n is an independent sample of $X_{\sigma_{B(0,1)}}$ under $\mathbb{P}_0$. From the two constructions below in addition with induction, one has that ρ_n and $X_{\mathcal{I}(n)}$ has the same distribution starting at x, for all $n \geq 0$ and $x \in D$. □

Let

$$N = \min\{n \in \mathbb{N} : \rho_n \notin D\}. \tag{4.11}$$

This random variable describes the quantity of balls $\widetilde{B}_n$ that the process $(X_t)_{t\geq 0}$ exits before exits the domain D. The following Theorem ensures that N is almost surely finite.

Theorem 4.2 ([25], Theorem 5.4) *Let D be a open and bounded set. Therefore for all $x \in D$, there exists a constant $\widetilde{p} = \widetilde{p}(\alpha, d) > 0$ independent of x and D, and a random variable Γ*

such that $N \leq \Gamma$ $\mathbb{P}_x$-a.s., where

$$\mathbb{P}_x(\Gamma = k) = (1-\widetilde{p})^{k-1}\widetilde{p}, \quad k \in \mathbb{N}. \tag{4.12}$$

Remark 4.1.1 Although the random variable Γ has the same distribution for each $x \in D$, it is not the same random variable for each $x \in D$.

Remark 4.1.2 This theorem implies that

$$\mathbb{P}_x(N > n) \leq \mathbb{P}_x(\Gamma > n) = (1-\widetilde{p})^n, \qquad n \in \mathbb{N}.$$

The definition of $\mathcal{I}(n)$ and N in (4.10) and (4.11) imply that the total time of $(X_t)_{t\geq 0}$ that takes to exit N balls $\widetilde{B}_1, ..., \widetilde{B}_N$ is equal to the time of $(X_t)_{t\geq 0}$ that takes to exit D. More precisely

Lemma 4.1.2 For $x \in D$, let X_t be an isotropic α-stable process. Therefore, a.s.

$$\mathcal{I}(N) = \sigma_D.$$

PROOF. For the inequality $\geq$, note by definition of N that

$$X_{\mathcal{I}(N)} \notin D.$$

Recall that σ_D is the infimum time $t \geq 0$ such that $X_t \notin D$, then

$$\mathcal{I}(N) \geq \sigma_D.$$

For $\leq$ suppose by contradiction that $\sigma_D < \mathcal{I}(N)$. If $\sigma_D < \mathcal{I}(N-1)$, then

$$X_{\mathcal{I}(N-1)} \notin D.$$

This is a contradiction with the definition of N, because $N-1$ is a natural less than N satisfying the above condition. Therefore $\mathcal{I}(N-1) \leq \sigma_D$ and this implies that there exists $t^* \geq 0$ such that

$$\mathcal{I}(N) > \sigma_D = \mathcal{I}(N-1) + t^*,$$

Using the definition of $\mathcal{I}(n)$, for $n \in \mathbb{N}$ and the supposition $\sigma_D < \mathcal{I}(N)$, one has

$$t^* < \tau_N,$$

but

$$X_{\sigma_D} = X_{\mathcal{I}(N-1)+t^*} \notin D,$$

therefore, from the definition of τ_N in (4.9),

$$\tau_N \leq t^*,$$

a contradiction. Therefore $\mathcal{I}(N-1) \leq \sigma_D$ and we can conclude that

$$\mathcal{I}(N) = \sigma_D.$$

□

Remark 4.1.3 From the relation between $X_{\mathcal{I}(n)}$ and ρ_n for $n \in \mathbb{N}$, it follows that

$$\mathbb{E}_x\left[\rho_N\right] = \mathbb{E}_x\left[X_{\mathcal{I}(N)}\right] = \mathbb{E}_x\left[X_{\sigma_D}\right].$$

Remark 4.1.3 give us a relation between $(X_t)_{t\geq 0}$ and $(\rho_n)_{n\geq 0}$. Figure **??** shows in an example the relation between WoS and isotropic α-stable processes, exiting a bounded domain D.

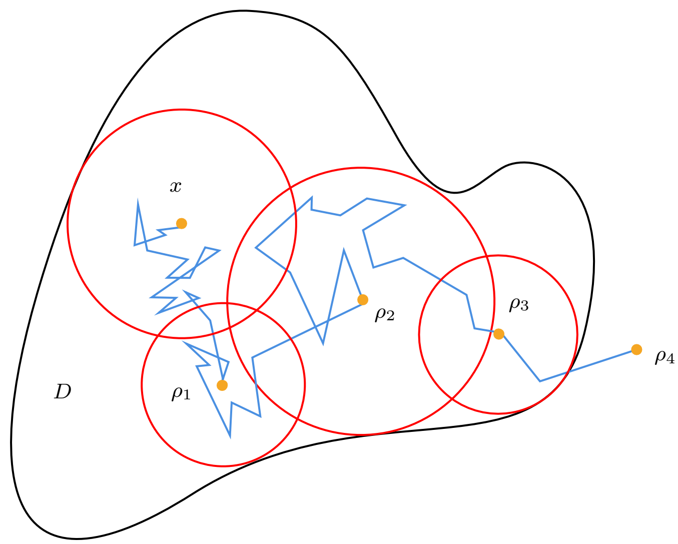

Figure 4.1: Illustration of isotropic α-stable and WoS processes starting at x exiting a domain D. The blue line represents the α-stable process $(X_t)_{t\geq 0}$, the orange dots are the WoS process $(\rho_n)_{n\geq 0}$ and the red balls are given by Definition 4.1. In this case $N = 4$ and $\rho_4 = X_{\sigma_D}$.

Chapter 5

Stochastic representation of the Fractional Laplacian

5.1. Stochastic Representation

Recall Problem (1.2). The following theorem gives an stochastic representation of the solution of problem (1.2) from the process $(X_t)_{t\geq 0}$. The proof of this Theorem can be found in [25]

Theorem 5.1 ([25], Theorem 6.1) *Let $d \geq 2$ and assume that D is a bounded domain in $\mathbb{R}^d$. Additionally, assume* (Hg-0) *and* (Hf-0). *Then there exist a* **unique continuous solution** *for* (1.2) *in $L^1_\alpha(\mathbb{R}^d)$, given by the explicit formula*

$$u(x) = \mathbb{E}_x \left[g(X_{\sigma_D}) \right] + \mathbb{E}_x \left[\int_0^{\sigma_D} f(X_s) ds \right], \tag{5.1}$$

valid for any $x \in D$.

The previous representation can be expressed in terms of the WoS process. For this define the expected occupation measure of the stable process prior to exiting a ball of radius $r > 0$ centered in $x \in \mathbb{R}^d$ as follows:

$$V_r(x, dy) := \int_0^\infty \mathbb{P}_x \left(X_t \in dy, t < \sigma_{B(x,r)} \right) dt, \qquad x \in \mathbb{R}^d, \quad |y| < 1, \quad r > 0. \tag{5.2}$$

We have the following result for $V_1(0, dy)$.

Theorem 5.2 ([25], Theorem 6.2) *The measure $V_1(0, dy)$ is given for $|y| < 1$, by*

$$V_1(0, dy) = 2^{-\alpha} \pi^{-d/2} \frac{\Gamma(d/2)}{\Gamma(\alpha/2)^2} |y|^{\alpha - d} \left(\int_0^{|y|^{-2}-1} (u+1)^{-d/2} u^{\alpha/2-1} du \right) dy. \tag{5.3}$$

Denote $V_r(x, f(\cdot)) = \int_{|y-x|<r} f(y) V_r(x, dy)$ for a bounded measurable function f. $V_r(x, f(\cdot))$ defines the expected value of f under the measure $V_r(x, dy)$ over the ball $B(x, r)$. An important property of this expected value is the following: for $r > 0$ and $x \in \mathbb{R}^d$

$$V_r(x, f(\cdot)) = V_1(0, f(x + r\cdot)). \tag{5.4}$$

The proof of this property can be found in [25]. The following Lemma uses the WoS process and the Theorem 5.2. Recall that $(\rho_n)_{n=1,\ldots,N}$ represents the WoS process defined in Chapter 4, and $r_n = \mathrm{dist}(\rho_n, \partial D)$.

Lemma 5.1 ([25], Lema 6.3) *For $x \in D$, $g \in L^1_\alpha(D^c)$ and $f \in C^{\alpha+\varepsilon_0}(\overline{D})$ we have the representation*

$$u(x) = \mathbb{E}_x\left[g(\rho_N)\right] + \mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha V_1(0, f(\rho_{n-1} + r_n \cdot))\right]. \tag{5.5}$$

Remark 5.1.1 Recall that ρ_n and $X_{\mathcal{I}(n)}$ are equal on law under $\mathbb{P}_x$ for all $n \in \mathbb{N}$. Therefore we can write

$$u(x) = \mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] + \mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha V_1\left(0, f\left(X_{\mathcal{I}(n-1)} + r_n \cdot\right)\right)\right]. \tag{5.6}$$

5.2. Equivalent representations of non-homogeneous solution

Consider again the problem (1.2). Remember from Remark 5.1.1 that its solution can be written as

$$u(x) = \mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] + \mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha V_1(0, f(X_{\mathcal{I}(n-1)} + r_n \cdot))\right].$$

Notice also that from the definition of $V_1(0, f(\cdot))$, it can be expressed as the expectation of f under the measure $V_1(0, dy)$ on $B(0,1)$. This measure is not necessarily a probability measure, so we are going to normalize the measure $V_1(0, dy)$. For this define for all $d \geq 2$, $d \in \mathbb{N}$ and $\alpha \in (0,2)$

$$\kappa_{d,\alpha} = \int_{B(0,1)} V_1(0, dy).$$

In the following Lemma we prove that $\kappa_{d,\alpha}$ is positive and finite

Lemma 5.2.1 For all $d \geq 2$ and $\alpha \in (0,2)$, we have that $0 < \kappa_{d,\alpha} < +\infty$.

PROOF. Notice first from Theorem 4.1 that

$$\kappa_{d,\alpha} = \widetilde{c}_{d,\alpha} \int_{B(0,1)} |y|^{\alpha-d} \left(\int_0^{|y|^{-2}-1} (u+1)^{-d/2} u^{\alpha/2-1} du\right) dy,$$

where $\widetilde{c}_{d,\alpha} = 2^{-\alpha}\pi^{-d/2}\dfrac{\Gamma(d/2)}{\Gamma(\alpha/2)^2}$. Now we work with the interior integral. With a change of variables $u = \dfrac{1-t}{t}$ and integral properties one has:

$$\begin{aligned}
&\int_1^{|y|^2} t^{d/2}\left(\frac{1-t}{t}\right)^{\alpha/2-1}(-t^{-2})dt \\
&\quad= \int_0^1 t^{d/2-\alpha/2-1}(1-t)^{\alpha/2-1}dt - \int_0^{|y|^2} t^{d/2-\alpha/2-1}(1-t)^{\alpha/2-1}dt.
\end{aligned}$$

For $z, w > 0$, $x \in [0,1]$ let $B(z,w)$ and $I(x;z,w)$ be the Beta and the Incomplete Beta functions respectively, defined as

$$B(z,w) := \int_0^1 u^{z-1}(1-u)^{w-1}du,$$
$$I(x;z,w) := \frac{1}{B(z,w)} \int_0^x u^{z-1}(1-u)^{w-1}du.$$

For further details of these functions the reader can consult [13]. Notice that $\kappa_{d,\alpha}$ can be written in terms of $B(z,w)$ and $I(x;z,w)$. Indeed

$$\kappa_{d,\alpha} = \widetilde{c}_{d,\alpha} B\left(\frac{d}{2} - \frac{\alpha}{2}, \frac{\alpha}{2}\right) \int_{B(0,1)} |y|^{\alpha-d} \left(1 - I\left(|y|^2; \frac{d}{2} - \frac{\alpha}{2}, \frac{\alpha}{2}\right)\right) dy.$$

Note by property of Beta function that

$$B\left(\frac{d}{2} - \frac{\alpha}{2}, \frac{\alpha}{2}\right) = \frac{\Gamma\left(\frac{d}{2} - \frac{\alpha}{2}\right)\Gamma\left(\frac{\alpha}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}.$$

The Gamma function is well defined and positive on $(0, \infty)$. If $d > \alpha$ then

$$0 < B\left(\frac{d}{2} - \frac{\alpha}{2}, \frac{\alpha}{2}\right) < +\infty.$$

On the other hand side, note by the definition of $I(x;z,w)$ that for $x < 1$,

$$0 \leq I(x;z,w) < \frac{1}{B(z,w)} \int_0^1 u^{z-1}(1-u)^{w-1}du = 1,$$

Then for all $|y|^2 < 1$,

$$0 < 1 - I\left(|y|^2; \frac{d}{2} - \frac{\alpha}{2}, \frac{\alpha}{2}\right) \leq 1.$$

Therefore in $\kappa_{d,\alpha}$ we are integrating the multiplication of two positive functions over a set of positive measure. This implies that

$$0 < \kappa_{d,\alpha} \leq \widetilde{c}_{d,\alpha} B\left(\frac{d}{2} - \frac{\alpha}{2}, \frac{\alpha}{2}\right) \int_{B(0,1)} |y|^{\alpha-d} dy.$$

The above integral can be calculated using a change of variables in spherical coordinates, and his value is finite. Finally we conclude that

$$0 < \kappa_{d,\alpha} < +\infty.$$

□

Now we are able to define a probability measure μ on $B(0,1)$ given by

$$\mu(dy) := \kappa_{d,\alpha}^{-1} V_1(0, dy).$$

Therefore, for any bounded measurable function f we have

$$\begin{aligned} V_1(0, f(X_{\mathcal{I}(n-1)} + r_n \cdot)) &= \int_{B(0,1)} f(X_{\mathcal{I}(n-1)} + r_n y) V_1(0, dy) \\ &= \kappa_{d,\alpha} \int_{B(0,1)} f(X_{\mathcal{I}(n-1)} + r_n y) \mu(dy) \\ &= \kappa_{d,\alpha} \mathbb{E}^{(\mu)} \left[f(X_{\mathcal{I}(n-1)} + r_n \cdot) \right]. \end{aligned} \tag{5.7}$$

where $\mathbb{E}^{(\mu)}$ correspond to the expectation over the probability measure μ on $B(0,1)$. With this representation, we can rewrite the solution of (1.2) as

$$u(x) = \mathbb{E}_x \left[g\left(X_{\mathcal{I}(N)}\right) \right] + \mathbb{E}_x \left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} \mathbb{E}^{(\mu)} \left[f\left(X_{\mathcal{I}(n-1)} + r_n \cdot\right) \right] \right]. \tag{5.8}$$

From the construction of $\kappa_{d,\alpha}$, the following properties are valid

Lemma 5.2.2 One has

1.
$$\mathbb{E}_x \left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} \right] = \mathbb{E}_x \left[\sigma_D \right],$$

2.
$$\mathbb{E}_x \left[\left| \sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} \right|^2 \right] \leq \mathbb{E}_x \left[\sigma_D^2 \right].$$

PROOF.

1. Notice from the definition of $V_1(0, f(\cdot))$, with $f \equiv 1$ that

$$\kappa_{d,\alpha} = V_1\left(0, 1\left(X_{\mathcal{I}(n-1)} + r_n \cdot\right)\right).$$

It follows from (5.4) that

$$r_n^\alpha \kappa_{d,\alpha} = V_{r_n}\left(X_{\mathcal{I}(n-1)}, 1(\cdot)\right).$$

Moreover

$$\begin{aligned} \mathbb{E}_x \left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} \right] &= \mathbb{E}_x \left[\sum_{n=1}^{N} V_{r_n}\left(X_{\mathcal{I}(n-1)}, 1(\cdot)\right) \right] \\ &= \mathbb{E}_x \left[\int_0^{\sigma_D} 1(X_s) ds \right] = \mathbb{E}_x \left[\sigma_D \right]. \end{aligned}$$

2. From the definition of $V_r(x, f(\cdot))$ with $f \equiv 1$, it follows that

$$V_{r_n}\left(X_{\mathcal{I}(n-1)}, 1(\cdot)\right) = \mathbb{E}_{X_{\mathcal{I}(n-1)}} \left[\int_0^{\sigma_{B(X_{\mathcal{I}(n-1)}, r_n)}} 1(X_t) dt \right] = \mathbb{E}_{X_{\mathcal{I}(n-1)}} \left[\sigma_{B(X_{\mathcal{I}(n-1)}, r_n)} \right].$$

By definition of τ_n, $n \in \mathbb{N}$, (4.9) and Markov property one has

$$\mathbb{E}_{X_{\mathcal{I}(n-1)}} \left[\sigma_{B(X_{\mathcal{I}(n-1)}, r_n)} \right] = \mathbb{E}_0 \left[\tau_n \right].$$

Then, Jensen inequality implies

$$\mathbb{E}_x\left[\left|\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\right|^2\right] = \mathbb{E}_x\left[\left|\sum_{n=1}^{N} \mathbb{E}_0\left[\tau_n\right]\right|^2\right] \leq \mathbb{E}_x\left[\mathbb{E}_0\left[\left|\sum_{n=1}^{N} \tau_n\right|^2\right]\right].$$

Finally, by tower property

$$\mathbb{E}_x\left[\left|\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\right|^2\right] \leq \mathbb{E}_x\left[\mathcal{I}(N)^2\right] = \mathbb{E}_x\left[\sigma_D^2\right].$$

□

Chapter 6

Approximation of solutions of the Fractional Dirichlet problem using DNNs: the boundary data case

As usual, Problem (1.2) can be decomposed in two subproblems, that will be treated in a separate way. In this Chapter we first deal with the homogeneous case.

6.1. Homogeneous Fractional Laplacian

We consider (1.2) with $f \equiv 0$, namely,

$$\begin{cases} (-\Delta)^{\alpha/2}u(x) &= 0 \qquad \text{for } x \in D, \\ u(x) &= g(x) \quad \text{for } x \in D^c. \end{cases} \tag{6.1}$$

Note that under (Hg-0), one has from (5.1)

$$u(x) = \mathbb{E}_x\left[g(X_{\sigma_D})\right], \qquad x \in D. \tag{6.2}$$

The main idea of this section is approximate the solution (6.2) by a deep neural network with ReLu activation, with an accurateness $\varepsilon > 0$. For this we going to assume that g can be approximated by a ReLu DNN satisfying several hypotheses. These hypotheses are expressed in the following assumption.

Recall that $|||\cdot|||$ represents the maximum number of hidden layers dimensions introduced in (3.5), $\mathcal{R}$ is the realization of a DNN as in (3.3), and $\mathcal{D}$ was introduced in (3.4).

Assumptions 1 Let $d \geq 2$. Let $g : D^c \to \mathbb{R}$ satisfying (Hg-0). Let $\delta_g \in (0,1)$, $a, b \geq 1$, $p \in (1, \alpha)$ and $B > 0$. Then there exists a *ReLu DNN* $\Phi_g \in \mathbf{N}$ with

1. $\mathcal{R}(\Phi_g) : D^c \to \mathbb{R}$ is continuous, and

2. The following are satisfied:

$$|g(y) - (\mathcal{R}(\Phi_g))(y)| \leq \delta_g B d^p (1+|y|)^p, \quad \forall y \in D^c. \tag{Hg-1}$$
$$|(\mathcal{R}(\Phi_g))(y)| \leq B d^p (1+|y|)^p, \quad \forall y \in D^c. \tag{Hg-2}$$
$$|||\mathcal{D}(\Phi_g)||| \leq B d^b \delta_g^{-a}, \tag{Hg-3}$$

Remark 6.1.1 We use the hypotheses presented in [23] for the approximation of function defined over non bounded sets.

In addition to the previous assumptions, we will require *structural properties* related to the domain D itself.

Assumptions 2 Let $\alpha \in (1,2)$, $a, b \geq 1$ and $B > 0$. Suppose that D bounded domain enjoys the following structure.

1. For any $\delta_{\text{dist}} \in (0,1)$, the function $x \mapsto \text{dist}(x, \partial D)$ can be approximated by a ReLu DNN $\Phi_{\text{dist}} \in \mathbf{N}$ such that

$$\sup_{x \in D} |\text{dist}(x, \partial D) - (\mathcal{R}(\Phi_{\text{dist}}))(x)| \leq \delta_{\text{dist}}, \tag{HD-1}$$

and

$$|||\mathcal{D}(\Phi_{\text{dist}})||| \leq B d^b \lceil \log\left(\delta_{\text{dist}}^{-1}\right) \rceil^a. \tag{HD-2}$$

2. For all $\delta_\alpha \in (0,1)$ there exists a ReLu DNN $\Phi_\alpha \in \mathbf{N}$ such that

$$\sup_{x \in [0, \text{diam}(D)]} |(\mathcal{R}(\Phi_\alpha))(x) - x^\alpha| \leq \delta_\alpha. \tag{HD-3}$$

and

$$|||\mathcal{D}(\Phi_\alpha)||| \leq B d^b \delta_\alpha^{-a}. \tag{HD-4}$$

Moreover, $\mathcal{R}(\Phi_\alpha)$ is a L_α-Lipschitz function, $L_\alpha > 0$, for $|x| \leq \text{diam}(D)$.

Remark 6.1.2 Notice that Assumption (HD-3) is assured by Hornik's Theorem [22]. Also, (HD-2) may seem too demanding because of the log term, but actually this is the situation in the case of a ball, see [17].

In the next proposition prove the existence of a ReLu DNN such that the Dirichlet problem without source is well approximated.

Proposition 6.1 *Let $\alpha \in (1,2)$, $L_g > 0$ and*

$$p, s \in (1, \alpha) \text{ such that } s < \tfrac{\alpha}{p} \quad \text{and} \quad q \in \left[s, \tfrac{\alpha}{p}\right). \tag{6.3}$$

Suppose that the function g satisfies (Hg-0) *and Assumptions 1. Suppose additionally that D satisfies Assumptions 2. Then for all $\varepsilon \in (0,1)$ there exists a ReLu DNN $\Psi_{1,\varepsilon}$ that satisfies*

1. *Proximity in $L^q(D)$:*

$$\left(\int_D \left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] - (\mathcal{R}(\Psi_{1,\varepsilon}))(x)\right|^q dx\right)^{\frac{1}{q}} \leq \varepsilon. \tag{6.4}$$

2. *Realization:* $\mathcal{R}(\Psi_{1,\varepsilon})$ *has the following structure: there exists* $M \in \mathbb{N}$, $\overline{N}_i \in \mathbb{N}$, $Y_{i,n}$ *i.i.d. copies of* $X_{\sigma_{B(0,1)}}$ *under* $\mathbb{P}_0$, *for* $i = 1, ..., M$, $n = 1, ..., \overline{N}_i$ *such that for all* $x \in D$,

$$\mathcal{R}(\Psi_{1,\varepsilon})(x) = \frac{1}{M} \sum_{i=1}^{M} \left(\mathcal{R}(\Phi_g) \circ \mathcal{R}(\Phi^i_{\overline{N}_i}) \right)(x) , \tag{6.5}$$

where $\Phi^i_{\overline{N}_i}$ *is a ReLu DNN approximating* $X^i_{\mathcal{I}(\overline{N}_i)} = X^i_{\mathcal{I}(\overline{N}_i)}(x, Y_{i,1}, ..., Y_{i,\overline{N}_i})$.

3. *Bounds: There exists* $\widetilde{B} > 0$ *such that*

$$|||\mathcal{D}(\Psi_{1,\varepsilon})||| \leq \widetilde{B}|D|^{\frac{1}{q}\left(2a+ap+\frac{s}{s-1}(1+2a+ap)\right)} d^{b+2ap+2ap^2+\frac{ps}{s-1}(1+2a+ap)} \varepsilon^{-a-\frac{s}{s-1}(1+2a+ap)}. \tag{6.6}$$

Remark 6.1.3 The hypotheses (6.3) are non empty if $\alpha \in (1, 2)$. This requirement is standard in the literature devoted to the Fractional Laplacian, where some proofs are highly dependent on the cases $\alpha \in (0, 1]$ versus $\alpha \in (1, 2)$.

Remark 6.1.4 The condition $\alpha \in (1, 2)$ is very important. In particular, if $\alpha \in (0, 1)$ the hypotheses (6.3) are empty, and moreover, processes that we are working with not necessarily have finite expectation and there are not guarantee on the convergence of the ReLu DNNs.

6.2. Proof of Proposition 6.1: existence

The proof will be divided in several steps. As explained before, we follow the ideas in [17], with several changes due to the nonlocal character of the treated equation.

Let s, p and q be as in (6.3).

Step 1. Preliminaries. Let $(\rho_n)_{n\in\mathbb{N}}$ be a WoS process introduced in Definition 4.1 starting from $x \in D$. Let also $\mathcal{I}(N)$ and N be defined as in (4.10) and (4.11). Recall that from (5.1),

$$u(x) = \mathbb{E}_x[g(\rho_N)] = \mathbb{E}_x[g(X_{\mathcal{I}(N)})]. \tag{6.7}$$

From the construction of $X_{\mathcal{I}(N)}$, one has that $X_{\mathcal{I}(N)} \in D^c$ and it depends on N i.i.d. copies of $X_{\sigma_{B(0,1)}}$, namely

$$X_{\mathcal{I}(N)} = X_{\mathcal{I}(N)}(x, Y_1, ..., Y_N), \tag{6.8}$$

where each Y_n, $n = 1, \ldots, N$, is an independent copy of $X_{\sigma_{B(0,1)}}$ under $\mathbb{P}_0$.

Let $M \in \mathbb{N}$. Consider M copies of $X_{\mathcal{I}(N)}$ starting from $x \in D$, as described in (6.8). We denote such copies as

$$X^i_{\mathcal{I}(N_i)} = X^i_{\mathcal{I}(N_i)}(x, Y_{i,1}, ..., Y_{i,N_i}), \tag{6.9}$$

with $Y_{i,n}$, $i = 1, ..., M$, $n = 1, ..., N_i$ i.i.d. copies of $X_{\sigma_{B(0,1)}}$ under $\mathbb{P}_0$, and where each N_i is an i.i.d. copy of N. Notice that for each copy, N_i can be different (as a random variable).

With this in mind, and following [17], we introduce the *Monte Carlo operator*

$$E_M(x) := \frac{1}{M}\sum_{i=1}^{M} (\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(N_i)}\right), \tag{6.10}$$

where $\mathcal{R}(\Phi_g)$ denotes the realization as a continuous function of the DNN $\Phi_g \in \mathbf{N}$ that approximates g in Assumption 1. Notice that $E_M(x)$ may not be a DNN in the general case.

Our main objective in the following steps is to obtain suitable bounds on the difference between the expectation of $g(X_{\mathcal{I}(N)})$ and $E_M(x)$, in a certain sense to be determined. Step 2 controls the difference between $\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right]$ and the intermediate term $\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right]$. Notice that this last term is not necessarily a DNN, because of the quantity $X_{\mathcal{I}(N)}$.

Step 2. Define

$$J_1 := \left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] - \mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right]\right|.$$

Notice that by Jensen inequality and hypothesis (Hg-1) one has

$$J_1 \leq \mathbb{E}_x\left[\left|g\left(X_{\mathcal{I}(N)}\right) - (\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right|\right] \leq Bd^p\delta_g\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p\right].$$

Recall that we have an expression for $\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^\beta\right]$ with $\beta < \alpha$ from Corollary 4.1. The idea is to find a bound for $\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p\right]$ in terms of (4.5).

Let $R > 1$ be large enough to have $D \subset B^* := B(x,R)$. The right hand side of the previous inequality is going to be separated in two terms: the case where $\sigma_D = \sigma_{B^*}$, and otherwise. Notice that $\sigma_D > \sigma_{B^*}$ is not possible. We obtain:

$$J_1 \leq Bd^p\delta_g\Big(\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p \mathbf{1}_{\{\sigma_D=\sigma_{B^*}\}}\right] + \mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p \mathbf{1}_{\{\sigma_D<\sigma_{B^*}\}}\right]\Big).$$

In the case of the equality, the processes $X_{\mathcal{I}(N)}$ and $X_{\sigma_{B^*}}$ are equal on law under $\mathbb{P}_x$ from Lemma 4.1.2 and Remark 4.1.3. Then the Markov property and the scaling property of the process (see (4.2) and (4.3)) can be used to get

$$\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p \mathbf{1}_{\{\sigma_D=\sigma_{B^*}\}}\right] = \mathbb{E}_0\left[\left(1+\left|x+RX_{\sigma_{B(0,1)}}\right|\right)^p\right].$$

On the other hand note that if $\sigma_D < \sigma_{B^*}$ then $X_{\mathcal{I}(N)} \in B^* \setminus D$. Therefore

$$\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p \mathbf{1}_{\{\sigma_D<\sigma_{B^*}\}}\right] \leq \sup_{y\in B^*\setminus D}(1+|y|)^p.$$

We conclude

$$J_1 \leq Bd^p\delta_g\left(\mathbb{E}_0\left[\left(1+\left|x+RX_{\sigma_{B(0,1)}}\right|\right)^p\right] + \sup_{y\in B^*\setminus D}(1+|y|)^p\right).$$

Using the Minkowski inequality and the fact that the sets D and $B^* \setminus D$ are bounded, one has

$$J_1 \leq Bd^p\delta_g \left(\left(1+|x|+R\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^p\right]^{\frac{1}{p}}\right)^p + \sup_{y\in B^*\setminus D}(1+|y|)^p \right)$$
$$\leq Bd^p\delta_g \left(\left(K_1+R\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^p\right]^{\frac{1}{p}}\right)^p + K_2^p \right),$$

where K_1 and K_2 are constants such that for all $x \in D$, $y \in B^* \setminus D$

$$1+|x| \leq K_1 \quad \text{and} \quad 1+|y| \leq K_2. \tag{6.11}$$

By Corollaries 2.1 and 4.1 one has

$$\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^p\right]^{\frac{1}{p}} = K(\alpha,p)^{\frac{1}{p}} < \infty \quad \iff \quad p<\alpha.$$

Therefore, from the choice of p, one has that J_1 is finite and bounded as follows:

$$J_1 \leq Bd^p\delta_g \left(\left(K_1 + RK(\alpha,p)^{\frac{1}{p}}\right)^p + K_2^p \right), \tag{6.12}$$

with $K(\alpha,p) < +\infty$ defined in (4.5).

Step 3. In this step we control the difference between the intermediate term $\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right]$ previously introduced in Step 2, and the Monte Carlo $E_M(x)$ (6.10). Define

$$J_2 := \left\|\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right] - E_M(x)\right\|_{L^q(\Omega,\mathbb{P}_x)}.$$

In order to bound this term, we are going to use Corollary 2.2. First of all notice from (Hg-2) that

$$\mathbb{E}_x\left[\left|(\mathcal{R}(\Phi_g))(X_{\mathcal{I}(N)})\right|\right] < Bd^p\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p\right].$$

Note by Step 2 that

$$\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^p\right] \leq \left(K_1 + RK(\alpha,p)^{\frac{1}{p}}\right)^p + K_2^p < +\infty,$$

where K_1 and K_2 are defined in (6.11). Therefore one can conclude that

$$\mathbb{E}_x\left[\left|(\mathcal{R}(\Phi_g))(X_{\mathcal{I}(N)})\right|\right] < \infty.$$

Then for all $i \in \{1,...,M\}$, $(\mathcal{R}(\Phi_g))(X^i_{\mathcal{I}(N_i)}) \in L^1(\Omega,\mathbb{P}_x)$. For s as in (6.3), Corollary 2.2 ensures that for all $q \in [s,\infty)$ (and in particular for all q as in (6.3)), one has

$$J_2 \leq \frac{\Theta_{q,s}}{M^{1-\frac{1}{s}}}\left\|\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right] - (\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right\|_{L^q(\Omega,\mathbb{P}_x)}. \tag{6.13}$$

Now we bound the norm on the right hand side of (6.13). By Minkowski's inequality one has

$$
\begin{aligned}
&\left\|\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right]-(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right\|_{L^q(\Omega,\mathbb{P}_x)}\\
&\quad\leq\left\|\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right]\right\|_{L^q(\Omega,\mathbb{P}_x)}+\left\|(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right\|_{L^q(\Omega,\mathbb{P}_x)}\\
&\quad\leq 2\mathbb{E}_x\left[\left|(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right|^q\right]^{\frac{1}{q}}.
\end{aligned}
$$

Now using hypothesis (Hg-2) and the same results in previous Step to obtain

$$
\begin{aligned}
&\left\|\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right]-(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right\|_{L^q(\Omega,\mathbb{P}_x)}\\
&\quad\leq 2Bd^p\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^{pq}\right]^{\frac{1}{q}}\\
&\quad\leq 2Bd^p\left(\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^{pq}\mathbf{1}_{\{\sigma_D=\sigma_{B^*}\}}\right]^{\frac{1}{q}}+\mathbb{E}_x\left[\left(1+\left|X_{\mathcal{I}(N)}\right|\right)^{pq}\mathbf{1}_{\{\sigma_D<\sigma_{B^*}\}}\right]^{\frac{1}{q}}\right),
\end{aligned}
$$

where we recall that B^* is a ball in $\mathbb{R}^d$ centered in x with radious $R>1$ large enough such that $D\subset B^*$. Then using the scaling property of X and Minkowski inequality, we have

$$
\begin{aligned}
&\left\|\mathbb{E}_x\left[(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right]-(\mathcal{R}(\Phi_g))\left(X_{\mathcal{I}(N)}\right)\right\|_{L^q(\Omega,\mathbb{P}_x)}\\
&\quad\leq 2Bd^p\left(\mathbb{E}_0\left[\left(1+\left|x+RX_{\sigma_{B(0,1)}}\right|\right)^{pq}\right]^{\frac{1}{q}}+\sup_{y\in B^*\setminus D}(1+|y|)^{pq}\right)\\
&\quad\leq 2Bd^p\left(\left(K_1+R\mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^{pq}\right]^{\frac{1}{pq}}\right)^p+K_2^p\right).
\end{aligned}
\tag{6.14}
$$

Therefore, by (6.13), (6.14) and Corollary 4.1 we have that J_2 is finite and bounded as follows:

$$
J_2\leq\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}}Bd^p\left(\left(K_1+RK(\alpha,pq)^{\frac{1}{pq}}\right)^p+K_2^p\right).
\tag{6.15}
$$

Step 4. Thanks to Steps 2 and 3 now it is possible to bound the difference

$$
\left\|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right]-E_M(x)\right\|_{L^q(\Omega,\mathbb{P}_x)}.
$$

Indeed, first notice from Jensen inequality in (4.5) that for $1<q<\frac{\alpha}{p}$ (see (6.3)),

$$
K(\alpha,p)^{\frac{1}{p}}\leq K(\alpha,pq)^{\frac{1}{pq}}<+\infty.
$$

Condition $q<\frac{\alpha}{p}$ is necessary in order to have $K(\alpha,pq)$ finite (see Corollary (4.1)). It follows from (6.12), (6.15) and Minkowski's inequality that

$$
\begin{aligned}
&\left\|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right]-E_M(x)\right\|_{L^q(\Omega,\mathbb{P}_x)}\leq J_1+J_2\\
&\quad\leq\left(\delta_g+\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}}\right)Bd^p\left(\left(K_1+RK(\alpha,pq)^{\frac{1}{pq}}\right)^p+K_2^p\right).
\end{aligned}
\tag{6.16}
$$

Define

$$C := \left(\left(K_1 + RK(\alpha, pq)^{\frac{1}{pq}} \right)^p + K_2^p \right) < \infty. \tag{6.17}$$

Note that the choice of R depends on the starting point x in order to have $D \subset B(x, R)$. If we choose e.g. $R = 2\,\mathrm{diam}(D)$, it follows that for all $x \in D$, $D \subset B(x, 2\,\mathrm{diam}(D))$ and then C is uniform w.r.t. $x \in D$. Fubini and (6.16) implies that

$$\mathbb{E}_x \left[\int_D \left| \mathbb{E}_x \left[g\left(X_{\mathcal{I}(N)} \right) \right] - E_M(x) \right|^q dx \right] \leq \left(\delta_g + \frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}} \right)^q |D| B^q d^{pq} C^q. \tag{6.18}$$

In the following steps we are going to control two quantities that help us to obtain bounds for the random variables N_i and $|Y_{i,n}|$, for all $i = 1, ..., M$, $n = 1, ..., N_i$. Although similar to the steps followed in [17], here we need additional estimates because of the non continuous nature of the Lévy jump processes.

Step 5. In order to bound the following expectation

$$\mathbb{E}_x \left[\left| \mathbb{E}_x[N] - \frac{1}{M} \sum_{i=1}^{M} N_i \right|^q \right],$$

we are going to use Corollary 2.2. Notice by Theorem 4.2 that for all $x \in D$ there exists a geometric random variable Γ with parameter $\widetilde{p} = \widetilde{p}(\alpha, d) > 0$ such that

$$\mathbb{E}_x \left[|N| \right] \leq \mathbb{E}_x \left[\Gamma \right] = \frac{1}{\widetilde{p}} < \infty,$$

and then for all $i \in \{1, ..., M\}$, $N_i \in L^1(\Omega, \mathbb{P}_x)$. For s as in (6.3), Corollary 2.2 implies for all q as in (6.3) that

$$\mathbb{E}_x \left[\left| \mathbb{E}_x[N] - \frac{1}{M} \sum_{i=1}^{M} N_i \right|^q \right] \leq \left(\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}} \right)^q \mathbb{E}_x \left[|N|^q \right] \leq \left(\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}} \right)^q \mathbb{E}_x \left[|N|^2 \right], \tag{6.19}$$

where we used that $q < 2$ and then $\mathbb{E}_x \left[| \cdot |^q \right] \leq \mathbb{E}_x \left[| \cdot |^2 \right]$. Recall that

$$\mathbb{E}_x \left[|N|^2 \right] \leq \mathbb{E}_x \left[\Gamma^2 \right] = \frac{2 - \widetilde{p}}{\widetilde{p}^2} < \infty, \tag{6.20}$$

and therefore, it holds from (6.19) and (6.20) that

$$\mathbb{E}_x \left[\left| \mathbb{E}_x[N] - \frac{1}{M} \sum_{i=1}^{M} N_i \right|^q \right] \leq \left(\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}} \right)^q \frac{2 - \widetilde{p}}{\widetilde{p}^2}. \tag{6.21}$$

Step 6. Finally, we want to estimate

$$\mathbb{E}_x \left[\left| \mathbb{E}_x \left[\sum_{n=1}^{N} |Y_n| \right] - \frac{1}{M} \sum_{i=1}^{M} \sum_{n=1}^{N_i} |Y_{i,n}| \right|^q \right],$$

where $Y_{i,n}$ were introduced in (6.9). As in the previous step, we use the Corollary 2.2. First of all, it follows from the independence of $(Y_n)_{n=1}^k$ and N for fixed $k \in \mathbb{N}$ (Y_n and X are independent), and the law of total expectation that

$$\begin{aligned}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|\right] &= \sum_{k\geq 1}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right| \;\middle|\; N=k\right]\mathbb{P}_x(N=k)\\ &= \sum_{k\geq 1}\mathbb{E}_0\left[\left|\sum_{n=1}^{k}|Y_n|\right|\right]\mathbb{P}_x(N=k).\end{aligned}$$

Recall that $(Y_n)_{n=1}^k$ are i.i.d. with the same distribution as $X_{\sigma_{B(0,1)}}$. Triangle inequality ensures that

$$\begin{aligned}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|\right] &\leq \sum_{k\geq 1}\sum_{n=1}^{k}\mathbb{E}_0\left[|Y_n|\right]\mathbb{P}_x(N=k)\\ &= \mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|\right]\sum_{k\geq 1}k\mathbb{P}_x(N=k)\\ &= K(\alpha,1)\mathbb{E}_x[N].\end{aligned}$$

Then for all $i \in \{1,...,M\}$, $\sum_{n=1}^{N_i}|Y_{i,n}| \in L^1(\Omega,\mathbb{P}_x)$. Moreover, with similar arguments

$$\begin{aligned}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|^q\right] &= \sum_{k\geq 1}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|^q \;\middle|\; N=k\right]\mathbb{P}_x(N=k)\\ &= \sum_{k\geq 1}\mathbb{E}_0\left[\left|\sum_{n=1}^{k}|Y_n|\right|^q\right]\mathbb{P}_x(N=k).\end{aligned}$$

Recall from the bounds of q that appear in (6.3), one has that $q \in (1,2)$ and the function $|\cdot|^q$ is convex. This implies that, for all $k \in \mathbb{N}$

$$\left|\sum_{n=1}^{k}\frac{|Y_n|}{k}\right|^q \leq \sum_{n=1}^{k}\frac{|Y_n|^q}{k}.$$

Therefore

$$\left|\sum_{n=1}^{k}|Y_n|\right|^q \leq k^{q-1}\sum_{n=1}^{k}|Y_n|^q.$$

Replacing this on the previous estimate one has

$$\begin{aligned}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|^q\right] &\leq \sum_{k\geq 1}\sum_{n=1}^{k}k^{q-1}\mathbb{E}_0\left[|Y_n|^q\right]\mathbb{P}_x(N=k)\\ &= \mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|^q\right]\sum_{k\geq 1}k^q\mathbb{P}_x(N=k)\\ &= K(\alpha,q)\mathbb{E}_x[N^q].\end{aligned} \tag{6.22}$$

For s as in (6.3), Corollary 2.2 implies for all q as in (6.3) that

$$\mathbb{E}_x\left[\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\right]\le\left(\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}}\right)^q\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|^q\right].$$

Therefore it follows from (6.20) and (6.22) that

$$\mathbb{E}_x\left[\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\right]\le\left(\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}}\right)^q K(\alpha,q)\frac{2-\widetilde{p}}{\widetilde{p}^2}. \tag{6.23}$$

Step 7. Coupling the bounds obtained in (6.18), (6.21) and (6.23), it holds that

$$\begin{aligned}\mathbb{E}_x\Bigg[\int_D\left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right]-E_M(x)\right|^q dx+\left|\mathbb{E}_x[N]-\frac{1}{M}\sum_{i=1}^{M}N_i\right|^q\\+\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\Bigg]\\\le\left(\delta_g+\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}}\right)^q|D|B^q d^{pq}C^q+\left(\frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}}\right)^q(1+K(\alpha,q))\frac{2-\widetilde{p}}{\widetilde{p}^2}.\quad=:\mathrm{error}_g^q\end{aligned} \tag{6.24}$$

Using now that $\mathbb{E}(Z)\le c<+\infty$, we summarize the following result.

Lemma 6.2.1 There exists $\overline{N}_i\in\mathbb{N}$, $Y_{i,n}$ i.i.d. copies of $X_{\sigma_{B(0,1)}}$ under $\mathbb{P}_0$, $i=1,...,M$, $n=1,...,\overline{N}_i$ such that

$$\begin{aligned}&\int_D\left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right]-\frac{1}{M}\sum_{i=1}^{M}(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right)\right|^q dx\\&+\left|\mathbb{E}_x[N]-\frac{1}{M}\sum_{i=1}^{M}\overline{N}_i\right|^q+\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\\&\le\mathrm{error}_g^q.\end{aligned} \tag{6.25}$$

With a slight abuse of notation, we redefine E_M from (6.10) as

$$E_M(x)=\frac{1}{M}\sum_{i=1}^{M}(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right). \tag{6.26}$$

Step 8. We are going to prove now that $X^i_{\mathcal{I}(\overline{N}_i)}$ can be approximated by a ReLu DNN. Let $\delta_{\mathrm{dist}}\in(0,1)$. Recall that from (HD-1) there exists $\Phi_{\mathrm{dist}}\in\mathbf{N}$ ReLu DNN such that for all $x\in D$

$$|(\mathcal{R}(\Phi_{\mathrm{dist}}))(x)-\mathrm{dist}(x,\partial D)|\le\delta_{\mathrm{dist}}.$$

Define $(\Phi_{i,n})_{i=1,\dots,M,n=1,\dots,\overline{N}_i}\in\mathbf{N}$ as follows: for $x\in D$

$$(\mathcal{R}(\Phi_{i,1}))(x)=x+Y_{i,1}(\mathcal{R}(\Phi_{\mathrm{dist}}))(x), \tag{6.27}$$

and for all $n = 2, ..., \overline{N}_i$, $x \in D$

$$(\mathcal{R}(\Phi_{i,n}))(x) = (\mathcal{R}(\Phi_{i,n-1}))(x) + Y_{i,n}\left(\mathcal{R}(\Phi_{\text{dist}}) \circ \mathcal{R}(\Phi_{i,n-1})\right)(x). \tag{6.28}$$

In the Section 6.3 we will see that $(\Phi_{i,n})_{i=1,...,M,n=1,...,\overline{N}_i}$ is indeed a ReLu DNN. Note that, for $x \in D$, $i = 1, ..., M$,

$$\left|X^i_{\mathcal{I}(1)} - (\mathcal{R}(\Phi_{i,1}))(x)\right| \leq |Y_{i,1}| \left|(\mathcal{R}(\Phi_{\text{dist}}))(x) - \text{dist}(x, \partial D)\right| \leq \delta_{\text{dist}} \sum_{n=1}^{\overline{N}_i} |Y_{i,n}|,$$

and for all $n = 2, ..., \overline{N}_i$, by triangle inequality

$$\begin{aligned}\left|X^i_{\mathcal{I}(n)} - (\mathcal{R}(\Phi_{i,n}))(x)\right| &\leq \left|X^i_{\mathcal{I}(n-1)} - (\mathcal{R}(\Phi_{i,n-1}))(x)\right| \\ &+ |Y_{i,n}| \left|\text{dist}\left(X^i_{\mathcal{I}(n-1)}, \partial D\right) - \text{dist}\left((\mathcal{R}(\Phi_{i,n-1}))(x), \partial D\right)\right| \\ &+ |Y_{i,n}| \left|\text{dist}\left((\mathcal{R}(\Phi_{i,n-1}))(x), \partial D\right) - \left(\mathcal{R}(\Phi_{\text{dist}}) \circ \mathcal{R}(\Phi_{i,n-1})\right)(x)\right|.\end{aligned}$$

Using the hypothesis on Φ_{dist} and the fact that the function $x \to \text{dist}(x, \partial D)$ is 1-Lipschitz one has

$$\left|X^i_{\mathcal{I}(n)} - (\mathcal{R}(\Phi_{i,n}))(x)\right| \leq \left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right) \left|X^i_{\mathcal{I}(n-1)} - (\mathcal{R}(\Phi_{i,n-1}))(x)\right| + \delta_{\text{dist}} \sum_{n=1}^{\overline{N}_i} |Y_{i,n}|.$$

By the previous recursion one obtain that for all $i = 1, ..., M$

$$\begin{aligned}\left|X^i_{\mathcal{I}(\overline{N}_i)} - \left(\mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right| &\leq \left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}|\right) \delta_{\text{dist}} \sum_{i=1}^{\overline{N}_i} \left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{i-1} \\ &\leq \left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}|\right) \delta_{\text{dist}} \frac{\left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\overline{N}_i} - 1}{\left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right) - 1} \\ &\leq \delta_{\text{dist}} \left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\overline{N}_i}.\end{aligned}$$

Step 9. With the ReLu DNNs defined in Step 8, we are able to find a ReLu DNN that approximates $E_M(x)$. Define $\Phi^i_g \in \mathbf{N}$ as follows

$$\left(\mathcal{R}(\Phi^i_g)\right)(x) = \left(\mathcal{R}(\Phi_g) \circ \mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x), \tag{6.29}$$

valid for $x \in D$. Notice from Lemma 3.2 that Φ^i_g is indeed a ReLu DNN. For full details see Section 6.1. By triangle inequality one has

$$\begin{aligned}&\left|(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right) - \left(\mathcal{R}(\Phi^i_g)\right)(x)\right| \\ &\quad \leq \left|(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right) - g\left(X^i_{\mathcal{I}(\overline{N}_i)}\right)\right| + \left|g\left(X^i_{\mathcal{I}(\overline{N}_i)}\right) - \left(g \circ \mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right| \\ &\qquad + \left|\left(g \circ \mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x) - \left(\mathcal{R}(\Phi^i_g)\right)(x)\right|.\end{aligned}$$

We use the hypothesis (Hg-1) and the assumption that g is L_g-Lipschitz to obtain

$$\begin{aligned}&\left|(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right)-\left(\mathcal{R}(\Phi^i_g)\right)(x)\right|\\&\quad\leq Bd^p\delta_g\left(\left(1+\left|X^i_{\mathcal{I}(\overline{N}_i)}\right|\right)^p+\left(1+\left|\left(\mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right|\right)^p\right)+L_g\left|X^i_{\mathcal{I}(\overline{N}_i)}-\left(\mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right|.\end{aligned}$$

By triangle inequality one has

$$\left|\left(\mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right|\leq\left|X^i_{\mathcal{I}(\overline{N}_i)}-\left(\mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right|+\left|X^i_{\mathcal{I}(\overline{N}_i)}\right|.$$

With the previous estimate and using that $(\cdot)^p$ is a convex function, we obtain

$$\begin{aligned}&\left|(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right)-\left(\mathcal{R}(\Phi^i_g)\right)(x)\right|\\&\leq Bd^p\delta_g\left(1+\left|X^i_{\mathcal{I}(\overline{N}_i)}\right|\right)^p+L_g\left|X^i_{\mathcal{I}(\overline{N}_i)}-\left(\mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right|\\&\quad+2^{p-1}Bd^p\delta_g\left(\left(1+\left|X^i_{\mathcal{I}(\overline{N}_i)}\right|\right)^p+\left|X^i_{\mathcal{I}(\overline{N}_i)}-\left(\mathcal{R}(\Phi_{i,\overline{N}_i})\right)(x)\right|^p\right).\end{aligned}$$

Notice that from (4.1),

$$\left|X^i_{\mathcal{I}(\overline{N}_i)}\right|\leq|x|+\operatorname{diam}(D)\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|.$$

Therefore, in addition to Step 6, one has

$$\begin{aligned}\left|(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right)-\left(\mathcal{R}(\Phi^i_g)\right)(x)\right|&\leq 3Bd^p\delta_g\left(1+|x|+\operatorname{diam}(D)\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|\right)^p\\&+L_g\delta_{\text{dist}}\left(1+\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|\right)^{\overline{N}_i}+2Bd^p\delta_g\delta_{\text{dist}}^p\left(1+\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|\right)^{p\overline{N}_i}.\end{aligned}\tag{6.30}$$

Now define for $\varepsilon\in(0,1)$ the ReLu DNN $\Psi_{1,\varepsilon}$ such that it satisfies for $x\in D$

$$(\mathcal{R}(\Psi_{1,\varepsilon}))(x)=\frac{1}{M}\sum_{i=1}^{M}\left(\mathcal{R}(\Phi^i_g)\right)(x).$$

This is the requested DNN. Section 6.3 shows that $\Psi_{1,\varepsilon}$ is a ReLu DNN. From the bound obtained in (6.30), we have that

$$\begin{aligned}|E_M(x)-(\mathcal{R}(\Psi_{1,\varepsilon}))(x)|&\leq\frac{1}{M}\sum_{i=1}^{M}\left|(\mathcal{R}(\Phi_g))\left(X^i_{\mathcal{I}(\overline{N}_i)}\right)-\left(\mathcal{R}(\Phi^i_g)\right)(x)\right|\\&\leq 3Bd^p\delta_g\left(K_1+\operatorname{diam}(D)\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|\right)^p+L_g\delta_{\text{dist}}\left(1+\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|\right)^{\sum_{i=1}^{M}\overline{N}_i}\\&\quad+2Bd^p\delta_g\delta_{\text{dist}}^p\left(1+\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|\right)^{p\sum_{i=1}^{M}\overline{N}_i}.\end{aligned}$$

Step 10. We want to bound error_g. Using that $\frac{1}{q} < 1$ one has

$$\begin{aligned}\text{error}_g &\leq \left(\delta_g + \frac{2\Theta_{q,s}}{M^{1-\frac{1}{s}}}\right)|D|^{\frac{1}{q}}Bd^pC + 2\frac{\Theta_{q,s}}{M^{1-\frac{1}{s}}}\left(1+K(\alpha,q)^{\frac{1}{q}}\right)\left(\frac{2-\widetilde{p}}{\widetilde{p}^2}\right)^{\frac{1}{q}}. \\ &= 2\frac{\Theta_{q,s}}{M^{1-\frac{1}{s}}}\left(|D|^{\frac{1}{q}}Bd^pC + \left(\frac{2-\widetilde{p}}{\widetilde{p}^2}\right)^{\frac{1}{q}}\left(1+K(\alpha,q)^{\frac{1}{q}}\right)\right) + \delta_g|D|^{\frac{1}{q}}Bd^pC\end{aligned}$$

Denote

$$C_1 = 2\Theta_{q,s}\left(|D|^{\frac{1}{q}}Bd^pC + \left(\frac{2-\widetilde{p}}{\widetilde{p}^2}\right)^{\frac{1}{q}}\left(1+K(\alpha,q)^{\frac{1}{q}}\right)\right), \quad \text{and} \quad C_2 = |D|^{\frac{1}{q}}Bd^pC. \tag{6.31}$$

Note that C_1 and C_2 are polynomial on the dimension d. Then

$$\text{error}_g \leq \frac{C_1}{M^{1-\frac{1}{s}}} + C_2\delta_g. \tag{6.32}$$

In adition, thanks to Step 5, one has

$$\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}| \leq M\left(\text{error}_g + \mathbb{E}_x\left[\sum_{n=1}^{N}|Y_{i,n}|\right]\right) \leq M\left(\text{error}_g + K(\alpha,1)\frac{1}{\widetilde{p}}\right).$$

Define $C_3 := K(\alpha,1)/\widetilde{p}$, then

$$\begin{aligned}\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}| &\leq M^{\frac{1}{s}}C_1 + M\delta_g C_2 + MC_3 \\ &\leq M^{\frac{1}{s}}C_1 + M(C_2+C_3).\end{aligned}$$

On the other hand side define $C_4 = \frac{1}{\widetilde{p}}$, then

$$\begin{aligned}\sum_{i=1}^{M}\overline{N}_i &\leq M(\text{error}_g + \mathbb{E}_x[N]) \leq M^{\frac{1}{s}}C_1 + M\delta_g C_2 + \frac{M}{\widetilde{p}} \\ &\leq M^{\frac{1}{s}}C_1 + M(C_2+C_4).\end{aligned}$$

Step 11. Using the auxiliary Lemma 6.2.1 and (6.32), it follows that

$$\left(\int_D \left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] - E_M(x)\right|^q dx\right)^{\frac{1}{q}} \leq \frac{C_1}{M^{1-\frac{1}{s}}} + C_2\delta_g.$$

In addition, from Step 9 and (6.32) one has

$$\begin{aligned}
&\left(\int_D |E_M(x) - (\mathcal{R}(\Psi_1))(x)|^q \, dx\right)^{\frac{1}{q}} \\
&\quad \leq 3|D|^{\frac{1}{q}} B d^p \delta_g \left(K_1 + \operatorname{diam}(D)\left(M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)\right)^p \\
&\qquad + |D|^{\frac{1}{q}} L_g \delta_{\text{dist}} \left(1 + M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)^{M^{\frac{1}{s}} C_1 + M(C_2 + C_4)} \\
&\qquad + 2|D|^{\frac{1}{q}} B d^p \delta_g \delta_{\text{dist}}^p \left(1 + M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)^{p\left(M^{\frac{1}{s}} C_1 + M(C_2 + C_4)\right)}.
\end{aligned}$$

Therefore, Minkowski inequality implies that

$$\begin{aligned}
&\left(\int_D \left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] - (\mathcal{R}(\Psi_{1,\varepsilon}))(x)\right|^q dx\right)^{\frac{1}{q}} \\
&\leq \left(\int_D \left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] - E_M(x)\right|^q dx\right)^{\frac{1}{q}} + \left(\int_D |E_M(x) - (\mathcal{R}(\Psi_1))(x)|^q \, dx\right)^{\frac{1}{q}} \\
&\leq \frac{C_1}{M^{1-\frac{1}{s}}} + C_2 \delta_g + 3|D|^{\frac{1}{q}} B d^p \delta_g \left(K_1 + \operatorname{diam}(D)\left(M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)\right)^p \\
&\quad + |D|^{\frac{1}{q}} L_g \delta_{\text{dist}} \left(1 + M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)^{M^{\frac{1}{s}} C_1 + M(C_2 + C_4)} \\
&\quad + 2|D|^{\frac{1}{q}} B d^p \delta_g \delta_{\text{dist}}^p \left(1 + M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)^{p\left(M^{\frac{1}{s}} C_1 + M(C_2 + C_4)\right)}.
\end{aligned} \tag{6.33}$$

for $\varepsilon \in (0,1)$ let $M \in \mathbb{N}$ large enough such that

$$M = \left\lceil \left(\frac{5C_1}{\varepsilon}\right)^{\frac{s}{s-1}} \right\rceil$$

, and $\delta_{\text{dist}} \in (0,1)$ small enough such that

$$\delta_{\text{dist}} = \frac{\varepsilon}{5|D|^{\frac{1}{q}} L_g} \left(1 + M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)^{-\left(M^{\frac{1}{s}} C_1 + M(C_2 + C_4)\right)}.$$

Let

$$C_5 = \max\left\{C_2, 3|D|^{\frac{1}{q}} B d^p \left(K_1 + \operatorname{diam}(D)\left(M^{\frac{1}{s}} C_1 + M(C_2 + C_3)\right)\right)^p, \frac{2|D|^{\frac{1}{q}} B d^p}{5^p |D|^{\frac{p}{q}} L_g^p}\right\}, \tag{6.34}$$

and consider $\delta_g \in (0,1)$ small enough such that

$$\delta_g = \frac{\varepsilon}{5C_5}.$$

Therefore each term of (6.33) can be bounded by $\varepsilon/5$. Then

$$\left(\int_D \left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] - (\mathcal{R}(\Psi_{1,\varepsilon}))(x)\right|^q dx\right)^{\frac{1}{q}} \leq \varepsilon.$$

This allos us to conclude that 6.2 can be approximated in $L^q(D)$ by a DNN $\Psi_{1,\varepsilon}$ with accurateness $\varepsilon \in (0,1)$.

6.3. Proof of Proposition 6.1: Quantification of DNNs

In this Section we will prove that $\Psi_{1,\varepsilon}$ is in fact a ReLu DNN wich does not suffer of the curse of dimensionality.

Step 12. We now use the Definitions and Lemmas of Chapter 3 to study $\Psi_{1,\varepsilon}$. Let

$$\beta_{\text{dist}} = \mathcal{D}(\Phi_{\text{dist}}) \quad \text{and} \quad H_{\text{dist}} = \dim(\beta_{\text{dist}}) - 2.$$

And we will verify by induction that for all $i = 1, ..., M$ $n = 1, ..., \overline{N}_i$, $\Phi_{i,n}$ (defined in 6.27 and 6.28) is a ReLu DNN that satisfy

$$\mathcal{D}(\Phi_{i,n}) = \bigodot_{m=1}^{n} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}), \tag{6.35}$$

where

$$\widetilde{\beta}_{\text{dist}} = (\beta_{\text{dist},0}, ..., \beta_{\text{dist},H_{\text{dist}}}, d) \in \mathbb{N}^{H_{\text{dist}}+2}.$$

If 6.35 is true, then from (6.35) and the definition of the operator $\odot$ is easy to see that

$$|||\mathcal{D}(\Phi_{i,n})||| \leq 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||, \quad \text{and} \quad \dim(\mathcal{D}(\Phi_{i,n})) = (H_{\text{dist}} + 1)n + 1. \tag{6.36}$$

For $n = 1$, recall the definition of $\Phi_{i,1}$ from (6.27). By Lemma 3.5 one has that

$$Y_{i,1}\mathcal{R}(\Phi_{\text{dist}}) \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \widetilde{\beta}_{\text{dist}}\right\}\right).$$

By Lemma 3.1, the identity function can be represented by a ReLu DNN with $H_{\text{dist}} + 2$ number of layers. Therefore by Lemma 3.3 it follows that $\mathcal{R}(\Phi_{i,1}) \in C(D, \mathbb{R}^d)$ and

$$\mathcal{D}(\Phi_{i,1}) = d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}, \qquad \dim(\mathcal{D}(\Phi_{i,1})) = H_{\text{dist}} + 2.$$

Moreover

$$|||\mathcal{D}(\Phi_{i,1})||| \leq 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||.$$

Now suppose that for $n = 2, ..., \overline{N}_i - 1$ that (6.35) is valid. Recall the definition of $\Phi_{i,n}$ from (6.28). Notice that $\mathcal{R}(\Phi_{i,n+1})$ can be written as

$$\mathcal{R}(\Phi_{i,n+1}) = \mathcal{R}(\widetilde{\Phi}_{i,n+1}) \circ \mathcal{R}(\Phi_{i,n}).$$

where $\widetilde{\Phi}_{i,n} \in \mathbf{N}$ is a ReLu DNN that satisfies

$$\left(\mathcal{R}(\widetilde{\Phi}_{i,n})\right)(x) = x + Y_{i,n}\left(\mathcal{R}(\Phi_{\text{dist}})\right)(x).$$

By the same arguments as in the case $n = 1$, it follows for all $n = 2, ..., \overline{N}_i$ that

$$\mathcal{D}(\widetilde{\Phi}_{i,n}) = d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}, \quad \dim(\mathcal{D}(\widetilde{\Phi}_{i,n})) = H_{\text{dist}} + 2,$$

and

$$\left\|\left|\mathcal{D}(\widetilde{\Phi}_{i,n})\right|\right\| \leq 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||.$$

Therefore from the inductive hypothesis (6.35) and Lemma 3.2, $\Phi_{i,n+1}$ is a ReLu DNN that satisfies

$$\mathcal{D}(\Phi_{i,n+1}) = (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}) \odot \left(\bigodot_{m=1}^{n} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}})\right) = \bigodot_{m=1}^{n+1} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}).$$

Then the claim for $\Phi_{i,n}$ is proved for any $i = 1, ..., M$, $n = 1, ..., \overline{N}_i$. Recall that 6.36 is valid too. Therefore

$$\mathcal{D}(\Phi_{i,\overline{N}_i}) = \bigodot_{m=1}^{\overline{N}_i} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}).$$

Moreover

$$\left\|\left|\mathcal{D}(\Phi_{i,\overline{N}_i})\right|\right\| \leq 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||, \quad \text{and} \quad \dim(\mathcal{D}(\Phi_{i,\overline{N}_i})) = (H_{\text{dist}} + 1)\overline{N}_i + 1.$$

Let $\beta_g = \mathcal{D}(\Phi_g)$ and $H_g = \dim(\beta_g) - 2$. By Lemma 3.2 one has that

$$\mathcal{D}(\Phi_g^i) = \beta_g \odot \left(\bigodot_{m=1}^{\overline{N}_i} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}})\right), \qquad \dim(\mathcal{D}(\Phi_g^i)) = (H_{\text{dist}} + 1)\overline{N}_i + H_g + 2.$$

Moreover

$$\left\|\left|\mathcal{D}(\Phi_g^i)\right|\right\| \leq \max\{|||\mathcal{D}(\Phi_g)|||, 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||\}.$$

Recall that $\overline{N}_i$ not necessarily be the same for $i = 1, ..., M$. Now we need that for all $i = 1, ..., M$, Φ_g^i have the same number of layers to use Lemma 3.3. For any $i = 1, ..., M$ define

$$H_i = (H_{\text{dist}} + 1)\left(\sum_{j=1}^{M} \overline{N}_j - \overline{N}_i\right) - 1.$$

By Lemma 3.1, The identity function can be represented by a ReLu DNN with H_i hidden layers. Recall the definition of Φ_g^i in (6.29). Using Lemma 3.2 we have that

$$\mathcal{D}(\Phi_g^i) = \mathfrak{n}_{H_i+2} \odot \beta_g \odot \left(\bigodot_{m=1}^{\overline{N}_i} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}})\right),$$

and

$$\dim(\mathcal{D}(\Phi_g^i)) = (H_{\text{dist}} + 1)\sum_{i=1}^{M} \overline{N}_i + H_g + 2.$$

Now we use Lemma 3.3 to conclude that

$$\mathcal{D}(\Psi_{1,\varepsilon}) = \boxplus_{i=1}^{M} \left(\mathfrak{n}_{H_i+2} \odot \beta_g \odot \left(\bigodot_{m=1}^{\overline{N}_i} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}})\right)\right),$$

and

$$\dim(\mathcal{D}(\Psi_{1,\varepsilon})) = (H_{\text{dist}} + 1)\sum_{i=1}^{M} \overline{N}_i + H_g + 2.$$

In addition

$$\begin{aligned} |||\mathcal{D}(\Psi_{1,\varepsilon})||| &\le \sum_{i=1}^{M} \max\{|||\mathcal{D}(\Phi_g)|||, 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||\} \\ &\le M(|||\mathcal{D}(\Phi_g)||| + 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||). \end{aligned} \tag{6.37}$$

Notice fron (6.31) that the constants C_1 and C_2 are bounded by a multiple of $|D|^{\frac{1}{q}} d^p$. Therefore, by choice of M,

$$M \le B_1 |D|^{\frac{s}{q(s-1)}} d^{\frac{ps}{s-1}} \varepsilon^{-\frac{s}{s-1}}, \tag{6.38}$$

where $B_1 > 0$ is a generic constant. With (6.38) and the bound of C_1 and C_2, we have that C_5 defined in (6.34) is bounded by a multiple of

$$|D|^{\frac{1}{q}\left(1+p+\frac{ps}{s-1}\right)} d^{p+p^2+\frac{p^2 s}{s-1}} \varepsilon^{-\frac{ps}{s-1}}.$$

By the choice of δ_g we have, for $B_2 > 0$ a generic constant that

$$\delta_g^{-a} \le B_2 |D|^{\frac{a}{q}\left(1+p+\frac{ps}{s-1}\right)} d^{ap+ap^2+\frac{ap^2 s}{s-1}} \varepsilon^{-a-\frac{aps}{s-1}}. \tag{6.39}$$

For δ_{dist} we estimate $\log\left(\delta_{\text{dist}}^{-1}\right)$ as indicates Assumption 2. By the choice of δ_{dist} and properties of log function, we have that

$$\log\left(\delta_{\text{dist}}^{-1}\right) \le 5|D|^{\frac{1}{q}} L_g \varepsilon^{-1} + (M^{\frac{1}{s}} C_1 + M(C_2 + C_4))(1 + M^{\frac{1}{s}} C_1 + M(C_2 + C_3)).$$

Therefore

$$\lceil \log\left(\delta_{\text{dist}}^{-1}\right) \rceil^a \le B_3 |D|^{\frac{2a}{q}\left(1+\frac{s}{s-1}\right)} d^{2ap\left(1+\frac{s}{s-1}\right)} \varepsilon^{-a-\frac{2as}{s-1}}, \tag{6.40}$$

where $B_3 > 0$ is a generic constant. Assumptions 1 and 2, in addition with (6.37) implies that

$$|||\mathcal{D}(\Psi_{1,\varepsilon})||| \le B_4 d^b M(\delta_g^{-a} + \lceil \log\left(\delta_{\text{dist}}^{-1}\right) \rceil^a),$$

where $B_4 > 0$ is a generic constant. Therefore, from (6.38), (6.39) and (6.40) we conclude that there exists $\widetilde{B} > 0$ such that

$$|||\mathcal{D}(\Psi_{1,\varepsilon})||| \le \widetilde{B} |D|^{\frac{1}{q}\left(2a+ap+\frac{s}{s-1}(1+2a+ap)\right)} d^{b+2ap+ap^2+\frac{ps}{s-1}(1+2a+ap)} \varepsilon^{-a-\frac{s}{s-1}(1+2a+ap)}.$$

Note also that this implies from Remark 3.1.2 that $\Psi_{1,\varepsilon}$ overcomes the curse of dimensionality. This completes the proof of Proposition 6.1.

Chapter 7

Approximation of solutions of the Fractional Dirichlet problem using DNNs: the source case

7.1. Non-homogeneous Fractional Laplacian

In the previous Chapter we have proved the the solution (6.2) of the fractional Dirichlet Problem without source can be approximated by a ReLu DNN. In this Section we focus in the term

$$\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[f\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right]. \tag{7.1}$$

We will prove that (7.1) can be approximated by a ReLu DNN that does not suffer of the curse of dimensionality. Notice that (7.1) corresponds to the extra term in the solution (1.3) of the fractional Dirichlet Problem with source (1.2). In order to do this approximation, the following assumption will be introduced

Assumptions 3 Let $d \geq 2$. Let $f : D \to \mathbb{R}$ a function satisfying (Hf-0). Let $\delta_f \in (0,1)$, $a, b \geq 1$ and $B > 0$. Then there exists a ReLu DNN $\Phi_f \in \mathbf{N}$ with

1. $\mathcal{R}(\Phi_f) : D \to \mathbb{R}$ is $\widetilde{L}_f$-Lipschitz continuous, $\widetilde{L}_f > 0$, and
2. The following are satisfied:

$$|f(x) - (\mathcal{R}(\Phi_f))(x)| \leq \delta_f, \qquad x \in D. \tag{Hf-1}$$

$$|||\mathcal{D}(\Phi_f)||| \leq Bd^b\delta_f^{-a}. \tag{Hf-2}$$

Remark 7.1.1 If Φ_f satisfies Assumptions 3, then it holds for all $x \in D$ that

$$|(\mathcal{R}(\Phi_f))(x)| \leq |f(x) - (\mathcal{R}(\Phi_f))(x)| + |f(x)| \leq \delta_f + \|f\|_{L^\infty(D)}.$$

Then

$$\|\mathcal{R}(\Phi_f)\|_{L^\infty(D)} \leq \delta_f + \|f\|_{L^\infty(D)}. \tag{7.2}$$

The main result of this Section is the following proposition, that ensures the existence of a ReLu DNN such that (7.1) is well approximated

Proposition 7.1 *Let $\alpha \in (1,2)$, $L_f > 0$ and*

$$p, s \in (1,\alpha) \text{ such that } s < \tfrac{\alpha}{p}, \quad \text{and} \quad q \in \Big[s, \tfrac{\alpha}{p}\Big). \tag{7.3}$$

Suppose that f is a function satisfying (Hf-0) *and Assumptions 3. Suppose additionally that D satisfies Assumptions 2.*

Then for all $\widetilde{\varepsilon} \in (0,1)$, there exists a ReLu DNN $\Psi_{2,\widetilde{\varepsilon}}$ such that

1. *Proximity in $L^q(D)$:*

$$\left(\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha V_1(0, f(X_{\mathcal{I}(n-1)} + r_n \cdot))\right] - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x)\right|^q\right)^{\frac{1}{q}} \leq \widetilde{\varepsilon}. \tag{7.4}$$

2. *Realization: $\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})$ has the following structure: there exist $M_1, M_2 \in \mathbb{N}$, $\overline{N}_i \in \mathbb{N}$, $Y_{i,n}$ i.i.d copies of $X_{\sigma_{B(0,1)}}$ under $\mathbb{P}_0$, $v_{i,j,n}$ i.i.d copies with law μ under $B(0,1)$, for $i = 1, ..., M_1$, $j = 1, ..., M_2$, $n = 1, ..., \overline{N}_i$ such that for all $x \in D$,*

$$\left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x) = \frac{1}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{\overline{N}_i} \kappa_{d,\alpha}\left(\mathcal{R}(\Upsilon)\right)\Big(\left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x), \left(\mathcal{R}(\Phi_f^{i,n})\right)(x)\Big), \tag{7.5}$$

where for all $y \in D$

$$\left(\mathcal{R}(\Phi_r^{i,n})\right)(y) = \left(\mathcal{R}(\Phi_{\text{dist}}) \circ \mathcal{R}(\Phi_{i,n-1})\right)(y), \tag{7.6}$$

$$\left(\mathcal{R}(\Phi_f^{i,n})\right)(y) = \frac{1}{M_2}\sum_{j=1}^{M_2}\left(\mathcal{R}(\Phi_f) \circ \left(\mathcal{R}(\Phi_{i,n}) + v_{i,j,n}\mathcal{R}(\Phi_r^{i,n})\right)\right)(y), \tag{7.7}$$

and $\mathcal{R}(\Phi_{i,n})$ is a Relu DNN that approximates $X^i_{\mathcal{I}(n)}$, for $i = 1, ..., M_1$, $n = 1, ..., \overline{N}_i$.

3. *Bounds: there exists $\widetilde{B} > 0$ such that*

$$\left|\left|\left|\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})\right|\right|\right| \leq \widetilde{B}|D|^{\frac{1}{q}\left(1+2a+\frac{2s}{s-1}(1+a)\right)} d^b \widetilde{\varepsilon}^{-a-\frac{2s}{s-1}(1+a)}. \tag{7.8}$$

7.2. Proof of Proposition 7.1: existence

As in the proof of Proposition 6.1, this proof will be divided in several steps. Let s, p and q as in (7.3).

Step 1. Let $(\rho_n)_{n\in\mathbb{N}}$ the WoS process starting at $x \in D$. Recall that for all $n = 1, ..., N$ the process $X_{\mathcal{I}(n)}$ depends of the point $x \in D$ and n copies of $X_{\sigma_{B(0,1)}}$, namely

$$X_{\mathcal{I}(n)} = X_{\mathcal{I}(n)}(x, Y_1, ..., Y_n), \tag{7.9}$$

where Y_k, $k = 1, ..., n$ are i.i.d. copies of $X_{\sigma_{B(0,1)}}$ under $\mathbb{P}_0$. Let $M_1 \in \mathbb{N}$. Consider M_1 copies of $X_{\mathcal{I}(n)}$ starting at $x \in D$, as described in (7.9). We denote such copies as

$$X^i_{\mathcal{I}(n)} = X^i_{\mathcal{I}(n)}(x, Y_{i,1}, ..., Y_{i,n}),$$

where $Y_{i,k}$, $i = 1, ..., M_1$, $n = 1..., N_i$, $k = 1, ..., n$ are i.i.d. copies of $X_{\sigma_{B(0,1)}}$ under $\mathbb{P}_0$, and each N_i is an i.i.d. copy of N. Recall that N_i not necessarily be the same (as a random variable).

Let $M_2 \in \mathbb{N}$. for all $n = 1, ..., N$ let $(v_{j,n})_{j=1}^{M_2}$ be M_2 copies of a random variable v with distribution μ over $B(0,1)$. For all $n = 1, ..., N$ and $\chi \in L^2(B(0,1), \mu)$ define the Monte Carlo operator

$$E^n_{M_2}(\chi(\cdot)) = \frac{1}{M_2} \sum_{j=1}^{M_2} \chi(v_{j,n}), \tag{7.10}$$

and we will refer to E_{M_2} when the copies of v in 7.10 do not depend on n. Additionally define the operator

$$E_{M_1}(x) = \frac{1}{M_1} \sum_{i=1}^{M_1} \sum_{n=1}^{N_i} (r^i_n)^\alpha \kappa_{d,\alpha} E_{M_2} \left((\mathcal{R}(\Phi_f)) \left(X^i_{\mathcal{I}(n-1)} + r^i_n \cdot \right) \right), \tag{7.11}$$

where

$$r_{i,n} = \operatorname{dist}\left(X^i_{\mathcal{I}(n-1)}, \partial D \right), \tag{7.12}$$

and $\mathcal{R}(\Phi_f)$ denotes the realization as a Lipschitz continuous funtion of the DNN $\Phi_f \in \mathbf{N}$ that approximates f in Assumption (3). Note that E_{M_1} not necessarily be a DNN.

We want to establish suitable bounds of the difference between (7.1) and the operator $E_{M_1}(x)$. For this, in the next step we work for all $n = 1, ..., N$ with the term

$$E^n_{M_2} \left((\mathcal{R}(\Phi_f)) \left(X_{\mathcal{I}(n-1)} + r_n \cdot \right) \right).$$

Step 2. Notice by Remark 7.1.1 that for all $n = 1, ..., N$ that

$$\mathbb{E}^{(\mu)} \left(\left| (\mathcal{R}(\Phi_f)) \left(X_{\mathcal{I}(n-1)} + r_n \cdot \right) \right| \right) \leq \delta_f + \|f\|_{L^\infty(D)}. \tag{7.13}$$

Then for all $j = 1, ..., M_2$, $(\mathcal{R}(\Phi_f)) \left(X_{\mathcal{I}(n-1)} + r_n v_{j,n} \right) \in L^1(B(0,1), \mu)$. For s as in (7.3) it follows from Corollary 2.2 that for all q as in (7.3)

$$\begin{aligned} &\left\| \mathbb{E}^{(\mu)} \left[(\mathcal{R}(\Phi_f)) \left(X_{\mathcal{I}(n-1)} + r_n \cdot \right) \right] - E_{M_2} \left((\mathcal{R}(\Phi_f)) \left(X_{\mathcal{I}(n-1)} + r_n \cdot \right) \right) \right\|_{L^q(B(0,1),\mu)} \\ &\leq \frac{2\Theta_{q,s}}{M_2^{1-\frac{1}{s}}} \mathbb{E}^{(\mu)} \left[(\mathcal{R}(\Phi_f)) \left(X_{\mathcal{I}(n-1)} + r_n \cdot \right)^q \right]^{\frac{1}{q}}. \end{aligned}$$

From Remark 7.1.1 it follows that

$$\mathbb{E}^{(\mu)} \left(\left| (\mathcal{R}(\Phi_f)) \left(X_{\mathcal{I}(n-1)} + r_n \cdot \right) \right|^q \right)^{\frac{1}{q}} \leq \delta_f + \|f\|_{L^\infty(D)}.$$

Therefore

$$\left\|\mathbb{E}^{(\mu)}\left[(\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]-E_{M_2}\left((\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right)\right\|_{L^q(B(0,1),\mu)} \leq \frac{2\Theta_{q,s}\left(\delta_f+\|f\|_{L^\infty(D)}\right)}{M_2^{1-\frac{1}{s}}}.$$

Then for any $n=1,...,N$ there exists $v_{j,n}$, $j=1,...,M_2$, i.i.d random variables with distribution μ such that

$$\left|\mathbb{E}^{(\mu)}\left[(\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]-\frac{1}{M_2}\sum_{j=1}^{M_2}(\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}+r_n v_{j,n}\right)\right| \leq \frac{2\Theta_{q,s}\left(\delta_f+\|f\|_{L^\infty(D)}\right)}{M_2^{1-\frac{1}{s}}}.$$

We redefine $E^n_{M_2}$ with the random variables $v_{j,n}$ found for all $n=1,...,N$.

In the two next steps we control the difference between (7.1) and E_{M_1} with the intermediate term

$$\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} E^n_{M_2}\left(\mathcal{R}(\Phi_f)\right)\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right],$$

Step 3. Define

$$J_3=\left\|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[(\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right] - \mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} E^n_{M_2}\left((\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right)\right]\right\|_{L^q(\Omega,\mathbb{P}_x)}.$$

From Step 2 we have

$$J_3 \leq \frac{2\Theta_{q,s}}{M_2^{1-\frac{1}{s}}}\left(\delta_f+\|f\|_{L^\infty(D)}\right)\left\|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha}\right]\right\|_{L^q(\Omega,\mathbb{P}_x)}.$$

Using Lemma 5.2.2 it follows that

$$J_3 \leq \frac{2\Theta_{q,s}}{M_2^{1-\frac{1}{s}}}\left(\delta_f+\|f\|_{L^\infty(D)}\right)\mathbb{E}_x[\sigma_D].$$

Step 4. Define

$$J_4=\left\|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} E^n_{M_2}\left(\mathcal{R}(\Phi_f)\right)\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]-E_{M_1}(x)\right\|_{L^q(\Omega,\mathbb{P}_x)}.$$

Recall Remark 7.1.1. Then

$$\mathbb{E}_x\left[\left|\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} E^n_{M_2}\left(\mathcal{R}(\Phi_f)\right)\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right|\right] \leq (\delta_f + \|f\|_{L^\infty(D)})\mathbb{E}_x[\sigma_D] < \infty.$$

This implies that

$$\sum_{n=1}^{N} r_{i,n}^\alpha \kappa_{d,\alpha} E^n_{M_2}\left(\mathcal{R}(\Phi_f)\right)\left(X^i_{\mathcal{I}(n-1)}+r_{i,n}\cdot\right) \in L^1(\Omega,\mathbb{P}_x).$$

Then using Corollary 2.2 we have for s as in (7.3) it holds for all q as in (7.3) that

$$J_4 \leq \frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\mathbb{E}_x\left[\left(\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} E^n_{M_2}\left(\left(\mathcal{R}(\phi_f)\right)\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right)\right)^q\right]^{\frac{1}{q}}. \tag{7.14}$$

Remark 7.1.1 implies that

$$J_4 \leq \frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\left(\delta_f + \|f\|_{L^\infty(D)}\right)\mathbb{E}_x\left[\left(\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha}\right)^q\right]^{\frac{1}{q}}$$

By Lemma 5.2.2 and Jensen inequality with $(\cdot)^{\frac{2}{q}}$, $q<2$, we have

$$\mathbb{E}_x\left[\left(\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha}\right)^q\right]^{\frac{1}{q}} \leq \mathbb{E}_x\left[\left(\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha}\right)^2\right]^{\frac{1}{2}} \leq \mathbb{E}_x[\sigma_D^2]^{\frac{1}{2}}.$$

Therefore

$$J_4 \leq \frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\left(\delta_f + \|f\|_{L^\infty(D)}\right)\mathbb{E}_x[\sigma_D^2]^{\frac{1}{2}}. \tag{7.15}$$

Step 5. With the bounds obtained in Steps 3 and 4, we have by Minkowski inequality that

$$\begin{aligned}\left\|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[\left(\mathcal{R}(\Phi_f)\right)\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right] - E_{M_1}(x)\right\|_{L^q(\Omega,\mathbb{P}_x)} &\leq J_3 + J_4 \\ &\leq 2\Theta_{q,s}\left(\delta_f + \|f\|_{L^\infty(D)}\right)\left(\frac{\mathbb{E}_x[\sigma_D]}{M_2^{1-\frac{1}{s}}} + \frac{\mathbb{E}_x[|\sigma_D|^2]^{\frac{1}{2}}}{M_1^{1-\frac{1}{s}}}\right).\end{aligned} \tag{7.16}$$

Fubini and (7.16) implies that

$$\begin{aligned}&\mathbb{E}_x\left[\int_D\left|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[\left(\mathcal{R}(\Phi_f)\right)\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right] - E_{M_1}(x)\right|^q dx\right] \\ &\leq 2^q|D|\Theta_{q,s}^q\left(\delta_f + \|f\|_{L^\infty(D)}\right)^q\left(\frac{\mathbb{E}_x[\sigma_D]}{M_2^{1-\frac{1}{s}}} + \frac{\mathbb{E}_x[\sigma_D^2]^{\frac{1}{2}}}{M_1^{1-\frac{1}{s}}}\right)^q\end{aligned} \tag{7.17}$$

On the other hand side, from hypothesis (Hf-1) and Lemma 5.2.2 one has

$$\mathbb{E}_x\left[\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[(\mathcal{R}(\phi_f))\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right]\right.\right.$$
$$\left.\left.-\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[f\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right]\right|^q dx\right] \le \delta_f^q|D|\mathbb{E}_x[\sigma_D]^q.$$

Therefore, using that $(\cdot)^q$ is a convex function and (7.17) it follows that

$$\begin{aligned}
&\mathbb{E}_x\left[\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[f\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right]-E_{M_1}(x)\right|^q dx\right]\\
&\le 2^{q-1}\delta_f^q|D|\mathbb{E}_x[\sigma_D]^q+2^{q-1}2^q|D|\Theta_{q,s}^q\left(\delta_f+\|f\|_{L^\infty(D)}\right)^q\left(\frac{\mathbb{E}_x[\sigma_D]}{M_2^{1-\frac{1}{s}}}+\frac{\mathbb{E}_x[\sigma_D^2]^{\frac{1}{2}}}{M_1^{1-\frac{1}{s}}}\right)^q \qquad (7.18)\\
&\le 2\delta_f^q|D|\mathbb{E}_x[\sigma_D]^q+2^{q+1}|D|\Theta_{q,s}^q\left(\delta_f+\|f\|_{L^\infty(D)}\right)^q\left(\frac{\mathbb{E}_x[\sigma_D]}{M_2^{1-\frac{1}{s}}}+\frac{\mathbb{E}_x[\sigma_D^2]^{\frac{1}{2}}}{M_1^{1-\frac{1}{s}}}\right)^q,
\end{aligned}$$

where we use that $2^{q-1}<2$ since $q<2$.

Step 6. In order to bound the following expectation

$$\mathbb{E}_x\left[\left|\mathbb{E}_x[N]-\frac{1}{M_1}\sum_{i=1}^{M_1}N_i\right|^q\right],$$

we use Corollary 2.2. Notice by Theorem 4.2 that for all $x\in D$ there exists a geometric random variable Γ with parameter $\widetilde{p}=\widetilde{p}(\alpha,d)>0$ such that

$$\mathbb{E}_x\left[|N|\right]\le\mathbb{E}_x\left[\Gamma\right]=\frac{1}{\widetilde{p}}<\infty,$$

and then for all $i\in\{1,...,M_1\}$, $N_i\in L^1(\mathbb{P}_x,|\cdot|)$. For s as in (7.3), Corollary 2.2 implies for all q as in (7.3) that

$$\mathbb{E}_x\left[\left|\mathbb{E}_x[N]-\frac{1}{M_1}\sum_{i=1}^{M_1}N_i\right|^q\right]\le\left(\frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\right)^q\mathbb{E}_x\left[|N|^q\right]\le\left(\frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\right)^q\mathbb{E}_x\left[|N|^2\right], \qquad (7.19)$$

where we used that $q<2$ and then $\mathbb{E}_x\left[|\cdot|^q\right]\le\mathbb{E}_x\left[|\cdot|^2\right]$. Recall that

$$\mathbb{E}_x\left[|N|^2\right]\le\mathbb{E}_x\left[\Gamma^2\right]=\frac{2-\widetilde{p}}{\widetilde{p}^2}<\infty, \qquad (7.20)$$

and therefore, it holds from (7.19) and (7.20) that

$$\mathbb{E}_x\left[\left|\mathbb{E}_x[N]-\frac{1}{M_1}\sum_{i=1}^{M_1}N_i\right|^q\right]\le\left(\frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\right)^q\frac{2-\widetilde{p}}{\widetilde{p}^2}. \qquad (7.21)$$

Step 7. We want to estimate

$$\mathbb{E}_x\left[\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\right]$$

As in the previous step, we use the Corollary 2.2. First of all, it follows from the independence of $(Y_n)_{n=1}^k$ and N for fixed $k \in \mathbb{N}$ (Y_n and X are independent), and law of total expectation that

$$\begin{aligned}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|\right] &= \sum_{k\geq 1}\mathbb{E}_0\left[\left|\sum_{n=1}^{N}|Y_n|\right|\,\middle|\, N=k\right]\mathbb{P}_x(N=k)\\ &= \sum_{k\geq 1}\mathbb{E}_0\left[\left|\sum_{n=1}^{k}|Y_n|\right|\right]\mathbb{P}_x(N=k).\end{aligned}$$

Recall that $(Y_n)_{n=1}^k$ are i.i.d. with the same distribution as $X_{\sigma_{B(0,1)}}$. Triangle inequality ensures that

$$\begin{aligned}\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|\right] &\leq \sum_{k\geq 1}\sum_{n=1}^{k}\mathbb{E}_0\left[|Y_n|\right]\mathbb{P}_x(N=k)\\ &= \mathbb{E}_0\left[\left|X_{\sigma_{B(0,1)}}\right|\right]\sum_{k\geq 1}k\mathbb{P}_x(N=k)\\ &= K(\alpha,1)\mathbb{E}_x[N].\end{aligned}$$

and then for all $i \in \{1,...,M_1\}$, $\sum_{n=1}^{N}|Y_n| \in L^1(\mathbb{P}_x,|\cdot|)$. Moreover, with similar arguments it holds that

$$\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_{i,n}|\right|^q\right] \leq K(\alpha,q)\mathbb{E}_x[N^q]. \tag{7.22}$$

For s as in (7.3), Corollary 2.2 implies for all q as in (7.3) that

$$\mathbb{E}_x\left[\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\right] \leq \left(\frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\right)^q\mathbb{E}_x\left[\left|\sum_{n=1}^{N}|Y_n|\right|^q\right].$$

And therefore it follows from (7.20) and (7.22) that

$$\mathbb{E}_x\left[\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\right] \leq \left(\frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\right)^q K(\alpha,q)\frac{2-\widetilde{p}}{\widetilde{p}^2}. \tag{7.23}$$

Step 8. It follows from (7.18), (7.21) and (7.23) that

$$\begin{aligned}
&\mathbb{E}_x\left[\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[\left(\mathcal{R}(\Phi_f)\right)\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right]-E_{M_1}(x)\right|^q dx\right.\\
&\qquad\left.+\left|\mathbb{E}_x\left[N\right]-\frac{1}{M_1}\sum_{i=1}^{M_1}N_i\right|^q+\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{N_i}|Y_{i,n}|\right|^q\right]\\
&\leq 2\delta_f^q|D|\mathbb{E}_x[\sigma_D]^q+2^{q+1}|D|\Theta_{q,s}^q\left(\delta_f+\|f\|_{L^\infty(D)}\right)^q\left(\frac{\mathbb{E}_x[\sigma_D]}{M_2^{1-\frac{1}{s}}}+\frac{\mathbb{E}_x[\sigma_D^2]^{\frac{1}{2}}}{M_1^{1-\frac{1}{s}}}\right)^q\\
&\qquad+\left(\frac{2\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\right)^q(1+K(\alpha,q))\frac{2-\widetilde{p}}{\widetilde{p}^2}=:\mathrm{error}_f^q.
\end{aligned}\tag{7.24}$$

Using now that $\mathbb{E}(Z)\leq c<\infty$ we have the following result

Lemma 7.2.1 This implies that there exists $\overline{N}_i\in\mathbb{N}$, $Y_{i,n}$ i.i.d. copies of $X_{\sigma_{B(0,1)}}$, $v_{i,j,n}$ i.i.d. random variables with law μ, $i=1,...,M_1$, $j=1,...,M_2$, $n=1,...,\overline{N}_i$ such that

$$\begin{aligned}
&\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^{\alpha}\kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[f\left(X_{\mathcal{I}(n-1)}+r_n\cdot\right)\right]\right]-E_{M_1}(x)\right|^q dx\\
&\quad+\left|\mathbb{E}_x\left[N\right]-\frac{1}{M_1}\sum_{i=1}^{M_1}\overline{N}_i\right|^q+\left|\mathbb{E}_x\left[\sum_{n=1}^{N}|Y_n|\right]-\frac{1}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|\right|^q\\
&\leq \mathrm{error}_f^q.
\end{aligned}\tag{7.25}$$

Here, E_{M_1} will be redefined from (7.11) according to the copies found in Lemma 7.2.1, with the Monte Carlo operator $E_{M_2}^{i,n}$ defined from the copies $(v_{i,j,n})_{j=1}^{M_2}$.

Step 9. As similar in Step 8 of Proposition 6.1, we can see that for all $i=1,...,M_1$ $n=1,...,\overline{N}_i$ $X^i_{\mathcal{I}(n-1)}$ can be approximated by a ReLu DNN $\Phi_{i,n-1}\in\mathbf{N}$ which satisfies for all $x\in D$

$$\left|X^i_{\mathcal{I}(n-1)}-\left(\mathcal{R}(\Phi_{i,n-1})\right)(x)\right|\leq\delta_{\mathrm{dist}}\left(\sum_{n=1}^{\overline{N}_i}|Y_{i,n}|+1\right)^{\overline{N}_i}.$$

We can find now a DNN that approximates $r_{i,n}$ in (7.12). Indeed, define $\Phi_r^{i,n}\in\mathbf{N}$ as follows:

$$\left(\mathcal{R}(\Phi_r^{i,n})\right)(x)=\left(\mathcal{R}(\Phi_{\mathrm{dist}})\circ\mathcal{R}(\Phi_{i,n-1})\right)(x),$$

valid for $x\in D$. In Section 7.3 we show that $\Phi_r^{i,n}$ is in fact a ReLu DNN. For $x\in D$ we have by triangle inequality that

$$\begin{aligned}
\left|r_{i,n}-\left(\mathcal{R}(\Phi_r^{i,n})\right)(x)\right|&\leq\left|r_{i,n}-\mathrm{dist}\left(\left(\Phi_{i,n-1}\right)(x),\partial D\right)\right|\\
&\quad+\left|\mathrm{dist}\left(\left(\Phi_{i,n-1}\right)(x),\partial D\right)-\left(\mathcal{R}(\Phi_r^{i,n})\right)(x)\right|
\end{aligned}$$

Hypothesis (HD-1) and the fact that the function $x \mapsto \text{dist}(x, \partial D)$ is 1-Lipschitz implies

$$\begin{aligned}\left|r_{i,n} - \left(\mathcal{R}(\Phi_r^{i,n})\right)(x)\right| &\le \left|X^i_{\mathcal{I}(n-1)} - \left(\mathcal{R}(\Phi_{i,n-1})\right)(x)\right| + \delta_{\text{dist}} \\ &\le \delta_{\text{dist}} \left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\overline{N}_i} + \delta_{\text{dist}}.\end{aligned}$$

Step 10. We will find now a DNN that approximates

$$E^{i,n}_{M_2}\left(\left(\mathcal{R}(\Phi_f)\right)\left(X^i_{\mathcal{I}(n-1)} + r_{i,n}\cdot\right)\right). \tag{7.26}$$

Define the DNN $\Phi_f^{i,n} \in \mathbf{N}$ as follows: for $x \in D$

$$\left(\mathcal{R}(\Phi_f^{i,n})\right)(x) = \frac{1}{M_2}\sum_{j=1}^{M_2}\left(\mathcal{R}(\Phi_f) \circ \left(\mathcal{R}(\Phi_{i,n-1}) + v_{i,j,n}\mathcal{R}(\Phi_r^{i,n})\right)\right)(x).$$

In Section 7.3 we will prove that $\Phi_f^{i,n}$ is a ReLu DNN. We now use the assumption that $\mathcal{R}(\Phi_f)$ is a $\widetilde{L}_f$-Lipschitz function to obtain

$$\begin{aligned}&\left|E^{i,n}_{M_2}\left(\left(\mathcal{R}(\Phi_f)\right)\left(X^i_{\mathcal{I}(n-1)} + r_{i,n}\cdot\right)\right) - \left(\mathcal{R}(\Phi_f^{i,n})\right)(x)\right| \\ &\quad\le \frac{\widetilde{L}_f}{M_2}\sum_{j=1}^{M_2}\left(\left|X^i_{\mathcal{I}(n-1)} - \left(\mathcal{R}(\Phi_{i,n-1})\right)(x)\right| + |v_{i,j,n}|\left|r_{i,n} - \left(\mathcal{R}(\Phi_r^{i,n})\right)(x)\right|\right).\end{aligned}$$

Notice that for all $i = 1, ..., M_1$, $j = 1, ..., M_2$, $n = 1, ..., \overline{N}_i$ one has $|v_{i,j,n}| \le 1$ ($v_{i,j,n}$ is a random variable on $B(0,1)$). Therefore, it follows from Step 9 that

$$\begin{aligned}&\left|E^{i,n}_{M_2}\left(\left(\mathcal{R}(\Phi_f)\right)\left(X^i_{\mathcal{I}(n-1)} + r_{i,n}\cdot\right)\right) - \left(\mathcal{R}(\Phi_f^{i,n})\right)(x)\right| \\ &\quad\le \widetilde{L}_f\delta_{\text{dist}}\left(2\left(\sum_{n=1}^{\overline{N}_i}|Y_{i,n}| + 1\right)^{\overline{N}_i} + 1\right).\end{aligned}$$

Step 11. We want to approximate the multiplication between $r^\alpha_{i,n}$ and (7.26). For all $i = 1, ..., M_1$, $n = 1, ..., \overline{N}_i$ define the DNN $\Upsilon_{i,n} \in \mathbf{N}$ as

$$\left(\mathcal{R}(\Upsilon_{i,n})\right)(x) = \left(\mathcal{R}(\Upsilon)\right)\left(\left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x), \left(\mathcal{R}(\Phi_f^{i,n})\right)(x)\right),$$

valid for $x \in D$. In Section 7.3 we show that $\Upsilon_{i,n}$ is a ReLu DNN. Note by triangle inequality that for all $x \in D$

$$\begin{aligned}
&\left| r_{i,n}^{\alpha} E_{M_2}^{i,n} \left(\left(\mathcal{R}(\Phi_f)\right) \left(X_{\mathcal{I}(n-1)}^{i} + r_{i,n} \cdot \right) \right) - \left(\mathcal{R}(\Upsilon_{i,n})\right)(x) \right| \\
&\quad \leq \left| r_{i,n}^{\alpha} \left(E_{M_2}^{i,n} \left(\left(\mathcal{R}(\Phi_f)\right) \left(X_{\mathcal{I}(n-1)}^{i} + r_{i,n} \cdot \right) \right) - \left(\mathcal{R}(\Phi_f^{i,n})\right)(x) \right) \right| \\
&\qquad + \left| \left(\mathcal{R}(\Phi_f^{i,n})\right)(x) \left(r_{i,n}^{\alpha} - \left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x) \right) \right| \\
&\qquad + \left| \left(\mathcal{R}(\Phi_f^{i,n})\right)(x) \left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x) - \left(\mathcal{R}(\Upsilon_{i,n})\right)(x) \right|
\end{aligned}$$

For all $i = 1, ..., M_1$, $n = 1, ..., M_1$ one has $r_n^i < \operatorname{diam}(D)$. From Step 9, the first term can be bounded as

$$\begin{aligned}
&\left| r_{i,n}^{\alpha} \left(E_{M_2}^{i,n} \left(\left(\mathcal{R}(\Phi_f)\right) \left(X_{\mathcal{I}(n-1)}^{i} + r_{i,n} \cdot \right) \right) - \left(\mathcal{R}(\Phi_f^{i,n})\right)(x) \right) \right| \\
&\qquad \leq \operatorname{diam}(D)^{\alpha} \widetilde{L}_f \delta_{\text{dist}} \left(2 \left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1 \right)^{\overline{N}_i} + 1 \right).
\end{aligned}$$

For the second term of the inequality, note that

$$\left| \left(\mathcal{R}(\Phi_f^{i,n})\right)(x) \right| \leq \|\mathcal{R}(\Phi_f)\|_{L^\infty(D)} \leq \delta_f + \|f\|_{L^\infty(D)}.$$

Also, by triangle inequality

$$\begin{aligned}
\left| r_{i,n}^{\alpha} - \left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x) \right| &\leq \left| r_{i,n}^{\alpha} - \left(\mathcal{R}(\Phi_\alpha)\right)(r_{i,n}) \right| \\
&\quad + \left| \left(\mathcal{R}(\Phi_\alpha)\right)(r_{i,n}) - \left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x) \right|.
\end{aligned}$$

From the Hypothesis HD-3 and the fact that $\mathcal{R}(\Phi_\alpha)$ is L_α-Lipschitz one has

$$\left| r_{i,n}^{\alpha} - \left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x) \right| \leq \delta_\alpha + L_\alpha \left| r_{i,n} - \left(\mathcal{R}(\Phi_r^{i,n})\right)(x) \right|$$

And by Step 9 it follows that

$$\begin{aligned}
&\left| \left(\mathcal{R}(\Phi_f^{i,n})\right)(x) \left(r_{i,n}^{\alpha} - \left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x) \right) \right| \\
&\qquad \leq \left(\delta_f + \|f\|_{L^\infty(D)} \right) \left(\delta_\alpha + L_\alpha \delta_{\text{dist}} \left(\left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1 \right)^{\overline{N}_i} + 1 \right) \right).
\end{aligned}$$

Finally, by Lemma 3.6 for all $\delta_\Upsilon \in \left(0, \frac{1}{2}\right)$ the third term can be bounded by

$$\left| \left(\mathcal{R}(\Phi_f^{i,n})\right)(x) \left(\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n})\right)(x) - \left(\mathcal{R}(\Upsilon_{i,n})\right)(x) \right| \leq \delta_\Upsilon.$$

with κ from the Lemma 3.6 equal to

$$\kappa = \max \left\{ 1 + \|f\|_{L^\infty(D)}, 1 + L_\alpha \left(\left(\sum_{i=1}^{\overline{N}_i} |Y_{i,n}| + 1 \right)^{\overline{N}_i} + 1 \right) + \operatorname{diam}(D)^{\alpha} \right\}$$

Therefore

$$\begin{aligned}
&\left|r_{i,n}^{\alpha} E_{M_2}^{i,n}\left((\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}^i + r_{i,n}\cdot\right)\right) - (\mathcal{R}(\Upsilon_{i,n}))(x)\right| \\
&\quad \leq \operatorname{diam}(D)^{\alpha} L_f \delta_{\text{dist}} \left(2\left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\overline{N}_i} + 1\right) \\
&\qquad + \left(\delta_f + \|f\|_{L^\infty(D)}\right)\left(\delta_\alpha + L_\alpha \delta_{\text{dist}}\left(\left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\overline{N}_i} + 1\right)\right) + \delta_\Upsilon. \qquad (7.27) \\
&\quad \leq \delta_\Upsilon + \delta_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right) \\
&\qquad + \delta_{\text{dist}}\left(\operatorname{diam}(D)^{\alpha} L_f + L_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\left(2\left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\overline{N}_i} + 1\right).
\end{aligned}$$

Step 12. For $\widetilde{\varepsilon} \in (0,1)$ define the DNN $\Psi_{2,\widetilde{\varepsilon}} \in \mathbf{N}$ as follows: for any $x \in D$

$$\left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x) = \frac{1}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{\overline{N}_i} \kappa_{d,\alpha}\left(\mathcal{R}\left(\Upsilon_{i,n}\right)\right)(x).$$

This is the requested DNN. See Section 7.3 for the proof that $\Psi_{2,\widetilde{\varepsilon}}$ is indeed a ReLu DNN. By triangle inequality and (7.27) we have

$$\begin{aligned}
&\left|E_{M_1}(x) - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x)\right| \\
&\quad \leq \frac{\kappa_{d,\alpha}}{M_1}\sum_{i=1}^{M_1}\sum_{n=1}^{\overline{N}_i}\left|r_{i,n}^{\alpha} E_{M_2}^{i,n}\left((\mathcal{R}(\Phi_f))\left(X_{\mathcal{I}(n-1)}^i + r_{i,n}\cdot\right)\right) - (\mathcal{R}(\Upsilon_{i,n}))(x)\right|. \\
&\quad \leq \frac{\kappa_{d,\alpha}}{M_1}\left(\delta_\Upsilon + \delta_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\sum_{i=1}^{M_1}\overline{N}_i \\
&\qquad + \frac{\kappa_{d,\alpha}}{M_1}\delta_{\text{dist}}\left(\operatorname{diam}(D)^{\alpha} L_f + L_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\sum_{i=1}^{M_1}\overline{N}_i\left(2\left(\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\overline{N}_i} + 1\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\left|E_{M_1}(x) - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x)\right| \\
&\quad \leq \frac{\kappa_{d,\alpha}}{M_1}\left(\delta_\Upsilon + \delta_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\sum_{i=1}^{M_1}\overline{N}_i \\
&\qquad + \kappa_{d,\alpha}\delta_{\text{dist}}\left(\operatorname{diam}(D)^{\alpha} L_f + L_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\left(\sum_{i=1}^{M_1}\overline{N}_i\right)\ell,
\end{aligned}$$

with $\ell := \left(2\left(\sum_{i=1}^{M_1}\sum_{n=1}^{\overline{N}_i} |Y_{i,n}| + 1\right)^{\sum_{i=1}^{M_1}\overline{N}_i} + 1\right)$.

Step 13. We want to bound error_f. Notice that

$$\mathrm{error}_f \leq 2^{\frac{1}{q}}|D|^{\frac{1}{q}}\delta_f\mathbb{E}_x[\sigma_D] + 2^{1+\frac{1}{q}}|D|^{\frac{1}{q}}\Theta_{q,s}\left(\delta_f + \|f\|_{L^\infty(D)}\right)\left(\frac{\mathbb{E}_x\left[\sigma_D\right]}{M_2^{1-\frac{1}{s}}} + \frac{\mathbb{E}_x\left[\sigma_D^2\right]^{\frac{1}{2}}}{M_1^{1-\frac{1}{s}}}\right)$$
$$+ \frac{2^{1+\frac{1}{q}}\Theta_{q,s}}{M_1^{1-\frac{1}{s}}}\left(1+K(\alpha,q)\right)^{\frac{1}{q}}\left(\frac{2-\widetilde{p}}{\widetilde{p}^2}\right)^{\frac{1}{q}}.$$

Consider now $M \in \mathbb{N}$ and let $M = M_1 = M_2$. Define the constant $\widetilde{C}_1$ as

$$\widetilde{C}_1 = 2^{1+\frac{1}{q}}\Theta_{q,s}\left(|D|^{\frac{1}{q}}\left(1+\|f\|_{L^\infty(D)}\right)\left(\mathbb{E}_x\left[\sigma_D\right] + \mathbb{E}_x\left[\sigma_D^2\right]^{\frac{1}{2}}\right) + \left(1+K(\alpha,q)\right)^{\frac{1}{q}}\left(\frac{2-\widetilde{p}}{\widetilde{p}^2}\right)^{\frac{1}{q}}\right),$$

and the constant $\widetilde{C}_2$ as

$$\widetilde{C}_2 = 2^{\frac{1}{q}}|D|^{\frac{1}{q}}\mathbb{E}_x\left[\sigma_D\right].$$

Therefore

$$\mathrm{error}_f \leq \frac{\widetilde{C}_1}{M^{1-\frac{1}{s}}} + \widetilde{C}_2\delta_f. \tag{7.28}$$

In addition

$$\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}| \leq M\left(\mathrm{error}_f + \mathbb{E}_x\left[\sum_{n=1}^{N}|Y_{i,n}|\right]\right) \leq M\left(\mathrm{error}_f + K(\alpha,1)\frac{1}{\widetilde{p}}\right).$$

Recall that $C_3 = K(\alpha,1)\frac{1}{\widetilde{p}}$. Then

$$\sum_{i=1}^{M}\sum_{n=1}^{\overline{N}_i}|Y_{i,n}| \leq M^{\frac{1}{s}}\widetilde{C}_1 + M\left(\delta_f\widetilde{C}_2 + C_3\right).$$
$$\leq M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_3).$$

Recall that $C_4 = \frac{1}{\widetilde{p}}$, therefore

$$\sum_{i=1}^{M}\overline{N}_i \leq M(\mathrm{error}_f + \mathbb{E}_x[N]) \leq M^{\frac{1}{s}}\widetilde{C}_1 + M\left(\delta_f\widetilde{C}_2 + C_4\right)$$
$$\leq M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_4).$$

From the Step 12 and the estimates of this step it follows that

$$\left|E_{M_1}(x) - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x)\right|$$
$$\leq \kappa_{d,\alpha}\left(\delta_\Upsilon + \delta_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\left(M^{\frac{1}{s}-1}\widetilde{C}_1 + \widetilde{C}_2 + C_4\right)$$
$$+ \kappa_{d,\alpha}\delta_{\mathrm{dist}}\left(\mathrm{diam}(D)^\alpha L_f + L_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\left(M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_4)\right)\widetilde{\ell},$$

where $\widetilde{\ell} := \left(2\left(M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_3) + 1\right)^{M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2+C_4)} + 1\right)$.

Step 14. Lemma 7.2.1, the inequality (7.28), Step 13 and Minkowski inequality ensure that

$$\begin{aligned}
&\left(\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^N r_n^\alpha \kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[f\left(X_{\mathcal{I}(n-1)} + r_n\cdot\right)\right]\right] - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x)\right|^q dx\right)^{\frac{1}{q}} \\
&\leq \left(\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^N r_n^\alpha \kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[f\left(X_{\mathcal{I}(n-1)} + r_n\cdot\right)\right]\right] - E_{M_1}(x)\right|^q dx\right)^{\frac{1}{q}} \\
&\quad + \left(\int_D \left|E_{M_1}(x) - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x)\right|^q dx\right)^{\frac{1}{q}} \\
&\leq \frac{\widetilde{C}_1}{M^{1-\frac{1}{s}}} + \widetilde{C}_2\delta_f + |D|^{\frac{1}{q}}\kappa_{d,\alpha}\left(\delta_\Upsilon + \delta_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\left(M^{\frac{1}{s}-1}\widetilde{C}_1 + \widetilde{C}_2 + C_4\right) \\
&\quad + |D|^{\frac{1}{q}}\kappa_{d,\alpha}\delta_{\text{dist}}\left(\text{diam}(D)^\alpha L_f + L_\alpha\left(\delta_f + \|f\|_{L^\infty(D)}\right)\right)\left(M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_4)\right)\widetilde{\ell}.
\end{aligned}$$

For $\widetilde{\varepsilon} \in (0,1)$, let $M \in \mathbb{N}$ large enough such that

$$M = \left\lceil \left(\frac{5\widetilde{C}_1}{\widetilde{\varepsilon}}\right)^{\frac{s}{s-1}} \right\rceil,$$

and from the choice of M let $\delta_\Upsilon \in \left(0, \frac{1}{2}\right)$ $\delta_{\text{dist}}, \delta_\alpha \in (0,1)$ small enough such that

$$\begin{aligned}
\delta_\Upsilon &= \frac{\widetilde{\varepsilon}}{5|D|^{\frac{1}{q}}\kappa_{d,\alpha}}\left(M^{\frac{1}{s}-1}\widetilde{C}_1 + \widetilde{C}_2 + C_4\right)^{-1}, \\
\delta_\alpha &= \frac{\widetilde{\varepsilon}}{5|D|^{\frac{1}{q}}\kappa_{d,\alpha}}\left(1 + \|f\|_{L^\infty(D)}\right)^{-1}\left(M^{\frac{1}{s}-1}\widetilde{C}_1 + \widetilde{C}_2 + C_4\right)^{-1}, \\
\delta_{\text{dist}} &= \frac{\widetilde{\varepsilon}}{5|D|^{\frac{1}{q}}\kappa_{d,\alpha}\widetilde{\ell}}\left(\text{diam}(D)^\alpha L_f + L_\alpha\left(1 + \|f\|_{L^\infty(D)}\right)\right)^{-1}\left(M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_4)\right)^{-1}.
\end{aligned}$$

Finally we choose $\delta_f \in (0,1)$ small enough such that

$$\delta_f = \frac{\widetilde{\varepsilon}}{5\widetilde{C}_2}.$$

Therefore

$$\left(\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^N r_n^\alpha \kappa_{d,\alpha}\mathbb{E}^{(\mu)}\left[f\left(X_{\mathcal{I}(n-1)} + r_n\cdot\right)\right]\right] - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)(x)\right|^q dx\right)^{\frac{1}{q}} \leq \widetilde{\varepsilon}.$$

We conclude that for all $\widetilde{\varepsilon} \in (0,1)$ there exists $\Psi_{2,\widetilde{\varepsilon}} \in \mathbf{N}$ that approximates (7.1) with accuracy $\widetilde{\varepsilon}$.

7.3. Proof of Proposition 7.1: quantification of DNNs

In this Section we will prove that $\Psi_{2,\widetilde{\varepsilon}}$ is a ReLu DNN such that overcomes the curse of dimensionality.

Step 15. We now study the DNN $\Psi_{2,\widetilde{\varepsilon}}$ with the Definitions and Lemmas of Chapter 3. Let

$$\beta_{\text{dist}} = \mathcal{D}(\Phi_{\text{dist}}) \quad \text{and} \quad H_{\text{dist}} = \dim(\beta_{\text{dist}}) - 2.$$

Recall from Step 12 in the proof of Proposition 6.1 that for all $i = 1, ..., M$

$$\mathcal{D}(\Phi_{i,1}) = d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}, \qquad \dim(\mathcal{D}(\Phi_{i,1})) = H_{\text{dist}} + 2,$$

where

$$\widetilde{\beta}_{\text{dist}} = (\beta_{\text{dist},0}, ..., \beta_{\text{dist},H_{\text{dist}}}, d) \in \mathbb{N}^{H_{\text{dist}}+2},$$

and for all $n = 2, ..., \overline{N}_i$

$$\mathcal{D}(\Phi_{i,n}) = \mathop{\odot}_{m=1}^{n} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}), \quad \dim(\mathcal{D}(\Phi_{i,n})) = (H_{\text{dist}} + 1)n + 1,$$

with

$$|||\mathcal{D}(\Phi_{i,n})||| \leq 2d + |||\mathcal{D}(\Phi_{\text{dist}})|||.$$

Denote

$$\beta_f = \mathcal{D}(\Phi_f) \quad \text{and} \quad H_f = \dim(\beta_f) - 2.$$

Define $\widetilde{\Phi}_{i,j,n} \in \mathbf{N}$ as follows:

$$\mathcal{R}(\widetilde{\Phi}_{i,j,n}) = x + v_{i,j,n}\mathcal{R}(\Phi_{\text{dist}})(x).$$

As similar as in the case of $\Phi_{i,1}$, we have that $\widetilde{\Phi}_{i,j,n}$ is a ReLu DNN such that

$$\mathcal{D}(\widetilde{\Phi}_{i,j,n}) = d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}, \quad \dim(\mathcal{D}(\widetilde{\Phi}_{i,j,n})) = H_{\text{dist}} + 2.$$

Moreover

$$\left|\left|\left|\mathcal{D}(\widetilde{\Phi}_{i,j,n})\right|\right|\right| \leq 2d + |||\mathcal{D}(\Phi_{dist})|||.$$

Using the DNN $\widetilde{\Phi}_{i,j,n}$ we have that

$$\mathcal{R}(\Phi_{i,n-1}) + v_{i,j,n}\mathcal{R}(\Phi_r^{i,n}) = \mathcal{R}(\widetilde{\Phi}_{i,j,n}) \circ \mathcal{R}(\Phi_{i,n-1}).$$

Therefore, by Lemma 3.2 it follows that

$$\mathcal{R}(\Phi_f) \circ \left(\mathcal{R}(\Phi_{i,n}) + v_{i,j,n}\mathcal{R}(\Phi_r^{i,n})\right) \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \beta_f \odot \left(\mathop{\odot}_{m=1}^{n} \left(d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}\right)\right)\right\}\right).$$

with $(H_{\text{dist}} + 1)n + H_f + 2$ the total number of layers. Note this ReLu DNN is continuous from D to $\mathbb{R}$. Let

$$\beta_\alpha = \mathcal{D}(\Phi_\alpha), \quad \text{and} \quad H_\alpha = \dim(\beta_\alpha) - 2.$$

For $n = 1, ..., \overline{N}_i$ we compound the previous DNN with the identity with $(H_{\text{dist}}+1)(\sum_{i=1}^{M_1} \overline{N}_i -$

$n) + H_\alpha + 1$ layers to obtain a DNN $\widehat{\Phi}_{i,j,n} \in \mathbf{N}$ such that

$$\mathcal{R}(\widehat{\Phi}_{i,j,n}) = \mathcal{R}(\Phi_f) \circ \mathcal{R}(\widetilde{\Phi}_{i,j,n}) \circ \mathcal{R}(\Phi_{i,n}),$$

with

$$\mathcal{D}(\widehat{\Phi}_{i,j,n}) = \left(\mathfrak{n}_{(H_{\text{dist}}+1)(\sum_{i=1}^{M_1} \overline{N}_i - n) + H_\alpha + 1}\right) \odot \beta_f \odot \left(\bigodot_{m=1}^{n} \left(d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}\right)\right),$$

and

$$\dim(\mathcal{D}(\widehat{\Phi}_{i,j,n})) = (H_{\text{dist}} + 1) \sum_{i=1}^{M_1} \overline{N}_i + H_\alpha + H_f + 2.$$

Note from the definition of DNN $\Phi_f^{i,n}$ that

$$\mathcal{R}(\Phi_f^{i,n}) = \sum_{j=1}^{M_2} \mathcal{R}(\widehat{\Phi}_{i,j,n}).$$

Therefore, Lemma 3.3 implies that $\Phi_f^{i,n}$ is a ReLu DNN with

$$\mathcal{D}(\Phi_f^{i,n}) = \boxplus_{j=1}^{M_2} \left(\mathfrak{n}_{(H_{\text{dist}}+1)(\sum_{i=1}^{M_1} \overline{N}_i - n) + H_\alpha + 1} \odot \beta_f \odot \left(\bigodot_{m=1}^{n} \left(d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}\right)\right)\right),$$

and

$$\dim(\mathcal{D}(\Phi_f^{i,n})) = (H_{\text{dist}} + 1) \sum_{i=1}^{M_1} \overline{N}_i + H_\alpha + H_f + 2.$$

Moreover

$$\left\|\!\left|\mathcal{D}(\Phi_f^{i,n})\right|\!\right\| \leq \sum_{j=1}^{M_2} \max\{\|\!|\mathcal{D}(\Phi_f)|\!\|, 2d + \|\!|\mathcal{D}(\Phi_{\text{dist}})|\!\|\} = M_2 \max\{\|\!|\mathcal{D}(\Phi_f)|\!\|, 2d + \|\!|\mathcal{D}(\Phi_{\text{dist}})|\!\|\}.$$

On the other hand side, note that

$$\mathcal{D}(\Phi_r^{i,n}) = \beta_{\text{dist}} \odot \mathcal{D}(\Phi_{i,n-1}), \quad \dim(\mathcal{D}(\Phi_r^{i,n})) = (H_{\text{dist}} + 1)n + 1,$$

and

$$\left\|\!\left|\mathcal{D}(\Phi_r^{i,n})\right|\!\right\| \leq \max\{2d, \|\!|\mathcal{D}(\Phi_{\text{dist}})|\!\|, 2d + \|\!|\mathcal{D}(\Phi_{\text{dist}})|\!\|\} = 2d + \|\!|\mathcal{D}(\Phi_{\text{dist}})|\!\|.$$

Therefore by Lemma 3.2

$$\mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n}) \in \mathcal{R}\left(\left\{\Phi \in \mathbf{N} : \mathcal{D}(\Phi) = \beta_\alpha \odot \beta_{\text{dist}} \odot \left(\bigodot_{m=1}^{n} \left(d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}\right)\right)\right\}\right),$$

with $(H_{\text{dist}} + 1)n + H_\alpha + 2$ number of layers. Like before, we compound the previous DNN with the identity with $(H_{\text{dist}} + 1)(\sum_{i=1}^{M_1} \overline{N}_i - n) + H_f + 1$ to obtain, by Lemma 3.2 a DNN $\widehat{\Phi}_{i,n} \in \mathbf{N}$ such that

$$\mathcal{R}(\widehat{\Phi}_{i,n}) = \mathcal{R}(\Phi_\alpha) \circ \mathcal{R}(\Phi_r^{i,n}),$$

with

$$\mathcal{D}(\widehat{\Phi}_{i,n}) = \mathfrak{n}_{(H_{\text{dist}}+1)(\sum_{i=1}^{M_1} \overline{N}_i - n) + H_f + 1} \odot \beta_\alpha \odot \beta_{\text{dist}} \odot \left(\bigodot_{m=1}^{n} \left(d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}}\right)\right),$$

and

$$\dim(\mathcal{D}(\widehat{\Phi}_{i,n})) = (H_{\text{dist}}+1)\sum_{i=1}^{M_1}\overline{N}_i + H_\alpha + H_f + 2.$$

Moreover

$$\left|\!\left|\!\left|\mathcal{D}(\widehat{\Phi}_{i,n})\right|\!\right|\!\right| \leq \max\{|\!|\!|\mathcal{D}(\Phi_\alpha)|\!|\!|, 2d + |\!|\!|\mathcal{D}(\Phi_{\text{dist}})|\!|\!|\}.$$

Define $H \in \mathbb{N}$ as

$$H = (H_{\text{dist}}+1)\sum_{i=1}^{M_1}\overline{N}_i + H_\alpha + H_f.$$

We now realize a parallelization between the DNNs $\widehat{\Phi}_{i,n}$ and $\Phi_f^{i,n}$. By Lemma 3.4, there exists a ReLu DNN $\overline{\Phi}_{i,n} \in \mathbf{N}$ such that

$$\mathcal{R}(\overline{\Phi}_{i,n}) = (\mathcal{R}(\widehat{\Phi}_{i,n}), \mathcal{R}(\Phi_f^{i,n})),$$

with

$$\mathcal{D}(\overline{\Phi}_{i,n}) = \mathcal{D}(\widehat{\Phi}_{i,n}) \boxplus \mathcal{D}(\Phi_f^{i,n}) + e_{H+2},$$

and

$$\dim(\mathcal{D}(\overline{\Phi}_{i,n})) = (H_{\text{dist}}+1)\sum_{i=1}^{M_1}\overline{N}_i + H_\alpha + H_f + 2,$$

where

$$e_{H+2} = (0,...,0,1) \in \mathbb{R}^{H+2}.$$

Moreover,

$$\left|\!\left|\!\left|\mathcal{D}(\overline{\Phi}_{i,n})\right|\!\right|\!\right| \leq \left|\!\left|\!\left|\mathcal{D}(\widehat{\Phi}_{i,n})\right|\!\right|\!\right| + \left|\!\left|\!\left|\mathcal{D}(\Phi_f^{i,n})\right|\!\right|\!\right|,$$

and thus

$$\left|\!\left|\!\left|\mathcal{D}(\overline{\Phi}_{i,n})\right|\!\right|\!\right| \leq |\!|\!|\mathcal{D}(\Phi_\alpha)|\!|\!| + M_2|\!|\!|\mathcal{D}(\Phi_f)|\!|\!| + (M_2+1)(2d + |\!|\!|\mathcal{D}(\Phi_{\text{dist}})|\!|\!|).$$

Let

$$\beta_\Upsilon = \mathcal{D}(\Upsilon), \quad \text{and} \quad H_\Upsilon = \dim(\beta_\Upsilon) + 2.$$

Therefore, by Lemma 3.2 it follows that

$$\mathcal{D}(\Upsilon_{i,n}) = \beta_\Upsilon \odot \left(\mathcal{D}(\widehat{\Phi}_{i,n}) \boxplus \mathcal{D}(\Phi_f^{i,n}) + e_{H+2}\right),$$

and

$$\dim(\mathcal{D}(\Upsilon_{i,n})) = H + H_\Upsilon + 3.$$

Moreover

$$\begin{aligned}|\!|\!|\mathcal{D}(\Upsilon_{i,n})|\!|\!| &\leq \max\{|\!|\!|\mathcal{D}(\Upsilon)|\!|\!|, \left|\!\left|\!\left|\mathcal{D}(\overline{\Phi}_{i,n})\right|\!\right|\!\right|\}\\ &\leq |\!|\!|\mathcal{D}(\Upsilon)|\!|\!| + |\!|\!|\mathcal{D}(\Phi_\alpha)|\!|\!| + M_2|\!|\!|\mathcal{D}(\Phi_f)|\!|\!| + (M_2+1)(2d + |\!|\!|\mathcal{D}(\Phi_{\text{dist}})|\!|\!|).\end{aligned}$$

Finally, from Lemma 3.3 it follows that

$$\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}}) = \boxplus_{i=1}^{M_1} \boxplus_{n=1}^{\overline{N}_i} \left(\beta_\Upsilon \odot \left(\mathcal{D}(\widehat{\Phi}_{i,n}) \boxplus \mathcal{D}(\Phi_f^{i,n}) + e_{H+2}\right)\right),$$

and

$$\dim(\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})) = H + H_\Upsilon + 3.$$

Moreover

$$\left\|\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})\right\| \leq \Big(|||\mathcal{D}(\Upsilon)|||+|||\mathcal{D}(\Phi_\alpha)|||+M_2|||\mathcal{D}(\Phi_f)|||+(M_2+1)(2d+|||\mathcal{D}(\Phi_{\text{dist}})|||)\Big)\sum_{i=1}^{M_1}\overline{N}_i. \quad (7.29)$$

Notice from the definition of $\widetilde{C}_1$ and $\widetilde{C}_2$ that both constants are multiple of $|D|^{\frac{1}{q}}$. Therefore, by choice of M_1 and M_2 we have

$$M_2 = M_1 \leq B_1|D|^{\frac{s}{q(s-1)}}\widetilde{\varepsilon}^{-\frac{s}{s-1}}, \quad (7.30)$$

where $B_1 > 0$ is a generic constant. The choice of δ_f and the constant $\widetilde{C}_2$ implies that

$$\delta_f^{-a} \leq B_2|D|^{\frac{a}{q}}\widetilde{\varepsilon}^{-a}, \quad (7.31)$$

for some constant $B_2 > 0$. From the choice of δ_Υ, δ_α and (7.30) it follows that

$$\log\Big(\delta_\Upsilon^{-1}\Big) \leq \delta_\Upsilon^{-1} \leq B_3|D|^{\frac{2}{q}}\widetilde{\varepsilon}^{-1}, \quad (7.32)$$

and

$$\delta_\alpha^{-a} \leq B_4|D|^{\frac{2a}{q}}\widetilde{\varepsilon}^{-a}. \quad (7.33)$$

where $B_3, B_4 > 0$ are a generic constant, and from the choice of δ_{dist} and properties of Logarithm function we have

$$\log\Big(\delta_{\text{dist}}^{-1}\Big) \leq 5|D|^{\frac{1}{q}}\kappa_{d,\alpha}\Big(\text{diam}(D)^\alpha L_f + L_\alpha\Big(1+\|f\|_{L^\infty(D)}\Big)\Big)\Big(M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_4)\Big)\widetilde{\varepsilon}^{-1}$$
$$+ 4\Big(M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_4)\Big)\Big(M^{\frac{1}{s}}\widetilde{C}_1 + M(\widetilde{C}_2 + C_3) + 1\Big).$$

Therefore from (7.30)

$$\lceil\log\Big(\delta_{\text{dist}}^{-1}\Big)\rceil^a \leq B_5|D|^{\frac{2a}{q}\left(1+\frac{s}{s-1}\right)}\widetilde{\varepsilon}^{-a-\frac{2as}{s-1}}, \quad (7.34)$$

for some $B_5 > 0$ generic. Note also that

$$\sum_{i=1}^{M_1}\overline{N}_i \leq B_6|D|^{\frac{1}{q}\left(1+\frac{s}{s-1}\right)}\widetilde{\varepsilon}^{-\frac{s}{s-1}}, \quad (7.35)$$

with $B_6 > 0$. Finally from Assumptions 2, 3 and inequalities (7.29), (7.35) we got

$$\left\|\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})\right\|$$
$$\leq B_6|D|^{\frac{1}{q}\left(1+\frac{s}{s-1}\right)}\widetilde{\varepsilon}^{-\frac{s}{s-1}}\Big(\log\Big(\delta_\Upsilon^{-1}\Big) + \delta_\alpha^{-a} + M_2 B d^b \delta_f^{-a} + (M_2+1)(2d + Bd^b\lceil\log\Big(\delta_{\text{dist}}^{-1}\Big)\rceil^a)\Big).$$

Finally, from inequalities (7.30), (7.31), (7.32), (7.33) and (7.34) we conclude that there exists $\widetilde{B} > 0$ such that

$$\left\|\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})\right\| \leq \widetilde{B}|D|^{\frac{1}{q}\left(1+2a+\frac{2s}{s-1}(1+a)\right)}d^b\widetilde{\varepsilon}^{-a-\frac{2s}{s-1}(1+a)}.$$

This completes the proof of Proposition 7.1.

Chapter 8

Proof of the Main Result

This final Chapter is devoted to the proof of Theorem 1.7.1. Gathering Propositions 6.1 and 7.1, Theorem 1.7.1 is finally proved.

Step 1. Let $\alpha \in (1,2)$, $p, s \in (1,\alpha)$ such that $s < \frac{\alpha}{p}$ and $q \in \left[s, \frac{\alpha}{p}\right)$. Let Assumptions (Hg-0) and (Hf-0) be satisfied. Recall from Theorem 5.1 and Lemma 5.1 that the solution u of (1.2) takes the form of (5.8), namely

$$u(x) = \mathbb{E}_x \left[g\left(X_{\mathcal{I}(N)} \right) \right] + \mathbb{E}_x \left[\sum_{n=1}^{N} r_n^{\alpha} \kappa_{d,\alpha} \mathbb{E}^{(\mu)} \left[f\left(X_{\mathcal{I}(n-1)} + r_n \cdot \right) \right] \right], \qquad x \in D.$$

From Propositions 6.1 and 7.1 for all $\varepsilon, \widetilde{\varepsilon} \in (0,1)$ there exist ReLu DNNs $\Psi_{1,\varepsilon}$ and $\Psi_{2,\widetilde{\varepsilon}}$ that satisfy (6.4) and (7.4). For the right approximation of the ReLu DNNs, δ_{dist} and M will be defined as

$$\delta_{\text{dist}} \leq \min\{\ell_1, \ell_2\}, \qquad M \geq \max\left\{ \left\lceil \left(\frac{5C_1}{\varepsilon} \right)^{\frac{s}{s-1}} \right\rceil, \left\lceil \left(\frac{5\widetilde{C}_1}{\widetilde{\varepsilon}} \right)^{\frac{s}{s-1}} \right\rceil \right\},$$

where

$$\ell_1 := \frac{\varepsilon}{5|D|^{\frac{1}{q}} L_g} \left(1 + M^{\frac{1}{s}} C_1 + M(C_2 + C_3) \right)^{-\left(M^{\frac{1}{s}} C_1 + M(C_2 + C_4) \right)},$$

and

$$\ell_2 := \frac{\widetilde{\varepsilon}}{5|D|^{\frac{1}{q}} \widetilde{\ell}} \left(\operatorname{diam}(D)^{\alpha} L_f + L_{\alpha} \left(1 + \|f\|_{L^{\infty}(D)} \right) \right)^{-1} \left(M^{\frac{1}{s}} \widetilde{C}_1 + M(\widetilde{C}_2 + C_4) \right)^{-1}.$$

Recall that the constants in ℓ_1 and ℓ_2 are defined in Propositions 6.1 and 7.1. Let $\epsilon \in (0,1)$ and define the ReLu DNN Ψ_{ϵ} that satisfies for all $x \in D$

$$\left(\mathcal{R}(\Psi_{\epsilon}) \right)(x) = \left(\mathcal{R}(\Psi_{1,\varepsilon}) \right)(x) + \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}}) \right)(x),$$

where $\varepsilon = \widetilde{\varepsilon} = \frac{\epsilon}{2}$. From Minkowski inequality one has

$$\begin{aligned}
&\left(\int_D |u(x) - (\mathcal{R}(\Psi_\epsilon))(x)|^q \, dx\right)^{\frac{1}{q}} \\
&\quad \leq \left(\int_D \left|\mathbb{E}_x\left[g\left(X_{\mathcal{I}(N)}\right)\right] - (\mathcal{R}(\Psi_{1,\varepsilon}))(x)\right|^q dx\right)^{\frac{1}{q}} \\
&\quad\quad + \left(\int_D \left|\mathbb{E}_x\left[\sum_{n=1}^{N} r_n^\alpha \kappa_{d,\alpha} \mathbb{E}^{(\mu)}\left[f(X_{\mathcal{I}(n-1)} + r_n \cdot)\right]\right] - \left(\mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})\right)\right|^q dx\right)^{\frac{1}{q}} \\
&\quad \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

Step 2 We now study the ReLu DNN Ψ_ϵ. For $i = 1, ..., M$ Let $\overline{N}_{i,1}, \overline{N}_{i,2}$ the random variables $\overline{N}_i$ found in the Propositions 6.1 and 7.1, respectively. Recall that the DNN $\Psi_{1,\varepsilon}$ satisfy

$$\mathcal{D}(\Psi_{1,\varepsilon}) = \mathop{\boxplus}_{i=1}^{M} \left(\mathfrak{n}_{H_i+2} \odot \beta_g \odot \left(\mathop{\odot}_{m=1}^{\overline{N}_{i,1}} (d\mathfrak{n}_{H_{\text{dist}}+2} \boxplus \widetilde{\beta}_{\text{dist}})\right)\right),$$

where H_i is defined as

$$H_i = (H_{\text{dist}} + 1)\left(\sum_{j=1}^{M} \overline{N}_{j,1} - \overline{N}_{i,1}\right) - 1.$$

and

$$\dim(\mathcal{D}(\Psi_{1,\varepsilon})) = (H_{\text{dist}} + 1)\sum_{i=1}^{M} \overline{N}_{i,1} + H_g + 2.$$

Recall also that $\Psi_{2,\widetilde{\varepsilon}}$ satisfy

$$\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}}) = \mathop{\boxplus}_{i=1}^{M_1} \mathop{\boxplus}_{n=1}^{\overline{N}_{i,2}} \left(\beta_\Upsilon \odot \left(\mathcal{D}(\widehat{\Phi}_{i,n}) \boxplus \mathcal{D}(\Phi_f^{i,n}) + e_{H+2}\right)\right),$$

where $e_{H+2} = (0, ..., 0, 1) \in \mathbb{R}^{H+2}$ with H defined as

$$H = (H_{\text{dist}} + 1)\sum_{i=1}^{M_1} \overline{N}_{i,2} + H_\alpha + H_f,$$

and

$$\dim(\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})) = H + H_\Upsilon + 3.$$

To use Lemma 3.3, the ReLu DNNs $\Psi_{1,\varepsilon}$ and $\Psi_{2,\widetilde{\varepsilon}}$ must have the same number of layers. We compound each DNN by a suitable ReLu DNN that represents the identity function. Define then the ReLu DNN $\overline{\Psi}_{1,\varepsilon}$ that satisfy $\mathcal{R}(\overline{\Psi}_{1,\varepsilon}) = \mathcal{R}(\Psi_{1,\varepsilon})$ with

$$\mathcal{D}(\overline{\Psi}_{1,\varepsilon}) = \mathfrak{n}_{H+H_\Upsilon+3} \odot \mathcal{D}(\Psi_{1,\varepsilon}),$$

and Define the ReLu DNN $\overline{\Psi}_{2,\widetilde{\varepsilon}}$ that satisfy $\mathcal{R}(\overline{\Psi}_{2,\widetilde{\varepsilon}}) = \mathcal{R}(\Psi_{2,\widetilde{\varepsilon}})$ with

$$\mathcal{D}(\overline{\Psi}_{2,\widetilde{\varepsilon}}) = \mathfrak{n}_{(H_{\text{dist}}+1)\sum_{i=1}^{M}\overline{N}_{i,1}+H_g+2} \odot \mathcal{D}(\Psi_{2,\widetilde{\varepsilon}}),$$

Therefore we have that $\dim(\mathcal{D}(\overline{\Psi}_{1,\varepsilon})) = \dim(\mathcal{D}(\overline{\Psi}_{2,\widetilde{\varepsilon}}))$. Moreover

$$\dim(\mathcal{D}(\overline{\Psi}_{1,\varepsilon})) = (H_{\text{dist}}+1)\sum_{i=1}^{M}(\overline{N}_{i,1}+\overline{N}_{i,2}) + H_\alpha + H_f + H_g + H_\Upsilon + 4.$$

Therefore we can use Lemma 3.3 to obtain that Ψ_ϵ is a ReLu DNN such that

$$\mathcal{D}(\Psi_\epsilon) = \mathcal{D}(\overline{\Psi}_{1,\varepsilon}) \boxplus \mathcal{D}(\overline{\Psi}_{2,\widetilde{\varepsilon}}),$$

and

$$\dim(\mathcal{D}(\Psi_\epsilon)) = (H_{\text{dist}}+1)\sum_{i=1}^{M}(\overline{N}_{i,1}+\overline{N}_{i,2}) + H_\alpha + H_f + H_g + H_\Upsilon + 4.$$

Moreover

$$|||\mathcal{D}(\Psi_\epsilon)||| \leq |||\mathcal{D}(\Psi_{1,\varepsilon})||| + \left|\left|\left|\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})\right|\right|\right|.$$

Recall from Propositions 6.1 and 7.1 that there exists $\widetilde{B} > 0$ such that

$$|||\mathcal{D}(\Psi_{1,\varepsilon})||| \leq \widetilde{B}|D|^{\frac{1}{q}\left(2a+ap+\frac{s}{s-1}(1+2a+ap)\right)} d^{b+2ap+2ap^2+\frac{ps}{s-1}(1+2a+ap)} \varepsilon^{-a-\frac{s}{s-1}(1+2a+ap)}.$$

and

$$\left|\left|\left|\mathcal{D}(\Psi_{2,\widetilde{\varepsilon}})\right|\right|\right| \leq \widetilde{B}|D|^{\frac{1}{q}\left(1+2a+\frac{2s}{s-1}(1+a)\right)} d^b \widetilde{\varepsilon}^{-a-\frac{2s}{s-1}(1+a)}.$$

Therefore

$$|||\mathcal{D}(\Psi_\epsilon)||| \leq \widehat{B}|D|^{\frac{1}{q}\left(1+2a+ap+\frac{s}{s-1}(2+2a+ap)\right)} d^{b+2ap+ap^2+\frac{ps}{s-1}(1+2a+ap)} \epsilon^{-a-\frac{s}{s-1}(2+2a+ap)},$$

where $\widehat{B} > 0$ is a generic constant. Theorem 1.7.1 can be concluded choosing $\eta > 0$ as the maximum between $\frac{1}{q}\left(1+2a+ap+\frac{s}{s-1}(2+2a+ap)\right)$, $b+2ap+ap^2+\frac{ps}{s-1}(1+2a+ap)$ and $\frac{s}{s-1}(2+2a+ap)$.

Chapter 9

Conclusions

9.1. Findings

In this work we establish DNNs approximations for the fractional Dirichlet Problem for $\alpha \in (1,2)$, with arbitrary accuracy, and avoiding the curse of dimensionality.

As stated before, for the approximation of the solutions of Problem (1.2) we followed the ideas presented in the work of Grohs and Herrmann [17], with several changes due the non local nature of the fractional Laplacian. In particular:

1. The non local problem has the boundary condition g defined on the complement of the domain D and the local problem has g defined on ∂D. This variation changes the way to approximates g by DNNs. The Assumption 1 is classical in the literature for functions defined in unbounded sets (see, e.g. [23]).

2. The isotropic α-stable process associated to the fractional Laplacian has no second moment, therefore the approximation can not be approximated in $L^2(D)$, but in $L^q(D)$ for some suitable $q < 2$.

3. The process associated to the local case is a Brownian motion that is continuous, then the norm of the process exiting the unit ball centered at the origin is equal to 1, i.e, $|X_{\sigma_{B(0,1)}}| = 1$. In the non local case the isotropic α-stable is a pure jump process, therefore $|X_{\sigma_{B(0,1)}}| > 1$. This is the reason because in our proof we approximate the copies of this random variable to have that the copies of $X_{\sigma_{B(0,1)}}$ exits near the domain D.

4. The last notable difference is about the sum that appears in (1.4): The value of r_n is raised to the power of α, then we need an extra hypothesis for the approximation of the function $(\cdot)^\alpha$ by DNNs. In the local case r_n is squared, then can be approximated by DNNs using the Lemma 3.6 stated in Chapter 3.

Finally, just before finishing this work, we learned about the research by Changtao Sheng et al. [30], who have showed numerical simulations of the Problem (1.2) using similar Monte Carlo methods. However, our methods are radically different in terms of the main goal, which is here to approximate the solution by DNNs in a rigorous fashion.

9.2. Perspectives

For the proof of main theorem and the Assumptions used in this work, the condition $\alpha > 1$ was required. We do not discard that in the case of $\alpha \in (0, 1)$ there exists a DNN that approximates the solution with arbitrary accuracy, but unfortunately here we do not deal with this case. We expect to be able to treat this case in a future work.

We also remark that the solution approximated is the unique continuous solution for the fractional Dirichlet Problem (1.2), but there exist blow up solutions for a similar problem, studied in [1]. In order to obtain these solutions, an additional boundary condition h is required on ∂D. The condition g in Problem (1.2) therefore has support on $D^c \setminus \partial D$, and the solution studied in [1] has a representation from the Green function and the boundary conditions. A possible extension of this thesis is to prove that the solutions exposed in [1] can be approximated by deep neural networks overcoming the curse of dimensionality.

A similar way to represent solutions to the Problem (1.2) is found in the work of Gulian and Pang [20]. They worked with the spectral fractional Laplacian, and thanks to stochastic calculus results (see, e.g. [3, 8]), the processes described in that work and the isotropic α-stable processes are similar. In addition, in that paper they found a Feynman-Kac formula for the parabolic generalized problem for the associated fractional Laplacian.

In a possible extension we could see if the parabolic generalized problem can be adapted to our setting, i.e. the solution of the parabolic case can be approximated by DNNs that overcome the curse of dimensionality. Even better, following the results by Topp and collaborators, we may extend these results for a fractional Problem with more general fractional operators.

Bibliografía

[1] N. Abatangelo. *Large s-harmonic functions and boundary blow-up solutions for the fractional Laplacian.* arXiv preprint arXiv:1310.3193, 2013.

[2] G. Allaire. *Analyse numérique et optimisation.* Editions de l'Ecole Polytechnique, Paris, 2006.

[3] D. Applebaum. *Lévy Processes and Stochastic Calculus, 2nd ed.* Cambridge Studies in Advanced Mathematics. doi:10.1017/CBO9780511809781, 2009.

[4] G. Acosta and J. P. Borthagaray. *A fractional Laplace equation: regularity of solutions and finite element approximations.* SIAM Journal on Numerical Analysis, 55(2), 472-495, 2017.

[5] C. Beck, S. Becker, P. Grohs, N. Jaafari and A. Jentzen. *Solving the Kolmogorov PDE by Means of Deep Learning.* ournal of Scientific Computing, vol. 88, no. 3, July 2021. https://doi.org/10.1007/s10915-021-01590-0.

[6] C. Beck and A. Jentzen. *Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations.* Journal of Nonlinear Science, 29(4), 1563-1619, 2019.

[7] J. Berner, P. Grohs and A. Jentzen. *Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations.* SIAM Journal on Mathematics of Data Science, 2(3), 631-657, 2020.

[8] J. Bertoin. *Lévy Processes.* Volume 121 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge. ISBN 0-521-56243-0, 1996.

[9] R. Blumenthal, R. Getoor and D. Ray. *On the Distribution of First Hits for the Symmetric Stable Processes.* Transactions of the American Mathematical Society, 99(3), 540–554, 1961. https://doi.org/10.2307/1993561.

[10] A. Bonito, J. P. Borthagaray, R. H. Nochetto, E. Otárola and A. J. Salgado. *Numerical methods for fractional diffusion.* Computing and Visualization in Science, 19(5), 19-46, 2018.

[11] L. Caffarelli, and L. Silvestre, *An extension problem related to the fractional Laplacian*, Comm. Partial Differential Equations 32 (2007), no. 7-9, 1245–1260.

[12] S. Cox, M. Hutzenthaler, A. Jentzen, J. van Neerven and T. Welti. *Convergence in Hölder norms with applications to Monte Carlo methods in infinite dimensions.* IMA Journal of Numerical Analysis, 41(1), 493-548, 2021.

[13] J, Dutka. *The incomplete Beta function—a historical profile.* Archive for history of exact sciences, 11-29, 1981.

[14] D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei. *Deep Neural Network Approximation Theory*. IEEE Transactions on Information Theory, vol. 67, no. 5, May 2021, pp. 2581–623. https://doi.org/10.1109/tit.2021.3062161.

[15] R. Getoor. *First Passage Times for Symmetric Stable Processes in Space*. Transactions of the American Mathematical Society, 101(1), 75–90, 1961. https://doi.org/10.2307/1993412

[16] L. Gonon and C. Schwab. *Deep ReLU neural networks overcome the curse of dimensionality for partial integrodifferential equations*. arXiv preprint arXiv:2102.11707, 2021.

[17] P. Grohs and L. Herrmann. *Deep neural network approximation for high-dimensional elliptic PDEs with boundary conditions*. IMA Journal of Numerical Analysis, 2021. https://doi.org/10.1093/imanum/drab031.

[18] P. Grohs, F. Hornung, A. Jentzen, and P. Von Wurstemberger. *A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations*. arXiv preprint arXiv:1809.02362, 2018.

[19] P. Grohs, A. Jentzen and D. Salimova. *Deep neural network approximations for Monte Carlo algorithms.* arXiv preprint arXiv:1908.10828, 2019.

[20] M. Gulian, and G. Pang. *Stochastic solution of elliptic and parabolic boundary value problems for the spectral fractional Laplacian.* arXiv preprint arXiv:1812.01206, 2018.

[21] M. Gulian„ M. Raissi, P. Perdikaris and G. Karniadakis. *Machine Learning of Space-Fractional Differential Equations.* SIAM Journal on Scientific Computing, vol. 41, no. 4, Jan. 2019, pp. A2485–509. https://doi.org/10.1137/18m1204991.

[22] Kurt Hornik et al., *Approximation Capabilities of Multilayer Feedforward Networks*, Neural Networks, Vol. 4, pp. 251-257. 1991.

[23] M. Hutzenthaler, A. Jentzen, Thomas Kruse, Tuan Anh Nguyen, *A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations*, Partial Differ. Equ. Appl. 1, 10 (2020).

[24] A. Jentzen, D. Salimova, and T. Welti. *A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients.* Communications in Mathematical Sciences, vol. 19, no. 5, 2021, pp. 1167–205. https://doi.org/10.4310/cms.2021.v19.n5.a1.

[25] A. E Kyprianou, A. Osojnik and T. Shardlow, *Unbiased 'walk-on-spheres' Monte Carlo methods for the fractional Laplacian*, IMA Journal of Numerical Analysis, Volume 38, Issue 3, July 2018, pp 1550–1578, https://doi.org/10.1093/imanum/drx042.

[26] A. Kyprianou and J. Pardo. *Stable Lévy Processes via Lamperti-Type Representations* (Institute of Mathematical Statistics Monographs). Cambridge: Cambridge University Press, 2021.

[27] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: isoperimetry and processes* (Vol. 23). Springer Science & Business Media, 1991.

[28] A. Lischke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, ... and G. E. Karniadakis. *What is the fractional Laplacian?*. arXiv preprint arXiv:1801.09767, 2018.

[29] R. Schilling. *An Introduction to Lévy and Feller Processes.* Advanced Courses in

Mathematics-CRM Barcelona 2014. arXiv preprint arXiv:1603.00251, 2016.

[30] C. Sheng, B. Su and C. Xu. *Efficient Monte Carlo method for integral fractional Laplacian in multiple dimensions*. arXiv preprint, arXiv:2204.08860, 2022.

[31] D. J. Smith, and M. K. Vamanamurthy. *How Small Is a Unit Ball?*. Mathematics Magazine, 62(2), 101–107, 1989. https://doi.org/10.2307/2690391.

[32] W. E, J. Han and A. Jentzen. *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*. Commun. Math. Stat. 5(4), 349–380, 2017.

[33] W.E, J. Han and A. Jentzen. *Solving high-dimensional partial differential equations using deep learning*. Proc. Natl. Acad. Sci. 115(34), 8505–8510, 2018.