OPTIMAL CONTROL PROBLEMS CONSTRAINED BY FRACTIONAL PDES AND APPLICATION TO DEEP NEURAL NETWORKS

by

Deepanshu Verma
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Mathematics

Committee:

______________________________	Dr. Harbir Antil, Committee Chair
______________________________	Dr. Mahamadi Warma, Committee Member
______________________________	Dr. Carlos Rautenberg, Committee Member
______________________________	Dr. Andrei Draganescu, Committee Member
______________________________	Dr. Denis Ridzal, Committee Member
______________________________	Dr. David Walnut, Department Chair
______________________________	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
______________________________	Dr. Fernando Miralles-Wilhelm, Dean, College of Science

Date: ______________________________ Summer Semester 2021
George Mason University
Fairfax, VA

Optimal Control Problems Constrained by Fractional PDEs
and Application to Deep Neural Networks

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Deepanshu Verma
Master of Science
Indian Institute of Technology Bombay, India, 2018
Bachelor of Science
SGTB Khalsa College, Delhi University, India, 2015

Director: Dr. Harbir Antil
Director, Center for Mathematics and Artificial Intelligence (CMAI)
Associate Professor, Department of Mathematical Sciences

Summer Semester 2021
George Mason University
Fairfax, VA

Dedication

For my parents, brother and wife..!

Acknowledgments

First and foremost, I would like to express my deep gratitude for my supervisor Dr. Harbir Antil for his valuable guidance and constant support throughout my study. With his adept knowledge, caring advise and gentle words of encouragement, he made sure that I stayed on track during these years. I thank him for always finding time despite a busy schedule to discuss and review my work multiple times thereby ensuring that I was able to achieve his towering standards. He taught me the true meaning of perseverance.

I would like to thank my dissertation committee members, Dr. Mahamadi Warma, Dr. Carlos Rautenberg, Dr. Andrei Draganescu and Dr. Denis Ridzal, for their support and guidance throughout this study.

I am grateful to Dr. Howard C. Elman, Dr. Rainald Löhner, Dr. Neela Nataraj, Dr. Carlos Rautenberg and Dr. Mahamadi Warma, for giving me the opportunity to collaborate with them and expand the horizon of my research area. I must also thank Dr. Thomas Brown, Dr. Akwum Onwunta, Dr. Ratna Khatri and Rafael Arndt for their collaboration and fruitful mathematical discussions during the study.

I would also like to express my gratitude to the faculty and staff at Department of Mathematical Sciences, George Mason University and Department of Mathematics, IIT Bombay for imparting me with the knowledge I have been using extensively.

I would also like to thank all my colleagues and friends at Department of Mathematical Sciences, George Mason University and Department of Mathematics, IIT Bombay who contributed to my mental well-being and achieving this goal.

Special thanks to my parents, brother and wife for their love and support.

I also acknowledge the support from Office of the Provost at George Mason University for the Presidential Fellowship, NSF grants DMS-1818772, DMS-1913004, DMS-1521590 and the Air Force Office of Scientific Research Award NO: FA9550-19-1-0036.

Table of Contents

List of Tables

List of Figures

Abstract

OPTIMAL CONTROL PROBLEMS CONSTRAINED BY FRACTIONAL PDES AND APPLICATION TO DEEP NEURAL NETWORKS

Deepanshu Verma, PhD

George Mason University, 2021

Dissertation Director: Dr. Harbir Antil

Motivated by several applications in geophysics, imaging, machine learning, elasticity, finance, anomalous diffusion, etc., this thesis develops algorithms to solve optimal control problems constrained by fractional partial differential equations (FPDEs). It also introduces a novel fractional derivative based Deep Neural Networks (DNNs) to efficiently tackle inverse problems. Fractional operators have recently emerged as an excellent modeling alternative to their classical counterparts. This success can be attributed to the facts that these operators allow long range interactions (nonlocal), they can account for memory information and finally they enforce less smoothness than their classical counterparts. By exploiting all these features, the thesis develops several novel mathematical tools and algorithms which are of wider interest. For example, the notions of weak and very-weak solutions to fractional elliptic and parabolic problems have been introduced. Existence, and higher regularity, of solutions to fractional Dirichlet problems with measure-valued datum has been established. Moreau-Yosida regularization based algorithms have been introduced to solve fractional state constrained optimal control problems.

The thesis begins by introducing, a new notion of optimal control. In particular, in the parabolic setting, we establish that it is possible to have an external optimal control. Recall

that the classical models only allow control placement either on the boundary or inside the domain. A complete analysis of the Dirichlet and Robin optimal control problems, with constraints on the control, has been provided and the presented numerical results confirms the theoretical findings.

In addition to the control constraints, obstacle type constraints on the PDE solution naturally arise in many different applications. To tackle this, the thesis introduces novel state constrained optimal control problems with fractional PDEs as constraints. One of the key challenges here is that the Lagrange multiplier corresponding to the state constraints is a signed Radon measure, this results in a low regularity for the adjoint solution. A complete analysis for this problem has been provided. This is followed by a Moreau-Yosida regularization based algorithm to solve both the elliptic and parabolic optimal control problems. Here convergence analysis (with rates) of the regularized solutions to the original one is established. Next, a finite element method is introduced and discretization error estimates are established in the elliptic setting. Theoretical results are substantiated by numerical experiments.

Finally, in recent years, deep learning has emerged as the method of choice for classification problems. However, its role in physics based modeling and inverse problems has been limited. Some of the key challenges include, vanishing and exploding gradients. The thesis introduces a new DNN which allows connectivity between different layers. The main novelty is the modeling of DNN using fractional time derivatives instead of the standard one. The resulting fractional-DNN is then applied to learn the parameter-to-solution map in parameterized PDEs. Subsequently, this approximation is employed to solve Bayesian inverse problems. A speedup of over 100 times is observed in comparison to the existing approaches.

Chapter 1: Introduction

Optimal control of fractional PDEs has recently received a lot of attention. We refer to [18] for the optimal control of fractional semilinear PDEs involving both spectral and integral fractional Laplacians with distributed control, see also [48] for such a control of an integral operator. We refer to [11] for the boundary control with the spectral fractional Laplacian and [6] for the exterior optimal control of fractional elliptic PDEs. See [14] for the optimal control of quasi-linear fractional PDEs where the control is in the coefficients. We also refer to [4] for a multigrid method for optimal control problems with linear fractional PDEs (with spectral fractional Laplacian) as constraints.

Fractional operators have recently emerged as an excellent modeling alternative to their classical counterparts. This success can be attributed to the facts that these operators allow long range interactions (nonlocal), they can account for memory information and finally they enforce less smoothness than their classical counterparts. For example, article [135] derives a fractional Helmholtz equation using the first principle arguments. The authors also show a direct qualitative match between the numerical simulations and real data for a problem in Geophysical Electromagnetics. Other examples include dislocations in crystals [51], Lévy process in finance [122, 136], image segmentation and denoising [3, 13, 16, 66], phase field modeling [3], fractional diffusion maps (data analysis) [15], fractional deep neural networks [17], porous media flow [127], etc.

There are various ways to define fractional Laplacian, we refer to [98] where several of these definitions have been shown to be equivalent in unbounded domains. In this thesis, we will operate in bounded domains $\Omega \subset \mathbb{R}^N$ with boundary $\partial\Omega$. We shall focus on two fractional operators. The first one is the *integral fractional Laplacian* defined, for $0 < s < 1$, as,

$$(-\Delta)^s u(x) = C_{N,s}\,\text{P.V.} \int_{\mathbb{R}^N} \frac{u(x)-u(y)}{|x-y|^{N+2s}}\,dy \tag{1.1}$$

where P.V. denotes the Cauchy principal value. The article [126] establishes that the fractional Laplacian in (1.1) arise due to the continuum limit of discrete long-jump random walks.

The second fractional operator is the Caputo fractional derivative defined for $0 < \gamma < 1$ as

$$\partial_t^\gamma u(t) = \frac{1}{\Gamma(1-\gamma)} \int_0^t \frac{u'(r)}{(t-r)^\gamma} \mathrm{d}r. \tag{1.2}$$

Indeed both $(-\Delta)^s$ and ∂^γ are nonlocal operators. For the first operator, we need information about the underlying function in the entire $\mathbb{R}^N$ and for the latter we need information about the function from the beginning of time, to compute the derivative at a given time.

1.1 Outline

The thesis is divided into several interconnected chapters. In Chapter 2, we introduce the notations and provide some preliminary results. Chapter 3 discusses the well-posedness results for elliptic and parabolic fractional PDEs. The focus of Chapter 4 is on external optimal control of fractional parabolic PDEs with control constraints. Chapter 5 provides a complete analysis of optimal control problems constrained by fractional elliptic PDEs with both control and state constraints. In Chapter 6, we extend these results to the parabolic case and introduce Moreau-Yosida regularization based algorithms to handle state constraints. Finally, Chapter 7 introduces novel fractional time derivative based Deep Neural Networks to efficiently solve parameterized PDEs and Bayesian inverse problems.

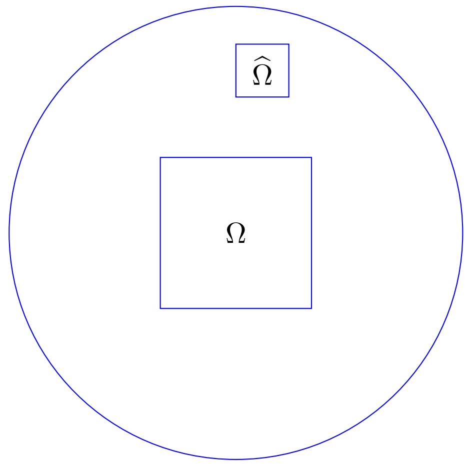

Figure 1.1: External Source identification of a diffusion process in Ω with control supported in $\widehat{\Omega}$, disjoint from Ω.

1.2 External Optimal Control Problems

Many realistic applications require a source or control to be placed outside the observation domain Ω, where the PDE is solved. This situation can arise, for example in (i) Magnetotellurics [135], which is a method for inferring the earth's subsurface electrical conductivity from the measurements at the Earth's surface; (ii) Magnetic Drug Targeting [9, 10], where chemotherapeutic agents are delivered to their desired targets like tumors, by using magnetic nanoparticles and an external magnetic field.

This is in contrast to the traditional approaches, which require the source/control to be placed either inside of the domain Ω or at the boundary, $\partial\Omega$. For instance, consider the source/control identification problem for the Poisson equation

$$-\Delta u = f \quad \text{in } \Omega, \quad u = z \quad \text{on } \partial\Omega, \tag{1.3}$$

in the scenario illustrated in Figure 1.1. The above equation prohibits the placement of control/source in $\widehat{\Omega}$ because the Laplacian, Δ, is a local operator. In other words, in order to compute Δ at a point it suffices to know the values of the function in an arbitrarily small neighborhood. But for the scenario in Figure 1.1 we want the opposite to happen. That is, an operator of non-local nature like the fractional Laplacian, $(-\Delta)^s$, in (1.1) is required.

Non-local counterpart of (1.3) corresponding to $(-\Delta)^s$ is the non-local diffusion equation, for $0 < s < 1$,

$$(-\Delta)^s u = f \quad \text{in } \Omega, \quad u = z \quad \text{on } \mathbb{R}^N \setminus \Omega. \tag{1.4}$$

Notice that in addition to placing the source f inside Ω, (1.4) allows us to place the source/control z in the exterior of Ω and hence the optimal control problems with PDE constraint of type (1.4) is known as the exterior optimal problem. The article [6] introduced the notion of exterior optimal control problem with elliptic fractional PDEs of type (1.4), as constraints.

In Chapter 4, we extend the elliptic case in [6] to the parabolic setting. We begin this chapter by defining the fractional parabolic exterior optimal control problem for the Dirichlet, Robin and Neumann exterior conditions. These problems require dealing with the nonlocal normal derivative. We create a functional analytic framework to show well-posedness and derive the first order optimality conditions for these problems. We present an approach on how to approximate, with convergence rates, the fractional parabolic exterior Dirichlet problem by the fractional parabolic Robin problem.

1.3 Fractional Optimal Control Problems with State Constraints

Next, we consider the optimal control problems where the solution to the state equation (1.4) fulfills obstacle type constraints. These problems occur naturally in several real life applications. For example, in financial mathematics the obstacle problem involving fractional Laplacian arises as a pricing model for American style options, see [46]. In elasticity, the Signorini problem of finding the configuration of an elastic membrane in equilibrium that stays above some rigid surface can be rewritten as an obstacle problem which also involves fractional Laplacian, see [61, 118]. Our list of applications is not complete, there are several other problems where fractional operators appear, see for example, [107, 119, 123].

Motivated by the above applications, we introduce the notion of optimal control problems governed by fractional elliptic PDEs with both state and control constraints in Chapter 5. We establish well-posedness of the optimal control problem and derive the first order optimality conditions. We emphasize that the classical case was considered by E. Casas [38] but almost none of the existing results are applicable to our fractional case. It turns out that the adjoint equation obtained in the optimality conditions is a fractional partial differential equation with a measure as the right-hand side datum. We study the Sobolev regularity of solutions to such equations. As an application of the regularity result of the adjoint equation, we establish the Sobolev regularity of the optimal control. In addition, even weaker controls can be used under this setup.

1.4 Moreau-Yosida Regularization for Optimal Control of Fractional PDEs

Chapter 6 focuses on the analysis and implementation of optimal control of both fractional elliptic and parabolic PDEs with state as well as control constraints. The key challenge again is handling of the state constraints. We use the Moreau-Yosida regularization to implement the state constraint optimal control problem in both elliptic and parabolic cases. We establish convergence, along with rate, of the regularized optimal control problems to the original ones. The spatial discretization is carried out using a finite element method and the discretization error estimates are provided in the elliptic setting. Several illustrative numerical examples in both elliptic and parabolic setting have been provided.

1.5 Fractional Operators in Deep Learning and Bayesian Inverse Problems

In the final chapter, we plan to focus on the novel DNNs to approximate parameterized PDEs and Bayesian inverse problems. Deep learning has recently shown remarkable impact in classification problems such as, in healthcare to detect cancer and classification of images.

Deep learning is a subset of machine learning where the goal is to identify (learn) an approximate map

$$\widehat{\Phi}(\boldsymbol{\xi};\boldsymbol{\theta}) = \mathbf{u}, \tag{1.5}$$

where $\boldsymbol{\theta}$ denotes the parameters that needs to be identified. Moreover, $(\boldsymbol{\xi}, \mathbf{u})$ denotes the input-output pairs. The above mentioned learning process aims at transforming the given input $\boldsymbol{\xi}$ through multiple layers. This includes affine transformations and application of nonlinear activation functions. A popular class of DNNs is the Residual Neural Networks (ResNets) [79, 82]. Here the learning (training) problem can itself be understood as an optimal control problem where one needs to minimize a loss (cost) functional, subject to the ResNet as constraints. The ultimate goal is to identify parameters $\boldsymbol{\theta}$.

In [77,79], authors show that a ResNet can be thought of as a forward Euler discretization of a continuous ODE,

$$\begin{aligned} d_t\widehat{\Phi}(t) &= \sigma(W(t)\widehat{\Phi}(t) + b(t)), \quad t \in (0, T], \\ \widehat{\Phi}(0) &= \boldsymbol{\xi}, \end{aligned} \tag{1.6}$$

where the trainable parameters are $\boldsymbol{\theta} = (W, b)$ and σ is a nonlinear activation function. Therefore, the training problem can be cast as an optimal control problem of the dynamical system (1.6). We will revisit and discuss this in section 7.3.1. DNNs suffer from the so-called vanishing gradient issue [128], which means losing relevant gradient information while propagating through the layers. While ResNets help to some extent, DenseNets introduced in [90] are better in overcoming this issue. However, DenseNet is an adhoc method and lacks mathematical rigor. In our recent work [17], we introduced a rigorous mathematical framework which enable connectivity between all the layers for classification problems. This is achieved by replacing the standard time derivative in (1.6) by the Caputo fractional time derivative as defined in (1.2). In Chapter 7, we introduce a novel fractional derivative based DNN to approximate parameterized PDEs. Subsequently, we extend our approach to Bayesian inverse problems. We observe a speedup of over 100 times over the existing approaches.

1.6 Publications

The content of this thesis has appeared in the following four publications where I led the entire discussion. However, I received guidance from Dr. Harbir Antil, Dr. Thomas S. Brown, Dr. Howard C. Elman, Dr. Akwum Onwunta and Dr. Mahamadi Warma.

1. H. Antil, D. Verma, and M. Warma, *External optimal control of fractional parabolic PDEs*, ESAIM Control Optim. Calc. Var. **26** (2020).

2. H. Antil, D. Verma, and M. Warma, *Optimal control of fractional elliptic PDEs with state constraints and characterization of the dual of fractional-order Sobolev spaces*, J. Optim. Theory Appl. **186** (2020), no. 1, 1–23.

3. H. Antil, T. S. Brown, D. Verma and M. Warma, *Optimal control of fractional PDEs with state and control constraints*, (Submitted 2020).

4. H. Antil, H. C. Elman, A. Onwunta, and D. Verma, *Novel deep neural networks for solving Bayesian statistical inverse*, (Submitted to SIAM Data Science 2021).

I also co-led article five with Dr. Thomas S. Brown (George Mason University) and article six with Rafael Arndt (George Mason University). Here we received guidance from Dr. Harbir Antil, Dr. Rainald Löhner, Dr. Carlos R. Rautenberg and Dr. Fumiya Togashi.

5. T. S. Brown, H. Antil, R. Löhner, F. Togashi, and D. Verma, *Novel DNNs for stiff ODEs with applications to chemically reacting flows*, (Accepted in CFDML 2021).

6. H. Antil, R. Arndt, C. N. Rautenberg and D. Verma, *Non-diffusive variational problems with distributional and weak gradient constraints*, (Submitted 2021).

Chapter 2: Notation and Preliminaries

The purpose of this chapter is to introduce the notations and some preliminary results. Unless otherwise stated, $\Omega \subset \mathbb{R}^N$ $(N \geq 1)$ is an arbitrary bounded open set, $0 < s < 1$ and $1 \leq p < \infty$. For a function u defined in $\mathbb{R}^N$ (or in Ω), we shall denote by $D_{s,p}u$, the function defined in $\mathbb{R}^N \times \mathbb{R}^N$ (or in $\Omega \times \Omega$) by $D_{s,p}u[x,y] := \dfrac{u(x)-u(y)}{|x-y|^{\frac{N}{p}+s}}$. Then, we define the Sobolev space

$$W^{s,p}(\Omega) := \Big\{u \in L^p(\Omega) : \ D_{s,p}u \in L^p(\Omega \times \Omega)\Big\}$$

and we endow it with the norm $\|u\|_{W^{s,p}(\Omega)} := \left(\int_\Omega |u|^p\, dx + \|D_{s,p}u\|^p_{L^p(\Omega\times\Omega)}\right)^{\frac{1}{p}}$. We let $W_0^{s,p}(\Omega) := \overline{\mathcal{D}(\Omega)}^{W^{s,p}(\Omega)}$, where $\mathcal{D}(\Omega)$ is the space of test functions.

If $s = 1$ and $p = 2$, then we shall denote $W^{1,2}(\Omega) := \{u \in L^2(\Omega) : \ |\nabla u| \in L^2(\Omega)\}$ and $W_0^{1,2}(\Omega) := \overline{\mathcal{D}(\Omega)}^{W^{1,2}(\Omega)}$ by $H^1(\Omega)$ and $H_0^1(\Omega)$, respectively.

Since Ω is assumed to be bounded, we have the following continuous embeddings:

$$W_0^{s,2}(\Omega) \hookrightarrow \begin{cases} L^p(\Omega), & \text{if } \ N \geq 2s, \\ C^{0,s-\frac{N}{2}}(\overline{\Omega}), & \text{if } \ N < 2s, \end{cases} \tag{2.1}$$

with $p = 2^\star := \dfrac{2N}{N-2s}$ if $N > 2s$, and $p \in [1,\infty[$ arbitrary if $N = 2s$.

A complete characterization of $W_0^{s,p}(\Omega)$ for arbitrary bounded open sets is given in [133]. By [73, Theorem 1.4.2.4, p.25] (see also, [29,133]) if Ω has a Lipschitz continuous boundary and $\frac{1}{p} < s < 1$, then

$$\|u\|_{W_0^{s,p}(\Omega)} = \|D_{s,p}u\|_{L^p(\Omega\times\Omega)} \tag{2.2}$$

defines an equivalent norm on $W_0^{s,p}(\Omega)$. In that case, we shall use this norm for $W_0^{s,p}(\Omega)$.

To study the Fractional Dirichlet problems, we need to consider the following function space:

$$\widetilde{W}_0^{s,p}(\Omega) := \Big\{u \in W^{s,p}(\mathbb{R}^N) : \ u = 0 \ \text{ on } \ \mathbb{R}^N \setminus \Omega\Big\}.$$

Let $\Omega \subset \mathbb{R}^N$ be a bounded open set with a Lipschitz continuous boundary. By [62, Theorem 6], $\mathcal{D}(\Omega)$ is dense in $\widetilde{W}_0^{s,p}(\Omega)$. Moreover, for every $0 < s < 1$, we have the following norm on $\widetilde{W}_0^{s,p}(\Omega)$,

$$\|u\|^p_{\widetilde{W}_0^{s,p}(\Omega)} := \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{|u(x)-u(y)|^p}{|x-y|^{N+sp}}\,dxdy = \|D_{s,p}u\|^p_{L^p(\Omega\times\Omega)} + \int_\Omega |u|^p\vartheta(x)\,dx, \tag{2.3}$$

where $\vartheta(x) := 2\displaystyle\int_{\mathbb{R}^N\setminus\Omega} \frac{1}{|x-y|^{N+sp}}\,dy, \ \ x \in \Omega$.

Remark 2.1. The following conditions hold.

1. The embeddings (2.1) hold with $W_0^{s,2}(\Omega)$ replaced by $\widetilde{W}_0^{s,2}(\Omega)$.

2. Let p satisfy

$$p > \frac{N}{2s} \quad \text{if} \ \ N > 2s, \quad p > 1 \quad \text{if} \ \ N = 2s, \quad p = 1 \quad \text{if} \ \ N < 2s, \tag{2.4}$$

and $p' := \frac{p}{p-1}$. Then it is easy to see that $\widetilde{W}_0^{s,2}(\Omega) \hookrightarrow L^{p'}(\Omega)$.

We next state an important result for $\widetilde{W}_0^{s,p}(\Omega)$ taken from [73, Chapter 1].

Theorem 2.1. Let $\Omega \subset \mathbb{R}^N$ be a bounded open set with a Lipschitz continuous boundary and $1 < p < \infty$. If $\frac{1}{p} < s < 1$, then $\widetilde{W}_0^{s,p}(\Omega) = W_0^{s,p}(\Omega)$ with equivalent norms.

Under the hypotheses of Theorem 2.1, we have that if $\frac{1}{p} < s < 1$, then

$$\|u\|_{\widetilde{W}_0^{s,p}(\Omega)} = \|D_{s,p}u\|_{L^p(\Omega\times\Omega)}. \tag{2.5}$$

In other words, the integral involving the function $\vartheta(x)$ in (2.3) is not relevant.

If $0 < s < 1$, $p \in]1,\infty[$ and $p' := \frac{p}{p-1}$, then the spaces $W^{-s,p'}(\mathbb{R}^N)$ and $\widetilde{W}^{-s,p'}(\Omega)$ is defined as the dual of $W^{s,p}(\mathbb{R}^N)$ and $\widetilde{W}_0^{s,p}(\Omega)$, respectively. Moreover, $\langle\cdot,\cdot\rangle$ shall denote their duality pairing whenever it is clear from the context.

To study the fractional Robin problem we shall need the following Sobolev space introduced in [50]. For $\kappa \in L^1(\mathbb{R}^N \setminus \Omega)$ fixed, we let

$$W^{s,2}_{\Omega,\kappa} := \Big\{u : \mathbb{R}^N \to \mathbb{R} \ \text{ measurable and } \ \|u\|_{W^{s,2}_{\Omega,\kappa}} < \infty\Big\},$$

where

$$\|u\|_{W^{s,2}_{\Omega,\kappa}} := \left(\|u\|^2_{L^2(\Omega)} + \||\kappa|^{\frac{1}{2}}u\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} + \int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{|u(x)-u(y)|^2}{|x-y|^{N+2s}}dxdy\right)^{\frac{1}{2}}. \tag{2.6}$$

Let μ be the measure on $\mathbb{R}^N \setminus \Omega$ given by $d\mu = |\kappa|dx$. With this setting, the norm in (2.6) can be rewritten as

$$\|u\|_{W^{s,2}_{\Omega,\kappa}} := \left(\|u\|^2_{L^2(\Omega)} + \|u\|^2_{L^2(\mathbb{R}^N\setminus\Omega,\mu)} + \int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{|u(x)-u(y)|^2}{|x-y|^{N+2s}}dxdy\right)^{\frac{1}{2}}. \tag{2.7}$$

If $\kappa = 0$, then we shall let $W^{s,2}_{\Omega,0} = W^{s,2}_{\Omega}$. The following result has been proved in [50, Proposition 3.1].

Proposition 2.1. Let $\kappa \in L^1(\mathbb{R}^N \setminus \Omega)$. Then $W^{s,2}_{\Omega,\kappa}$ is a Hilbert space.

Next, for a Banach space $\mathbb{X}$, we shall denote the vector-valued Banach spaces

$$H^1_{0,0}((0,T);\mathbb{X}) := \Big\{u \in H^1((0,T);\mathbb{X}) :\ u(0,\cdot) = 0\Big\},$$

$$H^1_{0,T}((0,T);\mathbb{X}) := \Big\{u \in H^1((0,T);\mathbb{X}) :\ u(T,\cdot) = 0\Big\},$$

and

$$H^1_{0}((0,T);\mathbb{X}) := \Big\{u \in H^1((0,T);\mathbb{X}) :\ u(0,\cdot) = u(T,\cdot) = 0\Big\}.$$

We notice that the continuous embedding $H^1((0,T);\mathbb{X}) \hookrightarrow C([0,T];\mathbb{X})$ holds, so that, for $u \in H^1((0,T);\mathbb{X})$, the values $u(0,\cdot)$ and $u(T,\cdot)$ make sense.

We now define the fractional Laplacian. Consider the space

$$\mathbb{L}^1_s(\mathbb{R}^N) := \left\{u : \mathbb{R}^N \to \mathbb{R} \ \text{ measurable and } \int_{\mathbb{R}^N} \frac{|u(x)|}{(1+|x|)^{N+2s}}\, dx < \infty\right\}.$$

Then, for a function $u \in \mathbb{L}^1_s(\mathbb{R}^N)$ and $\varepsilon > 0$ we let

$$(-\Delta)^s_\varepsilon u(x) := C_{N,s} \int_{\{y\in\mathbb{R}^N, |y-x|>\varepsilon\}} \frac{u(x)-u(y)}{|x-y|^{N+2s}} dy, \quad x \in \mathbb{R}^N,$$

where the normalization constant $C_{N,s}$ is given by

$$C_{N,s} := \frac{s2^{2s}\Gamma\left(\frac{2s+N}{2}\right)}{\pi^{\frac{N}{2}}\Gamma(1-s)}, \tag{2.8}$$

and Γ is the usual Euler Gamma function (see e.g. [28, 31, 34–36, 49, 132, 133]). The **fractional Laplacian** $(-\Delta)^s$ is then defined for $u \in \mathbb{L}^1_s(\mathbb{R}^N)$ by the formula

$$(-\Delta)^s u(x) = C_{N,s}\text{P.V.}\int_{\mathbb{R}^N} \frac{u(x)-u(y)}{|x-y|^{N+2s}}dy = \lim_{\varepsilon\downarrow 0}(-\Delta)^s_\varepsilon u(x), \quad x \in \mathbb{R}^N, \tag{2.9}$$

provided that the limit exists for a.e. $x \in \mathbb{R}^N$. It has been shown in [32, Proposition 2.2] that for $u \in \mathcal{D}(\Omega)$, we have

$$\lim_{s\uparrow 1^-}\int_{\mathbb{R}^N} u(-\Delta)^s u\, dx = \int_{\mathbb{R}^N} |\nabla u|^2 dx = -\int_{\mathbb{R}^N} u\Delta u\, dx = -\int_{\Omega} u\Delta u\, dx.$$

This is where the constant $C_{N,s}$ plays a crucial role.

Next, we define the operator $(-\Delta)^s_D$ on $L^2(\Omega)$ as follows.

$$D((-\Delta)^s_D) := \Big\{u|_\Omega :\ u \in \widetilde{W}^{s,p}_0(\Omega) \text{ and } (-\Delta)^s u \in L^2(\Omega)\Big\},\ \ (-\Delta)^s_D(u|_\Omega) := ((-\Delta)^s u)|_\Omega, \tag{2.10}$$

where $D((-\Delta)^s_D)$ denotes the domain of $(-\Delta)^s_D$. Then $(-\Delta)^s_D$ is the realization in $L^2(\Omega)$ of $(-\Delta)^s$ with the Dirichlet exterior condition $u = 0$ in $\mathbb{R}^N \setminus \Omega$. The following result is well-known (see e.g. [27, 45, 116]).

Proposition 2.2. The operator $(-\Delta)^s_D$ has a compact resolvent and $-(-\Delta)^s_D$ generates a strongly continuous submarkovian semigroup $(e^{-t(-\Delta)^s_D})_{t\geq 0}$ on $L^2(\Omega)$. The operator $(-\Delta)^s_D$ can be also viewed as a bounded operator from $\widetilde{W}^{s,p}_0(\Omega)$ into $\widetilde{W}^{-s,2}(\Omega)$. In this case $-(-\Delta)^s_D$ also generates a strongly continuous semigroup $(e^{-t(-\Delta)^s_D})_{t\geq 0}$ on $\widetilde{W}^{-s,2}(\Omega)$.

Next, for $u \in W^{s,2}_\Omega$ we define the nonlocal normal derivative $\mathcal{N}_s$ as follows:

$$\mathcal{N}_s u(x) := C_{N,s}\int_\Omega \frac{u(x)-u(y)}{|x-y|^{N+2s}}\, dy, \quad x \in \mathbb{R}^N \setminus \overline{\Omega}. \tag{2.11}$$

We shall call $\mathcal{N}_s$ the *interaction operator*. Notice that the origin of the term "interaction" goes back to [53]. Clearly $\mathcal{N}_s$ is a nonlocal operator and it is well defined as shown in the following result (see e.g. [68, Lemma 3.2]).

Lemma 2.1. The interaction operator $\mathcal{N}_s$ maps $W^{s,2}(\mathbb{R}^N)$ into $W^{s,2}_{\mathrm{loc}}(\mathbb{R}^N \setminus \Omega)$.

Despite the fact that $\mathcal{N}_s$ is defined in $\mathbb{R}^N \setminus \Omega$, it is still known as the "normal" derivative. This is due to its similarity with the classical normal derivative (see e.g. [6, Proposition 2.2]).

We conclude this chapter by stating the integration by parts formula for the fractional Laplacian.

Proposition 2.3 (**The integration by parts formula for** $(-\Delta)^s$**).** Assume that Ω has a Lipschitz continuous boundary. Let $u \in W^{s,2}_{\Omega}$ be such that $(-\Delta)^s u \in L^2(\Omega)$ and $\mathcal{N}_s u \in L^2(\mathbb{R}^N \setminus \Omega)$. Then, for every $v \in W^{s,2}_{\Omega} \cap L^2(\mathbb{R}^N \setminus \Omega)$, we have

$$\frac{C_{N,s}}{2} \iint_{\mathbb{R}^{2N} \setminus (\mathbb{R}^N \setminus \Omega)^2} \frac{(u(x)-u(y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy = \int_{\Omega} v(-\Delta)^s u \, dx + \int_{\mathbb{R}^N \setminus \Omega} v \mathcal{N}_s u \, dx, \tag{2.12}$$

where $\mathbb{R}^{2N} \setminus (\mathbb{R}^N \setminus \Omega)^2 = (\Omega \times \Omega) \cup (\Omega \times (\mathbb{R}^N \setminus \Omega)) \cup ((\mathbb{R}^N \setminus \Omega) \times \Omega)$.

The proof of the preceding proposition is included in [50, Lemma 3.2] for smooth functions. The version given here is obtained by using an approximation argument (see e.g. [103, Proposition 3.7]).

Remark 2.2. If $u, v \in \widetilde{W}^{s,2}_0(\Omega)$ in Proposition 2.3, then the integration by parts formula is given by

$$\frac{C_{N,s}}{2} \iint_{\mathbb{R}^{2N}} \frac{(u(x)-u(y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy = \int_{\Omega} v(-\Delta)^s u \, dx.$$

Chapter 3: Fractional Partial Differential Equations

The purpose of this chapter is to introduce various notions of solutions to fractional elliptic and parabolic PDEs. Fractional order operators have recently emerged as a modeling alternative in various branches of science and technology. These are non-local operators that allow long-range interactions and are known to capture multi-scale behavior. Among different non-local operators, many studies are available for fractional Laplacian $(-\Delta)^s$ with $s \in (0,1)$. The article [126], for instance shows that $(-\Delta)^s$ arise from the continuum limit of discrete long-jump random walks. Besides nonlocality, another key feature of $(-\Delta)^s$ is that it is less strict in terms of regularity requirements, see for instance [113, Remark 7.2]. This latter feature further makes $(-\Delta)^s$ attractive for modeling problems with sharp transitions across interfaces such as phase field models, imaging etc.

The focus of this chapter is on showing the well-posedness of fractional PDEs with exterior data (Dirichlet, Robin or Neumann) or with non-smooth interior source (Radon measure). In all these cases, the notion of very-weak solutions will be established for both elliptic and parabolic PDEs. In addition, we provide conditions under which the very-weak solutions turn into weak solutions. Finally, we provide sharp conditions on the data under which the solution is shown to be continuous.

3.1 Fractional Elliptic PDEs

This section is dedicated to study the solutions to fractional elliptic PDEs. We consider Dirichlet and Robin problems in Sections 3.1.1 and 3.1.2, respectively. The results also hold for Neumann problem after minor modifications.

The non-homogeneous exterior value problem in the elliptic setting has been studied in [6]. For completeness, we only mention the statement of results in this section and refer the reader to [6] for details. Our main work starts from Theorem 3.2 where we establish the continuity of weak solutions to the Dirichlet problem with L^p-datum followed by well-posedness with measure valued datum.

3.1.1 Fractional Elliptic Problem with Dirichlet Conditions

We begin by writing the general form of fractional Dirichlet exterior value problem:

$$\begin{cases} (-\Delta)^s u = f & \text{in } \Omega, \\ u = g & \text{in } \mathbb{R}^N \setminus \Omega. \end{cases} \tag{3.1}$$

Next, we state the notion of weak solutions.

Definition 3.1 (**Weak solution to (non-homogeneous) elliptic Dirichlet problem).** Let $f \in \widetilde{W}^{-s,2}(\Omega)$, $g \in W^{s,2}(\mathbb{R}^N \setminus \Omega)$ and $\mathcal{Z} \in W^{s,2}(\mathbb{R}^N)$ be such that $\mathcal{Z}|_{\mathbb{R}^N\setminus\Omega} = g$. A function $u \in W^{s,2}(\mathbb{R}^N)$ is said to be a weak solution to (3.1) if $u - \mathcal{Z} \in \widetilde{W}_0^{s,2}(\Omega)$ and

$$\frac{C_{N,s}}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{(u(x) - u(y))(v(x) - v(y))}{|x - y|^{N+2s}} \, dxdy = \langle f, v \rangle,$$

for every $v \in \widetilde{W}_0^{s,2}(\Omega)$.

The existence and uniqueness of a weak solution u to (3.1) and the continuous dependence of u on the data f and g have been considered in [75], see also [68,130]. More precisely we have the following result.

Proposition 3.1. Let $f \in \widetilde{W}^{-s,2}(\Omega)$ and $g \in W^{s,2}(\mathbb{R}^N \setminus \Omega)$. Then there exists a unique weak solution u to (3.1) in the sense of Definition 3.1. In addition there is a constant $C > 0$ such that

$$\|u\|_{W^{s,2}(\mathbb{R}^N)} \leq C \left(\|f\|_{\widetilde{W}^{-s,2}(\Omega)} + \|g\|_{W^{s,2}(\mathbb{R}^N\setminus\Omega)} \right). \tag{3.2}$$

Remark 3.1. For the homogeneous exterior value problem, that is, when $g = 0$, the weak solution u to (3.1) lies in $\widetilde{W}_0^{s,2}(\Omega)$ and Proposition 3.1 holds.

Even though Proposition 3.1 is typically sufficient, we need the existence of solutions (in some notion) to the fractional Dirichlet problem (3.1) when datum in the exterior or interior is less regular, that is, when $g \in L^2(\mathbb{R}^N \setminus \Omega)$ or $f \in \mathcal{M}(\Omega)$, where $\mathcal{M}(\Omega)$ is the space of Radon measures on Ω. In order to tackle this situation the notion of very-weak solutions for (3.1) is needed, see for instance [6].

Definition 3.2 (**Very-weak solution to elliptic (non-homogeneous) Dirichlet problem).** Let $g \in L^2(\mathbb{R}^N \setminus \Omega)$ and $f \in \widetilde{W}^{-s,2}(\Omega)$. A function $u \in L^2(\mathbb{R}^N)$ is said to be a very-weak solution to (3.1) if the identity

$$\int_{\Omega} u(-\Delta)^s v \, dx = \langle f, v \rangle - \int_{\mathbb{R}^N \setminus \Omega} z \mathcal{N}_s v \, dx, \tag{3.3}$$

holds for every $v \in V := \{v \in \widetilde{W}_0^{s,2}(\Omega) \ : \ (-\Delta)^s v \in L^2(\Omega)\}$.

We now state the result [6, Theorem 3.5] which yields the existence and uniqueness of a very-weak solution u to (3.1) and the continuous dependence of u on the data f and g.

Theorem 3.1. Let $f \in \widetilde{W}^{-s,2}(\Omega)$ and $g \in L^2(\mathbb{R}^N \setminus \Omega)$. Then there exists a unique very-weak solution u to (3.1) according to Definition 3.2 that fulfills

$$\|u\|_{L^2(\Omega)} \leq C \left(\|f\|_{\widetilde{W}^{-s,2}(\Omega)} + \|g\|_{L^2(\mathbb{R}^N \setminus \Omega)} \right), \tag{3.4}$$

for a constant $C > 0$. In addition, if $g \in W^{s,2}(\mathbb{R}^N \setminus \Omega)$, then the following assertions hold.

1. Every weak solution of (3.1) is also a very-weak solution.

2. Every very-weak solution of (3.1) that belongs to $W^{s,2}(\mathbb{R}^N)$ is also a weak solution.

Next, we shift our focus on the fractional Dirichlet (homogeneous) exterior problem. Throughout the remainder of this subsection, we consider the problem:

$$\begin{cases} (-\Delta)^s u = f & \text{in } \Omega, \\ u = 0 & \text{in } \mathbb{R}^N \setminus \Omega. \end{cases} \tag{3.5}$$

The purpose is to establish continuity, up to the boundary of Ω, of weak solutions to (3.5) when $f \in L^p(\Omega)$ and to study the well-posedness of (3.5) with $f \in \mathcal{M}(\Omega) = (C_0(\Omega))^\star$. The latter space $C_0(\Omega)$ denotes the space of all continuous functions on $\overline{\Omega}$. Note that a function $u \in C_0(\Omega)$ can be extended to function in $C_0(\mathbb{R}^n)$ by taking $u = 0$ outside of Ω. Recall that $\mathcal{M}(\Omega)$ is the space of all Radon measures and we have

$$\langle \mu, v \rangle_{(C_0(\Omega))^\star, C_0(\Omega)} = \int_\Omega v \, d\mu, \quad \mu \in \mathcal{M}(\Omega), \quad v \in C(\overline{Q}).$$

In addition, we have the following norm on the space $\mathcal{M}(\Omega)$:

$$\|\mu\|_{\mathcal{M}(\Omega} = \sup_{v \in C_0(\Omega),\, |v| \le 1} \int_\Omega v \, d\mu.$$

We first show the continuity of weak solutions to (3.5). We recall that the paper [112] proves the optimal Hölder C^s-regularity of u under the condition that the datum $f \in L^\infty(\Omega)$. However, in our setting we only assume $f \in L^p(\Omega)$.

Theorem 3.2. Let Ω be a bounded Lipschitz domain satisfying the exterior cone condition. Assume that $f \in L^p(\Omega)$ with p as in (2.4). Then, every weak solution u of (3.5) belongs to $C_0(\Omega)$ and there is a constant $C = C(N, s, p, \Omega) > 0$ such that

$$\|u\|_{C_0(\Omega)} \le C \|f\|_{L^p(\Omega)}. \tag{3.6}$$

Proof. Let $f \in L^p(\Omega)$ and let $\{f_n\}_{n\in\mathbb{N}} \subset L^\infty(\Omega)$ be a sequence such that $\|f_n - f\|_{L^p(\Omega)} \to 0$ as $n \to \infty$. For each $n \in \mathbb{N}$, let u_n solve the following Dirichlet problem:

$$(-\Delta)^s u_n = f_n \quad \text{in } \Omega, \qquad u_n = 0 \quad \text{in } \mathbb{R}^N \setminus \Omega. \tag{3.7}$$

By [112, Proposition 1.1], $u_n \in C^s(\mathbb{R}^N)$. Next, subtracting (3.5) from (3.7), we obtain

$$(-\Delta)^s (u_n - u) = f_n - f \quad \text{in } \Omega, \qquad (u_n - u) = 0 \quad \text{in } \mathbb{R}^N \setminus \Omega.$$

Since $(f_n - f) \in L^p(\Omega)$, applying [18, Theorem 3.7], we get that $\|u_n - u\|_{L^\infty(\Omega)} \leq C\|f_n - f\|_{L^p(\Omega)} \to 0$ as $n \to \infty$. Thus, $\|u_n - u\|_{L^\infty(\Omega)} \to 0$ as $n \to \infty$. Since $u_n \in C_0(\Omega)$, it follows that $u \in C_0(\Omega)$. $\square$

In Corollary 5.2, we shall reduce the assumed $L^p(\Omega)$-regularity requirement on f in Theorem 3.2.

Next, we study the well-posedness of the elliptic PDE (3.5) when the right-hand side is given by a Radon measure, that is, when $f \in \mathcal{M}(\Omega)$.

Definition 3.3 (Very-weak solutions: Dirichlet problem with measure data). Let p be as in (2.4) and $\frac{1}{p} + \frac{1}{p'} = 1$. Let $f \in \mathcal{M}(\Omega)$. A function $u \in L^{p'}(\Omega)$ is said to be a very-weak solution to (3.5), if for every $v \in V := \{v \in C_0(\Omega) \cap \widetilde{W}_0^{s,2}(\Omega) : \ (-\Delta)^s v \in L^p(\Omega)\}$ we have

$$\int_\Omega u(-\Delta)^s v \, dx = \int_\Omega v \, df. \tag{3.8}$$

The following theorem proves the existence and uniqueness of the very-weak solution to (3.5) in the sense of Definition 3.3.

Theorem 3.3. Let Ω be a bounded Lipschitz domain satisfying the exterior cone condition. Let $f \in \mathcal{M}(\Omega)$ be a measure, p as in (2.4) and $p' := \frac{p}{p-1}$. Then, there exists a unique $u \in L^{p'}(\Omega)$ that solves (3.5) according to Definition 3.3, and there is a constant $C =$

$C(N, s, p, \Omega) > 0$ such that

$$\|u\|_{L^{p'}(\Omega)} \leq C\|f\|_{\mathcal{M}(\Omega)}. \tag{3.9}$$

Proof. For a given $\xi \in L^p(\Omega)$, we begin by considering the following auxiliary problem:

$$(-\Delta)^s v = \xi \quad \text{in } \Omega, \qquad v = 0 \quad \text{in } \mathbb{R}^N \setminus \Omega. \tag{3.10}$$

Since $L^p(\Omega) \hookrightarrow L^{p'}(\Omega) \hookrightarrow \widetilde{W}^{-s,2}(\Omega)$ (by Remark 2.1), it follows that there is a unique $v \in \widetilde{W}_0^{s,2}(\Omega)$ satisfying (3.10). By Theorem 3.2, $v \in C_0(\Omega)$.

Consider a mapping $\Xi : L^p(\Omega) \to C_0(\Omega), \quad \xi \mapsto \Xi\xi := v$. Notice that, Ξ is linear and continuous (by Theorem 3.2). Let us define $u := \Xi^* f$. Then, $u \in L^{p'}(\Omega)$. We show that u solves (3.5). Notice that,

$$\int_\Omega u\xi \; dx = \int_\Omega u(-\Delta)^s v \; dx = \int_\Omega (\Xi^* f)\xi \; dx = \int_\Omega v \; df, \tag{3.11}$$

for every $v \in V$. Thus, we have constructed a function $u \in L^{p'}(\Omega)$ that solves (3.5), according to Definition 3.7. Next, we show uniqueness of solutions. Assume that (3.5) has two solutions u_1 and u_2 with the same right hand side datum f. Then, it follows from (3.8) that

$$\int_\Omega (u_1 - u_2)(-\Delta)^s v \; dx = 0,$$

for every $v \in V$. It follows from the fundamental lemma of the calculus of variations that $u_1 = u_2$ a.e. in Ω and we have shown the uniqueness of solutions. It remains to prove the required estimate (3.9). From (3.11), we have that

$$\left|\int_\Omega u\xi \; dx\right| \leq \|\mu\|_{\mathcal{M}(\Omega)}\|v\|_{C_0(\Omega)} \leq C\|\mu\|_{\mathcal{M}(\Omega)}\|\xi\|_{L^p(\Omega)}, \tag{3.12}$$

where in the last estimate we have used Theorem 3.2. Then, dividing both sides of (3.12) by $\|\xi\|_{L^p(\Omega)}$ and taking the supremum over $\xi \in L^p(\Omega)$, we obtain (3.9). $\square$

The regularity of solutions to (3.5) given in Theorem 3.3 will be improved in Corollary 5.3.

3.1.2 Fractional Elliptic Problem with Robin Conditions

In this section, we consider the fractional elliptic Robin exterior value problem given by

$$\begin{cases} (-\Delta)^s u & = f \quad \text{in } \Omega, \\ \mathcal{N}_s u + \kappa u & = \kappa g \quad \text{in } \mathbb{R}^N \setminus \Omega, \end{cases} \tag{3.13}$$

where $\mathcal{N}_s$ is the interaction operator defined in (2.11) and $\kappa \in L^1(\mathbb{R}^N \setminus \Omega) \cap L^\infty(\mathbb{R}^N \setminus \Omega)$ is non-negative. We mention that the result in this subsection is taken from [6, Section 3.2]. Therefore we only state the result here and refer the reader to [6] for details. Next, we introduce the notion of weak solutions to the fractional elliptic Robin problem.

Definition 3.4. Let $g \in L^2(\mathbb{R}^N \setminus \Omega, \mu)$ and $f \in (W^{s,2}_{\Omega,\kappa})^\star$. A function $u \in W^{s,2}_{\Omega,\kappa}$ is said to be a weak solution of (3.13) if the identity

$$\begin{aligned} \iint_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} & \frac{(u(x)-u(y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy + \int_{\mathbb{R}^N\setminus\Omega} \kappa uv\, dx \\ & = \langle f, v\rangle_{(W^{s,2}_{\Omega,\kappa})^\star, W^{s,2}_{\Omega,\kappa}} + \int_{\mathbb{R}^N\setminus\Omega} \kappa g v\, dx, \end{aligned} \tag{3.14}$$

holds for every $v \in W^{s,2}_{\Omega,\kappa}$.

We next state the result [6, Proposition 3.9] concerning the existence and uniqueness (upto a constant) of the weak solution to Robin problem.

Theorem 3.4. Let $\kappa \in L^1(\mathbb{R}^N \setminus \Omega) \cap L^\infty(\mathbb{R}^N \setminus \Omega)$. Then, for every $g \in L^2(\mathbb{R}^N \setminus \Omega, \mu)$ and $f \in (W^{s,2}_{\Omega,\kappa})^\star$, there exists a weak solution $u \in W^{s,2}_{\Omega,\kappa}$ of (3.13).

3.2 Fractional Parabolic PDEs

In this section, we investigate fractional parabolic PDEs and extend the results from elliptic case to the parabolic case. The elliptic and parabolic problems are fundamentally different and the results presented here are novel in the context of parabolic (non-stationary) problems. In many situations, the techniques used in the elliptic case either cannot be directly used or they must be carefully adapted to the parabolic case. Definition 3.6 introduces the notion of weak-solution to the non-homogeneous fractional parabolic Dirichlet problem. Notice that we need an additional regularity (H^1-in time) to establish the notion of weak solution which is different to the elliptic case. In Definition 3.7, we introduce the notion of very-weak solutions to the fractional parabolic problem which requires integration-by-parts in both space and time. The duality argument proving the existence and uniqueness of solutions in the parabolic case is more involved (cf. Theorem 3.6) when compared to the elliptic case. In Definition 3.9, we introduce the notion of weak solutions to the Robin problem whose existence and uniqueness is shown in Theorem 3.9 by using the notion of integrated semigroups. Note that the concept of integrated semigroups is not needed in the elliptic case.

Theorem 3.7 involves boundedness of weak solutions to parabolic PDEs yielding continuity as a corollary. Such a result for the parabolic problem is new and is not an obvious consequence of the elliptic case (see Section 3.1.1) and cannot be obtained using classical properties of parabolic problems. In Definition 3.8, we introduce the notion of very-weak solutions to the parabolic problem with measure valued right-hand side and initial data followed by existence and uniqueness of such solutions in Theorem 3.8.

3.2.1 Fractional Parabolic Problem with Dirichlet Conditions

We begin by discussing the following fractional parabolic Dirichlet (non-homogeneous) exterior value problem:

$$\begin{cases} \partial_t u + (-\Delta)^s u = 0 & \text{in } Q := (0,T) \times \Omega, \\ u = z & \text{in } \Sigma := [0,T) \times (\mathbb{R}^N \setminus \Omega), \\ u(0,\cdot) = 0 & \text{in } \Omega. \end{cases} \tag{3.15}$$

Let us consider first the following auxiliary problem:

$$\begin{cases} \partial_t w + (-\Delta)^s w & = f \quad \text{in } Q, \\ w & = 0 \quad \text{in } \Sigma, \\ w(0,\cdot) & = 0 \quad \text{in } \Omega, \end{cases} \tag{3.16}$$

that is, a fractional parabolic equation with a nonzero right-hand side but a zero exterior condition. Notice that (3.16) can be rewritten as the following Cauchy problem:

$$\begin{cases} \partial_t w + (-\Delta)^s_D w = f & \text{in } Q, \\ w(0,\cdot) = 0 & \text{in } \Omega, \end{cases} \tag{3.17}$$

where we recall that $(-\Delta)^s_D$ is the operator defined in (2.10). Throughout this subsection $\langle \cdot, \cdot \rangle$ will denote the duality pairing between $\widetilde{W}_0^{s,2}(\Omega)$ and $\widetilde{W}^{-s,2}(\Omega)$.

We now introduce our notion of weak solutions to (3.16).

Definition 3.5 (**Weak solution to (homogeneous) parabolic Dirichlet problem).** Let $f \in L^2((0,T); \widetilde{W}^{-s,2}(\Omega))$. A function

$w \in \mathbb{U}_0 := L^2((0,T);\widetilde{W}_0^{s,2}(\Omega)) \cap H^1_{0,0}((0,T);\widetilde{W}^{-s,2}(\Omega))$ is said to be a weak solution to (3.16) if

$$\langle \partial_t w(t,\cdot), v\rangle + \frac{C_{N,s}}{2}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{(w(t,x)-w(t,y))(v(x)-v(y))}{|x-y|^{N+2s}}\,dxdy = \langle f(t,\cdot), v\rangle,$$

for every $v \in \widetilde{W}^{-s,2}(\Omega)$ and almost every $t \in (0,T)$.

Remark 3.2. We state the following facts.

1. A weak solution to (3.16) belongs to $C([0,T], L^2(\Omega))$ (see e.g. [99, Remark 9]).
2. If $f \in L^2((0,T);L^2(\Omega))$, then it has been shown in [27] (by using semigroup theory) that a weak solution to (3.16) enjoys the following regularity:

$$u \in C([0,T]; D((-\Delta)^s_D)) \cap H^1_{0,0}((0,T);L^2(\Omega)).$$

The existence and uniqueness of weak solutions to (3.16) were shown in [99, Theorem 26].

Proposition 3.2 (**Weak solutions to** (3.16))**.** Let $f \in L^2((0,T);\widetilde{W}_0^{s,2}(\Omega))$. Then there exists a unique weak solution $w \in \mathbb{U}_0$ to (3.16) in the sense of Definition 3.5 and is given by

$$w(t,x) = \int_0^t e^{-(t-\tau)(-\Delta)^s_D} f(\tau,x)\,d\tau, \tag{3.18}$$

where $(e^{-t(-\Delta)^s_D})_{t\geq 0}$ is the semigroup mentioned in Proposition 2.2. In addition, there is a constant $C>0$ such that

$$\|w\|_{\mathbb{U}_0} \leq C\|f\|_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))}. \tag{3.19}$$

We introduce next our notion of weak solutions to the non-homogeneous problem (3.15). Notice the higher regularity requirement on the datum z.

Definition 3.6 (**Weak solution to (non-homogeneous) parabolic Dirichlet problem).** Let $z \in H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$ and $\tilde{z} \in H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N))$ be such that $\tilde{z}|_{\Sigma} = z$. Then a function $u \in \mathbb{U} := L^2((0,T); W^{s,2}(\mathbb{R}^N)) \cap H^1_{0,0}((0,T); \widetilde{W}^{-s,2}(\Omega))$ is said to be a weak solution to (3.15) if $u - \tilde{z} \in \mathbb{U}_0$ and

$$\langle \partial_t u(t,\cdot), v \rangle + \frac{C_{N,s}}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{(u(t,x) - u(t,y))(v(x) - v(y))}{|x-y|^{N+2s}} \, dxdy = 0,$$

for every $v \in \widetilde{W}^{-s,2}(\Omega)$ and almost every $t \in (0,T)$.

Next, we show the well-posedness of (3.15).

Theorem 3.5 (**Weak solutions to** (3.15)**).** Let $z \in H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$. Then there exists a unique weak solution $u \in \mathbb{U}$ to (3.15). In addition, there is a constant $C > 0$ such that

$$\|u\|_{\mathbb{U}} \leq C \|z\|_{H^1((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))}. \tag{3.20}$$

Proof. Before we proceed with the proof, we need some preparation. Let us first assume that z depends only on the spatial variable x. Now consider the s-Harmonic extension $\tilde{z} \in W^{s,2}(\mathbb{R}^N)$ of $z \in W^{s,2}(\mathbb{R}^N \setminus \Omega)$ that solves the following Dirichlet problem:

$$\begin{cases} (-\Delta)^s \tilde{z} = 0 & \text{in } \Omega, \\ \tilde{z} = z & \text{in } \mathbb{R}^N \setminus \Omega, \end{cases} \tag{3.21}$$

in a weak sense. That is, given $z \in W^{s,2}(\mathbb{R}^N \setminus \Omega)$, there exists a unique $\tilde{z} \in W^{s,2}(\mathbb{R}^N)$ such that $\tilde{z}|_{\mathbb{R}^N \setminus \Omega} = z$ and $\tilde{z}$ solves (3.21) in the sense that

$$\frac{C_{N,s}}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{(\tilde{z}(x) - \tilde{z}(y))(v(x) - v(y))}{|x-y|^{N+2s}} \, dxdy = 0 \quad \text{for all } v \in \widetilde{W}^{s,2}_0(\Omega),$$

and there is a constant $C > 0$ such that

$$\|\tilde{z}\|_{W^{s,2}(\mathbb{R}^N)} \leq C\|z\|_{W^{s,2}(\mathbb{R}^N\setminus\Omega)}. \tag{3.22}$$

The existence of a weak solution to (3.21) and the continuous dependence on the datum z have been shown in [75] (see also, [68, 130]) under the assumption that Ω has a Lipschitz continuous boundary. If z is a function of (x,t) and belongs to $H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$, then it follows from the above arguments that $\tilde{z} \in H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N))$.

Next, we show the existence of a unique solution to (3.15) by using a lifting argument. We define $w := u - \tilde{z}$. Then $w|_\Sigma = 0$. Moreover, a simple calculation shows that w fulfills

$$\begin{cases} \partial_t w + (-\Delta)^s w = -\partial_t \tilde{z} & \text{in } Q, \\ w = 0 & \text{in } \Sigma, \\ w(0,\cdot) = 0 & \text{in } \Omega. \end{cases} \tag{3.23}$$

Since $\partial_t z \in L^2((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$, it follows from the above discussion that $\partial_t \tilde{z} \in L^2((0,T); W^{s,2}(\mathbb{R}^N))$. Hence, using Proposition 3.2 we get that there exists a unique $w \in \mathbb{U}_0$ solving (3.23). Thus, the unique solution $u \in \mathbb{U}$ is given by $u = w + \tilde{z}$. It remains to show the estimate (3.20). Firstly, since $w = 0$ in Σ, it follows from (3.19) that there is a constant $C > 0$ such that

$$\|w\|_{\mathbb{U}} = \|w\|_{\mathbb{U}_0} \leq C\|\partial_t \tilde{z}\|_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))}. \tag{3.24}$$

Secondly, it follows from (3.22) that there is a constant $C > 0$ such that

$$\|\tilde{z}\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))} \leq C\|z\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))}. \tag{3.25}$$

Thirdly, using (3.24) and (3.25) we get that there is a constant $C > 0$ such that

$$\begin{aligned}
\|u\|_{\mathbb{U}} &= \|w + \tilde{z}\|_{\mathbb{U}} \le \|w\|_{\mathbb{U}} + \|\tilde{z}\|_{\mathbb{U}} \\
&\le C\left(\|\partial_t \tilde{z}\|_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))} + \|z\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))} + \|\tilde{z}\|_{H^1((0,T);\widetilde{W}^{-s,2}(\Omega))}\right) \\
&\le C\left(\|\partial_t \tilde{z}\|_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))} + \|z\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))} + \|\tilde{z}\|_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))}\right), \quad (3.26)
\end{aligned}$$

where in the last estimate we have used the fact that

$$\|\tilde{z}\|^2_{H^1((0,T);\widetilde{W}^{-s,2}(\Omega))} = \|\tilde{z}\|^2_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))} + \|\partial_t \tilde{z}\|^2_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))}.$$

Since $\tilde{z} \in L^2((0,T); W^{s,2}(\mathbb{R}^N))$, it follows from (3.22) that

$$\begin{aligned}
\|\tilde{z}\|_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))} \le& C\|\tilde{z}\|_{L^2((0,T);W^{-s,2}(\mathbb{R}^N))} \le C\|\tilde{z}\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))} \\
\le& C\|z\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))}. \quad (3.27)
\end{aligned}$$

Note that $\partial_t \tilde{z}$ is a solution of the Dirichlet problem (3.21) with z replaced with $\partial_t z$. This shows that $\partial_t \tilde{z} \in L^2((0,T); W^{s,2}(\mathbb{R}^N))$. Hence, using (3.22) again, we obtain that

$$\begin{aligned}
\|\partial_t \tilde{z}\|_{L^2((0,T);\widetilde{W}^{-s,2}(\Omega))} \le& C\|\partial_t \tilde{z}\|_{L^2((0,T);W^{-s,2}(\mathbb{R}^N))} \le C\|\partial_t \tilde{z}\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))} \\
\le& C\|\partial_t z\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))}. \quad (3.28)
\end{aligned}$$

Combining (3.27) and (3.28) we get from (3.26) that

$$\|u\|_{\mathbb{U}} \le C\left(\|z\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))} + \|\partial_t z\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))}\right).$$

That is, we have shown (3.20). □

Remark 3.3. Let $(\varphi_n)_{n\in\mathbb{N}}$ be the orthonormal basis of eigenfunctions of $(-\Delta)^s_D$ associated with the eigenvalues $(\lambda_n)_{n\in\mathbb{N}}$. If in Theorem 3.5 one assumes that $z \in H_0^1((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$, then it has been shown in [134, Theorem 18] that the unique weak solution u of (3.15) is given by

$$u(t,x) = -\sum_{n=1}^{\infty} \left(\int_0^t (z(t-\tau,\cdot), \mathcal{N}_s\varphi_n)_{L^2(\mathbb{R}^N\setminus\Omega)}\, e^{-\lambda_n\tau}\, d\tau \right) \varphi_n(x).$$

Our next goal is to reduce the regularity requirements on the datum z in both space and time. We shall call the resulting solution u to be a very-weak solution.

Definition 3.7 (Very-weak solution to (non-homogeneous) parabolic Dirichlet problem). Let $z \in L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega))$. A function $u \in L^2((0,T); L^2(\mathbb{R}^N))$ is said to be a very-weak solution to (3.15) if the identity

$$\int_Q u\left(-\partial_t v + (-\Delta)^s v\right)\, dxdt = -\int_\Sigma z\mathcal{N}_s v\, dxdt, \tag{3.29}$$

holds for every $v \in L^2((0,T); V) \cap H^1_{0,T}((0,T); L^2(\Omega))$, where $V := \{v \in \widetilde{W}_0^{s,2}(\Omega) : (-\Delta)^s v \in L^2(\Omega)\}$.

Throughout the remainder of the thesis, without any mention we shall let

$$V := \{v \in \widetilde{W}_0^{s,2}(\Omega) :\ (-\Delta)^s v \in L^2(\Omega)\}.$$

The following result shows the existence and uniqueness of a very-weak solution to (3.15) in the sense of Definition 3.7. We will prove this result using a duality argument (see e.g. [70] for the case $s = 1$).

Theorem 3.6. Let $z \in L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega))$. Then there exists a unique very-weak solution u to (3.15) according to Definition 3.7 that fulfills

$$\|u\|_{L^2((0,T);L^2(\Omega))} \le C\|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}, \tag{3.30}$$

for a constant $C > 0$. In addition, if $z \in H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$, then the following assertions hold.

1. Every weak solution of (3.15) is also a very-weak solution.

2. Every very-weak solution of (3.15) that belongs to $\mathbb{U}$ is also a weak solution.

Proof. For a given $\zeta \in L^2((0,T); L^2(\Omega))$, we begin by considering the following "dual" problem:

$$\begin{cases} -\partial_t v + (-\Delta)^s v & = \zeta \quad \text{in } Q, \\ v & = 0 \quad \text{in } \Sigma, \\ v(T, \cdot) & = 0 \quad \text{in } \Omega. \end{cases} \tag{3.31}$$

We notice that in (3.31), it is not required that $\zeta(T, \cdot) = 0$ in Ω. Using semigroup theory as in Proposition 3.2 (see also Remark 3.2), we have that the problem (3.31) has a unique weak solution $v \in L^2((0,T); V) \cap H^1_{0,T}((0,T); L^2(\Omega))$. Hence, $\partial_t v \in L^2(Q)$ and $(-\Delta)^s v \in L^2(Q)$.

Since $v \in L^2((0,T); V) \cap H^1_{0,T}((0,T); L^2(\Omega))$, we have that $\mathcal{N}_s v \in L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega))$. We define the mapping

$$\mathcal{M} : L^2((0,T); L^2(\Omega)) \to L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega)), \qquad \zeta \mapsto \mathcal{M}\zeta := -\mathcal{N}_s v.$$

We notice that $\mathcal{M}$ is linear and continuous because there is a constant $C > 0$ such that

$$\begin{aligned} \|\mathcal{M}\zeta\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))} = \|\mathcal{N}_s v\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))} &\leq C\|v\|_{L^2((0,T);\widetilde{W}^{s,2}_0(\Omega))} \\ &\leq C\|\zeta\|_{L^2((0,T);L^2(\Omega))}. \end{aligned}$$

Let $u := \mathcal{M}^* z$. Then we have

$$\int_Q u\zeta \, dxdt = \int_Q u \left(-\partial_t v + (-\Delta)^s v\right) \, dxdt = \int_Q (\mathcal{M}^* z)\zeta \, dxdt = -\int_\Sigma z\mathcal{N}_s v \, dxdt.$$

We have constructed a function $u \in L^2((0,T);L^2(\mathbb{R}^N))$ that solves (3.29). Next, we show the uniqueness of very-weak solutions. Assume that the system (3.15) has two very weak-solutions u_1 and u_2 with the same exterior value z. Then, it follows from (3.29) that

$$\int_Q (u_1 - u_2)\left(-\partial_t v + (-\Delta)^s v\right)\, dxdt = 0,$$

for every $v \in L^2((0,T);V) \cap H^1_{0,T}((0,T);L^2(\Omega))$. Using the fundamental lemma of the calculus of variations, we can deduce from the preceding identity that $u_1 = u_2$ a.e. in Q. Since $u_1 = u_2$ a.e in Σ, we can conclude that $u_1 = u_2$ a.e. in $(0,T) \times \mathbb{R}^N$. Thus, we have shown the uniqueness of solutions.

Finally, we notice that there is a constant $C > 0$ such that

$$\begin{aligned}\left|\int_Q u\zeta\, dxdt\right| &\leq \|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\|\mathcal{N}_s v\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\\ &\leq C\|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\|\zeta\|_{L^2((0,T);L^2(\Omega))}.\end{aligned}$$

Dividing both sides of the preceding estimate by $\|\zeta\|_{L^2((0,T);L^2(\Omega))}$ and taking the supremum over $\zeta \in L^2((0,T);L^2(\Omega))$, we obtain (3.30).

Next, we prove the last two assertions of the theorem.

Assume that $z \in H^1_{0,0}((0,T);W^{s,2}(\mathbb{R}^N \setminus \Omega))$.

(a) Let $u \in \mathbb{U} \hookrightarrow L^2((0,T);L^2(\mathbb{R}^N))$ be a weak solution to (3.15). It follows from the definition that $u = z$ on Σ and in particular, we have that

$$\langle \partial_t u(t,\cdot), v\rangle + \frac{C_{N,s}}{2}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\frac{(u(t,x)-u(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\, dxdy = 0, \tag{3.32}$$

for every $v \in L^2((0,T);V) \cap H^1_{0,T}((0,T);L^2(\Omega))$ and almost every $t \in (0,T)$. Since $v(t,\cdot) = 0$

in $\mathbb{R}^N \setminus \Omega$, we have that

$$\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{(u(t,x)-u(y,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\,dxdy$$
$$= \int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(u(t,x)-u(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\,dxdy. \tag{3.33}$$

Using (3.32), (3.33), the integration by parts formula (2.12) together with the fact that $u = z$ in $\mathbb{R}^N \setminus \Omega$, we get that

$$\begin{aligned}
0 &= \langle \partial_t u(t,\cdot), v\rangle + \frac{C_{N,s}}{2}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{(u(t,x)-u(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\,dxdy \\
&= \langle \partial_t u(t,\cdot), v(t,\cdot)\rangle + \int_\Omega u(t,x)(-\Delta)^s v(t,x)\,dx + \int_{\mathbb{R}^N\setminus\Omega} u(t,x)\mathcal{N}_s v(t,x)\,dx \\
&= \langle \partial_t u(t,\cdot), v(t,\cdot)\rangle + \int_\Omega u(t,x)(-\Delta)^s v(t,x)\,dx + \int_{\mathbb{R}^N\setminus\Omega} z(t,x)\mathcal{N}_s v(t,x)\,dx.
\end{aligned}$$

Integrating the previous identity by parts over $(0,T)$, we get that

$$-\int_0^T \langle u(t,\cdot), \partial_t v(t,\cdot)\rangle\,dt + \int_Q u(-\Delta)^s v\,dxdt + \int_\Sigma z\mathcal{N}_s v\,dxdt = 0.$$

Since $u(t,\cdot)$, $\partial_t v(t,\cdot) \in L^2(\Omega)$, it follows from the preceding identity that

$$\int_Q u\Big(-\partial_t v + (-\Delta)^s v\Big)\,dxdt = -\int_\Sigma z\mathcal{N}_s v\,dxdt$$

for every $v \in L^2((0,T);V) \cap H^1_{0,T}((0,T);L^2(\Omega))$. Thus, u is a very-weak solution of (3.15).

(b) Let u be a very-weak solution to (3.15) and assume that $u \in \mathbb{U}$. Then $u = z$ in Σ. Moreover, $z \in H^1_{0,0}((0,T);W^{s,2}(\mathbb{R}^N\setminus\Omega))$ and if $\tilde{z} \in H^1_{0,0}((0,T);W^{s,2}(\mathbb{R}^N))$ is such that $\tilde{z}|_\Sigma = z$, then clearly $u - \tilde{z} \in \mathbb{U}_0$. Since u is a very-weak solution to (3.15), it follows from

Definition 3.7 that for every $v \in L^2((0,T);V) \cap H^1_{0,T}((0,T);L^2(\Omega))$, we have

$$\int_Q u(-\partial_t v + (-\Delta)^s v)\, dx = -\int_\Sigma z\mathcal{N}_s v\, dx. \tag{3.34}$$

Since $u \in \mathbb{U}$ and $v = 0$ on Σ, using the integration by parts formula (2.12), we get that

$$\begin{aligned}
&\int_0^T \langle \partial_t u(t,\cdot),(t,\cdot)v\rangle\, dt + \int_0^T \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{(u(t,x)-u(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\, dxdydt \\
&= \int_0^T \langle \partial_t u(t,\cdot),v(t,\cdot)\rangle\, dt + \int_0^T \int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(u(t,x)-u(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\, dxdydt \\
&= \int_Q u\Big(\partial_t v + (-\Delta)^s v\Big)\, dxdt + \int_\Sigma u\mathcal{N}_s v\, dxdt \\
&= \int_Q u\Big(\partial_t v + (-\Delta)^s v\Big)\, dxdt + \int_\Sigma z\mathcal{N}_s v\, dxdt.
\end{aligned} \tag{3.35}$$

It follows form (3.34) and (3.35) that for every $v \in L^2((0,T);V) \cap H^1_{0,T}((0,T);L^2(\Omega))$ we have the identity

$$\int_0^T \langle \partial_t u(t,\cdot),v(t,\cdot)\rangle\, dt + \int_0^T \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{(u(t,x)-u(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\, dxdydt = 0. \tag{3.36}$$

Since V is dense in $\widetilde{W}_0^{s,2}(\Omega)$ and $L^2(\Omega)$ is dense in $\widetilde{W}^{-s,2}(\Omega)$, it follows that (3.36) remains true for every $v \in L^2((0,T);\widetilde{W}_0^{s,2}(\Omega)) \cap H^1_{0,T}((0,T);\widetilde{W}^{-s,2}(\Omega))$. Notice that $v(t,\cdot) \in \widetilde{W}_0^{s,2}(\Omega)$ for a.e. $t \in (0,T]$. As a result, we have that the following pointwise formulation

$$\langle \partial_t u(t,\cdot),v\rangle + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{(u(t,x)-u(t,y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy = 0, \tag{3.37}$$

holds for every $v \in \widetilde{W}_0^{s,2}(\Omega)$ and a.e. $t \in (0,T)$. We have shown that u is the unique weak solution to (3.15) according to Definition 3.6. $\square$

Let us now shift our focus to the problem (3.16) with non-zero source and zero exterior conditions : For source $z \in L^r((0,T); L^p(\Omega))$ with p and r satisfying

$$1 > \frac{N}{2ps} + \frac{1}{r}, \tag{3.38}$$

find u such that

$$\begin{cases} \partial_t u + (-\Delta)^s_D u = z & \text{in } Q, \\ u(0,\cdot) = 0 & \text{in } \Omega. \end{cases} \tag{3.39}$$

The goal is to establish the continuity of solutions to (3.39) when the datum z belongs to $L^r((0,T); L^p(\Omega))$ with p, r fulfilling (3.38) and to study the well-posedness of (3.39) with $z \in (C(\overline{Q}))^\star = \mathcal{M}(\overline{Q})$. The latter space $\mathcal{M}(\overline{Q})$ denotes the space of all Radon measures on $\overline{Q}$ such that

$$\langle \mu, v \rangle_{(C(\overline{Q}))^\star, C(\overline{Q})} = \int_{\overline{Q}} v \, d\mu, \quad \mu \in \mathcal{M}(\overline{Q}), \quad v \in C(\overline{Q}).$$

In addition, we have the following norm on this space:

$$\|\mu\|_{\mathcal{M}(\overline{Q})} = \sup_{v \in C(\overline{Q}),\, |v| \leq 1} \int_{\overline{Q}} v \, d\mu.$$

We begin by proving a boundedness result for solutions to problem (3.39). We remark that a similar result has been recently shown in [99, Theorem 29 and Corollary 3] with a technical proof. Here, we will provide a much simpler proof which follows directly using semigroup arguments.

Theorem 3.7 (*u* **is bounded).** Let $\Omega \subset \mathbb{R}^N$ $(N \geq 1)$ be an arbitrary bounded open set. Let $z \in L^r((0,T); L^p(\Omega))$ with p and r fulfilling (3.38). Then every weak solution $u \in \mathbb{U}_0$

to (3.39) belongs to $L^\infty(Q)$ and there is a constant $C > 0$ such that

$$\|u\|_{L^\infty(Q)} \le C\|z\|_{L^r((0,T);L^p(\Omega))}. \tag{3.40}$$

Proof. Let p and r fulfill (3.38). We remark that using the embedding (2.1), the representation (3.18), and semigroup theory, the result is trivial if $p = \infty$ or $r = \infty$. Thus, without any restriction we may assume that $1 \le r, p < \infty$ satisfy (3.38). We mention that if $1 \le r, p < \infty$ satisfy (3.38), then necessarily $1 < r < \infty$. Notice also that (3.38) implies that $p > \frac{N}{2s} + \frac{p}{r} > \frac{N}{2s}$. Thus, it follows from the embedding (2.1) that the submarkovian semigroup $(e^{-t(-\Delta)^s_D})_{t\ge 0}$ is ultracontractive in the sense that it maps $L^1(\Omega)$ into $L^\infty(\Omega)$. More precisely, following line by line the proof of [65, Theorem 2.16] or the proof of the abstract result in [108, Lemma 6.5] (see also [47, Chapter 2]), we get that for all $1 \le p \le q \le \infty$, there exists a constant $C > 0$ such that for every $f \in L^p(\Omega)$ and $t > 0$ we have the estimate:

$$\|e^{-t(-\Delta)^s_D} f\|_{L^q(\Omega)} \le Ce^{-\lambda_1\left(\frac{1}{p}-\frac{1}{q}\right)t} t^{-\frac{N}{2s}\left(\frac{1}{p}-\frac{1}{q}\right)}\|f\|_{L^p(\Omega)}, \tag{3.41}$$

where $\lambda_1 > 0$ denotes the first eigenvalue of the operator $(-\Delta)^s_D$.

Next, applying (3.41) with $q = \infty$ and using the representation (3.18) of the solution u, we get that

$$\|u(t,\cdot)\|_{L^\infty(\Omega)} \le C\int_0^t e^{-\lambda_1\frac{(t-\tau)}{p}}(t-\tau)^{-\frac{N}{2sp}}\|z(\tau,\cdot)\|_{L^p(\Omega)}d\tau. \tag{3.42}$$

Using Young's convolution inequality, we get from (3.42) that

$$\|u\|_{L^\infty(Q)} \le C\|z\|_{L^r((0,T),L^p(\Omega))}\left(\int_0^T e^{-\lambda_1\frac{rt}{p(r-1)}}t^{-\frac{Nr}{2sp(r-1)}}\,dt\right)^{\frac{r-1}{r}}. \tag{3.43}$$

If $1 > \frac{N}{2sp}\frac{r}{r-1}$, that is, if p and r fulfill (3.38), then the integral in the right-hand side of (3.43) is convergent. $\square$

Under the assumption that Ω has the exterior ball condition, we obtain Corollary 3.1 as a direct consequence of Theorem 3.7. As remarked earlier, the following result for the parabolic problem (3.39) is new and cannot be obtained by using classical properties of parabolic problems. In fact, using maximal regularity results for abstract Cauchy problems (see e.g. [52]), we get that for $z \in L^r((0,T); L^p(\Omega))$, the weak solution u to (3.39) belongs to $W^{1,r}((0,T); L^p(\Omega)) \cap L^r((0,T); D((-\Delta)_D^s))$. This result only shows that $u \in C([0,T]; L^p(\Omega)) \cap L^r((0,T); D(-\Delta)_D^s))$. The global continuity of solutions on $\overline{Q}$ cannot be derived from the above maximal regularity result.

Corollary 3.1 (**u is continuous**)**.** Let $\Omega \subset \mathbb{R}^N$ $(N \geq 1)$ be a bounded open set with a Lipschitz continuous boundary and satisfies the exterior ball condition. Let $z \in L^r((0,T); L^p(\Omega))$ with p and r fulfilling (3.38) and $p, r < \infty$. Then every weak solution $u \in \mathbb{U}_0$ to (3.39) belongs to $C(\overline{Q})$ and there is a constant $C > 0$ such that

$$\|u\|_{C(\overline{Q})} \leq C\|z\|_{L^r((0,T);L^p(\Omega))}. \tag{3.44}$$

Proof. We prove the result in two steps.

Step 1: Let $\lambda \geq 0$ be real number, $f \in L^2(\Omega)$ and consider the following elliptic Dirichlet problem:

$$\begin{cases} (-\Delta)^s w + \lambda w = f & \text{in } \Omega \\ w = 0 & \text{in } \mathbb{R}^N \setminus \Omega. \end{cases} \tag{3.45}$$

By a weak solution of (3.45) we mean a function $w \in \widetilde{W}_0^{s,2}(\Omega)$ such that the equality

$$\frac{C_{N,s}}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{(w(x) - w(y))(v(x) - v(y))}{|x-y|^{N+2s}}\, dxdy + \lambda \int_\Omega wv\, dx = \int_\Omega fv\, dx,$$

holds for every $v \in \widetilde{W}_0^{s,2}(\Omega)$.

The existence and uniqueness of weak solutions to the Dirichlet problem (3.45) are a direct consequence of the classical Lax-Milgram theorem.

In Theorem 3.2, we have shown that, if $f \in L^p(\Omega)$ with $p > \frac{N}{2s}$ (see also [113] for the case $p = \infty$), then every weak solution w of the Dirichlet problem (3.45) belongs to $C_0(\Omega) := \{w \in C(\overline{\Omega}) : \; w = 0 \;\text{ on }\; \partial\Omega\}$. Thus, the resolvent operator $R(\lambda, (-\Delta)^s_D)$ maps $L^p(\Omega)$ into $C_0(\Omega)$ for every $\lambda \geq 0$. In particular, this shows that for every $t > 0$, the operator $e^{-t(-\Delta)^s_D}$ maps $L^p(\Omega)$ ($p > \frac{N}{2s}$) into the space $C_0(\Omega)$, that is, the semigroup $(e^{-t(-\Delta)^s_D})_{t\geq 0}$ has the strong Feller property. In addition, we have the following result which is interesting in its own, independently of the application given in this proof.

Let $(-\Delta)^s_{D,c}$ be the part of the operator $(-\Delta)^s_D$ in $C_0(\Omega)$, that is,

$$\begin{cases} D((-\Delta)^s_{D,c}) := \Big\{ v \in D((-\Delta)^s_D) \cap C_0(\Omega) : \; ((-\Delta)^s_D v)|_{\Omega} \in C_0(\Omega) \Big\}, \\ (-\Delta)^s_{D,c} v = ((-\Delta)^s_D v)|_{\Omega}. \end{cases}$$

From the above properties of the resolvent operator and the semigroup, together with the fact that $D((-\Delta)^s_{D,c})$ is dense in $C_0(\Omega)$, we can deduce that the operator $-(-\Delta)^s_{D,c}$ generates a strongly continuous semigroup $(e^{-t(-\Delta)^s_{D,c}})_{t\geq 0}$ on $C_0(\Omega)$. Thus, for every $f \in C_0(\Omega)$ we have that the function

$$v(t,x) := \int_0^t e^{-(t-\tau)(-\Delta)^s_{D,c}} f(x) \, d\tau$$

belongs to $C([0,\infty), C_0(\Omega))$. We can then deduce that for every $z \in C([0,T]; C_0(\Omega))$, the unique weak solution $u \in \mathbb{U}_0$ to (3.39) given by

$$u(t,x) := \int_0^t e^{-(t-\tau)(-\Delta)^s_{D,c}} z(\tau, x) \, d\tau$$

belongs to $C(\overline{Q})$.

Step 2: Now, let p and r satisfy (3.38). Since the space $C([0,T];C_0(\Omega))$ is dense in $L^r((0,T);L^p(\Omega))$, we can construct a sequence $\{z_n\}_{n\in\mathbb{N}}$ such that

$$\{z_n\}_{n\in\mathbb{N}} \subset C([0,T];C_0(\Omega)) \quad \text{and} \quad z_n \to z \text{ in } L^r((0,T);L^p(\Omega)) \text{ as } n \to \infty.$$

Let $u_n \in \mathbb{U}_0$ be the weak solution to (3.39) with datum z_n. It follows from Step 1 that $u_n \in C(\overline{Q})$. Whence, subtracting the equations satisfied by (u,z) and (u_n,z_n) and using the estimate (3.40) from Theorem 3.7 we obtain that there is a constant $C > 0$ such that for every $n \in \mathbb{N}$ we have

$$\|u - u_n\|_{L^\infty(Q)} \leq C\|z - z_n\|_{L^r((0,T);L^p(\Omega))}.$$

As a result, we have that $u_n \to u$ in $L^\infty(Q)$ as $n \to \infty$. Since u is the uniform limit on Q of a sequence of continuous functions $\{u_n\}_{n\in\mathbb{N}}$ on $\overline{Q}$, it follows that u is also continuous on $\overline{Q}$. $\square$

Our next result shows the well-posedness of the parabolic PDE when the right-hand side and initial conditions are given by a Radon measure as can be seen in the following: For $z \in \mathcal{M}(\overline{Q})$, let $z_Q := z|_Q$ and $z_0 := z|_{\{0\}\times\overline{\Omega}}$, and consider

$$\begin{cases} \partial_t u + (-\Delta)^s)_D u &= z_Q \quad \text{in } Q, \\ u(0,\cdot) &= z_0 \quad \text{in } \overline{\Omega}. \end{cases} \tag{3.46}$$

However, prior to this result, we need to introduce the notion of very-weak solutions similar to the fractional elliptic problems, cf. Definition 3.3.

Definition 3.8 (Very-weak solutions). Let $z \in \mathcal{M}(\overline{Q})$ and p, r satisfy (3.38). A function

$u \in \Big(L^r((0,T);L^p(\Omega))\Big)^\star$ is said to be a very-weak solution to (3.46) if the identity

$$\int_Q u\left(-\partial_t v + (-\Delta)^s v\right)\, dxdt = \int_{\overline{Q}} v dz$$
$$= \int_Q v dz_Q(t,x) + \int_{\overline{\Omega}} v(0,x) dz_0(x),$$

holds for every $v \in \Big\{ C(\overline{Q}) \cap \mathbb{U}_0 \ : \ v(T,\cdot) = 0 \text{ in } \Omega,\ \left(-\partial_t + (-\Delta)^s\right) v \in L^r((0,T);L^p(\Omega)) \Big\}$.

Now, we prove the existence and uniqueness of very-weak solutions to (3.46).

Theorem 3.8. Let $1 \le p, r < \infty$ fulfill (3.38) and $z \in \mathcal{M}(\overline{Q})$. Then there is a unique very weak solution $u \in (L^r((0,T);L^p(\Omega)))^\star$ of (3.46). Moreover, there is a constant $C > 0$ such that

$$\|u\|_{L^r((0,T);L^p(\Omega))^\star} \le C \|z\|_{\mathcal{M}(\overline{Q})}. \tag{3.47}$$

Proof. We prove the theorem in three steps.

Step 1: Given $\zeta \in L^r((0,T);L^p(\Omega))$ where p, r fulfill (3.38), we begin by considering the following "dual" problem

$$\begin{cases} -\partial_t w + (-\Delta)^s w &= \zeta \quad \text{in } Q, \\ w &= 0 \quad \text{in } \Sigma, \\ w(T,\cdot) &= 0 \quad \text{in } \Omega. \end{cases} \tag{3.48}$$

After using semigroup theory as in Proposition 3.2, we can deduce that (3.48) has a unique weak solution $w \in \mathbb{U}_0$. In addition, from (3.48) we have that $\left(-\partial_t + (-\Delta)^s\right) w \in L^r((0,T);L^p(\Omega))$. It follows from Corollary 3.1 that $w \in C(\overline{Q})$. Thus w is a valid "test function" according to Definition 3.8.

Step 2: Towards this end, we define the map

$$\Xi : L^r((0,T); L^p(\Omega)) \to C(\overline{Q})$$

$$\zeta \mapsto \Xi\zeta =: w.$$

Due to Corollary 3.1, Ξ is linear and continuous.

We are now ready to construct a unique u. We set $u := \Xi^* z$, then $u \in (L^r((0,T); L^p(\Omega)))^\star$, moreover u solves (3.46) according to Definition 3.8. Indeed

$$\int_Q u\zeta \; dxdt = \int_Q u\left(-\partial_t w + (-\Delta)^s w\right) \; dxdt = \int_Q (\Xi^* z)\zeta \; dxdt = \int_{\overline{Q}} w dz, \tag{3.49}$$

that is, u is a solution of (3.46) according to Definition 3.8 and we have shown the existence.

Next, we prove the uniqueness of very-weak solution. Assume that (3.46) has two very weak solutions u_1 and u_2 with the same right hand side datum z. Then it follows from (3.49) that

$$\int_Q (u_1 - u_2)\left(-\partial_t v + (-\Delta)^s v\right) \; dxdt = 0, \tag{3.50}$$

for every $v \in \Big\{C(\overline{Q}) \cap \mathbb{U}_0 \;:\; v(T,\cdot) = 0 \text{ in } \Omega,\; \left(-\partial_t + (-\Delta)^s\right) v \in L^r((0,T); L^p(\Omega))\Big\}$. It follows from Step 1 that the mapping

$$\mathbb{U}_0 \cap C(\overline{Q}) \to L^r((0,T); L^p(\Omega)) : \quad v \mapsto \left(-\partial_t v + (-\Delta)^s v\right)$$

is surjective. Thus, we can deduce from (3.50) that

$$\int_Q (u_1 - u_2) w \; dxdt = 0,$$

for every $w \in L^r((0,T); L^p(\Omega))$. Exploiting the fundamental lemma of the calculus of variations we can conclude from the preceding identity that $u_1 - u_2 = 0$ a.e. in Ω and we have shown the uniqueness.

Step 3: It then remains to show the bound (3.47). It follows from (3.49) that

$$\left|\int_Q u\zeta \, dxdt\right| \le \|z\|_{\mathcal{M}(\overline{Q})}\|w\|_{C(\overline{Q})} \le C\|z\|_{\mathcal{M}(\overline{Q})}\|\zeta\|_{L^r((0,T);L^p(\Omega))}, \tag{3.51}$$

where in the last step we have used Corollary 3.1. Finally dividing both sides of the estimate (3.51) by $\|\zeta\|_{L^r((0,T);L^p(\Omega))}$ and taking the supremum over all functions $\zeta \in L^r((0,T);L^p(\Omega))$ we obtain the desired result. $\square$

3.2.2 Fractional Parabolic Problem with Robin Conditions

This section considers the parabolic Robin exterior value problem

$$\begin{cases} \partial_t u + (-\Delta)^s u = 0 & \text{in } Q, \\ \mathcal{N}_s u + \kappa u = \kappa z & \text{in } \Sigma, \\ u(0,\cdot) = 0 & \text{in } \Omega, \end{cases} \tag{3.52}$$

where N_s denotes the interaction operator given in (2.11) and $\kappa \in L^1(\mathbb{R}^N \setminus \Omega) \cap L^\infty(\mathbb{R}^N \setminus \Omega)$ is non-negative.

Let $W^{s,2}_{\Omega,\kappa}$ be the Banach space introduced in (2.6) and μ be the measure on $\mathbb{R}^N \setminus \Omega$ given by $d\mu = |\kappa|dx = \kappa dx$, since we have assumed that κ is non-negative. In this subsection $\langle \cdot, \cdot \rangle$ shall denote the duality pairing between $(W^{s,2}_{\Omega,\kappa})^\star$ and $W^{s,2}_{\Omega,\kappa}$. Next, we introduce our notion of weak solutions to the Robin problem.

Definition 3.9. Let $z \in L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega, \mu))$. A function $u \in L^2((0,T);W^{s,2}_{\Omega,\kappa}) \cap$

$H^1_{0,0}((0,T);(W^{s,2}_{\Omega,\kappa})^\star)$ is said to be a weak solution of (3.52) if the identity

$$\langle \partial_t u(t,\cdot), v\rangle + \int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(u(t,x)-u(t,y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy \\ + \int_{\mathbb{R}^N\setminus\Omega} \kappa(x)u(t,x)v(x)\, dx = \int_{\mathbb{R}^N\setminus\Omega} \kappa(x)z(t,x)v(x)\, dx, \tag{3.53}$$

holds for every $v \in W^{s,2}_{\Omega,\kappa}$ and almost every $t \in (0,T)$.

Throughout the following subsection, for $u, v \in W^{s,2}_{\Omega,\kappa}$ we shall denote

$$\mathcal{E}(u,v) := \frac{C_{N,s}}{2}\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(u(x)-u(y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy + \int_{\mathbb{R}^N\setminus\Omega} \kappa uv\, dx.$$

Next, we show the existence and uniqueness of weak solutions to (3.52).

Theorem 3.9. Let $\kappa \in L^1(\mathbb{R}^N\setminus\Omega)\cap L^\infty(\mathbb{R}^N\setminus\Omega)$ be non-negative. Then for every $z \in L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega,\mu))$, there exists a unique weak solution $u \in L^2((0,T);W^{s,2}_{\Omega,\kappa})\cap H^1_{0,0}((0,T);(W^{s,2}_{\Omega,\kappa})^\star)$ of (3.52).

Proof. We prove the result in several steps.

Step 1. Define the operator A in $L^2(\Omega)\times L^2(\mathbb{R}^N\setminus\Omega,\mu)$ as follows:

$$\begin{cases} D(A) := \Big\{(u,0) :\ u \in W^{s,2}_{\Omega,\kappa},\ (-\Delta)^s u \in L^2(\Omega),\ \mathcal{N}_s u \in L^2(\mathbb{R}^N\setminus\Omega,\mu)\Big\}, \\ A(u,0) = \left(-(-\Delta)^s u, -\mathcal{N}_s u - \kappa u\right). \end{cases}$$

Let $(f,g) \in L^2(\Omega)\times L^2(\mathbb{R}^N\setminus\Omega,\mu)$. We claim that $(u,0)\in D(A)$ with $-A(u,0)=(f,g)$ if and only if

$$\mathcal{E}(u,v) = \int_\Omega fv\, dx + \int_{\mathbb{R}^N\setminus\Omega} gv\, d\mu, \tag{3.54}$$

for all $v \in W^{s,2}_{\Omega,\kappa}$. Indeed, we have that $(u,0) \in D(A)$ with $-A(u,0) = (f,g)$ if and only if u is a weak solution of the following elliptic problem:

$$\begin{cases} (-\Delta)^s u = f & \text{in } \Omega, \\ \mathcal{N}_s u + \kappa u = \kappa g & \text{in } \mathbb{R}^N \setminus \Omega. \end{cases} \tag{3.55}$$

It has been shown in [6] (see also [103]) that u solves (3.55) if and only if (3.54) holds and the claim is proved.

Step 2. Firstly, let $\lambda > 0$ be a real number. We show that the operator $\lambda - A : D(A) \to L^2(\Omega) \times L^2(\mathbb{R}^N \setminus \Omega, \mu)$ is invertible. It is clear that for every $\lambda > 0$ there is a constant $\alpha > 0$ such that

$$\lambda \int_\Omega |u|^2 \, dx + \mathcal{E}(u,u) \geq \alpha \|u\|^2_{W^{s,2}_{\Omega,\kappa}} \tag{3.56}$$

for all $u \in W^{s,2}_{\Omega,\kappa}$. Hence, by Lax-Milgram's Theorem, for every $(f,g) \in L^2(\Omega) \times L^2(\mathbb{R}^N \setminus \Omega, \mu)$ there exists a unique $u \in W^{s,2}_{\Omega,\kappa}$ such that

$$\lambda \int_\Omega uv \, dx + \mathcal{E}(u,v) = \int_\Omega fv \, dx + \int_{\mathbb{R}^N \setminus \Omega} gv \, d\mu, \tag{3.57}$$

for all $v \in W^{s,2}_{\Omega,\kappa}$. By Step 1, this means that there is a unique $u \in W^{s,2}_{\Omega,\kappa}$ with $(u,0) \in D(A)$ and

$$(\lambda - A)(u,0) = (\lambda u, 0) - A(u,0) = (f,g).$$

We have shown that $\lambda - A : D(A) \to L^2(\Omega) \times L^2(\mathbb{R}^N \setminus \Omega, \mu)$ is a bijection for every $\lambda > 0$.

Secondly, assume now that $f \leq 0$ a.e. in Ω and $g \leq 0$ μ-a.e. in $\mathbb{R}^N \setminus \Omega$. Let the function $(u,0) := (\lambda - A)^{-1}(f,g)$ and set $v := u^+ := \max\{u, 0\}$. It follows from [133] that

$u^+ \in W^{s,2}_{\Omega,\kappa}$. Let $u^- := \max\{-u, 0\}$. Since

$$\begin{aligned}\Big(u^-(x) - u^-(y)\Big)\Big(u^+(x) - u^+(y)\Big) =& u^-(x)u^+(x) - u^-(x)u^+(y) - u^-(y)u^+(x) + u^-(y)u^+(y)\\ =& - u^-(x)u^+(y) - u^-(y)u^+(x) \leq 0,\end{aligned}$$

we have that $\mathcal{E}(u^-, u^+) \leq 0$. Hence,

$$\mathcal{E}(u, v) = \mathcal{E}(u^+ - u^-, u^+) = \mathcal{E}(u^+, u^+) - \mathcal{E}(u^-, u^+) \geq 0.$$

Then by (3.57), we have that

$$0 \leq \lambda \int_\Omega |u|^2\, dx + \mathcal{E}(u, u^+) = \int_\Omega f u^+\, dx + \int_{\mathbb{R}^N \setminus \Omega} g u^+\, d\mu \leq 0.$$

By (3.56) this implies that $u^+ = 0$, that is, $u \leq 0$ almost everywhere. We have shown that the resolvent $(\lambda - A)^{-1}$ is a positive operator. Since every positive linear operator is continuous (see e.g., [20]), we can deduce that $(\lambda - A)$ is in fact invertible.

Thirdly, we have in particular shown that the operator A is closed since $-A$ is the operator associated with the closed form $\mathcal{E}$. Hence, $D(A)$ endowed with the graph norm is a Banach space and by definition of A, we have that $D(A) \subset W^{s,2}_{\Omega,\kappa} \times \{0\}$. Since both of these spaces are continuously embedded into $L^2(\Omega) \times L^2(\mathbb{R}^N \setminus \Omega, \mu)$, we can deduce from the closed graph theorem that $D(A)$ is continuously embedded into $W^{s,2}_{\Omega,\kappa} \times \{0\}$.

Step 3. Now since $L^2(\Omega) \times L^2(\mathbb{R}^N \setminus \Omega, \mu)$ is a Banach lattice with an order continuous norm and by Step 2 the operator A is resolvent positive, it follows from [19, Theorem 3.11.7] that $-A$ generates a once integrated semigroup on $L^2(\Omega) \times L^2(\mathbb{R}^N \setminus \Omega, \mu)$. Hence, using the theory of integrated semigroups and abstract Cauchy problems studied in [19, Section 3.11] and proceeding as in [106, Section 2], we can deduce that for every $z \in$

$L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega,\mu))$, the problem (3.52) has a unique weak solution. $\square$

We conclude this section by showing that if z is more regular in the time variable, then the existence of weak solutions to (3.52) can easily be proved without using the theory of integrated semigroups.

Proposition 3.3. Let $\kappa \in L^1(\mathbb{R}^N\setminus\Omega)\cap L^\infty(\mathbb{R}^N\setminus\Omega)$. Then for every $z \in H^1_{0,0}((0,T);L^2(\mathbb{R}^N\setminus\Omega,\mu))$, there exists a unique weak solution $u \in L^2((0,T);W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,0}((0,T);(W^{s,2}_{\Omega,\kappa})^\star)$ of (3.52).

Proof. We proceed as in the proof of Theorem 3.5. Firstly, assume that $z \in L^2(\mathbb{R}^N \setminus \Omega,\mu)$ does not depend on the time variable. Let $\tilde{z}$ be the solution of the following elliptic Robin problem:

$$\begin{cases} (-\Delta)^s \tilde{z} = 0 & \text{in } \Omega, \\ \mathcal{N}_s \tilde{z} + \kappa \tilde{z} = \kappa z & \text{in } \mathbb{R}^N \setminus \Omega, \end{cases} \tag{3.58}$$

in the sense that $\tilde{z} \in W^{s,2}_{\Omega,\kappa}$ and

$$\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(\tilde{z}(x)-\tilde{z}(y))(v(x)-v(y))}{|x-y|^{N+2s}}\,dxdy + \int_{\mathbb{R}^N\setminus\Omega} \kappa\tilde{z}v\,dx = \int_{\mathbb{R}^N\setminus\Omega} \kappa z v\,dx, \tag{3.59}$$

for every $v \in W^{s,2}_{\Omega,\kappa}$. Under our assumptions, it has been shown in [6] that (3.58) has a unique solution $\tilde{z}$.

Secondly, assume that $z \in H^1_{0,0}((0,T);L^2(\mathbb{R}^N \setminus \Omega,\mu))$. Since in this case $\partial_t \tilde{z}$ will be a solution of (3.58) with z replaced by $\partial_t z$, we can deduce that (3.58) has a unique solution $\tilde{z} \in H^1_{0,0}((0,T);W^{s,2}_{\Omega,\kappa})$.

Consider the following parabolic problem with $w := u - \tilde{z}$:

$$\begin{cases} \partial_t w + (-\Delta)^s w = -\partial_t \tilde{z} & \text{in } Q, \\ \mathcal{N}_s w + \kappa w = 0 & \text{in } \Sigma, \\ w(0, \cdot) = 0 & \text{in } \Omega. \end{cases} \tag{3.60}$$

Let $(-\Delta)^s_R$ be the realization in $L^2(\Omega)$ of $(-\Delta)^s$ with the zero Robin exterior condition $\mathcal{N}_s w + \kappa w = 0$ in $\mathbb{R}^N \setminus \Omega$. We refer to [45] for a precise description of this operator. Then the parabolic problem (3.60) can be rewritten as the following Cauchy problem

$$\begin{cases} \partial_t w + (-\Delta)^s_R w = -\partial_t \tilde{z} & \text{in } Q, \\ w(0, \cdot) = 0 & \text{in } \Omega. \end{cases}$$

It has been shown in [45] (see also [103]) that the operator $-(-\Delta)^s_R$ generates a strongly continuous submarkovian semigroup $(e^{-t(-\Delta)^s_R})_{t \geq 0}$ in $L^2(\Omega)$. Hence, using semigroup theory, we can deduce that (3.60) has a unique weak solution w that belongs to $L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa})^\star)$ and is given by

$$w(t,x) = -\int_0^t e^{-(t-\tau)(-\Delta)^s_R} \partial_\tau \tilde{z}(\tau, x)\, dx.$$

It is clear that $u := w + \tilde{z}$ is the unique weak solution of (3.52). $\square$

Chapter 4: External Optimal Control of Fractional Parabolic PDEs

In this chapter, we extend the work in [6] on the elliptic (stationary) case to the parabolic (non-stationary) case. The previous chapters have laid the foundation to study this novel parabolic optimal control problem where the control lies in the exterior. As mentioned earlier, the need for these novel optimal control concepts stems from the fact that the classical PDE models only allow placing the control/source either on the boundary or in the interior where the PDE is satisfied. However, the nonlocal behavior of the fractional operator now allows placing the control/source in the exterior.

Let $\Omega \subset \mathbb{R}^N$, $N \geq 1$, be a bounded open set with boundary $\partial\Omega$. Consider the Banach spaces (Z_D, U_D) and (Z_R, U_R), where the subscripts D and R denote Dirichlet and Robin, respectively. The goal of this chapter is to study the following parabolic external optimal control (or source identification) problems:

- **Fractional parabolic Dirichlet exterior control (source identification) problem:** Given $\xi \geq 0$ a constant penalty parameter we consider the following minimization problem:

$$\min_{(u,z)\in(U_D,Z_D)} \Big(J(u) + \frac{\xi}{2}\|z\|^2_{Z_D} \Big), \tag{4.1a}$$

  subject to the fractional parabolic Dirichlet exterior value problem: Find $u \in U_D$ solving

$$\begin{cases} \partial_t u + (-\Delta)^s u = 0 & \text{in } Q := (0,T)\times\Omega, \\ u = z & \text{in } \Sigma := [0,T)\times(\mathbb{R}^N \setminus \Omega), \\ u(0,\cdot) = 0 & \text{in } \Omega, \end{cases} \tag{4.1b}$$

and the control constraints

$$z \in Z_{ad,D}, \tag{4.1c}$$

with $Z_{ad,D} \subset Z_D$ being a closed and convex subset. Here, $Z_D := L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega))$, $U_D := L^2((0,T); L^2(\Omega))$ and the functional J is assumed to be weakly lower-semicontinuous and satisfies suitable conditions. We refer to Section 4.1.1 for more details.

- **Fractional parabolic Robin exterior control (source identification) problem:** Given $\xi \geq 0$ a constant penalty parameter we consider the minimization problem

$$\min_{(u,z)\in(U_R,Z_R)} \Big(J(u) + \frac{\xi}{2}\|z\|^2_{Z_R} \Big), \tag{4.2a}$$

subject to the fractional parabolic Robin exterior value problem: Find $u \in U_R$ solving

$$\begin{cases} \partial_t u + (-\Delta)^s u = 0 & \text{in } Q, \\ \mathcal{N}_s u + \kappa u = \kappa z & \text{in } \Sigma, \\ u(0,\cdot) = 0 & \text{in } \Omega, \end{cases} \tag{4.2b}$$

and the control constraints

$$z \in Z_{ad,R}, \tag{4.2c}$$

with $Z_{ad,R} \subset Z_R$ being a closed and convex subset. Recall, $\mathcal{N}_s$ denotes the interaction operator as is given in (2.11), $\kappa \in L^1(\mathbb{R}^N \setminus \Omega) \cap L^\infty(\mathbb{R}^N \setminus \Omega)$ and is non-negative. The Banach spaces $Z_R := L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega, \mu))$,

$U_R := L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1((0,T); (W^{s,2}_{\Omega,\kappa})^\star)$ and the functional J is also weakly lower-semicontinuous and satisfies suitable conditions. We refer to Chapter 2 for the definition of the spaces involved and to Section 4.1.2 for further details on the functional J.

A widely used example of a functional J is as follows. Let $u_d \in L^2((0,T); L^2(\Omega))$ be given and consider the functional J defined by

$$J(u) := \frac{1}{2}\|u - u_d\|^2_{L^2((0,T);L^2(\Omega))}.$$

A typical example of a control constraint set, for instance, in the case of the Robin problem is as follows: given z_a, z_b with $z_a \le z_b$, we can take

$$Z_{ad,R} := \{z \in Z_R \ : \ z_a(t,x) \le z(t,x) \le z_b(t,x), \text{ a.e. in } \Sigma\}.$$

Nevertheless, our approach is not limited to these choices.

Notice that (4.2b) is a generalized exterior value problem and all the details (with minor modifications) transfer to the case when instead of $\mathcal{N}_s u + \kappa u = \kappa z$ in Σ we consider $\mathcal{N}_s u = \kappa z$ in Σ, where z denotes the control/source. The resulting optimal control problem is the parabolic Neumann exterior control problem.

The classical parabolic models, such as diffusion equations (with $s = 1$), are too restrictive. They only allow a source or a control placement either inside the domain Ω or on the boundary $\partial\Omega$. Notice that in both (4.1b) and (4.2b) the source/control z is placed in the exterior domain $\mathbb{R}^N \setminus \Omega$, disjoint from Ω. This is not possible for the classical models.

The key difficulties and novelties of this chapter are as follows:

(i) **Nonlocal normal derivative.** $\mathcal{N}_s u$ is the nonlocal normal derivative of u. This can be thought of as a restricted fractional Laplacian in $\mathbb{R}^N \setminus \Omega$. It is a very difficult object to handle both at the continuous and discrete levels. Indeed, the best known regularity result for $\mathcal{N}_s$ is given in Lemma 2.1 which says that $\mathcal{N}_s u \in W^{s,2}_{\text{loc}}(\mathbb{R}^N \setminus \Omega)$ whenever $u \in W^{s,2}(\mathbb{R}^N)$. Higher regularity results are currently unknown.

(ii) **Approximation of the Dirichlet problem by a Robin problem.** In the case of the parabolic Dirichlet problem (4.1), it is imperative to deal with $\mathcal{N}_s$. Indeed, we need to approximate the very-weak solution to the parabolic Dirichlet problem (4.1b) which requires computing $\mathcal{N}_s$ of the test functions (see (3.29)).

(iii) **Optimal control problems.** We establish the well-posedness of solutions to both parabolic Dirichlet and Robin control problems.

We remark that the parabolic case considered in this chapter is fundamentally different than the elliptic case considered in [6]. Section 4.1 deals with the Dirichlet and Robin fractional parabolic optimal control problems. The proofs in this section use standard calculus of variations technique, this is similar to the elliptic case, however one has to deal with both the space and time variables and the notion of solutions to the parabolic problems. Notice that in addition, now one has to solve the adjoint equation backward in time.

Section 4.2 deals with the approximation of the parabolic Dirichlet problem and the parabolic Dirichlet control problem by the parabolic Robin ones. In Theorem 4.5 we approximate the Dirichlet solutions by the Robin solutions. A similar result in the elliptic case was considered in [6]. The proof in the parabolic case to some extent is motivated by the elliptic case but one has to deal with both the space and time variables which requires a careful analysis and several changes in the previous arguments, for instance, the duality arguments are different. In Theorem 4.6 we present the approximation of the Dirichlet control/inverse problem by the Robin control/source problem. The arguments in this case are similar to the elliptic case after adapting it to the parabolic case (the proof has been omitted).

Finally, in Section 4.3 we present a numerical scheme to approximate the fractional parabolic state equation and the control/inverse problems. All the results presented here are completely new as such parabolic problems have not been considered before in the literature. The experiments illustrate the strength of the nonlocal approach over the local ones.

4.1 Exterior Optimal Control Problems

The purpose of this section is to study the Dirichlet and Robin optimal control problems (4.1) and (4.2), respectively. These are the subjects investigated in Sections 4.1.1 and 4.1.2, respectively.

4.1.1 Fractional Parabolic Dirichlet Exterior Optimal Control Problem

Due to Theorem 3.6, the control-to-state (solution) map

$$S : Z_D \to U_D, \quad z \mapsto Sz =: u,$$

is well-defined, linear and continuous. Furthermore, for $z \in Z_D$, we have $u := Sz \in L^2((0,T); L^2(\mathbb{R}^N))$. Let $J : U_D \to \mathbb{R}$ and consider the reduced functional

$$\mathcal{J} : Z_D \to \mathbb{R}, \;\; z \mapsto \mathcal{J}(z) := \Big(J(Sz) + \frac{\xi}{2}\|z\|^2_{Z_D} \Big).$$

Then we can write the reduced Dirichlet exterior parabolic optimal control problem as follows:

$$\min_{z \in Z_{ad,D}} \mathcal{J}(z). \tag{4.3}$$

Next, we state the well-posedness result for (4.1) and equivalently for (4.3).

Theorem 4.1. Let $Z_{ad,D}$ be a closed and convex subset of Z_D. Let either $\xi > 0$ with $J \geq 0$ or $Z_{ad,D}$ bounded and $J : U_D \to \mathbb{R}$ weakly lower-semicontinuous. Then there exists a solution $\bar{z}$ to (4.3) and equivalently to (4.1). If either J is convex and $\xi > 0$ or J is strictly convex and $\xi \geq 0$, then $\bar{z}$ is unique.

Proof. The proof is based on the direct method or the Weierstrass theorem [21, Theorem 3.2.1]. We sketch the proof here for completeness. For the functional $\mathcal{J} : Z_{ad,D} \to \mathbb{R}$, it is possible to construct a minimizing sequence $\{z_n\}_{n\in\mathbb{N}}$ (see [21, Theorem 3.2.1]) such that $\inf_{z\in Z_{ad,D}} \mathcal{J}(z) = \lim_{n\to\infty} \mathcal{J}(z_n)$. If $\xi > 0$ with $J \geq 0$ or $Z_{ad,D} \subset Z_D$ is bounded, then $\{z_n\}_{n\in\mathbb{N}}$ is a bounded sequence in Z_D which is a Hilbert space. As a result, we have that (up to a subsequence if necessary) $z_n \rightharpoonup \bar{z}$ (weak convergence) in Z_D as $n \to \infty$. Since $Z_{ad,D}$ is closed and convex, hence, is weakly closed, we have that $\bar{z} \in Z_{ad,D}$.

It remains to show that $(S\bar{z}, \bar{z})$ fulfills the state equation according to Definition 3.7 and $\bar{z}$ is a minimizer to (4.3). In order to show that $(S\bar{z}, \bar{z})$ fulfills the state equation, we need to focus on the identity

$$\int_Q u_n \left(-\partial_t v + (-\Delta)^s v\right)\, dxdt = -\int_\Sigma z_n \mathcal{N}_s v \; dxdt \tag{4.4}$$

for all $v \in L^2((0,T); V) \cap H^1_{0,T}((0,T); L^2(\Omega))$, as $n \to \infty$. Since (passing to a subsequence if necessary) $u_n := Sz_n \rightharpoonup S\bar{z} =: \bar{u}$ in U_D as $n \to \infty$, and $z_n \rightharpoonup \bar{z}$ in Z_D as $n \to \infty$, we can immediately take the limit in (4.4) as $n \to \infty$, and conclude that $(\bar{u}, \bar{z}) \in U_D \times Z_{ad,D}$ fulfills the state equation according to Definition 3.7.

Next, that $\bar{z}$ is the minimizer of (4.3) follows from the fact that $\mathcal{J}$ is weakly lower semicontinuous. Indeed, $\mathcal{J}$ is the sum of two weakly lower semicontinuous functions (recall that the norm is continuous and convex therefore weakly lower semicontinuous).

Finally, the uniqueness of $\bar{z}$ follows from the stated assumptions on J and ξ which leads to the strict convexity of the functional $\mathcal{J}$. $\square$

In order to derive the first order necessary optimality conditions, we need an expression of the adjoint operator S^*. We discuss this next. We notice that for every measurable set $E \subset \mathbb{R}^N$, we have that $L^2((0,T); L^2(E)) = L^2((0,T) \times E)$ with equivalent norms.

Lemma 4.1. The adjoint operator $S^* : U_D \to Z_D$ for the state equation (4.1b) is given by

$$S^* w = -\mathcal{N}_s p \in Z_D,$$

where $w \in U_D$ and $p \in L^2((0,T); \widetilde{W}_0^{s,2}(\Omega)) \cap H^1_{0,T}((0,T); \widetilde{W}^{-s,2}(\Omega))$ is the weak solution to the following adjoint problem:

$$\begin{cases} -\partial_t p + (-\Delta)^s p &= w \quad \text{in } Q, \\ p &= 0 \quad \text{in } \Sigma, \\ p(T, \cdot) &= 0 \quad \text{in } \Omega. \end{cases} \tag{4.5}$$

Proof. First of all, since S is linear and bounded, it follows that S^* is well-defined. Now for every $w \in U_D$ and $z \in Z_D$, we have that

$$(w, Sz)_{L^2((0,T);L^2(\Omega))} = (S^* w, z)_{L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega))}.$$

We notice that using semigroup theory (see e.g. Remark 3.2 and [27]) we have that $p \in L^2((0,T); V) \cap H^1_{0,T}((0,T); L^2(\Omega))$. Thus, $\partial_t p$, $(-\Delta)^s p \in L^2(Q)$. Next, testing the equation (4.5) with Sz which solves the state equation in the very-weak sense (cf. Definition 3.29) we get that

$$\begin{aligned} (w, Sz)_{L^2((0,T);L^2(\Omega))} &= (-\partial_t p + (-\Delta)^s p, Sz)_{L^2((0,T);L^2(\Omega))} \\ &= -(z, \mathcal{N}_s p)_{L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega))} = (z, S^* w)_{L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega))}. \end{aligned}$$

□

For the remainder of this section, we will assume that $\xi > 0$.

Theorem 4.2. Let $\mathcal{Z} \subset Z_D$ be open such that $Z_{ad,D} \subset \mathcal{Z}$ and let the assumptions of

Theorem 4.1 hold. Moreover, let $u \mapsto J(u) : U_D \to \mathbb{R}$ be continuously Fréchet differentiable with $J'(u) \in U_D$. If $\bar{z}$ is a minimizer of (4.3) over $Z_{ad,D}$, then the first order necessary optimality conditions are given by

$$(-\mathcal{N}_s \bar{p} + \xi \bar{z}, z - \bar{z})_{L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega))} \geq 0, \quad \forall z \in Z_{ad,D}, \tag{4.6}$$

where $\bar{p} \in L^2((0,T); \widetilde{W}_0^{s,2}(\Omega)) \cap H^1_{0,T}((0,T); \widetilde{W}^{-s,2}(\Omega))$ solves the adjoint equation

$$\begin{cases} -\partial_t \bar{p} + (-\Delta)^s \bar{p} = J'(\bar{u}) & \text{in } Q, \\ \bar{p} = 0 & \text{in } \Sigma, \\ \bar{p}(T, \cdot) = 0 & \text{in } \Omega, \end{cases} \tag{4.7}$$

with $\bar{u} := S\bar{z}$. Finally, (4.6) is equivalent to

$$\bar{z} = \mathcal{P}_{Z_{ad},D} \left(\xi^{-1} \mathcal{N}_s \bar{p} \right), \tag{4.8}$$

where $\mathcal{P}_{Z_{ad},D}$ is the projection onto the set $Z_{ad,D}$. Moreover, if J is convex, then (4.6) is a sufficient condition.

Proof. The statements are a direct consequence of the differentiability properties of J and the chain rule, combined with Lemma 4.1. Notice that, we have introduced the open set $\mathcal{Z}$ to properly define the derivative of $\mathcal{J}$. Let $h \in \mathcal{Z}$ be given. Then the directional derivative of $\mathcal{J}$ is given by

$$\begin{aligned} \mathcal{J}'(\bar{z})h &= (J'(S\bar{z}), Sh)_{L^2((0,T);L^2(\Omega))} + \xi(\bar{z}, h)_{L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega))} \\ &= (S^* J'(S\bar{z}) + \xi \bar{z}, h)_{L^2((0,T);L^2(\Omega))}, \end{aligned} \tag{4.9}$$

where we have used that $J'(S\bar{z}) \in \mathcal{L}(L^2((0,T); L^2(\Omega)), \mathbb{R}) = L^2((0,T); L^2(\Omega))$. Next from

Lemma 4.1, we have that

$$S^* J'(S\bar{z}) = -\mathcal{N}_s \bar{p},$$

where $\bar{p}$ solves (4.7). Recall that $\bar{p} \in L^2((0,T); \widetilde{W}_0^{s,2}(\Omega)) \cap H^1_{0,T}((0,T); \widetilde{W}^{-s,2}(\Omega))$ solving (4.7) also has the following regularity: $\partial_t \bar{p} \in L^2((0,T); L^2(\Omega))$ and this implies that $(-\Delta)^s \bar{p} \in L^2((0,T); L^2(\Omega))$. This implies that $\mathcal{N}_s \bar{p} \in L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega))$. Substituting this expression of $S^* J'(S\bar{z})$ in (4.9), we obtain that

$$\mathcal{J}'(\bar{z})h = (-\mathcal{N}_s \bar{p} + \xi \bar{z}, h)_{L^2((0,T);L^2(\Omega))}.$$

The remainder of the steps to obtain (4.6) are standard, see for instance [8, 88].

Finally, (4.8) follows by using [21, Theorem 3.3.5]. $\square$

4.1.2 Fractional Parabolic Robin Optimal Control Problem

Next, we shall focus on the Robin optimal control problem (4.2). Recall that

$$Z_R := L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega, \mu)), \quad U_R := L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa})^\star).$$

Recall also that $d\mu = \kappa dx$ with $\kappa \in L^1(\mathbb{R}^N \setminus \Omega) \cap L^\infty(\mathbb{R}^N \setminus \Omega)$ and is non-negative. Due to Theorem 3.9, the following control-to-state (solution) map

$$S : Z_R \to U_R, \quad z \mapsto Sz =: u,$$

is well-defined. In addition, S is linear and continuous. Owing to the continuous embedding $U_R \hookrightarrow L^2((0,T); L^2(\Omega))$, we can instead define

$$S : Z_R \to L^2((0,T); L^2(\Omega)).$$

Letting

$$\mathcal{J}(z) : Z_R \to \mathbb{R},\ z \mapsto \mathcal{J}(z) := \Big(J(Sz) + \frac{\xi}{2}\|z\|_{Z_R}^2\Big),$$

then the reduced Robin exterior parabolic optimal control problem is given by

$$\min_{z \in Z_{ad,R}} \mathcal{J}(z). \tag{4.10}$$

Throughout the following chapter, for $u, v \in W^{s,2}_{\Omega,\kappa}$ we shall denote

$$\mathcal{E}(u,v) := \frac{C_{N,s}}{2}\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(u(x)-u(y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy + \int_{\mathbb{R}^N\setminus\Omega} \kappa uv\, dx.$$

The following well-posedness result holds.

Theorem 4.3. Let $Z_{ad,R}$ be a convex and closed subset of Z_R and let either $\xi > 0$ with $J \geq 0$ or $Z_{ad,R} \subset Z_R$ bounded. If $J : L^2((0,T);L^2(\Omega)) \to \mathbb{R}$ is weakly lower-semicontinuous, then there exists a solution $\bar{z}$ to (4.10) and equivalently to (4.2). If either J is convex and $\xi > 0$ or J is strictly convex and $\xi \geq 0$, then $\bar{z}$ is unique.

Proof. The proof is similar to the proof of Theorem 4.1. We only discuss the part where $\{z_n\}_{n\in\mathbb{N}}$ is a minimizing sequence such that, passing to a subsequence if necessary, $z_n \rightharpoonup \bar{z}$ in $L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega,\mu))$ as $n \to \infty$. Let (Sz_n, z_n), $n \in \mathbb{N}$, be the solution of (4.2b). We need to show that there is a subsequence which converges to $(S\bar{z},\bar{z})$ in $L^2((0,T);W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,0}((0,T);(W^{s,2}_{\Omega,\kappa})^\star)$ as $n \to \infty$ and $(S\bar{z},\bar{z})$ solves (4.2b) in the weak sense (cf. Definition 3.9). Since $u_n := Sz_n \in L^2((0,T);W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,0}((0,T);(W^{s,2}_{\Omega,\kappa})^\star)$ solves (4.2b), we have that the identity

$$\langle \partial_t u_n(t,\cdot), v\rangle + \mathcal{E}(u_n(t,\cdot),v) = \int_{\mathbb{R}^N\setminus\Omega} z_n(t,x)v(x)\, d\mu, \tag{4.11}$$

holds for every $v \in W^{s,2}_{\Omega,\kappa}$ and a.e. $t \in (0,T)$, where $\mathcal{E}$ is as defined before the beginning of this theorem. We note that the mapping S is bounded due to Theorem 3.9. As a result, passing to a subsequence if necessary, we have that $Sz_n = u_n \rightharpoonup S\bar{z} = \bar{u}$ in $L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa})^\star)$ as $n \to \infty$. Then, taking the limit as $n \to \infty$ in (4.11) we get that

$$\langle \partial_t \bar{u}(t,\cdot), v\rangle + \mathcal{E}(\bar{u}(t,\cdot), v) = \int_{\mathbb{R}^N\setminus\Omega} \bar{z}(t,x)v(x)\, d\mu.$$

That is, $(S\bar{z}, \bar{z})$ solves (4.2b) in the weak sense (cf. Definition 3.9). □

As in the previous section, before we state the first order optimality conditions, we shall derive the expression of the adjoint operator S^*.

Lemma 4.2. The adjoint operator $S^* : L^2((0,T); L^2(\Omega)) \to Z_R$ is given by

$$(S^* w, z)_{Z_R} = \int_\Sigma pz\, d\mu dt \quad \forall z \in Z_R,$$

where $w \in L^2((0,T); L^2(\Omega))$ and $p \in L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,T}((0,T); (W^{s,2}_{\Omega,\kappa})^\star)$ is the weak solution to

$$\begin{cases} -\partial_t p + (-\Delta)^s p = w & \text{in } Q, \\ \mathcal{N}_s p + \kappa p = 0 & \text{in } \Sigma, \\ p(T,\cdot) = 0 & \text{in } \Omega. \end{cases} \tag{4.12}$$

Proof. Let $w \in L^2((0,T); L^2(\Omega))$ and $z \in Z_R$.
Since $Sz \in L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1((0,T); (W^{s,2}_{\Omega,\kappa})^\star) \subset L^2((0,T); L^2(\Omega))$, we can write

$$(w, Sz)_{L^2((0,T);L^2(\Omega))} = (S^* w, z)_{Z_R}.$$

Furthermore, testing (4.12) with $Sz = u$ we obtain that

$$\begin{aligned}(w, Sz)_{L^2((0,T);L^2(\Omega))} &= (-\partial_t p + (-\Delta)^s p, Sz)_{L^2((0,T);L^2(\Omega))} \\ &= \int_\Sigma zp \, d\mu dt = (S^* w, z)_{Z_R},\end{aligned}$$

where we have used the integration-by-parts in both space and time and the fact that $Sz = u$ solves the state equation according to Definition 3.9. $\square$

We conclude this section with the following first order optimality conditions result whose proof is similar to the Dirichlet case and is omitted for brevity. We shall assume that $\xi > 0$.

Theorem 4.4. Let $\mathcal{Z} \subset Z_R$ be open such that $Z_{ad,R} \subset \mathcal{Z}$ and let the assumptions of Theorem 4.3 hold. Let $u \mapsto J(u) : L^2((0,T); L^2(\Omega)) \to \mathbb{R}$ be continuously Fréchet differentiable with $J'(u) \in L^2((0,T); L^2(\Omega))$. If $\bar{z}$ is a minimizer of (4.10), then the first order necessary optimality conditions are given by

$$\int_\Sigma (\bar{p} + \xi \bar{z})(z - \bar{z}) \, d\mu dt \geq 0, \quad z \in Z_{ad,R} \tag{4.13}$$

where $\bar{p} \in L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,T}((0,T); (W^{s,2}_{\Omega,\kappa})^\star)$ solves the following adjoint equation:

$$\begin{cases} -\partial_t \bar{p} + (-\Delta)^s \bar{p} = J'(\bar{u}) & \text{in } Q, \\ \mathcal{N}_s \bar{p} + \kappa \bar{p} = 0 & \text{in } \Sigma, \\ \bar{p}(T, \cdot) = 0 & \text{in } \Omega. \end{cases} \tag{4.14}$$

Moreover, (4.13) is equivalent to

$$\bar{z} = \mathcal{P}_{Z_{ad,R}}(-\xi^{-1} \bar{p})$$

where $\mathcal{P}_{Z_{ad,R}}$ is the projection onto the set $Z_{ad,R}$. If J is convex, then (4.13) is also sufficient.

4.2 Approximation of the Dirichlet Exterior Value and Optimal Control Problems

Recall that the Dirichlet control problem requires approximations of the nonlocal normal derivative of the test functions (cf. (3.29)) and the solutions of the adjoint system (cf. (4.6)). The Nonlocal normal derivative $\mathcal{N}_s$ is a delicate object to handle both at the continuous level and at the discrete level. Indeed, the best known regularity result for $\mathcal{N}_s u$ is as given in Lemma 2.1. Moreover, a numerical approximation of this object is a daunting task. In order to circumvent the approximations of $\mathcal{N}_s u$ both in (3.29) and (4.6), in this section we propose to approximate the parabolic Dirichlet problem by the following parabolic Robin problem.

Let $n \in \mathbb{N}$. In this section we are interested in solutions u_n to the following parabolic Robin problem:

$$\begin{cases} \partial_t u_n + (-\Delta)^s u_n = 0 & \text{in } Q, \\ \mathcal{N}_s u_n + n\kappa u_n = n\kappa z & \text{in } \Sigma, \\ u_n(0,\cdot) = 0 & \text{in } \Omega, \end{cases} \tag{4.15}$$

that belong to the space $L^2((0,T); W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))^\star)$. Notice that $W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)$ is endowed with the norm

$$\|u\|_{W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N\setminus\Omega)} := \left(\|u\|^2_{W^{s,2}_{\Omega,\kappa}} + \|u\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} \right)^{\frac{1}{2}}, \quad u \in W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega). \tag{4.16}$$

Moreover, in our application we shall take κ such that its support supp$[\kappa]$ has a positive Lebesgue measure. Thus, we make the following assumption.

Assumption 1. We assume that $\kappa \in L^1(\mathbb{R}^N\setminus\Omega) \cap L^\infty(\mathbb{R}^N\setminus\Omega)$ and $\kappa > 0$ almost everywhere in $K := \text{supp}[\kappa] \subset \mathbb{R}^N \setminus \Omega$, and the Lebesgue measure $|K| > 0$.

It follows from Assumption 1 that $\displaystyle\int_{\mathbb{R}^N\setminus\Omega} \kappa \, dx > 0$.

We recall that a solution to (4.15) belongs to $L^2((0,T); W^{s,2}_{\Omega,\kappa}) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa})^\star)$ (this follows from Proposition 3.3). In order to show that this solution also belongs to $L^2((0,T); W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))^\star)$, we recall a result from [6, Lemma 6.2].

Lemma 4.3 (**[6, Lemma 6.2]**)**.** Assume that Assumption 1 holds. Then

$$\|u\|_W := \left(\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{|u(x)-u(y)|^2}{|x-y|^{N+2s}} \, dxdy + \int_{\mathbb{R}^N\setminus\Omega} |u|^2 \, dx \right)^{\frac{1}{2}} \tag{4.17}$$

defines an equivalent norm on $W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)$.

We are now ready to state the main result of this section whose proof is motivated by the elliptic case studied by the authors in [6].

Theorem 4.5 (**Approximation of weak solutions to the Dirichlet problem**)**.** Let Assumption 1 hold. Then the following assertions hold.

1. Let $z \in H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$ and

   $u_n \in L^2((0,T); W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))^\star)$ be the weak solution of (4.15). Let $u \in \mathbb{U}$ be the weak solution to the state equation (4.1b). Then there is a constant $C > 0$ (independent of n) such that

   $$\|u - u_n\|_{L^2((0,T);L^2(\mathbb{R}^N))} \le \frac{C}{n} \|u\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}. \tag{4.18}$$

   In particular, u_n converges strongly to u in $L^2((0,T); L^2(\Omega))$ as $n \to \infty$.

2. Let $z \in L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega))$ and

   $u_n \in L^2((0,T); W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))^\star)$ be the weak solution of (4.15). Then there is a subsequence that we still denote by $\{u_n\}_{n\in\mathbb{N}}$ and a $\tilde{u} \in L^2((0,T); L^2(\mathbb{R}^N))$ such that $u_n \rightharpoonup \tilde{u}$ in $L^2((0,T); L^2(\mathbb{R}^N))$ as $n \to \infty$, and $\tilde{u}$ satisfies

$$\int_Q \tilde{u}\Big(-\partial_t v + (-\Delta)^s v\Big)\, dxdt = -\int_\Sigma \tilde{u}\mathcal{N}_s v\, dxdt, \tag{4.19}$$

   for all $v \in L^2((0,T); V) \cap H^1_{0,T}((0,T); L^2(\Omega))$.

Proof. (a) We begin by discussing the well-posedness of (4.15). We first notice that under our assumption, we have that $W^{s,2}(\mathbb{R}^N \setminus \Omega) \hookrightarrow L^2(\mathbb{R}^N \setminus \Omega) \hookrightarrow L^2(\mathbb{R}^N \setminus \Omega, \mu)$. Now a weak solution $u_n \in L^2((0,T); W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))^\star)$ to (4.15) fulfills the identity

$$\begin{aligned}\langle \partial_t u_n(t,\cdot), v\rangle + \frac{C_{N,s}}{2}\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(u_n(t,x)-u_n(t,y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy\\ + n\int_{\mathbb{R}^N\setminus\Omega} u_n(t,x)v(x)\, d\mu = n\int_{\mathbb{R}^N\setminus\Omega} z(t,x)v(x)\, d\mu,\end{aligned} \tag{4.20}$$

for every $v \in W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)$ and almost every $t \in (0,T)$. For every $n \in \mathbb{N}$, the existence of a unique solution u_n to (4.15) follows by using the arguments of Proposition 3.3.

Next, we prove the estimate (4.18). For $v, w \in W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)$, we shall let

$$\mathcal{E}_n(v,w) := \frac{C_{N,s}}{2}\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(v(x)-v(y))(w(x)-w(y))}{|x-y|^{N+2s}}\, dxdy + n\int_{\mathbb{R}^N\setminus\Omega} vw\, d\mu. \tag{4.21}$$

It is not difficult to see (cf. [6, Eq. (6.17)]) that there is a constant $C > 0$ (independent of

n) such that

$$\frac{C_{N,s}}{2}\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2}\frac{|u_n(t,x)-u_n(t,y)|^2}{|x-y|^{N+2s}}\,dxdy+n\int_{\mathbb{R}^N\setminus\Omega}|u_n(t,x)|^2\,dx$$
$$\leq C\mathcal{E}_n(u_n(t,\cdot),u_n(t,\cdot)). \quad (4.22)$$

Next, let $u\in\mathbb{U}$ be the weak solution to the Dirichlet problem (4.1b) according to Definition 3.6 and let $v\in W^{s,2}_{\Omega,\kappa}\cap L^2(\mathbb{R}^N\setminus\Omega)$. Using the integration by parts formula (2.12) we get that

$$\begin{aligned}
&\langle\partial_t(u-u_n)(t,\cdot),v\rangle+\mathcal{E}_n((u-u_n)(t,\cdot),v)\\
&=\int_\Omega\Big(\partial_t(u-u_n)(t,x)+(-\Delta)^s(u-u_n)(t,x)\Big)v\,dx+\int_{\mathbb{R}^N\setminus\Omega}\mathcal{N}_s(u-u_n)(t,x)v(x)\,dx\\
&\quad+n\int_{\mathbb{R}^N\setminus\Omega}(u-u_n)(t,x)v(x)\,d\mu\\
&=\int_\Omega\left(\partial_t(u-u_n)(t,x)+(-\Delta)^s(u-u_n)(t,x)\right)v(x)\,dx+\int_{\mathbb{R}^N\setminus\Omega}v(x)\mathcal{N}_su(t,x)\,dx\\
&\quad-\int_{\mathbb{R}^N\setminus\Omega}\left(\mathcal{N}_su_n(t,x)+n\kappa(x)(u_n-z)\right)(t,x)v(x)\,dx\\
&=\int_{\mathbb{R}^N\setminus\Omega}v(x)\mathcal{N}_su(t,x)\,dx, \qquad (4.23)
\end{aligned}$$

where we have used that

$$\partial_t(u-u_n)+(-\Delta)^s(u-u_n)=0\ \text{ in } Q\ \text{ and }\ \mathcal{N}_su_n+n\kappa(u_n-z)=0\ \text{ in }\ \Sigma,$$

which follows from the fact that u is a solution to the Dirichlet problem (4.1b) and u_n a solution of (4.15). Letting $v:=(u-u_n)(t,\cdot)$ in (4.23) and using (4.22), we can conclude that there is a constant $C > 0$ (independent of n) such that

$$
\begin{aligned}
&C\langle \partial_t(u-u_n)(t,\cdot),(u-u_n)(t,\cdot)\rangle + n\|(u-u_n)(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} \\
&\quad \le C\Big(\langle \partial_t(u-u_n)(t,\cdot),(u-u_n)(t,\cdot)\rangle + \mathcal{E}_n((u-u_n)(t,\cdot),(u-u_n)(t,\cdot))\Big) \\
&\quad = C\int_{\mathbb{R}^N\setminus\Omega}(u-u_n)(t,x)\mathcal{N}_s u(t,x)\,dx \\
&\quad \le C\|(u-u_n)(t,\cdot)\|_{L^2(\mathbb{R}^N\setminus\Omega)}\|\mathcal{N}_s u(t,\cdot)\|_{L^2(\mathbb{R}^N\setminus\Omega)} \\
&\quad \le C\|(u-u_n)(t,\cdot)\|_{L^2(\mathbb{R}^N\setminus\Omega)}\|u(t,\cdot)\|_{W^{s,2}(\mathbb{R}^N)} \\
&\quad \le \frac{n}{2}\|(u-u_n)(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} + \frac{C^2}{2n}\|u(t,\cdot)\|^2_{W^{s,2}(\mathbb{R}^N)}.
\end{aligned}
$$

Hence,

$$C\langle \partial_t(u-u_n)(t,\cdot),(u-u_n)(t,\cdot)\rangle + \frac{n}{2}\|(u-u_n)(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} \le \frac{C}{n}\|u(t,\cdot)\|^2_{W^{s,2}(\mathbb{R}^N)},$$

where we have replaced the constant C^2 by C. Since

$$\langle \partial_t(u-u_n)(t,\cdot),(u-u_n)(t,\cdot)\rangle = \frac{1}{2}\partial_t\|(u-u_n)(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)},$$

it follows that

$$\frac{C}{2}\partial_t\|(u-u_n)(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} + \frac{n}{2}\|(u-u_n)(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} \le \frac{C}{n}\|u(t,\cdot)\|^2_{W^{s,2}(\mathbb{R}^N)}.$$

Thus,

$$\frac{C}{2}\|(u-u_n)(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)}+\frac{n}{2}\int_0^t\|(u-u_n)(\tau,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)}\,d\tau$$
$$\leq\frac{C}{n}\int_0^t\|u(\tau,\cdot)\|^2_{W^{s,2}(\mathbb{R}^N)}\,d\tau$$

which implies that

$$\begin{cases}\|u-u_n\|_{L^\infty((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\leq\frac{C}{\sqrt{n}}\|u\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}\\ \|u-u_n\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\leq\frac{C}{n}\|u\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}.\end{cases}\tag{4.24}$$

In order to obtain (4.18), it remains to estimate $\|u-u_n\|_{L^2((0,T);L^2(\Omega))}$.

We notice that $L^2((0,T);L^2(\Omega))=L^2((0,T)\times\Omega)$ with equivalent norms and

$$\|u-u_n\|_{L^2((0,T);L^2(\Omega))}=\sup_{\eta\in L^2((0,T);L^2(\Omega))}\frac{\left|\int_0^T\int_\Omega(u-u_n)\eta\,dxdt\right|}{\|\eta\|_{L^2((0,T);L^2(\Omega))}}.\tag{4.25}$$

For any $\eta\in L^2((0,T);L^2(\Omega))$, let $w\in L^2((0,T);\widetilde{W}_0^{s,2}(\Omega))\cap H^1_{0,T}((0,T);\widetilde{W}^{-s,2}(\Omega))$ solve the following dual problem:

$$\begin{cases}-\partial_t w+(-\Delta)^s w &=\eta\quad\text{in } Q,\\ w &=0\quad\text{in }\Sigma,\\ w(T,\cdot) &=0\quad\text{in }\Omega.\end{cases}\tag{4.26}$$

It follows from Proposition 3.2 that there is a unique solution w to (4.26) that fulfills

$$\|w\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}\leq C\|\eta\|_{L^2((0,T);L^2(\Omega))}.\tag{4.27}$$

Notice that $w \in L^2((0,T); \widetilde{W}_0^{s,2}(\Omega))$ and using (4.23) we obtain that

$$
\begin{aligned}
&\int_0^T \int_\Omega (u-u_n)(-\partial_t w + (-\Delta)^s w)\, dxdt \\
&= \int_0^T \langle \partial_t(u-u_n), w\rangle \, dt - \int_0^T \int_{\mathbb{R}^N\setminus\Omega} (u-u_n)\mathcal{N}_s w \, dxdt \\
&\quad + \frac{C_{N,s}}{2}\int_0^T \int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{((u-u_n)(t,x)-(u-u_n)(t,y))(w(t,x)-w(t,y))}{|x-y|^{N+2s}}\, dxdydt \\
&= \int_0^T \langle \partial_t(u-u_n), w\rangle \, dt + \int_0^T \mathcal{E}_n(u-u_n, w)\, dt - \int_0^T \int_{\mathbb{R}^N\setminus\Omega} (u-u_n)\mathcal{N}_s w \, dxdt \\
&= \int_0^T \int_{\mathbb{R}^N\setminus\Omega} w\mathcal{N}_s u \, dxdt - \int_0^T \int_{\mathbb{R}^N\setminus\Omega} (u-u_n)\mathcal{N}_s w \, dxdt \\
&= -\int_0^T \int_{\mathbb{R}^N\setminus\Omega} (u-u_n)\mathcal{N}_s w \, dxdt.
\end{aligned}
$$

Using the preceding identity, (4.24) and (4.27), we obtain that

$$
\begin{aligned}
\left|\int_0^T \int_\Omega (u-u_n)(-\partial_t w + (-\Delta)^s w)\ \ dxdt\right| = &\left|\int_0^T \int_{\mathbb{R}^N\setminus\Omega} (u-u_n)\mathcal{N}_s w \, dxdt\right| \\
\le & \|u-u_n\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\|\mathcal{N}_s w\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))} \\
\le & \frac{C}{n}\|u\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}\|w\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))} \\
\le & \frac{C}{n}\|u\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}\|\eta\|_{L^2((0,T);L^2(\Omega))}.
\end{aligned} \tag{4.28}
$$

Using (4.25) and (4.28) we get that

$$
\|u-u_n\|_{L^2((0,T);L^2(\Omega))} \le \frac{C}{n}\|u\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}. \tag{4.29}
$$

Now the estimate (4.18) follows from (4.24) and (4.29). The proof of Part (a) is complete.

(b) Let $z \in L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))$. Using our assumption, we immediately notice that we have the continuous embedding $L^2(\mathbb{R}^N\setminus\Omega) \hookrightarrow L^2(\mathbb{R}^N\setminus\Omega,\mu)$. In addition, $\{u_n\}_{n\in\mathbb{N}}$ satisfies (4.20). Then proceeding similarly as in (4.22) we can deduce that

$$
\begin{aligned}
C\langle\partial_t u_n(t,\cdot),u_n(t,\cdot)\rangle + n\|u_n(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)} &\le C\Big(\langle\partial_t u_n(t,\cdot),u_n(t,\cdot)\rangle + \mathcal{E}_n(u_n,u_n)\Big)\\
&\le nC\|\kappa\|_{L^\infty(\mathbb{R}^N\setminus\Omega)}\|z(t,\cdot)\|_{L^2(\mathbb{R}^N\setminus\Omega)}\|u_n(t,\cdot)\|_{L^2(\mathbb{R}^N\setminus\Omega)},
\end{aligned}
$$

for almost every $t\in(0,T)$. Since $\langle\partial_t u_n(t,\cdot),u_n(t,\cdot)\rangle = \frac{1}{2}\partial_t\|u_n(t,\cdot)\|^2_{L^2(\mathbb{R}^N\setminus\Omega)}$, we have that

$$
\|u_n\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))} \le C\|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}. \tag{4.30}
$$

In order to show that $\|u_n\|_{L^2((0,T);L^2(\Omega))}$ is uniformly bounded, we can proceed as in (4.29), i.e., by using a duality argument. Let $\eta \in L^2((0,T);L^2(\Omega))$ and $w\in\mathbb{U}_0$ be the weak solution of (4.26). Then using (4.20) and taking $w\in L^2((0,T);V)\cap H^1_{0,T}((0,T);L^2(\Omega))$, we get that

$$
\begin{aligned}
\int_Q u_n\eta\,dxdt &= \int_Q u_n(-\partial_t w + (-\Delta)^s w)\,dxdt\\
&= \int_0^T\langle\partial_t u_n,w\rangle\,dt - \int_0^T\int_{\mathbb{R}^N\setminus\Omega} u_n\mathcal{N}_s w\,dxdt\\
&\quad + \frac{C_{N,s}}{2}\int_0^T\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2}\frac{(u_n(t,x)-u_n(t,y))(w(t,x)-w(t,y))}{|x-y|^{N+2s}}\,dxdydt\\
&= -\int_0^T\int_{\mathbb{R}^N\setminus\Omega} u_n\mathcal{N}_s w\,dxdt.
\end{aligned}
$$

Using the above identity, (4.30) and (4.27) we obtain that

$$\begin{aligned}\left|\int_0^T\int_\Omega u_n\eta\, dxdt\right| =&\left|\int_0^T\int_{\mathbb{R}^N\setminus\Omega} u_n\mathcal{N}_s w\, dxdt\right|\\ \le&\|u_n\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\|\mathcal{N}_s w\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\\ \le&C\|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\|w\|_{L^2((0,T);W^{s,2}(\mathbb{R}^N))}\\ \le&C\|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}\|\eta\|_{L^2((0,T);L^2(\Omega))}.\end{aligned} \tag{4.31}$$

Thus,

$$\|u_n\|_{L^2((0,T);L^2(\Omega))} \le C\|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}. \tag{4.32}$$

Combining (4.30)-(4.32) we get that

$$\|u_n\|_{L^2((0,T);L^2(\mathbb{R}^N))} \le C\|z\|_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}. \tag{4.33}$$

Therefore, the sequence $\{u_n\}_{n\in\mathbb{N}}$ is bounded in $L^2((0,T);L^2(\mathbb{R}^N)) = L^2((0,T)\times\mathbb{R}^N)$. Thus, passing to a subsequence if necessary, we have that u_n converges weakly to some $\tilde{u}$ in $L^2((0,T);L^2(\mathbb{R}^N))$ as $n\to\infty$.

It remains to show (4.19). Notice that $W_0^{s,2}(\overline{\Omega}) \hookrightarrow W_{\Omega,\kappa}^{s,2}\cap L^2(\mathbb{R}^N\setminus\Omega)$. Thus, by (4.20) we have that

$$\begin{aligned}&\int_0^T \langle \partial_t u_n(t,\cdot), v(t,\cdot)\rangle\, dt\\ &\qquad+\frac{C_{N,s}}{2}\int_0^T\int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2}\frac{(u_n(t,x)-u_n(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\, dxdydt = 0,\end{aligned} \tag{4.34}$$

for every $v\in L^2((0,T);V)\cap H^1_{0,T}((0,T);L^2(\Omega))$. Next, applying the integration by parts formula (2.12) we can deduce that

$$\int_0^T \langle \partial_t u_n(t,\cdot), v(t,\cdot)\rangle \, dt$$

$$+ \frac{C_{N,s}}{2}\int_0^T \int\int_{\mathbb{R}^{2N}\setminus(\mathbb{R}^N\setminus\Omega)^2} \frac{(u_n(t,x)-u_n(t,y))(v(t,x)-v(t,y))}{|x-y|^{N+2s}}\, dxdydt$$

$$= \int_Q u_n(-\partial_t v + (-\Delta)^s v)\, dxdt + \int_\Sigma u_n \mathcal{N}_s v \, dxdt. \tag{4.35}$$

Combining (4.34)-(4.35) we get the identity

$$\int_Q u_n(-\partial_t v + (-\Delta)^s v)\, dxdt + \int_\Sigma u_n \mathcal{N}_s v \, dxdt = 0. \tag{4.36}$$

Taking the limit as $n \to \infty$ in (4.36), we obtain that

$$\int_Q \tilde{u}(-\partial_t v + (-\Delta)^s v)\, dxdt + \int_\Sigma \tilde{u} \mathcal{N}_s v \, dxdt = 0,$$

for every $v \in L^2((0,T);V) \cap H^1_{0,T}((0,T);L^2(\Omega))$. Thus, we have shown (4.19). $\square$

Next, we show the approximation of the parabolic Dirichlet control problem (4.1). Let $Z_R := L^2((0,T);L^2(\mathbb{R}^N \setminus \Omega))$ and consider the following minimization problem:

$$\min_{(u,z)\in(U_R,Z_R)} \Big(J(u) + \frac{\xi}{2}\|z\|^2_{Z_R} \Big), \tag{4.37a}$$

subject to the fractional parabolic Robin exterior value problem: Find $u \in U_R$ solving

$$\begin{cases} \partial_t u + (-\Delta)^s u = 0 & \text{in } Q, \\ \mathcal{N}_s u + n\kappa u = n\kappa z & \text{in } \Sigma, \\ u(0,\cdot) = 0 & \text{in } \Omega, \end{cases} \tag{4.37b}$$

and the control constraints

$$z \in Z_{ad,R}. \tag{4.37c}$$

Theorem 4.6 (**Approximation of the parabolic Dirichlet control problem).** The problem (4.37) admits a minimizer $(z_n, u(z_n)) \in Z_{ad,R} \times L^2((0,T); W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \cap H^1_{0,0}((0,T); (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))^\star)$. If $Z_R = H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$ and $Z_{ad,R} \subset Z_R$ is bounded, then for any sequence $\{n_\ell\}_{\ell=1}^\infty$ with $n_\ell \to \infty$, there exists a subsequence still denoted by $\{n_\ell\}_{\ell=1}^\infty$, such that $z_{n_\ell} \rightharpoonup \tilde{z}$ in $H^1_{0,0}((0,T); W^{s,2}(\mathbb{R}^N \setminus \Omega))$ and $u(z_{n_\ell}) \to u(\tilde{z})$ in $L^2((0,T); L^2(\mathbb{R}^N))$ as $n_\ell \to \infty$, with $(\tilde{z}, u(\tilde{z}))$ solving the parabolic Dirichlet control problem (4.1) with $Z_{ad,D}$ replaced by $Z_{ad,R}$.

Proof. The proof is similar to the elliptic case studied in [6] with the obvious modifications and is omitted for brevity. $\square$

We conclude this section by writing the stationarity system corresponding to (4.37): Find

$$(z, u, p) \in Z_{ad,R} \times \Big(L^2((0,T); W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \cap H^1((0,T); (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))^\star)\Big)^2$$

with $u(0,\cdot) = p(T,\cdot) = 0$ in Ω such that

$$\begin{cases} \langle \partial_t u(t,\cdot), v\rangle + \mathcal{E}_n(u(t,\cdot), v) & = \displaystyle\int_{\mathbb{R}^N \setminus \Omega} n\kappa(x) z(t,x) v(x)\, dx, \text{ a.e. } t \in (0,T), \\ \langle -\partial_t p(t,\cdot), w\rangle + \mathcal{E}_n(w, p(t,\cdot)) & = \displaystyle\int_\Omega J'(u(t,x)) w(x)\, dx, \text{ a.e. } t \in (0,T), \\ \displaystyle\int_\Sigma (n\kappa(x) p(t,x) + \xi z(t,x))(\widetilde{z} - z)(t,x)\, dx & \geq 0, \end{cases} \tag{4.38}$$

for all $(\widetilde{z}, v, w) \in Z_{ad,R} \times (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega)) \times (W^{s,2}_{\Omega,\kappa} \cap L^2(\mathbb{R}^N \setminus \Omega))$. Here $\mathcal{E}_n$ is as in (4.21).

4.3 Numerical Approximations

In this section, we shall introduce the numerical approximation of all the problems we have considered so far. We remark that solving parabolic fractional PDEs is a delicate issue. One has to assemble the integrals with singular kernels and the resulting system matrices are dense. On the top of that, the optimal control problem requires solving the state equation forward in time and adjoint equation backward in time. This can be prohibitively expensive. The purpose of this section is simply to illustrate that the numerical results are in agreement with the theory and to show the benefits of the fractional optimal control problem.

The rest of the section is organized as follows: In subsection 4.3.1 we first focus on the approximations of the Robin problem (4.15). With the help of a numerical example, we illustrate the sharpness of Theorem 4.5. This is followed by a source identification problem in subsection 4.3.2. The numerical example presented in subsection 4.3.2 clearly indicates the strength and flexibility of nonlocal problems over the local ones.

4.3.1 Approximation of Parabolic Dirichlet Problems by Parabolic Robin Problems

We begin by introducing a discrete scheme for the parabolic Robin problem (4.15) and recall that we can approximate the parabolic Dirichlet problem by the parabolic Robin problem. Let $\widetilde{\Omega}$ be an open bounded set that contains Ω, the support of z, and the support of κ. We consider a conforming simplicial triangulation of Ω and $\widetilde{\Omega} \setminus \Omega$ such that the resulting partition remains admissible. Throughout the following, we will assume that the support of z and κ are contained in $\widetilde{\Omega} \setminus \Omega$. Let $\mathbb{V}_h$ (on $\widetilde{\Omega}$) be the finite element space of continuous piecewise linear functions. We use the backward-Euler to carry out the time discretization: Let K denote the number of time intervals, we set the time-step to be $\tau = T/K$. Then for $k = 1, \ldots, K$, the fully discrete approximation of (4.15) with nonzero right-hand side f and initial datum $u^{(0)} = u(0, \cdot)$ is given by: find $u_h^{(k)} \in \mathbb{V}_h$ such that

$$\int_{\Omega} u_h^{(k)} v \, dx + \tau \mathcal{E}_n(u_h^{(k)}, v) = \tau \langle f^{(k)}, v \rangle + \tau \int_{\widetilde{\Omega}\setminus\Omega} n\kappa z^{(k)} v \, dx \\ + \int_{\Omega} u_h^{(k-1)} v \, dx \quad \forall v \in \mathbb{V}_h, \tag{4.39}$$

where $\mathcal{E}_n$ is as in (4.21). The approximation of the double integral over $\mathbb{R}^{2N} \setminus (\mathbb{R}^N \setminus \Omega)^2$ is carried out using the approach of [1]. The remaining integrals are computed using quadrature which is accurate for polynomials of degree less than and equal to 4. All the implementations are carried in Matlab and we use the direct solver to solve the linear systems. We emphasize that all our spatial meshes are generated using Gmsh [67].

We next consider an example of a parabolic Dirichlet problem with nonzero exterior conditions. Let $\Omega = B_0(1/2) \subset \mathbb{R}^2$ and $T = 1$. We aim to find u solving

$$\begin{cases} \partial_t u + (-\Delta)^s u & = u_{\text{exact}} + e^t \quad \text{in } Q, \\ u(t, \cdot) & = u_{\text{exact}}(t, \cdot) \quad \text{in } \Sigma, \\ u(0, \cdot) & = u_{\text{exact}}(0, \cdot) \quad \text{in } \Omega. \end{cases} \tag{4.40}$$

The exact solution for this problem is given by

$$u_{\text{exact}}(t, x) = \frac{2^{-2s} e^t}{\Gamma(1+s)^2} \left(1 - |x|^2\right)_+^s .$$

We set $\widetilde{\Omega} = B_0(1.5)$ and approximate (4.40) by using (4.39). Moreover, we set $\kappa = 1$ on its support. We divide the time interval $(0, 1)$ into 1800 subintervals. For a fixed $s = 0.6$ and spatial Degrees of Freedom (DoFs) $= 6017$, we study the $L^2((0,T); L^2(\Omega))$ error $\|u_{\text{exact}} - u_h\|_{L^2((0,T);L^2(\Omega))}$ with respect to n in Figure 4.1 (left). We obtain a convergence rate of $1/n$, as predicted by Theorem 4.5(a).

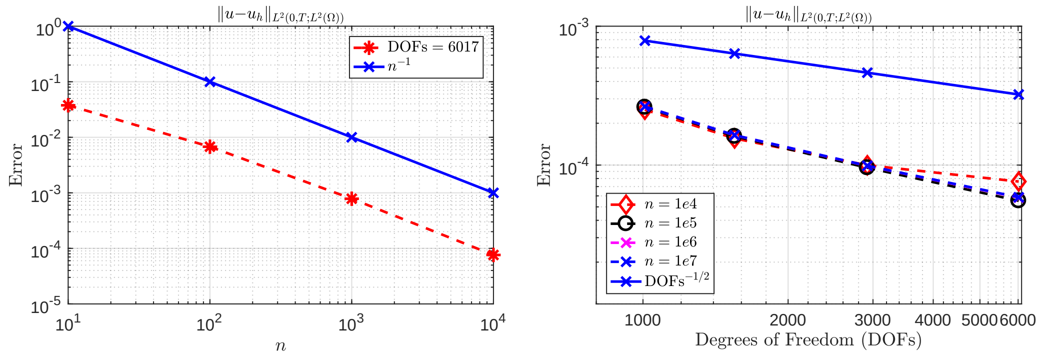

Figure 4.1: Left panel: Fix $s = 0.6$, Degrees of Freedom (DoFs) = 6017. The number of time intervals is 1800. The solid line denotes the reference line and the dotted line is the actual error. We observe that the error $\|u_{\text{exact}} - u_h\|_{L^2((0,T);L^2(\Omega))}$ with respect to n decays at the rate of $1/n$ as predicted by the estimate (4.18) in Theorem 4.5(a). Right panel: Let $s = 0.6$ and number of time intervals = 1800, be fixed. We have shown that the error with respect to spatial DoFs, for $n = 10^4, n = 10^5, n = 10^6$, and $n = 10^7$, behaves as $(\text{DoFs})^{-\frac{1}{2}}$.

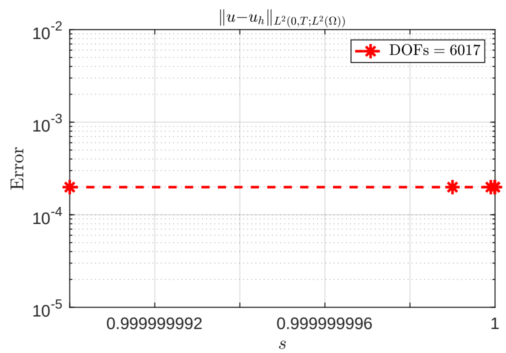

Figure 4.2: Behavior of $\|u_{\text{exact}} - u_h\|_{L^2(0,T;L^2(\Omega))}$ as $s \to 1$. We notice that the error remains stable.

In the right panel, in Figure 4.1, we have shown the error $\|u_{\text{exact}} - u_h\|_{L^2(0,T;L^2(\Omega))}$ for a fixed $s = 0.6$, but $n = 1e4, 1e5, 1e6, 1e7$, as a function of DoFs. We observe that the error remains stable with respect to n as we refine the spatial mesh. Moreover, the observed rate of convergence is $(\text{DoFs})^{-\frac{1}{2}}$.

For the same example, next we study the behavior of $\|u_{\text{exact}} - u_h\|_{L^2((0,T);L^2(\Omega))}$ as $s \to 1$ in Figure 4.2. We observe that the error remains stable. We conclude this section, with

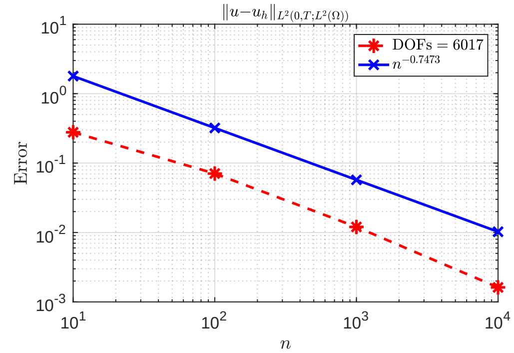

Figure 4.3: We use the non-smooth data given in (4.41) and show the error $\|u_{\text{exact}} - u_h\|_{L^2((0,T);L^2(\Omega))}$ with respect to n. We observe a rate of convergence less than predicted in Theorem 4.5 (a). This seems to indicate that the result of Theorem 4.5 (a) is sharp, since in this example we expect $u \notin L^2((0,T);W^{s,2}(\mathbb{R}^N))$.

another example where f and z are less regular than in the above example. We set

$$f(t,x) := \left(|0.1 - x_2|^{0.01} + |-0.1 - x_2|^{0.01}\right) e^t \quad \text{and} \tag{4.41}$$

$$z(t,x) := \left(|0.6 - x_2|^{0.01} + |-0.6 - x_2|^{0.01}\right) e^t. \tag{4.42}$$

Notice that in this case we do not have access to u_{exact}. Instead, we set u_{exact} to be the solution with $n = 10^8$. The error $\|u_{\text{exact}} - u_h\|_{L^2((0,T);L^2(\Omega))}$ with respect to n is shown in Figure 4.3. The example seems to give a convergence rate of $n^{-0.7473}$ which is lower than the rate we predicted in Theorem 4.5 (a). This appears to indicate that the result of Theorem 4.5 (a) are *sharp*, as in this example we expect $u \notin L^2((0,T);W^{s,2}(\mathbb{R}^N))$.

4.3.2 Parabolic Source/Control Identification Problems

After the validation in the previous example, we are now ready to consider a source/control identification problem where the source/control is located outside the domain Ω. The optimality system is as given in (4.38). The spatial discretization of all the optimization variables (u, z, p) is carried out using continuous piecewise linear finite elements and time

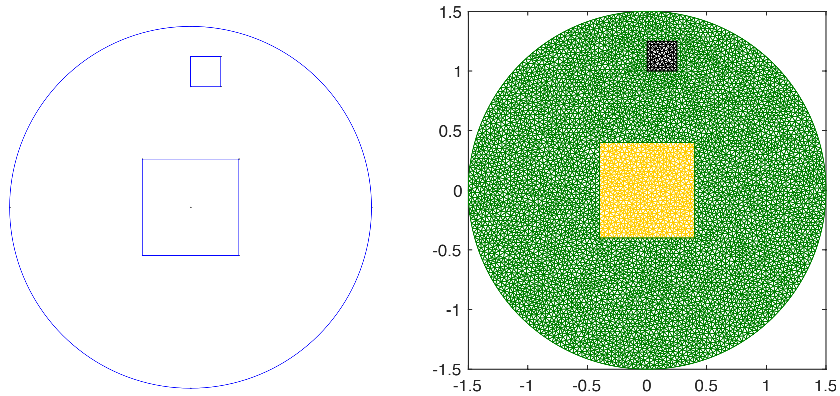

Figure 4.4: **Left panel:** The circle denotes $\widetilde{\Omega}$ and the larger square denotes the domain Ω. Moreover, the outer square inside $\widetilde{\Omega} \setminus \Omega$ is $\widehat{\Omega}$, i.e., the region where the source/control is supported. **Right panel**: A finite element mesh.

discretization using backward-Euler. We set the objective function to be

$$j(u, z) := J(u) + \frac{\xi}{2}\|z\|^2_{L^2((0,T);L^2(\mathbb{R}^N\setminus\Omega))}, \quad \text{with} \quad J(u) := \frac{1}{2}\|u - u_d\|^2_{L^2((0,T);L^2(\Omega))},$$

where $u_d : L^2((0,T); L^2(\Omega)) \to \mathbb{R}$ is the given data (observations). Moreover, we let $Z_{ad,R} := \{z \in L^2((0,T); L^2(\mathbb{R}^N \setminus \Omega)) : \ z \geq 0, \ \text{a.e. in } (0,T) \times \widehat{\Omega}\}$ where $\widehat{\Omega}$ is the support set of the control z that is contained in $\widetilde{\Omega} \setminus \Omega$. We solve the optimization problem using the projected-BFGS method with Armijo line search.

We consider the domain as given in Figure 4.4. The circle denotes $\widetilde{\Omega} = B_0(3/2)$ and the larger square denotes the domain $\Omega = [-0.4, 0.4]^2$. The smaller square, inside $\widetilde{\Omega} \setminus \Omega$, denoted by $\widehat{\Omega}$, is where the source/control is supported. The right panel shows a finite element mesh with DoFs $= 6103$.

We generate the data u_d as follows: for $z = 1$, we solve the state equation (first equation in (4.38)). We then add a normally distributed noise with mean zero and standard deviation 0.005. We call the resulting expression u_d. In addition, we set $\kappa = 1$ on its support and $n = 1e7$.

Next, we identify the source $\bar{z}_h$ by solving the optimality system (4.38) using the aforementioned optimization algorithm. For $\xi = 1e\text{-}8$, our results are shown in Figure 4.5. In the first two rows, we have plotted $\bar{z}_h$ (as a by-product of our optimization algorithm) for $s = 0.1$ at 4 time instances $t = 0.25, 0.3, 0.43, 0.58$. The third row shows $\bar{z}_h$ for $s = 0.8$ at only one of these four time instances since $\bar{z}_h$ is zero at the remaining three time instances. The zero $\bar{z}_h$ for all these time instances can be explained as follows: we know that when s approaches 1, the fractional Laplacian approaches the standard Laplacian $-\Delta$. The latter operator only imposes boundary conditions on $\partial\Omega$, but not exterior conditions as in the case of the fractional Laplacian.

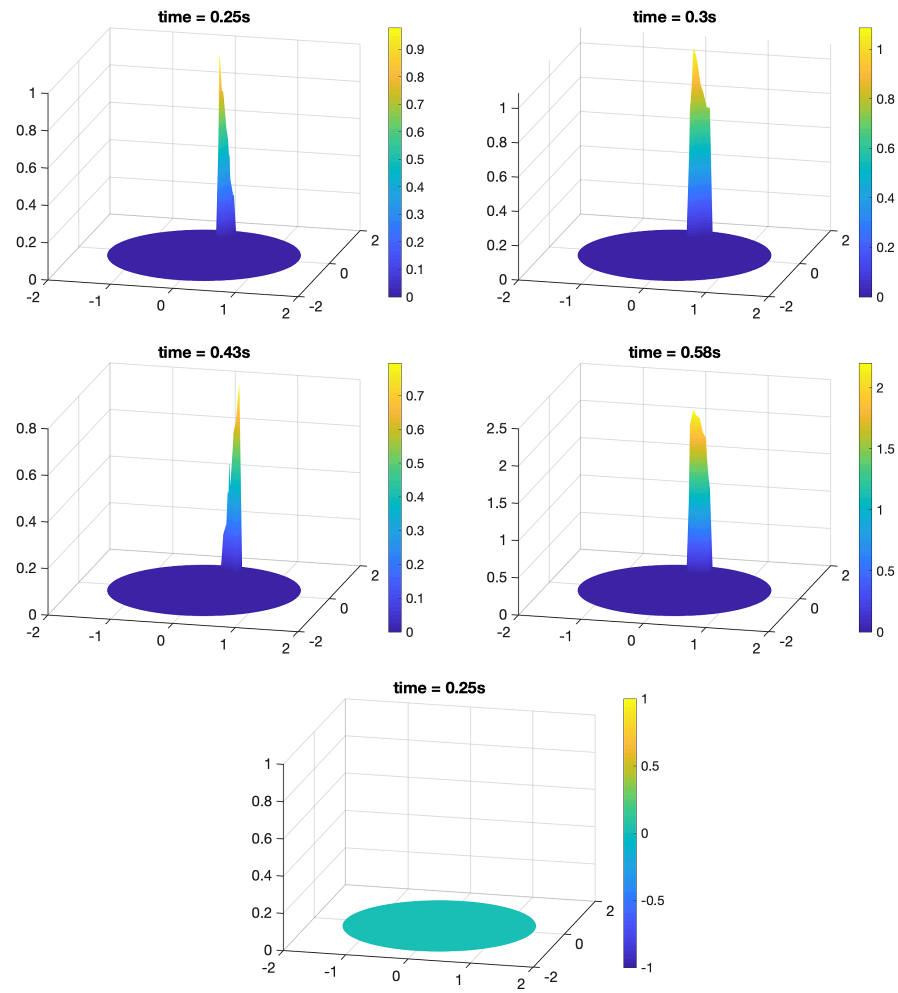

Figure 4.5: The first and second row show the source $\bar{z}_h$ for exponent $s = 0.1$ at 4 different time instances, $t = 0.25, 0.3, 0.43, 0.58$. The last row shows $\bar{z}_h$ for exponent $s = 0.8$ at $t = 0.25$. Notice that $\bar{z}_h \equiv 0$ at $t = 0.25$. For $s = 0.8$, we also obtain that $\bar{z}_h \equiv 0$ at $t = 0.3, 0.43, 0.58$ therefore we have omitted those plots. This comparison between $\bar{z}_h$ for $s = 0.1$ and $s = 0.8$ clearly indicates that we can recover the sources for smaller values of s but when s approaches 1, since the fractional Laplacian approaches the standard Laplacian, we cannot see the external source at all times, i.e., we obtain $\bar{z}_h \equiv 0$. Recall that, the standard Laplacian does not allow imposing exterior conditions.

Chapter 5: Optimal Control of Fractional Elliptic PDEs with State Constraints and Characterization of the Dual of Fractional Order Sobolev Spaces

As mentioned in the introduction, many real life applications, for example in finance, elasticity, etc., require constraints on the solution of the fractional PDEs. Therefore, in this chapter, we introduce and study fractional optimal control problems with both state and control constraints.

We remark that the case $s = 1$ is classical, see for instance, [38, 39, 41, 88, 125] and the references therein. Nevertheless, none of these existing works are directly applicable to the case of fractional state constraints as stated in (5.2b). For example, the characterization of the dual of integer order Sobolev spaces, which is needed to establish the regularity of solutions to the adjoint equation, was not known for fractional order Sobolev spaces. This adjoint regularity is then used to establish a higher regularity result for the optimal control.

This chapter is organized as follows. In Section 5.1, we introduce the problem under consideration and state the main difficulties and novelties. In Section 5.2, we show the well-posedness of the optimal control problem and derive the optimality conditions using some results from Section 3.1.1. In Section 5.3, we derive the characterization of the dual spaces of the fractional order Sobolev spaces. We conclude by giving higher regularity results for the associated adjoint and control variables in Section 5.4. In addition, we also discuss the case where we have weaker than L^p-controls.

5.1 Problem Formulation

Let $\Omega \subset \mathbb{R}^N$ ($N \geq 1$) be a bounded open set with boundary $\partial\Omega$. The main goal of this chapter is to introduce and study an optimal control problem with both control and state constraints:

$$\min_{(u,z)\in(U,Z)} J(u,z) \tag{5.1}$$

subject to the fractional elliptic PDE: find $u \in U$ solving

$$(-\Delta)^s u = z \quad \text{in } \Omega, \qquad u = 0 \quad \text{in } \mathbb{R}^N \setminus \Omega\,, \tag{5.2a}$$

as constraints and additional state constraints

$$u|_\Omega \in \mathcal{K} := \left\{ w \in C_0(\Omega) \ : \ w(x) \le u_b(x), \quad \forall x \in \overline{\Omega} \right\}, \tag{5.2b}$$

where $C_0(\Omega)$ is the space of all continuous functions in $\overline{\Omega}$ that vanish on $\partial\Omega$ and $u_b \in C(\overline{\Omega})$. Moreover, we also consider the control constraints

$$z \in Z_{ad} \subset L^p(\Omega) \tag{5.2c}$$

with Z_{ad} being a non-empty, closed, and convex set. In (5.2c), the real number p satisfies

$$p > \frac{N}{2s} \quad \text{if } \ N > 2s, \quad p > 1 \quad \text{if } \ N = 2s, \quad p = 1 \quad \text{if } \ N < 2s. \tag{5.3}$$

Notice that for $z \in L^p(\Omega)$, with p as in (5.3), we have that u solving (5.2a) belong to $L^\infty(\Omega)$ (see, Section 3.1, also [18]). We refer to Section 5.2 for more details and the precise assumptions on the functional J.

Main Difficulties and Novelties of the work.

1. **Equation with measure valued data.** The adjoint equation (see, Equation (5.8b)) associated with (5.2a) is a fractional PDE with a measure-valued datum. We showed in Theorem 3.3 the well-posedness of such PDEs in $L^{\frac{p}{p-1}}(\Omega)$ where p is as in (5.3). Here, we shall prove in Corollary 5.3 that solutions belong to $\widetilde{W}_0^{t,\frac{p}{p-1}}(\Omega)$ under suitable

assumptions on p and $0 < t < 1$.

2. **Characterization of the dual space $\widetilde{W}^{-s,p'}(\Omega)$.** Let $1 \le p < \infty$, $p' := \frac{p}{p-1}$ and let $\widetilde{W}^{-s,p'}(\Omega)$ denote the dual of $\widetilde{W}_0^{s,p}(\Omega)$ (see, Section 2). In Theorem 5.3, we shall show that if $1 \le p < \infty$ and $f \in \widetilde{W}^{-s,p'}(\Omega)$, then there is pair of functions $(f^0, f^1) \in L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ such that for every $v \in \widetilde{W}_0^{s,p}$, we have: $f(v) = \langle f, v\rangle = \int_\Omega f^0 v\ dx + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} f^1(x,y)\frac{v(x)-v(y)}{|x-y|^{\frac{N}{p}+s}}\ dxdy$. This characterization is one of the main novelties of this work.

3. **Higher regularity of solutions to the Dirichlet problem** (5.2a)**.** Using the above characterization of the dual spaces, we shall show in Corollary 5.2 that if $0 < t < s < 1$ and $2 < \frac{N}{s} < p \le \infty$, or $1 \le p' < 2$ and $\frac{1}{p'} < t < 1$, and $z \in \widetilde{W}^{-t,p}$, then weak solutions of the Dirichlet problem (5.2a) are also continuous up to the boundary of Ω. This is the first time that such a regularity result has been proved (with very weak right-hand side data) for the fractional Laplace operator.

4. **Higher regularity of the optimal control.** Using the above higher regularity of the adjoint variable, we shall establish the $W^{t,\frac{p}{p-1}}(\Omega)$-regularity of the optimal control. This higher regularity of the optimal control is crucial to establish some rates of convergence of the numerical methods.

5.2 The Optimal Control Problem

The purpose of this section is to study the existence of solutions to the optimal control problem (5.2) and establish the first order optimality conditions. Throughout this section, we shall assume that Ω is a bounded Lipschitz domain satisfying the exterior cone condition. Moreover, p is as in (5.3).

We begin by rewriting the optimal control problem (5.2). Let $(-\Delta)^s_D$ be the operator defined in (2.10). Then, the problem (5.2) can be rewritten as follows:

$$\min_{(u,z)\in(U,Z)} J(u,z)$$

subject to the constraints: (5.4)

$$(-\Delta)^s_D u = z \quad \text{in } \Omega, \qquad u|_\Omega \in \mathcal{K} \quad \text{and} \quad z \in Z_{ad}.$$

Next, we introduce the relevant function spaces. We let

$$Z := L^p(\Omega) \ \text{ and } \ U := \Big\{u \in \widetilde{W}_0^{s,2}(\Omega) \cap C_0(\Omega) : ((-\Delta)^s_D)u|_\Omega \in L^p(\Omega)\Big\}.$$

Then, U is a Banach space with the graph norm

$$\|u\|_U := \|u\|_{\widetilde{W}_0^{s,2}(\Omega)} + \|u\|_{C_0(\Omega)} + \|(-\Delta)^s_D u\|_{L^p(\Omega)}.$$

We let $Z_{ad} \subset Z$ a nonempty, closed, and convex set and $\mathcal{K}$ as in (5.2b), i.e.,

$$\mathcal{K} := \left\{w \in C_0(\Omega) \ : \ w(x) \le u_b(x), \quad \forall x \in \overline{\Omega}\right\}. \tag{5.5}$$

Notice that for every $z \in Z$, due to Theorem 3.2, there is a unique $u \in U$ that solves the state equation (5.2a).Thus, the control-to-state (solution) map, $S : Z \to U, \quad z \mapsto Sz =: u$, is well-defined, linear, and continuous. Since $U \hookrightarrow C_0(\Omega)$, we can consider the control-to-state map as $E \circ S : Z \to C_0(\Omega)$.

Next, we define the admissible control set as

$$\widehat{Z}_{ad} := \{z \in Z : \ z \in Z_{ad} \text{ and } (E \circ S)z \in \mathcal{K}\},$$

and as a result, the reduced minimization problem is given by

$$\min_{z \in \widehat{Z}_{ad}} \mathcal{J}(z) := J((E \circ S)z, z). \tag{5.6}$$

Next, we state the well-posedness result for (5.2) and equivalently for (5.6).

Theorem 5.1. Let Z_{ad} be a bounded, closed and convex subset of Z and $\mathcal{K}$ a convex and closed subset of $C_0(\Omega)$ such that $\widehat{Z}_{ad} \neq \emptyset$. If $J : L^2(\Omega) \times L^p(\Omega) \to \mathbb{R}$, with p as in (5.3), is weakly lower-semicontinuous, then there is a solution to (5.6).

Proof. The proof is based on the so-called direct method or the Weierstrass theorem [21, Theorem 3.2.1]. We provide some details for completeness. We can always construct a minimizing sequence $\{z_n\}_{n=1}^{\infty} \subset Z$ such that $\inf_{z \in Z_{ad}} \mathcal{J}(z) = \lim_{n\to\infty} \mathcal{J}(z_n)$. Since Z_{ad} is bounded, it follows that $\{z_n\}_{n=1}^{\infty}$ is a bounded sequence. Since Z is reflexive, we have that there exists a weakly convergent subsequence $\{z_n\}_{n=1}^{\infty}$ (not relabeled) such that $z_n \rightharpoonup \bar{z}$ in Z as $n \to \infty$. Next, since Z_{ad} is closed and convex, thus weakly closed, we have that $\bar{z} \in Z_{ad}$.

Next, we notice that $C_0(\Omega)$ is non-reflexive. However, we have that $u_n = Sz_n \in U \hookrightarrow C_0(\Omega)$ and $S \in \mathcal{L}(Z, C_0(\Omega))$. Thus, there is a subsequence $\{u_n\}$ (not-relabeled) that converges weakly* to $\bar{u}$ in $C_0(\Omega)$ as $n \to \infty$. Since $\mathcal{K}$ is also weakly closed, we have that $\bar{u} \in \mathcal{K}$.

Owing to the uniqueness of the limit and the assumption that $\widehat{Z}_{ad}$ is nonempty, we can deduce that $\bar{z} \in \widehat{Z}_{ad}$. Finally, it remains to show that $\bar{z}$ is a solution to (5.6). This follows from the weak lower-semicontinuity assumption on J. □

Before deriving the first order necessary optimality conditions, we make the following assumption.

Assumption 2 (Compatibility condition between $\mathcal{K}$ and Z_{ad}). There is a pair $(\hat{u}, \hat{z}) \in U \times Z$ that fulfills

$$(-\Delta)^s_D \hat{u} = \hat{z} \quad \text{in } \Omega, \qquad \hat{z} \in Z_{ad}, \qquad \hat{u}(x) < u_b(x) \quad \forall x \in \overline{\Omega}\,. \tag{5.7}$$

Notice that the last condition in Assumption 2, says that the state constraints in $\mathcal{K}$ are satisfied strictly. Assumption 2 is a compatibility condition between Z_{ad} and $\mathcal{K}$. For instance, in the absence of state constraints, it is immediately fulfilled. In addition, if $Z_{ad} = Z$, then again Assumption 2 is satisfied, see [88, pp. 87] for the classical case. But having both control and state constraints require a compatibility condition between $\mathcal{K}$ and Z_{ad} as otherwise the solution set might be empty. We need the state constraints to be strictly satisfied for the existence of Lagrange multipliers, see [125, pp. 340] for the classical case.

Using the definition of U, we have that $(-\Delta)^s_D : U \mapsto Z$ is a bounded operator and from Theorem 3.2, it is also surjective. We have the following first order necessary optimality conditions.

Theorem 5.2. Let $J : L^2(\Omega) \times L^p(\Omega) \to \mathbb{R}$, with p as in (5.3), be continuously Fréchet differentiable and assume that (5.7) holds. Let $(\bar{u}, \bar{z})$ be a solution to the optimization problem (5.2). Then, there exist a Lagrange multiplier $\bar{\mu} \in (C_0(\Omega))^\star$ and an adjoint variable $\bar{\xi} \in L^{p'}(\Omega)$ such that

$$(-\Delta)^s_D \bar{u} = \bar{z} \quad \text{in } \Omega, \tag{5.8a}$$

$$\langle \bar{\xi}, (-\Delta)^s_D v \rangle_{L^{p'}(\Omega), L^p(\Omega)} = (J_u(\bar{u}, \bar{z}), v)_{L^2(\Omega)} + \int_\Omega v \, d\bar{\mu}, \qquad \forall\, v \in U \tag{5.8b}$$

$$\langle \bar{\xi} + J_z(\bar{u}, \bar{z}), z - \bar{z} \rangle_{L^{p'}(\Omega), L^p(\Omega)} \geq 0, \qquad \forall\, z \in Z_{ad} \tag{5.8c}$$

$$\bar{\mu} \geq 0, \quad \bar{u}(x) \leq u_b(x) \text{ in } \Omega, \quad \text{and} \quad \int_\Omega (u_b - \bar{u}) \, d\mu = 0. \tag{5.8d}$$

Proof. We begin by checking the requirements for [88, Lemma 1.14]. We notice that $(-\Delta)^s_D : U \mapsto Z$ is bounded and surjective. Moreover, the condition (5.7) implies that the interior of the set $\mathcal{K}$ is nonempty. It remains to show the existence of a pair $(\hat{u}, \hat{z}) \in U \times Z_{ad}$ such that

$$(-\Delta)^s_D (\hat{u} - \bar{u}) - (\hat{z} - \bar{z}) = 0 \quad \text{in} \quad \Omega. \tag{5.9}$$

Since $(\bar{u}, \bar{z})$ solves the state equation, it follows from (5.9) that

$$(-\Delta)_D^s \hat{u} = \hat{z} \quad \text{in} \quad \Omega. \tag{5.10}$$

Notice that for every $\hat{z} \in Z_{ad}$, there is a unique $\hat{u}$ that solves (5.10) and, in particular, $(\hat{u}, \hat{z})$ works. Thus, the conditions of [88, Lemma 1.14] hold. Then, [88, Theorem 1.56] immediately implies that (5.8a)–(5.8c) hold. Instead of (5.8d), we obtain that

$$\bar{\mu} \in \mathcal{K}^\circ, \quad u(x) \leq u_b(x), \quad x \in \Omega, \quad \text{and} \quad \langle \bar{\mu}, \bar{u} \rangle_{C_0(\Omega)^*, C_0(\Omega)} = 0, \tag{5.11}$$

where $\mathcal{K}^\circ$ denotes the polar cone. Then, the equivalence between (5.11) and (5.8d) follows from a classical result in functional analysis (see, e.g., [88, pp. 88] for more details). $\square$

5.3 Characterization of the Dual of Fractional Order Sobolev Spaces

Given $0 < s < 1$, $1 \leq p < \infty$ and $p' := \frac{p}{p-1}$, the aim of this section is to give a complete characterization of the space $\widetilde{W}^{-s,p'}(\Omega)$. Recall that, $\widetilde{W}^{-s,p'}(\Omega)$ is defined as the dual of the space $\widetilde{W}_0^{s,p}(\Omega)$. Some of the arguments here are motivated by the classical case $s = 1$.

We start by stating this abstract result taken from [2, pp. 194].

Lemma 5.1. If X and W are two Banach spaces, then $X \times W$ endowed with the norm $\|(x, y)\|_{X \times W} := \|x\|_X + \|y\|_W$ is also a Banach space and the dual space $(X \times W)^\star$ is isometrically isomorphic to $X^\star \times W^\star$.

Let $1 \leq p < \infty$ and let $Y := L^p(\Omega) \times L^p(\mathbb{R}^N \times \mathbb{R}^N)$ be endowed with the norm $\|(v_1, v_2)\|_Y := \left(\|v_1\|_{L^p}^p + \|v_2\|_{L^p(\mathbb{R}^N \times \mathbb{R}^N)}^p \right)^{\frac{1}{p}}$. For $v \in \widetilde{W}_0^{s,p}(\Omega)$, we associate the vector

$Pv \in Y$ given by

$$Pv := (v, D_{s,p}v). \tag{5.12}$$

Since $\|Pv\|_Y = \|(v, D_{s,p}v)\|_Y = \|v\|_{\widetilde{W}_0^{s,p}(\Omega)}$, we have that P is an isometry and hence, injective. Therefore, $P : \widetilde{W}_0^{s,p}(\Omega) \mapsto Y$ is an isometric isomorphism of $\widetilde{W}_0^{s,p}(\Omega)$ onto its image $Z \subset Y$. Also, Z is a closed subspace of Y, because $\widetilde{W}_0^{s,p}(\Omega)$ is complete (isometries preserve completion).

Throughout this section without any mention, we shall let $Y := L^p(\Omega) \times L^p(\mathbb{R}^N \times \mathbb{R}^N)$.

Lemma 5.2. Let $1 \leq p < \infty$. Then, for every $f \in Y^\star$, there exists a unique $u = (u_1, u_2) \in L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ such that for every $v = (v_1, v_2) \in Y$, we have

$$f(v) = \int_\Omega u_1 v_1 \, dx + \int_{\mathbb{R}^N \times \mathbb{R}^N} u_2 v_2 \, dx \quad \text{and}$$

$$\|f\|_{Y^*} = \|u\|_{L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)} = \|u_1\|_{L^{p'}(\Omega)} + \|u_2\|_{L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)}.$$

Proof. Let $w \in L^p(\Omega)$. Then, $(w, 0) \in Y$. We define $f_1(w) := f(w, 0)$. Then, $f_1 \in (L^p(\Omega))^\star$. For arbitrary $w_1, w_2, w \in L^p(\Omega)$ and scalars α, β, we have

$$\begin{aligned} f_1(\alpha w_1 + \beta w_2) &= f(\alpha w_1 + \beta w_2, 0) = f(\alpha(w_1, 0) + \beta(w_2, 0)) \\ &= \alpha f((w_1, 0)) + \beta f((w_2, 0)) = \alpha f_1(w_1) + \beta f_1(w_2), \end{aligned}$$

and $|f_1(w)| = |f((w, 0))| \leq \|f\|_{Y^\star} \|(w, 0)\|_Y = \|f\|_{Y^\star} \|w\|_{L^p(\Omega)}$. Thus, $f_1 \in (L^p(\Omega))^\star = L^{p'}(\Omega)$.

Similarly, let $w \in L^p(\mathbb{R}^N \times \mathbb{R}^N)$. Then, $(0, w) \in Y$ and if we define $f_2(w) := f(0, w)$, we have $f_2 \in (L^p(\mathbb{R}^N \times \mathbb{R}^N))^\star = L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$.

Therefore, by the Riesz Representation theorem there exist a unique $u_1 \in L^{p'}(\Omega)$ and a unique $u_2 \in L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ such that $f(v_1, 0) = f_1(v_1) = \langle u_1, v_1 \rangle_{L^{p'}(\Omega), L^p(\Omega)}$ for every $v_1 \in L^p(\Omega)$ and $f(0, v_2) = f_2(v_2) = \langle u_2, v_2 \rangle_{L^{p'}(\mathbb{R}^N \times \mathbb{R}^N), L^p(\mathbb{R}^N \times \mathbb{R}^N)}$ for every $v_2 \in L^p(\mathbb{R}^N \times \mathbb{R}^N)$.

Now let $v := (v_1, v_2) \in Y$. We can write $v = (v_1, v_2) = (v_1, 0) + (0, v_2)$. Hence,

$$f(v) = f(v_1, 0) + f(0, v_2) = f_1(v_1) + f_2(v_2) = \int_\Omega u_1 v_1 \, dx + \int_{\mathbb{R}^N \times \mathbb{R}^N} u_2 v_2 \, dx.$$

Moreover,

$$\begin{aligned} |f(v)| &\le \|u_1\|_{L^{p'}(\Omega)} \|v_1\|_{L^p(\Omega)} + \|u_2\|_{L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)} \|v_2\|_{L^p(\mathbb{R}^N \times \mathbb{R}^N)} \\ &\le \|u\|_{L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)} \|v\|_Y. \end{aligned}$$

Therefore,

$$\|f\|_{Y^\star} \le \|u\|_{L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)}. \tag{5.13}$$

The proof of the first part is complete. It then remains to show that the norms in (5.13) are equal.

Let us first consider the case $1 < p < \infty$. Define

$$v_1(x) := \begin{cases} |u_1(x)|^{p'-2} \overline{u_1(x)}, & \text{if } u_1(x) \neq 0, \\ 0, & \text{if } u_1(x) = 0, \end{cases} \quad \text{and}$$

$$v_2(x, y) := \begin{cases} |u_2(x, y)|^{p'-2} \overline{u_2(x, y)}, & \text{if } u_2(x, y) \neq 0, \\ 0, & \text{if } u_2(x, y) = 0. \end{cases}$$

Then, for $v = (v_1, v_2)$ we have

$$
\begin{aligned}
|f(v)| &= |f(v_1, v_2)| = |f((v_1, 0) + (0, v_2))| = |f_1(v_1) + f_2(v_2)| \\
&= \Big| \langle u_1, v_1 \rangle_{L^{p'}(\Omega), L^p(\Omega)} + \langle u_2, v_2 \rangle_{L^{p'}(\mathbb{R}^N \times \mathbb{R}^N), L^p(\mathbb{R}^N \times \mathbb{R}^N)} \Big| \\
&= \|u_1\|^{p'}_{L^{p'}(\Omega)} + \|u_2\|^{p'}_{L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)} = \|u\|^{p'}_{L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)} \\
&= |\langle u, v \rangle_{Y^\star, Y}| = \|v\|_Y \|u\|_{Y^\star} = \|v\|_Y \|u\|_{L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)},
\end{aligned}
$$

where we have used the equality in Hölder's inequality, the equality holds because $|v_i|^p = |u_i|^{p'}$. Moreover, we have used the fact that $Y^\star \cong L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ due to Lemma 5.1.

Let us consider the case $p = 1$. Then, $Y = L^1(\Omega) \times L^1(\mathbb{R}^N \times \mathbb{R}^N)$ and we can set (due to Lemma 5.1) $Y^\star = L^\infty(\Omega) \times L^\infty(\mathbb{R}^N \times \mathbb{R}^N)$.

Notice that $\|u\|_{Y^\star} := \max \Big\{ \|u_1\|_{L^\infty(\Omega)}, \|u_2\|_{L^\infty(\mathbb{R}^N \times \mathbb{R}^N)} \Big\}$. To get the desired result, it is sufficient to show that $\|f\|_{Y^\star} \geq \|u\|_{Y^\star}$. Now, for any $\epsilon > 0$ and $k = 1$ there exists a measurable set $A \subset \Omega$ (or $\subset \mathbb{R}^N \times \mathbb{R}^N$ when $k = 2$) with finite, non zero measure such that $|u_k(x)| \geq \|u\|_{Y^\star} - \epsilon$, for almost every $x \in A$.

Next, we define $v_k(x) := \begin{cases} \frac{\overline{u_k(x)}}{|u_k(x)|}, & \text{for } x \in A \text{ and } u_k(x) \neq 0, \\ 0, & \text{elsewhere }. \end{cases}$

Set $v := (v_k, 0)$ if $k = 1$, otherwise set $v := (0, v_k)$. Then,

$$
\begin{aligned}
|f(v)| = \Big| \langle u_k, v_k \rangle_{Y^\star, Y} \Big| = \int_A |u_k(x)| \, dx &\geq \Big(\|u\|_{Y^\star} - \epsilon \Big) \|v\|_Y \\
&= \Big(\|u\|_{L^\infty(\Omega) \times L^\infty \mathbb{R}^N \times \mathbb{R}^N)} - \epsilon \Big) \|v\|_Y.
\end{aligned}
$$

Since ϵ is chosen arbitrarily, we have that the result follows from the definition of the operator norm. $\square$

Theorem 5.3. Let $1 \le p < \infty$ and $f \in \widetilde{W}^{-s,p'}(\Omega)$. Then, there exists $(f^0, f^1) \in L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ such that for all $v \in \widetilde{W}_0^{s,p}(\Omega)$,

$$\langle f, v\rangle_{\widetilde{W}^{-s,p'}(\Omega),\widetilde{W}_0^{s,p}(\Omega)} = \int_\Omega f^0 v\, dx + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} f^1(x,y) D_{s,p}v[x,y]\, dx\, dy, \tag{5.14}$$

$$\|f\|_{\widetilde{W}^{-s,p'}(\Omega)} = \inf\left\{\|(f^0, f^1)\|_{L^{p'}(\Omega)\times L^{p'}(\mathbb{R}^N\times\mathbb{R}^N)}\right\}, \tag{5.15}$$

where the infimum is taken over all $(f^0, f^1) \in L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ for which (5.14) holds. Moreover, if $1 < p < \infty$, then (f^0, f^1) is unique.

Proof. Define the linear functional $\widehat{L} : Z \to \mathbb{R}$, where $Z \subset Y$ is the range of P given in (5.12), by

$$\widehat{L}(Pv) = f(v), \quad v \in \widetilde{W}_0^{s,p}(\Omega)$$

$$\begin{tikzcd} \widetilde{W}_0^{s,p}(\Omega) \arrow[r, "P"] \arrow[dr, "f"'] & Z \subset Y \arrow[d, "\widehat{L}"] \\ & \mathbb{R}. \end{tikzcd}$$

Since P is an isometric isomorphism onto Z, it follows that $\widehat{L} \in Z^\star$ and

$$\|\widehat{L}\|_{Z^\star} = \sup_{\|Pv\|_Y = 1} |\langle \widehat{L}, Pv\rangle_{Y^\star, Y}| = \sup_{\|v\|_{\widetilde{W}_0^{s,p}} = 1} |\langle f, v\rangle_{\widetilde{W}^{-s,p'}(\Omega),\widetilde{W}_0^{s,p}(\Omega)}| = \|f\|_{W^{-s,p'}(\Omega)}.$$

By the Hahn-Banach theorem, there exists an $L \in Y^\star = L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ with $\|L\|_{Y^\star} = \|\widehat{L}\|_{Z^\star}$. Since $L \in Y^\star$, using Lemma 5.2, we have that there exists $(f^0, f^1) \in L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$ such that $L(v) = \displaystyle\int_\Omega f^0 v_1\, dx + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} f^1 v_2\, dxdy$ for every $v = (v_1, v_2) \in Y$. Notice that when $1 < p < \infty$, (f^0, f^1) is unique due to the uniform convexity of the Banach space $L^p(\Omega) \times L^p(\mathbb{R}^N \times \mathbb{R}^N)$.

Thus, for $v \in \widetilde{W}_0^{s,p}(\Omega)$ we have $Pv \in Y$. Using the definition of $\widehat{L}$ we get

$$f(v) = \widehat{L}(Pv) = L(Pv) = L(v, D_{s,p}v) = \int_\Omega f^0 v\, dx + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} f^1 D_{s,p}v\; dxdy,$$

which is (5.14), after noticing that $\langle f, v\rangle_{\widetilde{W}^{-s,p'}(\Omega),\widetilde{W}_0^{s,p}(\Omega)} = f(v)$. Moreover, we have $\|f\|_{\widetilde{W}^{-s,p'}(\Omega)} = \|\widehat{L}\|_{Z^\star} = \|L\|_{Y^\star} = \|(f^0, f^1)\|_{L^{p'}(\Omega)\times L^{p'}(\mathbb{R}^N\times\mathbb{R}^N)}$.

Now, for arbitrary $(g^0, g^1) \in L^{p'}(\Omega) \times L^{p'}(\mathbb{R}^N \times \mathbb{R}^N)$, for which (5.14) holds for all $v \in \widetilde{W}_0^{s,p}(\Omega)$, we can define L_g as $L_g(u) = \langle g^0, u_1\rangle_{L^{p'}(\Omega),L^p(\Omega)} + \langle g^1, u_2\rangle_{L^{p'}(\mathbb{R}^N\times\mathbb{R}^N),L^p(\mathbb{R}^N\times\mathbb{R}^N)}$, $\forall\ u \in Y$. Then, $L_g \in Y^\star$ and $L_g|_Z = \widehat{L}$ (due to (5.14)). As a result, $\|\widehat{L}\|_{Z^\star} \le \|L_g\|_{Y^\star}$. Thus, $\|f\|_{\widetilde{W}^{-s,p'}(\Omega)} \le \|g\|_{L^{p'}(\Omega)\times L^{p'}(\mathbb{R}^N\times\mathbb{R}^N)}$. □

In view of Theorem 2.1, we have the following result.

Corollary 5.1. Let $1 < p < \infty$, $\frac{1}{p} < s < 1$ and $f \in \widetilde{W}^{-s,p'}(\Omega)$. Then, there exists a unique $(f^0, f^1) \in L^{p'}(\Omega) \times L^{p'}(\Omega \times \Omega)$ such that for every $v \in \widetilde{W}_0^{s,p}(\Omega)$,

$$\langle f, v\rangle_{\widetilde{W}^{-s,p'}(\Omega),\widetilde{W}_0^{s,p}(\Omega)} = \int_\Omega f^0 v\, dx + \int_\Omega\int_\Omega f^1(x,y) D_{s,p}v[x,y]\; dx\; dy, \tag{5.16}$$

$$\|f\|_{\widetilde{W}^{-s,p'}(\Omega)} = \inf\left\{\|(f^0, f^1)\|_{L^{p'}(\Omega)\times L^{p'}(\Omega\times\Omega)}\right\}, \tag{5.17}$$

where the infimum is taken over all $(f^0, f^1) \in L^{p'}(\Omega) \times L^{p'}(\Omega \times \Omega)$ for which (5.16) holds.

5.4 Improved Regularity of State and Higher Regularity of Adjoint

In this section, we study the higher regularity properties of solutions to the Dirichlet problem (5.2a), with a right hand side $z \in \widetilde{W}^{-t,p}(\Omega)$, for suitable values of $p \in]1,\infty[$ and $0 < t < 1$.

5.4.1 Regularity of the State

Throughout the remainder of this section, for $u, v \in \widetilde{W}_0^{s,2}(\Omega)$, we shall let

$$\mathcal{E}(u,v) := \frac{C_{N,s}}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{(u(x)-u(y))(v(x)-v(y))}{|x-y|^{N+2s}}\, dxdy. \tag{5.18}$$

We start with the following theorem which can be viewed as the first main result of this section.

Theorem 5.4. Let $f_0 \in L^p(\Omega)$ with $p > \frac{N}{2s}$ and $f_1 \in L^q(\Omega \times \Omega)$ with $q > \frac{N}{s}$ if $N \geq 2s$ and $q \geq 2$ if $N < 2s$. Then, there exists a unique function $u \in \widetilde{W}_0^{s,2}(\Omega)$ satisfying

$$\mathcal{E}(u,v) = \int_\Omega f_0 v\, dx + \int_\Omega \int_\Omega f_1(x,y) D_{s,2}v[x,y]\, dxdy, \tag{5.19}$$

for every $v \in \widetilde{W}_0^{s,2}(\Omega)$. In addition, $u \in L^\infty(\Omega)$ and there is a constant $C > 0$ such that

$$\|u\|_{L^\infty(\Omega)} \leq C \left(\|f_0\|_{L^p(\Omega)} + \|f_1\|_{L^q(\Omega\times\Omega)} \right). \tag{5.20}$$

To prove the theorem, we need the following lemma taken from [97, Lemma B.1].

Lemma 5.3. Let $\Phi = \Phi(t)$ be a nonnegative, non-increasing function on a half line $t \geq k_0 \geq 0$ such that there are positive constants c, α and δ ($\delta > 1$) with

$$\Phi(h) \leq c(h-k)^{-\alpha}\Phi(k)^\delta \text{ for } h > k \geq k_0.$$

Then $\Phi(k_0+d)=0$ with $d^\alpha = c\Phi(k_0)^{\delta-1}2^{\alpha\delta/(\delta-1)}$.

Proof of Theorem 5.4. We prove the result in several steps.

Step 1: Firstly, we show that there is a unique $u \in \widetilde{W}_0^{s,2}(\Omega)$ satisfying (5.19). Indeed, recall that $\widetilde{W}_0^{s,2}(\Omega) \hookrightarrow L^{p'}(\Omega)$, by Remark 2.1. Notice also that if $N \geq 2s$, then $q \geq 2$. Since Ω is bounded, we have that, in all the cases, the continuous embedding $L^q(\Omega\times\Omega) \hookrightarrow L^2(\Omega\times\Omega)$ holds. Hence, using the classical Hölder inequality, we get that there is a constant $C>0$ such that

$$\left|\int_\Omega f_0 v\, dx + \int_\Omega\int_\Omega f_1(x,y)D_{s,2}v[x,y]\, dxdy\right| \leq C\left(\|f_0\|_{L^p(\Omega)} + \|f_1\|_{L^2(\Omega\times\Omega)}\right)\|v\|_{\widetilde{W}_0^{s,2}(\Omega)}.$$

Since the bilinear form $\mathcal{E}$ is continuous and coercive, it follows from the classical Lax-Milgram lemma that there is unique $u \in \widetilde{W}_0^{s,2}(\Omega)$ satisfying (5.19).

Step 2: Notice that, if $N < 2s$, then it follows from the embedding (2.1) that $u \in L^\infty(\Omega)$. We give the proof for the case $N > 2s$. The case $N = 2s$ follows with a simple modification of the case $N > 2s$.

Step 3: Let $u \in \widetilde{W}_0^{s,2}(\Omega)$ be the unique function satisfying (5.19). Let $k \geq 0$ be a real number and set $u_k := (|u|-k)^+\mathrm{sgn}(u)$. By [133, Lemma 2.7], we have that $u_k \in \widetilde{W}_0^{s,2}(\Omega)$ for every $k \geq 0$. Proceeding as in the proof of [12, Theorem 2.9], (see also [18, Proposition 3.10 and Section 3.3]), we get that for every $k \geq 0$,

$$\mathcal{E}(u_k,u_k) \leq \mathcal{E}(u_k,u) = \int_\Omega f_0 u_k\, dx + \int_\Omega\int_\Omega f_1(x,y)D_{s,2}u_k[x,y]\, dxdy. \tag{5.21}$$

Let $A_k := \{x \in \Omega:\ |u(x)| \geq k\}$. Then, it is clear that

$$u_k = \Big[(|u|-k)\mathrm{sign}(u)\Big]\chi_{A_k}. \tag{5.22}$$

Let $p_1 \in [1,\infty]$ be such that $\dfrac{1}{p} + \dfrac{1}{2^\star} + \dfrac{1}{p_1} = 1$, where we recall that $2^\star := \frac{2N}{N-2s}$. Since by assumption $p > \frac{N}{2s} = \frac{2^\star}{2^\star - 2}$, we have that

$$\frac{1}{p_1} = 1 - \frac{1}{2^\star} - \frac{1}{p} = \frac{2^\star}{2^\star} - \frac{1}{2^\star} - \frac{1}{p} > \frac{2^\star}{2^\star} - \frac{1}{2^\star} - \frac{2^\star - 2}{2^\star} = \frac{1}{2^\star} \Longrightarrow p_1 < 2^\star. \tag{5.23}$$

Using (5.22), the continuous embedding $\widetilde{W}_0^{s,2}(\Omega) \hookrightarrow L^{2^\star}(\Omega)$, and the Hölder inequality, we get that there is a constant $C > 0$ such that for every $k \geq 0$,

$$\int_\Omega f_0 u_k \, dx = \int_{A_k} f_0 u_k \, dx \leq \|f_0\|_{L^p(\Omega)} \|u_k\|_{\widetilde{W}_0^{s,2}(\Omega)} \|\chi_{A_k}\|_{L^{p_1}(\Omega)}. \tag{5.24}$$

Let $\delta_1 := \frac{2^\star}{p_1}$. Then $\delta_1 > 1$ by (5.23), but this not needed here. We have that for every $k \geq 0$,

$$\|\chi_{A_k}\|_{L^{p_1}(\Omega)} = |A_k|^{\frac{1}{p_1}} = \left(|A_k|^{\frac{1}{2^\star}}\right)^{\frac{2^\star}{p_1}} = \|\chi_{A_k}\|_{L^{2^\star}(\Omega)}^{\frac{2^\star}{p_1}} = \|\chi_{A_k}\|_{L^{2^\star}(\Omega)}^{\delta_1}. \tag{5.25}$$

Step 4: Next, let $q_1 \in [1, \infty]$ be such that $\dfrac{1}{q} + \dfrac{1}{2} + \dfrac{1}{q_1} = 1$. Since $q > \frac{N}{s} = 2\frac{N}{2s} = 2\frac{2^\star}{2^\star - 2}$, we have

$$\frac{1}{q_1} = 1 - \frac{1}{2} - \frac{1}{q} = 2\frac{2^\star}{2 \cdot 2^\star} - \frac{1}{2} - \frac{1}{q} > \frac{2 \cdot 2^\star}{2 \cdot 2^\star} - \frac{1}{2} - \frac{2^\star - 2}{2 \cdot 2^\star} = \frac{1}{2^\star} \Longrightarrow q_1 < 2^\star. \tag{5.26}$$

Using (5.22), the continuous embedding $\widetilde{W}_0^{s,2}(\Omega) \hookrightarrow L^{2^\star}(\Omega)$, and the Hölder inequality again, we can deduce that there is a constant $C > 0$ such that for every $k \geq 0$,

$$\begin{aligned}
\int_\Omega \int_\Omega f_1(x,y) D_{s,2} u_k[x,y] \, dxdy &= \int_{A_k} \int_{A_k} f_1(x,y) D_{s,2} u_k[x,y] \, dxdy \\
&+ \int_{A_k} \int_{\Omega \setminus A_k} f_1(x,y) D_{s,2} u_k[x,y] \, dxdy + \int_{\Omega \setminus A_k} \int_{A_k} f_1(x,y) D_{s,2} u_k[x,y] \, dxdy \\
&\leq C \|f_1\|_{L^q(\Omega \times \Omega)} \|u_k\|_{\widetilde{W}_0^{s,2}(\Omega)} \|\chi_{A_k}\|_{L^{q_1}(\Omega)}.
\end{aligned} \tag{5.27}$$

Let $\delta_2 := \frac{2^\star}{q_1}$. Then $\delta_2 > 1$ by (5.26), which is also not needed here. As in (5.25), for every $k \geq 0$,

$$\|\chi_{A_k}\|_{L^{q_1}(\Omega)} = \|\chi_{A_k}\|^{\delta_2}_{L^{2^\star}(\Omega)}. \tag{5.28}$$

Step 5: Let $\delta := \min\{\delta_1, \delta_2\} > 1$. It follows from (5.25) that

$\|\chi_{A_k}\|_{L^{p_1}(\Omega)} = \|\chi_{A_k}\|^{\delta_1}_{L^{2^\star}(\Omega)} = \|\chi_{A_k}\|^{\delta}_{L^{2^\star}(\Omega)}\|\chi_{A_k}\|^{\delta_1-\delta}_{L^{2^\star}(\Omega)} \leq \|\chi_\Omega\|^{\delta_1-\delta}_{L^{2^\star}(\Omega)}\|\chi_{A_k}\|^{\delta}_{L^{2^\star}(\Omega)}$ for every $k \geq 0$.

Similarly, it follows from (5.28) that $\|\chi_{A_k}\|_{L^{q_1}(\Omega)} \leq \|\chi_\Omega\|^{\delta_2-\delta}_{L^{2^\star}(\Omega)}\|\chi_{A_k}\|^{\delta}_{L^{2^\star}(\Omega)}$ for every $k \geq 0$.

We have shown that there is a constant $C > 0$ such that for every $k \geq 0$,

$$\max\{\|\chi_{A_k}\|_{L^{p_1}(\Omega)}, \|\chi_{A_k}\|_{L^{q_1}(\Omega)}\} \leq C\|\chi_{A_k}\|^{\delta}_{L^{2^\star}(\Omega)}. \tag{5.29}$$

Using (5.21), (5.24), (5.27), (5.29) and the fact that there is a constant $C > 0$ such that $C\|u_k\|_{\widetilde{W}^{s,2}_0(\Omega)} \leq \mathcal{E}(u_k, u_k)$, we get that there is a constant $C > 0$ such that for every $k \geq 0$,

$$\|u_k\|_{\widetilde{W}^{s,2}_0(\Omega)} \leq C\left(\|f_0\|_{L^p(\Omega)} + \|f_1\|_{L^q(\Omega\times\Omega)}\right)\|\chi_{A_k}\|^{\delta}_{L^{2^\star}(\Omega)}. \tag{5.30}$$

Using the continuous embedding $\widetilde{W}^{s,2}_0(\Omega) \hookrightarrow L^{2^\star}(\Omega)$ and (5.30), we get that there is a constant $C > 0$ such that for every $k \geq 0$,

$$\|u_k\|_{L^{2^\star}(\Omega)} \leq C\left(\|f_0\|_{L^p(\Omega)} + \|f_1\|_{L^q(\Omega\times\Omega)}\right)\|\chi_{A_k}\|^{\delta}_{L^{2^\star}(\Omega)}. \tag{5.31}$$

Step 6: Now let $h > k \geq 0$. Then, $A_h \subset A_k$ and in A_h, we have that $|u_k| \geq (h-k)$. Thus, it follows from (5.31) that there is a constant $C > 0$ such that for every $h > k \geq 0$,

$$\|\chi_{A_h}\|_{L^{2^\star}(\Omega)} \leq C(h-k)^{-1}\left(\|f_0\|_{L^p(\Omega)} + \|f_1\|_{L^q(\Omega\times\Omega)}\right)\|\chi_{A_k}\|^{\delta}_{L^{2^\star}(\Omega)}. \tag{5.32}$$

Let $\Phi(k) := \|\chi_{A_k}\|_{L^{2^\star}(\Omega)}$. It follows from (5.32) that for all $h > k \geq 0$, we have

$$\Phi(h) \leq C(h-k)^{-1} \left(\|f_0\|_{L^p(\Omega)} + \|f_1\|_{L^q(\Omega\times\Omega)} \right) \Phi(k)^{\delta}.$$

Applying Lemma 5.3 to the function Φ, we can deduce that there is a constant $C_1 > 0$ such that $\Phi(K) = 0$ with $K := C_1 C \left(\|f_0\|_{L^p(\Omega)} + \|f_1\|_{L^q(\Omega\times\Omega)} \right)$. We have shown (5.20) as needed. $\square$

The following theorem is the second main result of this section. Here, we reduce the regularity of the datum z, if one compares with [18, Theorem 3.7].

Theorem 5.5. Let $\Omega \subset \mathbb{R}^N$ be a bounded open set with a Lipschitz continuous boundary. Let $2 < \frac{N}{s} < p \leq \infty$ and $0 < \frac{p-1}{p} = \frac{1}{p'} < t < s < 1$. Then, for every $z \in \widetilde{W}^{-t,p}(\Omega)$, there is a unique solution $u \in \widetilde{W}_0^{s,2}(\Omega)$ of (5.2a). In addition, $u \in L^\infty(\Omega)$ and there is a constant $C > 0$ such that

$$\|u\|_{L^\infty(\Omega)} \leq C \|z\|_{\widetilde{W}^{-t,p}(\Omega)}. \tag{5.33}$$

Proof. We prove the result in several steps.

Step 1: Firstly, for $z \in \widetilde{W}^{-t,p}(\Omega)$, by a solution to (5.2a), we mean a function $u \in \widetilde{W}_0^{s,2}(\overline{\Omega})$ satisfying

$$\mathcal{E}(u,v) = \langle z, v \rangle_{\widetilde{W}^{-t,p}(\Omega), \widetilde{W}^{t,p'}(\Omega)}, \quad \forall\, v \in \widetilde{W}_0^{t,p'}(\Omega), \tag{5.34}$$

provided that the left and right hand side expressions make sense.

Step 2: Secondly, since $\frac{1}{p'} < t < 1$ and $z \in \widetilde{W}^{-t,p}(\Omega)$, it follows from Corollary 5.1 that there exists a pair of functions $(f^0, f^1) \in L^p(\Omega) \times L^p(\Omega \times \Omega)$ such that, for every

$v \in \widetilde{W}_0^{t,p'}(\Omega)$, we have

$$\langle z, v\rangle_{\widetilde{W}^{-t,p}(\Omega),\widetilde{W}^{t,p'}(\Omega)} = \int_\Omega f^0 v\,dx + \int_\Omega\int_\Omega f^1(x,y)D_{t,p'}v[x,y]\;dx\;dy. \tag{5.35}$$

Choose $(f^0, f^1) \in L^p(\Omega) \times L^p(\Omega\times\Omega)$ satisfying (5.35) and are such that

$$\|z\|_{\widetilde{W}^{-t,p}(\Omega)} = \|f^0\|_{L^p(\Omega)} + \|f^1\|_{L^p(\Omega\times\Omega)}. \tag{5.36}$$

Since $0 < t < s < 1$ and $2 > p'$, it follows that the continuous embedding $\widetilde{W}_0^{s,2}(\Omega) \hookrightarrow \widetilde{W}_0^{t,p'}(\Omega)$ holds. More precisely, there is a constant $C > 0$ such that

$$\left|D_{t,p'}v[x,y]\right| = \frac{|v(x)-v(y)|}{|x-y|^{\frac{p'}{N}+t}} = \frac{|v(x)-v(y)|}{|x-y|^{\frac{2}{N}+s}}|x-y|^{s-t+\frac{2-p'}{N}} \leq C\left|D_{s,2}v[x,y]\right|,$$

where we used that $s - t + \frac{2-p'}{N} > 0$. Thus, $\|D_{t,p'}v\|_{L^{p'}(\Omega\times\Omega)} \leq C\|D_{s,2}v\|_{L^2(\Omega\times\Omega)}$ for every $v \in \widetilde{W}_0^{s,2}(\Omega)$. Hence, (5.35) also holds for every $v \in \widetilde{W}_0^{s,2}(\Omega)$ and the expressions in (5.34) make sense.

Step 3: We claim that there is a unique $u \in \widetilde{W}_0^{s,2}(\Omega)$ satisfying (5.34). Indeed, let $v \in \widetilde{W}_0^{s,2}(\Omega)$. Using Step 2 and Remark 2.1, we get that there is a constant $C > 0$ such that for every $v \in \widetilde{W}_0^{t,p'}(\Omega)$,

$$\begin{aligned}
\left|\langle z, v\rangle_{\widetilde{W}^{-t,p}(\Omega),\widetilde{W}^{t,p'}(\Omega)}\right| &= \left|\int_\Omega f^0 v\,dx + \int_\Omega\int_\Omega f^1(x,y)D_{t,p'}v[x,y]\;dx\;dy\right| \\
&\leq \|f^0\|_{L^p(\Omega)}\|v\|_{L^{p'}(\Omega)} + \|f^1\|_{L^p(\Omega\times\Omega)}\|D_{t,p'}v\|_{L^{p'}(\Omega\times\Omega)} \\
&\leq C\left(\|f^0\|_{L^p(\Omega)} + \|f^1\|_{L^p(\Omega\times\Omega)}\right)\|v\|_{\widetilde{W}_0^{s,2}(\Omega)}.
\end{aligned}$$

We have shown that the right hand side of (5.35) defines a linear continuous functional on $\widetilde{W}_0^{s,2}(\Omega)$. Thus, the claim follows by applying the Lax-Milgram lemma.

Step 4: It follows from Step 3 that the unique function $u \in \widetilde{W}_0^{s,2}(\Omega)$ satisfying (5.34) is such that for every $v \in \widetilde{W_0}^{t,p'}(\Omega)$, we have

$$\mathcal{E}(u,v) = \langle z, v\rangle_{\widetilde{W}^{-t,p}(\Omega),\widetilde{W}_0^{t,p'}(\Omega)} \leq C\int_\Omega |f^0 v|\,dx + \int_\Omega\int_\Omega \left|f^1(x,y)D_{s,2}v[x,y]\right|\,dx\,dy.$$

Therefore, proceeding exactly as in the proof of Theorem 5.4, we get that $u \in L^\infty(\Omega)$ and there is a constant $C > 0$ such that $\|u\|_{L^\infty(\Omega)} \leq C\left(\|f^0\|_{L^p(\Omega)} + \|f^1\|_{L^p(\Omega\times\Omega)}\right) = C\|z\|_{\widetilde{W}^{-t,p}(\Omega)}$, where we have used (5.36). We have shown (5.33). □

We have the following regularity result as a corollary of Theorems 5.5 and 3.2.

Corollary 5.2. Let $\Omega \subset \mathbb{R}^N$ be a bounded Lipschitz domain satisfying the exterior cone condition. Let $2 < \frac{N}{s} < p \leq \infty$ and $0 < \frac{p-1}{p} = \frac{1}{p'} < t < s < 1$. Let $z \in \widetilde{W}^{-t,p}(\Omega)$ and let $u \in \widetilde{W}_0^{s,2}(\Omega)$ be the unique weak solution of (5.2a). Then, $u \in C_0(\Omega)$.

Proof. Let $z \in \widetilde{W}^{-t,p}(\Omega)$ and $\{z_n\}_{n\geq 1} \subset L^\infty(\Omega)$ a sequence such that $z_n \to z$ in $\widetilde{W}^{-t,p}(\Omega)$ as $n \to \infty$. Let $u_n \in \widetilde{W}_0^{s,2}(\Omega)$ satisfy $\mathcal{E}(u_n, v) = \langle z_n, v\rangle_{\widetilde{W}^{-t,p}(\Omega),\widetilde{W}^{t,p'}(\Omega)} = \int_\Omega z_n v\,dx$ for every $v \in \widetilde{W}_0^{s,2}(\Omega)$. It follows, from Theorem 3.2, that $u_n \in C_0(\Omega)$. Since $u_n - u \in \widetilde{W}_0^{s,2}(\Omega)$ and satisfies $\mathcal{E}(u_n - u, v) = \langle z_n - z, v\rangle_{\widetilde{W}^{-t,p}(\Omega),\widetilde{W}^{t,p'}(\Omega)}$ for every $v \in \widetilde{W}_0^{s,2}(\Omega)$, it follows, from Theorem 5.5, that $(u_n - u) \in L^\infty(\Omega)$ and there is a constant $C > 0$ (independent of n) such that $\|u_n - u\|_{L^\infty(\Omega)} \leq C\|z_n - z\|_{\widetilde{W}^{-t,p}(\Omega)}$. Since $u_n \in C_0(\Omega)$ and $z_n \to z$ in $\widetilde{W}^{-t,p}(\Omega)$ as $n \to \infty$, it follows from the preceding estimate that $u_n \to u$ in $L^\infty(\Omega)$ as $n \to \infty$. Thus, $u \in C_0(\Omega)$. □

Next, we improve the regularity of u solving (5.2a) with a measure μ as the right-hand side datum. Notice that such a result will immediately improve the regularity of the adjoint variable $\bar{\xi}$ solving (5.8b). Recall that the best result so far proved for solutions of (3.5) is given in Theorem 3.3.

Corollary 5.3. Let $\Omega \subset \mathbb{R}^N$ be a bounded Lipschitz domain satisfying the exterior cone condition. Let $2 < \frac{N}{s} < p \le \infty$ and $0 < \frac{p-1}{p} = \frac{1}{p'} < t < s < 1$. Let $\mu \in \mathcal{M}(\Omega)$. Then, there is a unique solution $u \in \widetilde{W}_0^{t,p'}(\Omega)$ to

$$(-\Delta)^s u = \mu \quad \text{in } \Omega, \qquad u = 0 \quad \text{in } \mathbb{R}^N \setminus \Omega,$$

and there is a constant $C > 0$ such that $\|u\|_{\widetilde{W}_0^{t,p'}(\Omega)} \le C\|\mu\|_{\mathcal{M}(\Omega)}$.

Proof. The proof follows exactly as the proof of Theorem 3.3 with the exception that, for the inequality (3.12), we use Corollary 5.2 to get $\left|\int_\Omega u\xi \, dx\right| \le \|\mu\|_{\mathcal{M}(\Omega)}\|v\|_{C_0(\Omega)} \le C\|\mu\|_{\mathcal{M}(\Omega)}\|\xi\|_{\widetilde{W}^{-t,p}(\Omega)}$. The preceding estimate implies that $u \in (\widetilde{W}^{-t,p}(\Omega))^\star = \widetilde{W}_0^{t,p'}(\Omega)$. □

Recall that the "strong form" of the adjoint equation (5.8b) is given by

$$(-\Delta)^s \bar{\xi} = J_u(\bar{u}, \bar{z}) + \bar{\mu} \quad \text{in } \Omega, \qquad \bar{\xi} = 0 \quad \text{in } \mathbb{R}^N \setminus \Omega. \tag{5.37}$$

Using Corollary 5.3 and the fact that $J_u(\bar{u}, \bar{z}) \in L^2(\Omega)$, we obtain the following regularity result.

Corollary 5.4 (Regularity of the adjoint variable). Let $\bar{\mu} \in \mathcal{M}(\Omega)$ and let $\bar{\xi}$ be the Lagrange multiplier given in Theorem 5.2. Then, under the conditions of Corollary 5.3, we have that $\bar{\xi} \in \widetilde{W}_0^{t,p'}(\Omega)$.

5.4.2 Regularity of Control

In this section, we apply the results obtained in the previous section to the optimal control problem. In the literature, a typical cost functional J is given by (cf. the monograph [7])

$$J(u, z) := \frac{1}{2}\|u - u_d\|^2_{L^2(\Omega)} + g(z)\,, \tag{5.38}$$

where $u_d \in L^2(\Omega)$ is given. When $g(z) := \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)}$, with given parameter $\alpha > 0$, then (5.8c) becomes

$$\bar{z} = \mathbb{P}_{Z_{ad}}\left(-\alpha^{-1}\bar{\xi}\right), \tag{5.39}$$

where $\mathbb{P}_{Z_{ad}}$ denotes the projection onto the set Z_{ad}. Recall that, $\bar{\xi}$ is the adjoint variable solving (5.8b). We emphasize that Z_{ad} is still the same as before, and with a choice of $J(\cdot,\cdot)$, all the assumptions in the previous results hold. Recall that the boundedness of Z_{ad} enforces $L^p(\Omega)$ regularity on the control z, in addition with the choice of $g(z) = \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)}$ in (5.38), we are enforcing $L^2(\Omega)$ regularity on the control z. As a result, p is always greater than equal to 2. Finally, (5.39) can improve the regularity of the optimal control from $L^p(\Omega)$ to $W^{t,p'}(\Omega)$.

Theorem 5.6 (Regularity of control)**.** Let the conditions of Corollary 5.4 hold and J be as in (5.38) with $g(z) := \frac{\alpha}{2}\|z\|^2_{L^2(\Omega)}$ with $\alpha > 0$. Given $a, b \in W^{t,p'}(\Omega)$ with $a < b$ a.e. in Ω, let $Z_{ad} := \{z \in L^p(\Omega) : a(x) \le z(x) \le b(x), \text{ a.e. in } \Omega\}$. Then the optimal control $\bar{z} \in W^{t,p'}(\Omega)$.

Proof. Under the assumption on Z_{ad}, the projection in (5.39) becomes
$\mathbb{P}_{Z_{ad}}(\bar{\xi}) := \max\{a, \min\{b, \bar{\xi}\}\}$. Since $\bar{\xi}$ in $\widetilde{W}^{t,p'}_0(\Omega)$, in particular, we have $\bar{\xi}|_\Omega \in W^{t,p'}(\Omega)$. Next since $b \in W^{t,p'}(\Omega)$, using [133, Lemma 2.7], we have that $v := \min\{b, \bar{\xi}\} \in W^{t,p'}(\Omega)$. Similarly, $\max\{a, v\} \in W^{t,p'}(\Omega)$. Thus from (5.39), we obtain that $\bar{z} \in W^{t,p'}(\Omega)$. $\square$

Remark 5.1. We notice that since $\bar{z} \in W^{t,p'}(\Omega)$ (by Theorem 5.6), we have that the regularity of the corresponding solution $\bar{u}$ to the state equation (5.2a) can also be improved. More precisely, by [75, Theorem 7.1 and p524], we have that $\bar{u} \in H^{s(t+2s)}_{p'}(\overline{\Omega}) \cap \widetilde{W}^{s,2}_0(\Omega)$. We refer to [74, Equation (2.9)] for the precise definition of the space $H^{s(t+2s)}_{p'}(\overline{\Omega})$.

5.4.3 Control in $\widetilde{\mathbf{W}}^{-\mathbf{t},\mathbf{p}}(\boldsymbol{\Omega})$ instead of $\mathbf{L}^{\mathbf{p}}(\boldsymbol{\Omega})$

Let Ω, s, t and p be as in Corollary 5.2. Then all the results obtained in Section 5.2 for the optimal control problem hold, with obvious modification of the proofs, if one considers the spaces

$$Z := \widetilde{W}^{-t,p}(\Omega) \quad \text{and} \quad U := \Big\{ u \in \widetilde{W}_0^{s,2}(\Omega) \cap C_0(\Omega) : ((-\Delta)_D^s u)|_\Omega \in \widetilde{W}^{-t,p}(\Omega) \Big\}.$$

Notice that, in this case, Z_{ad} is a closed and convex subset of $\widetilde{W}^{-t,p}(\Omega)$ instead of $L^p(\Omega)$. We further emphasize that even in this case, the adjoint variable still enjoys the higher regularity as given in Corollary 5.4. Furthermore, the result given in Theorem 5.6 remains valid if we replace $L^p(\Omega)$ by $\widetilde{W}^{-t,p}(\Omega)$.

5.5 Conclusions and Future Work

Summary. We have introduced a novel characterization of fractional order Sobolev spaces. For domains with exterior cone condition, we have shown continuity of solutions to fractional PDEs and we have established well-posedness of fractional PDEs with measure-valued data. These results are crucial to study the state (and control) constrained optimal control problems discussed in this work. They have helped to establish the well-posedness of the control problem, deriving the optimality conditions and the regularity of the optimal control.

Perspectives/Open Problems. It will be of interest to relax the aforementioned exterior cone condition to show the continuity of solutions to fractional PDEs. In Corollary 5.1, where under the assumption $\frac{1}{p} < s < 1$, we are able to replace the integral over $\mathbb{R}^N$ (cf. Theorem 5.3) by an integral over Ω, it will be interesting to extend this result to the full range of s, i.e., also when $0 < s \le \frac{1}{p}$.

Chapter 6: Optimal Control of Fractional PDEs with State and Control Constraints

In the previous chapter, we analyzed the theoretical framework for optimal control of fractional elliptic PDEs with state constraints. This chapter considers optimal control of both fractional elliptic and parabolic PDEs with both state and control constraints. The key challenge is how to handle the state constraints. We employ the Moreau-Yosida regularization to handle the state constraints in both elliptic and parabolic cases and establish convergence, with rate, of the regularized optimal control problems to the original ones.

We emphasize that the analysis of optimal control of parabolic PDEs with state constraints is not new, see for instance [40] for the classical case $s = 1$. Similarly to our observation in the elliptic case in Chapter 5, almost none of the existing works can be directly applied to our fractional setting.

For $\Omega \subset \mathbb{R}^N$ $(N \geq 1)$, a bounded open set with boundary $\partial\Omega$, we wish to study the following **fractional optimal control problems**:

$$\min_{(u,z)\in(U,Z)} J(u,z) \tag{6.1a}$$

subject to the fractional PDE: Find $u \in U$ solving

$$\begin{cases} \mathcal{A}_s u &= z \quad \text{in } Q, \\ u &= 0 \quad \text{in } \widehat{\Sigma}, \end{cases} \tag{6.1b}$$

with the state constraints

$$u|_Q \in \mathcal{K} := \{w \in \mathcal{C}(Q) \,:\, w \le u_b, \quad \text{in } \overline{Q}\} \tag{6.1c}$$

and the control constraints

$$z \in Z_{ad}. \tag{6.1d}$$

- **Elliptic case:** $\mathcal{A}_s = (-\Delta)^s$, $Q = \Omega$, $\widehat{\Sigma} := \Sigma := \mathbb{R}^N \setminus \Omega$. Moreover, $\mathcal{C}(Q) := C_0(\Omega)$ is the space of continuous functions in $\overline{\Omega}$ that vanish on $\partial\Omega$ and $u_b \in C(\overline{\Omega})$ such that $u_b \ge 0$ on $\partial\Omega$. By extending functions by zero outside Ω, we can identify $C_0(\Omega)$ with the space $\{u \in C_c(\mathbb{R}^N) \;:\; u = 0 \text{ in } \mathbb{R}^N \setminus \Omega\}$. Finally, $Z_{ad} \subset L^p(\Omega)$ is a non-empty, closed and convex set and p fulfills

$$1 > \frac{N}{2ps}\,. \tag{6.2}$$

- **Parabolic case:** $\mathcal{A}_s = \partial_t + (-\Delta)^s$, $Q = (0,T) \times \Omega$, $\widehat{\Sigma} = ((0,T) \times \Sigma) \cup (\{0\} \times \Sigma)$, with $\Sigma = \mathbb{R}^N \setminus \Omega$. Moreover, $\mathcal{C}(Q) := C(\overline{Q})$ is the space of continuous functions in $\overline{Q}$ and $u_b \in C(\overline{Q})$ with $u_b \ge 0$ on $(0,T) \times \partial\Omega$. Finally, $Z_{ad} \subset L^r((0,T), L^p(\Omega))$ is a non-empty, closed and convex set, and the real numbers p and r fulfill

$$1 > \frac{N}{2ps} + \frac{1}{r}\,. \tag{6.3}$$

The precise definition of the function spaces U and Z will be given in the forthcoming sections.

We note that the control problem with nonlocal PDE constraint in (6.1b) may look similar to Chapter 4 at first glance, however, the two problems are significantly different. First of all, the control is taken as a source in this chapter rather than in the exterior of the domain in Chapter 4. Secondly, here we consider both control and state constraints rather than just the control constraints in Chapter 4.

Due to the difficulties in solving such problems in practice, we introduce a regularized version of the optimal control problem using the Moreau-Yosida regularization. This enables us to create an algorithm to solve the elliptic and parabolic optimal control problems. While this approach is well-known, see [85, 92] for the classical elliptic case, there are no works in the literature for fractional problems. In the classical setting, we refer to [105], where Moreau-Yosida regularization for parabolic problems has been considered, but optimality conditions for the original problem are not discussed. In our work, we carry out the spatial discretization using finite element method and provide discretization error estimates in a particular elliptic setting.

The rest of the chapter is organized as follows. In Section 6.1, we study our parabolic optimal control problem, establish its well-posedness, and derive the first order optimality conditions. We refer to Chapter 5 for the analogous results in the elliptic case. We introduce the Moreau-Yosida regularized problem for both elliptic and parabolic cases in Section 6.2 and show convergence (with rate) of the regularized solutions to the original solution. Section 6.3 is devoted to finite element convergence analysis in a particular elliptic setting. We conclude the chapter with several illustrative numerical examples in Section 6.4.

6.1 Optimal Control Problem

The main goal of this section is to establish well-posdeness of the optimal control problem (6.1) in the parabolic setting and to derive the first order necessary optimality conditions. We start by equivalently rewriting the optimal control problem (6.1) in terms of $(-\Delta)^s_D$. Recall that $(-\Delta)^s_D$ is the realization of $(-\Delta)^s$ in $L^2(\Omega)$ with zero Dirichlet exterior conditions and it is a self-adjoint operator. In terms of $(-\Delta)^s_D$, (6.1) becomes

$$
\begin{aligned}
&\min_{(u,z)\in(U,Z)} J(u,z) \\
&\text{subject to} \\
&\partial_t u + (-\Delta)^s_D u = z \quad \text{in } Q, \quad u(0,\cdot) = 0 \quad \text{in } \Omega \\
&u|_Q \in \mathcal{K} \quad \text{and} \quad z \in Z_{ad}.
\end{aligned}
\tag{6.4}
$$

We next define the appropriate function spaces. Let

$$Z := L^r((0,T); L^p(\Omega)), \quad \text{with } p, r \text{ as in (6.3) but } 1 < p < \infty,\ 1 < r < \infty,$$

$$U := \{u \in \mathbb{U}_0 : (\partial_t + (-\Delta)^s_D)(u|_\Omega) \in L^r((0,T); L^p(\Omega))\}.$$

Here, U is a Banach space with the graph norm

$$\|u\|_U := \|u\|_{\mathbb{U}_0} + \|(\partial_t + (-\Delta)^s_D)(u|_\Omega)\|_{L^r((0,T);L^p(\Omega))}.$$

We let $Z_{ad} \subset Z$ to be a nonempty, closed, and convex set and $\mathcal{K}$ as in (6.1c). Notice that the spaces U and Z are reflexive, and under the assumption that r and p satisfy (6.3), it follows from Corollary 3.1 that $U \hookrightarrow C(\overline{Q})$. This is needed to show the existence of solution to (6.1).

Next using Corollary 3.1 we have that for every $z \in Z$ there is a unique $u \in U$ that solves (6.1b). As a result, the following control-to-state map

$$S : Z \to U, \quad z \mapsto Sz =: u$$

is well-defined, linear, and continuous. Due to the continuous embedding of U in $C(\overline{Q})$ we can in fact consider the control-to-state map as

$$E \circ S : Z \to C(\overline{Q}),$$

and we can define the admissible control set as

$$\widehat{Z}_{ad} := \{z \in Z \ : \ z \in Z_{ad}, \ (E \circ S)z \in \mathcal{K}\},$$

and thus the reduced minimization problem is given by

$$\min_{z \in \widehat{Z}_{ad}} \mathcal{J}(z) := J((E \circ S)z, z). \tag{6.5}$$

Towards this end, we are ready to state the well-posedness of (6.5) and equivalently (6.1).

Theorem 6.1. Let Z_{ad} be a closed, convex, bounded subset of Z and $\mathcal{K}$ a closed and convex subset of $C(\overline{Q})$ such that $\widehat{Z}_{ad}$ is nonempty. Moreover, let $J : L^2(Q) \times L^r((0,T); L^p(\Omega)) \to \mathbb{R}$ be weakly lower-semicontinuous. Then (6.5) has a solution.

Proof. The proof follows by using similar arguments as in the elliptic case, see Theorem 5.1 in Chapter 5 and has been omitted for brevity. □

Next, we derive the first order necessary conditions under the following Slater condition.

Assumption 3. There is some control $\widehat{z} \in Z_{ad}$ such that the corresponding state u fulfills the strict state constraints

$$u(t,x) < u_b(t,x), \quad \forall (t,x) \in \overline{Q}. \tag{6.6}$$

Under this assumption, we have the following first order necessary optimality conditions.

Theorem 6.2. Let $J : L^2(Q) \times L^r((0,T); L^p(\Omega)) \to \mathbb{R}$ be continuously Fréchet differentiable and let (6.6) hold. Let $(\bar{u}, \bar{z})$ be a solution to the optimization problem (6.1). Then there are Lagrange multipliers $\bar{\mu} \in \mathcal{M}(\overline{Q})$ and $\bar{\xi} \in (L^r((0,T);L^p(\Omega)))^\star$ such that

$$\partial_t \bar{u} + (-\Delta)^s_D \bar{u} = \bar{z}, \quad \text{in } Q, \quad \bar{u}(0,\cdot) = 0, \text{ in } \Omega, \tag{6.7a}$$

$$\begin{cases} -\partial_t \bar{\xi} + (-\Delta)^s_D \bar{\xi} & = J_u(\bar{u},\bar{z}) + \bar{\mu}_Q, \quad \text{in } Q, \\ \bar{\xi}(\cdot,T) & = \bar{\mu}_T, \quad \text{in } \overline{\Omega}, \end{cases} \tag{6.7b}$$

$$\langle \bar{\xi} + J_z(\bar{u},\bar{z}), z - \bar{z} \rangle_{(L^r((0,T);L^p(\Omega)))^\star, L^r((0,T);L^p(\Omega))} \geq 0, \qquad \forall\, z \in Z_{ad} \tag{6.7c}$$

$$\bar{\mu} \geq 0, \quad \bar{u}(x) \leq u_b(x) \text{ in } \overline{Q}, \quad \text{and} \quad \int_{\overline{Q}} (u_b - \bar{u})\, d\bar{\mu} = 0, \tag{6.7d}$$

where (6.7a) and (6.7b) are understood in the weak sense (see Definition 3.5) and very-weak sense (see Definition 3.8), respectively.

Proof. We will check the requirements of [88, Lemma 1.14] to complete the proof. Notice that $\partial_t + (-\Delta)^s_D : U \mapsto Z$ is bounded and surjective. In addition, we have that the interior of $\mathcal{K}$ is nonempty due to (6.6). In order to apply [88, Lemma 1.14], the only thing that remains to be shown is the existence of $(\hat{u},\hat{z}) \in U \times Z$ such that

$$\partial_t(\hat{u} - \bar{u}) + (-\Delta)^s_D(\hat{u} - \bar{u}) - (\hat{z} - \bar{z}) = 0 \ \text{ in } \ Q, \quad (\hat{u} - \bar{u}) = 0 \ \text{ in } \ \Omega. \tag{6.8}$$

Notice that for simplicity we have suppressed the initial condition. We recall that $(\bar{u},\bar{z})$ solves the state equation, as a result we obtain that

$$\partial_t \hat{u} + (-\Delta)^s_D \hat{u} = \hat{z} \ \text{ in } \ Q, \quad \hat{u} = 0 \ \text{ in } \ \Omega \tag{6.9}$$

Since for every $\hat{z} \in Z_{ad}$, there is a unique $\hat{u}$ that solves (6.9), in particular $(\hat{u},\hat{z})$ works. Then using [88, Lemma 1.14] we obtain (6.7a)–(6.7c). Moreover, (6.7d) follows from [88, Lemma 1.14] and the discussions given in [88, Page 88]. $\square$

6.2 Moreau-Yosida Regularization of Optimal Control Problem

The purpose of this section is to study the regularized optimal control problem using the well-known Moreau-Yosida regularization and to show that the regularized problem is an approximation to the original problem. The Moreau-Yosida regularized optimal control problem is given by

$$\min J^{\gamma}(u,z) := J(u,z) + \frac{1}{2\gamma}\|(\hat{\mu} + \gamma(u - u_b))_+\|^2_{L^2(Q)}, \tag{6.10a}$$

subject to the fractional PDE: Find $u \in U$ solving

$$\begin{cases} \mathcal{A}_s u &= z \quad \text{in } Q, \\ u &= 0 \quad \text{in } \widehat{\Sigma}, \end{cases} \tag{6.10b}$$

with

$$z \in Z_{ad}, \tag{6.10c}$$

where $0 \leq \hat{\mu} \in L^2(Q)$ is a shift parameter that can be taken to be zero, and $\gamma > 0$ denotes the regularization parameter. Here, $(\cdot)_+$ denotes $\max\{0, \cdot\}$. More information about this can be found in [92]. From hereon, we will use a particular cost functional J, given by,

$$J(u,z) := \frac{1}{2}\|u - u_d\|^2_{L^2(Q)} + \frac{\alpha}{2}\|z\|^2_{L^2(Q)},$$

with $\alpha > 0$ and we choose the relevant function spaces as follows:

1. **Elliptic case.**

$$Z := L^p(\Omega), \quad \text{with } p \text{ as in (6.2) but } 2 \le p < \infty,$$

$$U := \{u \in \widetilde{W}_0^{s,2}(\Omega) : (-\Delta)_D^s(u|_\Omega) \in L^p(\Omega)\}.$$

2. **Parabolic case.**

$$Z := L^r((0,T); L^p(\Omega)), \quad \text{with } p, r \text{ as in (6.3) but } 2 \le p < \infty,\ 2 \le r < \infty,$$

$$U := \{u \in \mathbb{U}_0 : (\partial_t + (-\Delta)_D^s)(u|_\Omega) \in L^r((0,T); L^p(\Omega))\}.$$

Then, again using the control-to-state mapping S, (6.10) can be reduced to

$$\min_{z \in Z_{ad}} \mathcal{J}^\gamma(z) := \mathcal{J}(z) + \frac{1}{2\gamma} \|(\hat{\mu} + \gamma((E \circ S)z - u_b))_+\|_{L^2(Q)}^2. \tag{6.11}$$

Existence and uniqueness of solution to the problem (6.10) can be done using the direct method, see for instance Theorem 5.1 in Chapter 5. For the remainder of the section, we shall only focus on the parabolic case as the elliptic case follows after minor modifications.

Theorem 6.3. Let $(\bar{u}^\gamma, \bar{z}^\gamma)$ be a solution to the regularized optimization problem (6.10). Then there exists a Lagrange multiplier $\bar{\xi}^\gamma \in \mathbb{U}_0$ such that

$$\begin{cases} \partial_t \bar{u}^\gamma + (-\Delta)_D^s \bar{u}^\gamma &= \bar{z}^\gamma, \quad \text{in } Q, \\ \bar{u}^\gamma(0,\cdot) &= 0, \quad \text{in } \Omega, \end{cases} \tag{6.12a}$$

$$\begin{cases} -\partial_t \bar{\xi}^\gamma + (-\Delta)_D^s \bar{\xi}^\gamma &= \bar{u}^\gamma - u_d + (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+, \quad \text{in } Q, \\ \bar{\xi}^\gamma(T,\cdot) &= 0 \quad \text{in } \Omega, \end{cases} \tag{6.12b}$$

$$\langle \bar{\xi}^\gamma + \alpha \bar{z}^\gamma, z - \bar{z}^\gamma \rangle_{L^2(Q)} \ge 0, \quad \forall\, z \in Z_{ad}. \tag{6.12c}$$

Proof. The proof is similar to the proof of Theorem 5.2 in Chapter 5 and has been omitted for brevity. □

Next, let us begin the analysis to show that the regularized problem (6.10) is indeed an approximation to the original problem (6.1). Let us start by deriving a uniform bound on the regularization term. Observe that for $\gamma \geq 1$,

$$\begin{aligned} \mathcal{J}(\bar{z}^\gamma) \leq \mathcal{J}^\gamma(\bar{z}^\gamma) \leq \mathcal{J}^\gamma(\bar{z}) \leq \mathcal{J}(\bar{z}) + \frac{1}{2\gamma}\|\hat{\mu}\|^2_{L^2(Q)} \\ \leq \mathcal{J}(\bar{z}) + \frac{1}{2}\|\hat{\mu}\|^2_{L^2(Q)} =: C_{\bar{z}}. \end{aligned} \tag{6.13}$$

Using (6.13) we get that $\frac{1}{2\gamma}\|(\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+\|^2_{L^2(Q)}$ is uniformly bounded. Also from (6.13) we have that

$$\|(\bar{u}^\gamma - u_b)_+\|^2_{L^2(Q)} \leq \frac{2}{\gamma}\left(\mathcal{J}(\bar{z}) - \mathcal{J}(\bar{z}^\gamma) + \frac{1}{2\gamma}\|\hat{\mu}\|^2_{L^2(Q)}\right).$$

Thus

$$\|(\bar{u}^\gamma - u_b)_+\|_{L^2(Q)} \leq \omega(\gamma^{-1})\gamma^{-\frac{1}{2}}, \tag{6.14}$$

where

$$\omega(\gamma^{-1}) := 2 \max \left(\frac{1}{2\gamma}\|\hat{\mu}\|^2_{L^2(Q)}, (\mathcal{J}(\bar{z}) - \mathcal{J}(\bar{z}^\gamma))_+\right)^{1/2}.$$

Since $\bar{z}^\gamma \to \bar{z}$ strongly in $L^2(Q)$ as $\gamma \to \infty$ (by [86, Proposition 2.1]) and $\mathcal{J}$ is continuous, we obtain that $\omega(\gamma^{-1}) \downarrow 0$ as $\gamma \to \infty$. Moreover, from (6.13) we can deduce that $\mathcal{J}(\bar{z}^\gamma) \leq C_{\bar{z}}$ which yields

$$\max\left(\|\bar{u}^\gamma - u_d\|^2_{L^2(Q)}, \alpha\|\bar{z}^\gamma\|^2_{L^2(Q)}\right) \leq 2C_{\bar{z}}. \tag{6.15}$$

Then from (6.12b), along with (6.14) and (6.15), we obtain that there is a constant $C > 0$ independent of γ such that

$$\begin{aligned} \|\bar{\xi}^\gamma\|_{\mathbb{U}_0} &\le C\left(\|\bar{u}^\gamma - u_d\|_{L^2(Q)} + \|\hat{\mu}\|_{L^2(Q)} + \gamma\|(\bar{u}^\gamma - u_b)_+\|_{L^2(Q)}\right) \\ &\le C\left(2 + \omega(\gamma^{-1})\sqrt{\gamma})\right). \end{aligned}$$

We now estimate the distance between $(\bar{u}, \bar{z})$ and $(\bar{u}^\gamma, \bar{z}^\gamma)$.

Theorem 6.4. Let $(\bar{u}, \bar{z})$ and $(\bar{u}^\gamma, \bar{z}^\gamma)$ denote the solutions of (6.1) and (6.10), respectively. Then,

$$\begin{aligned} \alpha\|\bar{z} - \bar{z}^\gamma\|^2_{L^2(Q)} + \|\bar{u} - \bar{u}^\gamma\|^2_{L^2(Q)} + \gamma\|(\bar{u}^\gamma - u_b)_+\|^2_{L^2(Q)} \\ \le \frac{1}{\gamma}\|\hat{\mu}\|^2_{L^2(Q)} + \langle\bar{\mu}, \bar{u}^\gamma - u_b\rangle_{\mathcal{M}(\overline{Q}), C(\overline{Q})}, \end{aligned} \tag{6.16}$$

and hence,

$$\|(\bar{u}^\gamma - u_b)_+\|_{L^2(Q)} \le \sqrt{\frac{2}{\gamma}} \max\left(\frac{1}{\gamma}\|\hat{\mu}\|^2_{L^2(Q)}, \langle\bar{\mu}, (\bar{u}^\gamma - u_b)_+\rangle_{\mathcal{M}(\overline{Q}), C(\overline{Q})}\right)^{\frac{1}{2}}. \tag{6.17}$$

Proof. We start by using the optimality conditions (6.7c) and (6.12c) with $z = \bar{z}^\gamma$ and $z = \bar{z}$, respectively. This yields

$$\alpha\|\bar{z} - \bar{z}^\gamma\|^2_{L^2(Q)} \le \int_Q (\bar{z} - \bar{z}^\gamma)(\bar{\xi}^\gamma - \bar{\xi})\, dxdt. \tag{6.18}$$

Next, we use the state equations (6.7a) and (6.12a) to get that, for every $v \in (L^r((0,T); L^p(\Omega)))^\star$,

$$\begin{aligned} \langle\partial_t(\bar{u} - \bar{u}^\gamma) + &(-\Delta)^s_D(\bar{u} - \bar{u}^\gamma), v\rangle_{L^r((0,T);L^p(\Omega)),(L^r((0,T);L^p(\Omega)))^\star} \\ &= \langle\bar{z} - \bar{z}^\gamma, v\rangle_{L^r((0,T);L^p(\Omega)),(L^r((0,T);L^p(\Omega)))^\star}. \end{aligned} \tag{6.19}$$

Recall that both $\bar{\xi} \in (L^r((0,T);L^p(\Omega)))^\star$ and

$\bar{\xi}^\gamma \in L^2((0,T);\widetilde{W}_0^{s,2}(\Omega)) \hookrightarrow (L^r((0,T);L^p(\Omega)))^\star$ due to the fact that $(L^r((0,T);L^p(\Omega)))^\star \cong L^{r'}((0,T);L^{p'}(\Omega))$ ([91, Theorem 1.3.10], for example) and Remark 2.1 (iii), so we can set $v := \bar{\xi}^\gamma - \bar{\xi}$ in (6.19). Subsequently, substituting (6.19) in (6.18), and using (6.7b) and (6.12b), along with $\bar{u}, \bar{u}^\gamma \in U$, we obtain

$$\alpha\|\bar{z} - \bar{z}^\gamma\|^2_{L^2(Q)} \leq \langle \partial_t(\bar{u} - \bar{u}^\gamma) + (-\Delta)^s_D(\bar{u} - u^\gamma), \bar{\xi}^\gamma - \bar{\xi}\rangle_{L^r((0,T);L^p(\Omega)),(L^r((0,T);L^p(\Omega)))^\star}$$

$$= -\|\bar{u} - \bar{u}^\gamma\|^2_{L^2(Q)} + \int_Q (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+(\bar{u} - \bar{u}^\gamma)\, dxdt - \langle \bar{\mu}, \bar{u} - \bar{u}^\gamma\rangle_{\mathcal{M}(\overline{Q}),C(\overline{Q})}.$$

From here, we add and subtract u_b to the term $(\bar{u} - \bar{u}^\gamma)$ and note that $\langle \bar{\mu}, u_b - \bar{u}\rangle_{\mathcal{M}(\overline{Q}),C(\overline{Q})} = 0$ from (6.7d). This yields

$$\begin{aligned}
\alpha\|\bar{z} - \bar{z}^\gamma\|^2_{L^2(Q)} + \|\bar{u} - \bar{u}^\gamma\|^2_{L^2(Q)} &\leq \langle \bar{\mu}, \bar{u}^\gamma - u_b\rangle_{\mathcal{M}(\overline{Q}),C(\overline{Q})} \\
&\qquad + \int_Q (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+(\bar{u} - \bar{u}^\gamma + u_b - u_b)\, dxdt \\
&\leq \langle \bar{\mu}, \bar{u}^\gamma - u_b\rangle_{\mathcal{M}(\overline{Q}),C(\overline{Q})} \\
&\qquad + \int_Q (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+(u_b - \bar{u}^\gamma)\, dxdt,
\end{aligned} \tag{6.20}$$

where we have used that $\int_Q (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+(\bar{u} - u_b)\, dxdt \leq 0$.

Next, let

$$Q^+_\gamma(\hat{\mu}) := \{(t,x) \in Q \ : \ \hat{\mu} + \gamma(\bar{u}^\gamma - u_b) > 0\}. \tag{6.21}$$

Then

$$\int_{Q\setminus Q^+_\gamma(\hat{\mu})} (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+(u_b - \bar{u}^\gamma)\, dxdt = 0.$$

This allows us to write

$$\begin{aligned}
\int_Q (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+ (u_b - \bar{u}^\gamma) \, dxdt &= \int_{Q_\gamma^+(\hat{\mu})} (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))(u_b - \bar{u}^\gamma) \, dxdt \\
&= \gamma \int_{Q_\gamma^+(\hat{\mu})} (\bar{u}^\gamma - u_b)(u_b - \bar{u}^\gamma) \, dxdt \\
&\qquad + \int_{Q_\gamma^+(\hat{\mu})} \hat{\mu}(u_b - \bar{u}^\gamma) \, dxdt \\
&= -\gamma \int_{Q_\gamma^+(\hat{\mu})} (\bar{u}^\gamma - u_b)^2 \, dxdt + \int_{Q_\gamma^+(\hat{\mu})} \hat{\mu}(u_b - \bar{u}^\gamma) \, dxdt \\
&\leq -\gamma \|(\bar{u}^\gamma - u_b)_+\|^2_{L^2(Q)} + \frac{1}{\gamma}\|\hat{\mu}\|^2_{L^2(Q)}, \qquad (6.22)
\end{aligned}$$

where the first term in the last inequality follows from the fact that

$$\gamma \|(\bar{u}^\gamma - u_b)_+\|^2_{L^2(Q)} = \gamma \|\bar{u}^\gamma - u_b\|^2_{L^2(Q_\gamma^+(0))} \leq \gamma \|\bar{u}^\gamma - u_b\|^2_{L^2(Q_\gamma^+(\hat{\mu}))}, \qquad (6.23)$$

because $Q_\gamma^+(0) := \{(x,t) \in Q : \bar{u}^\gamma - u_b > 0\} \subseteq Q_\gamma^+(\hat{\mu})$ and the second term follows because $\bar{u}^\gamma - u_b > \dfrac{-\hat{\mu}}{\gamma}$ in $Q_\gamma^+(\hat{\mu})$. Finally, substituting (6.22) in (6.20), we obtain (6.16).

Next, we use that $(\bar{u}^\gamma - u_b)_+ \in C(\overline{Q})$. This follows from the fact that $\bar{u}^\gamma$ is the solution to the state equation (6.1b), hence, it belongs to $C(\overline{Q})$. This, together with the non negativity of $\bar{\mu} \in \mathcal{M}(\overline{Q})$ combined with (6.16), yields (6.17). $\square$

If we take $u_b = 0$ in $\Omega \times [0,T]$ then, under the additional assumption $u_d \geq 0$, we can improve the decay of the violation of the state constraints, that is, we can show that $\|(\bar{u}^\gamma)_+\|_{L^2(Q)} = \mathcal{O}(\gamma^{-1})$ as $\gamma \to \infty$. The results are analogous to [85, Theorems 2.2 and 2.3].

6.3 Finite Element Discretization and Convergence Analysis for the Elliptic Problem

In this section we consider the discretized version of the Moreau-Yosida regularized elliptic optimal control problem. Before we discuss the discretization, we present some preliminary assumptions and results. The notation in this section will be similar to that of the previous sections, but for the elliptic problem. For completeness, and so we may reference it in what follows, we include the following analogue to Theorem 6.3.

Theorem 6.5. Let $(\bar{u}^\gamma, \bar{z}^\gamma)$ be a solution to the regularized optimization problem (6.10). Then there exists a Lagrange multiplier $\bar{\xi}^\gamma \in \mathbb{U}_0$ such that

$$(-\Delta)_D^s \bar{u}^\gamma = \bar{z}^\gamma, \quad \text{in } \Omega, \tag{6.24a}$$

$$(-\Delta)_D^s \bar{\xi}^\gamma = \bar{u}^\gamma - u_d + (\hat{\mu} + \gamma(\bar{u}^\gamma - u_b))_+, \quad \text{in } \Omega, \tag{6.24b}$$

$$(\bar{\xi}^\gamma + \alpha \bar{z}^\gamma, z - \bar{z}^\gamma)_{L^2(\Omega)} \geq 0, \quad \forall\, z \in Z_{ad}. \tag{6.24c}$$

For the remainder of this work, we will require more structure on the admissible set of controls Z_{ad}. Specifically we will use

$$Z_{ad} = \{z \in Z : a \leq z \leq b \quad a.e. \text{ in } \Omega\}, \tag{6.25}$$

where $a < b$ are constants. We next state a result on the regularity of solutions to the regularized problem in Theorem 6.5 for bounded Lipschitz domains. This result will be used throughout the remainder of this section.

Proposition 6.1. Let Ω be a bounded Lipschitz domain and $(\bar{u}^\gamma, \bar{\xi}^\gamma, \bar{z}^\gamma)$ be the solution to (6.24) for a fixed γ, then we have

$$\bar{u}^\gamma \in W^{\sigma-\varepsilon,2}(\Omega), \qquad \bar{\xi}^\gamma \in W^{\sigma-\varepsilon,2}(\Omega), \qquad \bar{z}^\gamma \in W^{\tau,2}(\Omega),$$

for every $\varepsilon > 0$ where $\sigma = \min\{2s, s + 1/2\}$ and $\tau = \min\{1, 2s - \varepsilon\}$.

Proof. Due to $u_d, \hat{\mu} \in L^2(\Omega)$, the right-hand-side of (6.24b) is in $L^2(\Omega)$ and not better, and therefore, for a fixed γ, we have $\bar{\xi}^\gamma \in W^{\sigma-\varepsilon,2}(\Omega)$ for all $\varepsilon > 0$, where $\sigma = \min\{2s, s + 1/2\}$. The optimality condition (6.24c) is equivalent to

$$\bar{z}^\gamma = \mathbb{P}_{Z_{ad}}\left(-\alpha^{-1}\bar{\xi}^\gamma\right) \tag{6.26}$$

where $\mathbb{P}_{Z_{ad}}$ is the projection onto the set Z_{ad}. Then using the fact that the projection formula (6.26) is equivalent to: $\bar{z}^\gamma = \min\left\{b, \max\left\{a, -\alpha^{-1}\bar{\xi}^\gamma\right\}\right\}$, we deduce that $\bar{z}^\gamma \in W^{\min\{1,2s-\varepsilon\},2}(\Omega)$. For this result, when $\min\{1, 2s - \varepsilon\} = 1$, we refer to [97, Chapter II, Theorem 3.1], otherwise we refer to [133, Lemma 2.9]. To obtain the regularity on $\bar{u}^\gamma$, we refer to [30, Theorem 2.1]. □

Remark 6.1. We note that the regularity results above in Proposition 6.1 cannot be improved for Lipschitz domains. This can be seen by noting that, according to [30, Theorem 2.1] and the discussion that follows, we have optimal regularity for the adjoint variable $\bar{\xi}^\gamma$ based on the right-hand side. Using this we obtain the regularity for $\bar{z}^\gamma$. The only room for improvement in the regularity is on $\bar{u}^\gamma$. This requires an assumption that Ω has a smooth boundary, and even then the improvement is limited, see [30, 129].

We next discuss a partial a priori analysis of error due to spatial discretization. We assume we have a quasi-uniform family of triangulations of Ω, where each mesh, denoted $\mathcal{T}_h$, consists of triangles T such that $\cup_{T\in\mathcal{T}_h} = \overline{\Omega}$. Given $\mathcal{T}_h$, we define the finite element space for the state and adjoint variables as

$$U_h = \{u_h \in C_0(\Omega) : u_h\big|_T \in \mathcal{P}_1,\ \forall T \in \mathcal{T}_h\},$$

where $\mathcal{P}_1$ is the space of polynomials of degree at most one.

For a given $z \in Z$, there exists a unique solution to the the problem find $u_h \in U_h$ such that

$$\mathcal{E}(u_h, v_h) = (z, v_h)_{L^2(\Omega)} \quad \forall v_h \in U_h,$$

where

$$\mathcal{E}(u_h, v_h) := \frac{C_{N,s}}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{(u_h(x) - u_h(y))(v_h(x) - v_h(y))}{|x-y|^{N+2s}} \, dxdy. \tag{6.27}$$

We define the discrete solution operator $S_h : Z \to U_h$ whose action is given by $Z \ni z \mapsto S_h z =: u_h$, where u_h solves the above problem.

Next, we define the discrete control space as

$$Z_h := \{z_h \in Z : z_h\big|_T \in \mathcal{P}_0, \ \forall T \in \mathcal{T}_h\},$$

where $\mathcal{P}_0$ denotes space of piecewise constants on the triangulation $\mathcal{T}_h$. Now the admissible discrete control set is given by

$$Z_{ad,h} := Z_{ad} \cap Z_h,$$

where Z_{ad} is as defined in (6.25). Let Π_h defined by

$$(\Pi_h w)\big|_T = \frac{1}{|T|} \int_T w, \qquad \forall T \in \mathcal{T}_h,$$

denote the orthogonal projection onto Z_h. Notice that for $z \in Z$ we have

$$(z - \Pi_h z, w_h)_{L^2(\Omega)} = 0, \qquad \text{for all } w_h \in Z_h. \tag{6.28}$$

Moreover, notice that if $z \in Z_{ad}$, then $\Pi_h z \in Z_{ad,h}$.

Remark 6.2 (Other Control Discretizations)**.** In Z_h, we have chosen the simplest piecewise constant discretization for the control. One can also choose other discretizations such as variational discretization [87] or piecewise linear discretization [114, 115]. However, we emphasize that our proofs carry over to the variational discretization after minor modifications, in addition the piecewise constant case is easier to implement in higher dimensions. Moreover, to fully take advantage of the piecewise linear discretization of control, one needs higher regularity of adjoint and control (for instance, Lipschitz), such regularity results are currently not known for fractional PDEs.

Now the fully discrete version of the regularized problem (6.10) in reduced form is given by

$$\min_{z_h \in Z_{ad,h}} \mathcal{J}_h^\gamma(z_h) := \frac{1}{2}\|S_h z_h - u_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|z_h\|^2_{L^2(\Omega)} + \frac{1}{2\gamma}\|(\hat{\mu} + \gamma(S_h z_h - u_b))_+\|^2_{L^2(\Omega)}. \tag{6.29}$$

The following result is a discrete analogue of Theorem 6.5 stating the first order optimality conditions for (6.29).

Theorem 6.6. Let $(\bar{u}_h^\gamma, \bar{z}_h^\gamma)$ be a solution to the optimization problem (6.29). There exists a Lagrange multiplier (adjoint variable) $\bar{\xi}_h^\gamma \in U_h$ such that

$$\begin{aligned}
&\mathcal{E}(\bar{u}_h^\gamma, v_h) = (\bar{z}_h^\gamma, v_h)_{L^2(\Omega)} && \forall v_h \in U_h,\\
&\mathcal{E}(\bar{\xi}_h^\gamma, v_h) = \big(\bar{u}_h^\gamma - u_d, v_h\big)_{L^2(\Omega)} + \big((\hat{\mu} + \gamma(\bar{u}_h^\gamma - u_b))_+, v_h\big)_{L^2(\Omega)} && \forall\, v_h \in U_h,\\
&\big(\bar{\xi}_h^\gamma + \alpha \bar{z}_h^\gamma, z_h - \bar{z}_h^\gamma\big)_{L^2(\Omega)} \geq 0 && \forall\, z_h \in Z_{ad,h}.
\end{aligned}$$

In order to discuss the convergence of our discrete approximations to the actual solution, we assume that we have a sequence of meshes $\mathcal{T}_h$. The quantity we wish to study is $\|\bar{z} - \bar{z}_h^\gamma\|_{L^2(\Omega)}$, which we will split as

$$\|\bar{z} - \bar{z}_h^\gamma\|_{L^2(\Omega)} \leq \|\bar{z} - \bar{z}^\gamma\|_{L^2(\Omega)} + \|\bar{z}^\gamma - \bar{z}_h^\gamma\|_{L^2(\Omega)}, \tag{6.30}$$

where $\bar{z}$ solves the continuous non-regularized optimal control problem, $\bar{z}^\gamma$ solves (6.11), and $\bar{z}_h^\gamma$ solves (6.29). The estimate of the first term in (6.30) is provided in Theorem 6.4.

Before we obtain our first result for the convergence of $\|\bar{z}^\gamma - \bar{z}_h^\gamma\|_{L^2(\Omega)}$, we introduce some notation to make the computations in what follows more tractable. We define the quantity

$$f(u) := u - u_d + (\hat{\mu} + \gamma(u - u_b))_+,$$

so that the adjoint equation (6.24b) becomes

$$(-\Delta)_D^s \bar{\xi}^\gamma = f(\bar{u}^\gamma), \quad \text{in } \Omega.$$

In what follows, for simplicity, whenever we write the operator S, we really mean that we are using $(E \circ S)$, as defined in Section 6.1. Furthermore, we introduce the continuous and discrete adjoint problem operators as $R := S^*$ and $R_h := S_h^*$, and note that these operators are well defined, linear, and bounded. With this we are ready to state our convergence results.

Theorem 6.7. Let $\bar{z}^\gamma$ and $\bar{z}_h^\gamma$ denote the solution of (6.11) and (6.29) respectively. The discretization error can be bounded as

$$\begin{aligned} \|\bar{z}^\gamma - \bar{z}_h^\gamma\|_{L^2(\Omega)} \le \frac{C}{\alpha}\Big(&\|(R - R_h) f(S\bar{z}^\gamma)\|_{L^2(\Omega)} + \|\Pi_h \bar{\xi}^\gamma - \bar{\xi}^\gamma\|_{L^2(\Omega)} \\ &+ (1+\gamma)\|(S - S_h)\bar{z}^\gamma\|_{L^2(\Omega)} \\ &+ (1+\gamma+\alpha)\|\Pi_h \bar{z}^\gamma - \bar{z}^\gamma\|_{L^2(\Omega)}\Big), \end{aligned} \tag{6.31}$$

where C is a positive constant independent of γ and h.

Proof. In what follows all of the norms and inner products used are in $L^2(\Omega)$ unless noted otherwise. Using the notation introduced above, we begin by defining the auxiliary variables

$$\hat{u}_h^\gamma := S_h \bar{z}^\gamma,$$

$$\hat{\xi}_h^\gamma := R_h f(\bar{u}^\gamma) = R_h f(S\bar{z}^\gamma),$$

$$\widetilde{\xi}_h^\gamma := R_h f(\hat{u}_h^\gamma) = R_h f(S_h \bar{z}^\gamma).$$

Next, we recall the continuous and discrete optimality conditions

$$(\bar{\xi}^\gamma + \alpha \bar{z}^\gamma, z - \bar{z}^\gamma) \geq 0 \qquad \forall\, z \in Z_{ad},$$

$$(\bar{\xi}_h^\gamma + \alpha \bar{z}_h^\gamma, z_h - \bar{z}_h^\gamma) \geq 0 \qquad \forall\, z_h \in Z_{ad,h}.$$

We then perform the following: (i) replace z with $\bar{z}_h^\gamma$ in the first inequality and z_h with $\Pi_h \bar{z}^\gamma$ in the second inequality; (ii) add the two inequalities; (iii) introduce and rearrange terms appropriately to obtain

$$\begin{aligned} \alpha \|\bar{z}^\gamma - \bar{z}_h^\gamma\|^2 &\leq (\bar{\xi}^\gamma - \bar{\xi}_h^\gamma, \bar{z}_h^\gamma - \bar{z}^\gamma) + (\bar{\xi}_h^\gamma + \alpha \bar{z}_h^\gamma, \Pi_h \bar{z}^\gamma - \bar{z}^\gamma) \\ &= (a) \qquad\qquad\qquad\qquad + (b). \end{aligned}$$

We can further separate (a) into

$$\begin{aligned} (\bar{\xi}^\gamma - \bar{\xi}_h^\gamma, \bar{z}_h^\gamma - \bar{z}^\gamma) &= (\bar{\xi}^\gamma - \hat{\xi}_h^\gamma + \hat{\xi}_h^\gamma - \widetilde{\xi}_h^\gamma + \widetilde{\xi}_h^\gamma - \bar{\xi}_h^\gamma, \bar{z}_h^\gamma - \bar{z}^\gamma) \\ &= (\bar{\xi}^\gamma - \hat{\xi}_h^\gamma, \bar{z}_h^\gamma - \bar{z}^\gamma) + (\hat{\xi}_h^\gamma - \widetilde{\xi}_h^\gamma, \bar{z}_h^\gamma - \bar{z}^\gamma) \\ &\quad + (\widetilde{\xi}_h^\gamma - \bar{\xi}_h^\gamma, \bar{z}_h^\gamma - \bar{z}^\gamma) \\ &= (I) + (II) + (III). \end{aligned} \tag{6.32}$$

Before we look at these three terms, we note that we can bound

$$f(u) - f(v) = u - u_d + (\hat{\mu} + \gamma(u - u_b))_+ - (v - u_d + (\hat{\mu} + \gamma(v - u_b))_+)$$

$$\leq u - v + \gamma(u - v)_+, \tag{6.33}$$

since $(u)_+ - (v)_+ \leq (u - v)_+$. Now we bound

$$(I) = (Rf(\bar{u}^\gamma) - R_h f(\bar{u}^\gamma), \bar{z}_h^\gamma - \bar{z}^\gamma) \leq C\|(R - R_h)f(S\bar{z}^\gamma)\|\|\bar{z}_h^\gamma - \bar{z}^\gamma\|,$$

$$(II) = (R_h f(\bar{u}^\gamma) - R_h f(\hat{u}_h^\gamma), \bar{z}_h^\gamma - \bar{z}^\gamma) \leq C(1 + \gamma)\|R_h(S - S_h)\bar{z}^\gamma\|\|\bar{z}_h^\gamma - \bar{z}^\gamma\|.$$

For (III), we use (6.33) to obtain

$$(III) = (R_h f(\hat{u}_h^\gamma) - R_h f(\bar{u}_h^\gamma), \bar{z}_h^\gamma - \bar{z}^\gamma)$$

$$\leq \Big(R_h(S_h(\bar{z}^\gamma - \bar{z}_h^\gamma) + \gamma(S_h(\bar{z}^\gamma - \bar{z}_h^\gamma))_+), \bar{z}_h^\gamma - \bar{z}^\gamma\Big).$$

From here, we note that $(R_h S_h(\bar{z}^\gamma - \bar{z}_h^\gamma), \bar{z}_h^\gamma - \bar{z}^\gamma) \leq 0$, and so we need only consider

$$(III) \leq \gamma(R_h(S_h(\bar{z}^\gamma - \bar{z}_h^\gamma)_+), \bar{z}_h^\gamma - \bar{z}^\gamma)$$

$$= \gamma((S_h(\bar{z}^\gamma - \bar{z}_h^\gamma))_+, S_h(\bar{z}_h^\gamma - \bar{z}^\gamma))$$

$$= -\gamma((S_h(\bar{z}^\gamma - \bar{z}_h^\gamma))_+, S_h(\bar{z}^\gamma - \bar{z}_h^\gamma))$$

$$\leq 0,$$

and so it may be removed from the bound. Now, turning our attention to (b), we note that it can be rewritten as $(\bar{\xi}_h^\gamma, \Pi_h \bar{z}^\gamma - \bar{z}^\gamma)$ using (6.28). Using similar techniques on (b) as we did on (a), we have

$$
\begin{aligned}
(b) &= (\bar{\xi}_h^\gamma - \widetilde{\xi}_h^\gamma + \widetilde{\xi}_h^\gamma - \hat{\xi}_h^\gamma + \hat{\xi}_h^\gamma - \bar{\xi}^\gamma + \bar{\xi}^\gamma, \Pi_h \bar{z}^\gamma - \bar{z}^\gamma) \\
&= (\bar{\xi}^\gamma - \hat{\xi}_h^\gamma, \bar{z}^\gamma - \Pi_h \bar{z}^\gamma) + (\hat{\xi}_h^\gamma - \widetilde{\xi}_h^\gamma, \bar{z}^\gamma - \Pi_h \bar{z}^\gamma) \\
&\quad + (\widetilde{\xi}_h^\gamma - \bar{\xi}_h^\gamma, \bar{z}^\gamma - \Pi_h \bar{z}^\gamma) + (\Pi_h \bar{\xi}^\gamma - \bar{\xi}^\gamma, \bar{z}^\gamma - \Pi_h \bar{z}^\gamma) \\
&= (i) + (ii) + (iii) + (iv). \qquad (6.34)
\end{aligned}
$$

The quantities (i) and (ii) can be bounded similarly to (I) and (II) above with the substitution of $(\bar{z}^\gamma - \Pi_h \bar{z}^\gamma)$ for $(\bar{z}_h^\gamma - \bar{z}^\gamma)$. Focusing on the remaining terms we obtain

$$
\begin{aligned}
(iii) &= (R_h f(\hat{u}_h^\gamma) - R_h f(\bar{u}_h^\gamma), \bar{z}^\gamma - \Pi_h \bar{z}^\gamma) \\
&\leq C(1+\gamma) \| R_h S_h (\bar{z}^\gamma - \bar{z}_h^\gamma) \| \| \bar{z}^\gamma - \Pi_h \bar{z}^\gamma \| \\
(iv) &\leq \| \Pi_h \bar{\xi}^\gamma - \bar{\xi}^\gamma \| \| \bar{z}^\gamma - \Pi_h \bar{z}^\gamma \|.
\end{aligned}
$$

Before combining all of the bounds, we note that the operators S_h, R_h and $R_h S_h$ are uniformly bounded in h, that is, there exists some $\hat{C}$ independent of h such that

$$
\|S_h\| + \|R_h\| + \|R_h S_h\| \leq \hat{C} \qquad \forall h,
$$

where the above norms are the appropriate operator norms. Now combining all of our bounds we have

$$\alpha\|\bar z^\gamma - \bar z_h^\gamma\|^2 \le C\Bigg(\|(R-R_h)f(S\bar z^\gamma)\|\big(\|\bar z_h^\gamma - \bar z^\gamma\| + \|\bar z^\gamma - \Pi_h \bar z^\gamma\|\big)$$

$$+ (1+\gamma)\hat C\|(S-S_h)\bar z^\gamma\|\big(\|\bar z_h^\gamma - \bar z^\gamma\| + \|\bar z^\gamma - \Pi_h \bar z^\gamma\|\big)$$

$$+ (1+\gamma)\hat C\|\bar z^\gamma - \bar z_h^\gamma\|\|\bar z^\gamma - \Pi_h \bar z^\gamma\|$$

$$+ \|\Pi_h\bar\xi^\gamma - \bar\xi^\gamma\|\|\bar z^\gamma - \Pi_h \bar z^\gamma\|\Bigg).$$

Next, we make use of Young's inequality with constant $\alpha/(2C)$ to obtain the bound

$$\alpha\|\bar z^\gamma - \bar z_h^\gamma\|^2 \le C\Bigg(\frac{C}{\alpha}\|(R-R_h)f(S\bar z^\gamma)\|^2 + \frac{(1+\gamma)^2\hat C^2 C}{\alpha}\|(S-S_h)\bar z^\gamma\|^2$$

$$+ \frac{C}{\alpha}\|\Pi_h\bar\xi^\gamma - \bar\xi^\gamma\|^2 + \left(\frac{3\alpha}{4C} + \frac{(1+\gamma)^2\hat C^2 C}{\alpha}\right)\|\Pi_h\bar z^\gamma - \bar z^\gamma\|^2$$

$$+ \frac{3\alpha}{4C}\|\bar z_h^\gamma - \bar z^\gamma\|^2\Bigg).$$

Grouping terms with our quantity of interest on the left and dividing by the resulting constant leads to (6.31). □

We now use the regularity results in Proposition 6.1 to get the following rate of convergence for a fixed γ.

Corollary 6.1. Under the assumptions of Theorem 6.7 and for a fixed γ, we have

$$\|\bar z^\gamma - \bar z_h^\gamma\|_{L^2(\Omega)} \le \frac{C}{\alpha}\Big((h^{2\beta}|\log h|^{2(1+\kappa)} + h^{\beta+s-\varepsilon})(\|\bar z^\gamma\|_{L^2(\Omega)} + \|u_d\|_{L^2(\Omega)} + \|\hat\mu\|_{L^2(\Omega)})$$

$$+ h^{2\beta}|\log h|^{2(1+\kappa)}(1+\gamma)\|\bar z^\gamma\|_{L^2(\Omega)} + h^\tau(1+\gamma+\alpha)|\bar z^\gamma|_{W^{\tau,2}(\Omega)}\Big),$$

for every $\varepsilon > 0$, where $\beta = \min\{s, 1/2\}, \tau = \min\{1, 2s - \varepsilon\}$, $\kappa = 1$ if $s = 1/2$ and zero otherwise.

Proof. For a fixed γ, we use the regularity from Proposition 6.1, the L^2 error estimates from, [30, Proposition 3.8] and interpolation estimates (see [60, Proposition 1.135] for example) to obtain the result. $\square$

6.4 Numerical Experiments

In this section we present numerical experiments in both the elliptic and parabolic cases. Let $\Omega \subset \mathbb{R}^2$ be a disk of radius 1/2 centered at the origin. We truncate $\mathbb{R}^2 \backslash \Omega$ and bound the exterior of Ω with a circle of radius 3/2 centered at the origin. For these experiments, we create a triangular mesh of Ω and the exterior. We discretize the state u and the adjoint p using standard $\mathcal{P}_1$ Lagrangian finite elements and we discretize the control using piecewise constant elements. For details on the discretization of the fractional Laplace operator using $\mathcal{P}_1$, see [1]. For the parabolic problems, we set $T = 1$ and divide the time interval $(0, T)$ into subintervals of equal length $\tau = 0.01$. The time-stepping in fractional PDEs is carried out using backward Euler. All the examples are solved using BFGS with line search, without any control constraints. Notice, that control constraints can be easily incorporated, for instance, by using projected BFGS [95].

6.4.1 Elliptic Problem

In this case, we define our desired state to be

$$u_d(x, y) = \frac{2^{-2s}}{\Gamma(1+s)^2}(1/4 - (x^2 + y^2)_+)^s.$$

For $s = 0.2$, the desired state is shown in the top left panel in Figure 6.2. We take $u_b = 0.1$ as our state constraint. For this experiment, we used a sequence of 8 meshes of Ω and its exterior. A summary of these meshes for easy reference is included in Table 6.1. In

Mesh Number	Number of Nodes	Number of Elements
1	1009	2048
2	1549	3136
3	2920	5894
4	6017	12112
5	8028	16146
6	10819	21744
7	15565	31256
8	24155	48468

Table 6.1: A description of the meshes used in the experiment described in Section 6.4.1. The first column gives the reference number we use to refer to the mesh.

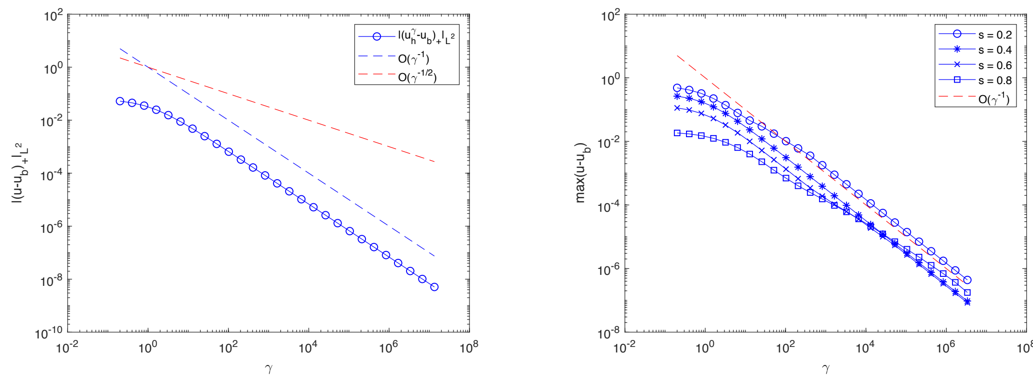

Figure 6.1: **Elliptic case.** Convergence of $\|(u-u_b)_+\|_{L^2(\Omega)}$ (left) and the violation of the state constraints in the max-norm ($\max(u-u_b)$) for given fractional exponent s (right) as γ increases.

the left panel in Figure 6.1 we show the convergence of $\|(u-u_b)_+\|_{L^2(\Omega)}$ for $s=0.4$ on the finest mesh as γ increases. We observe a convergence of $\mathcal{O}(\gamma^{-1})$ which is better than expected. However, this has also been documented in the literature (when $s=1$) and it can be rigorously established when $u_b=0$ and $u_d \geq 0$, see [85]. We also show the maximum error in $u-u_b$ with respect to γ for several values of the fractional exponent s in the right panel of Figure 6.1.

In Figure 6.2 we show the optimal state, control, and Lagrange multiplier for $s=0.2$ and $\gamma=419430.4$. We note that the control is a piecewise constant on the mesh. The

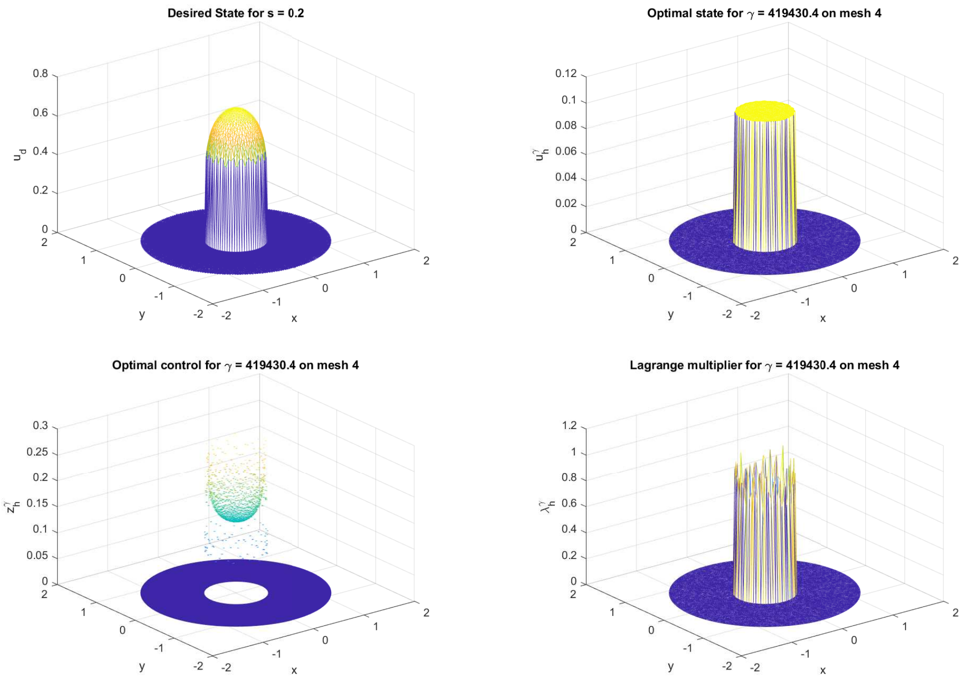

Figure 6.2: **Elliptic case.** The desired state (top left) optimal state (top right), control (bottom left) and Lagrange multiplier (bottom right) for $s = 0.2$ and $\gamma = 419430.4$.

optimal state in Figure 6.2 appears to be cleanly cut off at $u_b = 0.1$, complying with the state constraints and resulting in a cylindrical profile. Moreover, notice that the Lagrange multiplier corresponding to the inequality constraints is a measure (bottom right panel) as expected.

In Figure 6.3 we show some convergence results for $s = 0.2$ for the control and the state as γ increases. Since we do not have an exact solution for the control problem and the associated optimal state, obtaining error estimates for the optimal control and state is problematic. To obtain the plots in Figure 6.3 we solve the optimal control problem on the finest mesh (mesh 8), then project this solution to the coarse meshes. We then compute the L^2-error between solutions on the four coarsest meshes and the projected solution.

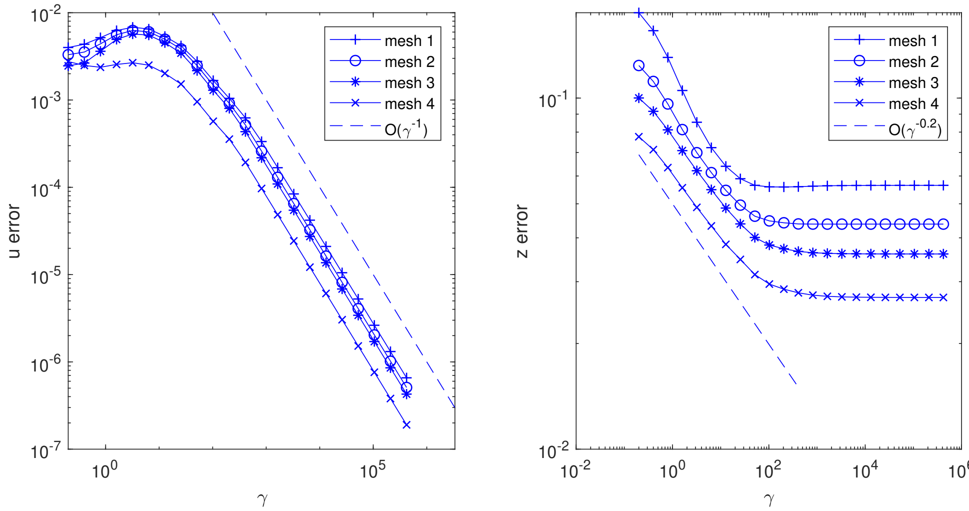

Figure 6.3: The error $\|u - u_h^\gamma\|_{L^2(\Omega)}$ (left) and $\|z - z_h^\gamma\|_{L^2(\Omega)}$ (right) with respect to γ for various values of h (different meshes). We recall that (u, z) are computed by solving the optimal control problem on mesh 8.

6.4.2 Parabolic Problem

To make what follows easier to read, we define

$$\widetilde{u}(x,y) = \frac{2^{-2s}}{\Gamma(1+s)^2}\left(1/4 - (x^2+y^2)_+\right)^s,$$

and subsequently define our desired state and state constraint respectively to be

$$u_d(x,y,t) = 10t^2\widetilde{u}, \qquad u_b(x,y,t) = \frac{1}{10}(1-t)^4\widetilde{u}.$$

For $s = 0.8$, Figure 6.4 shows the convergence of $\|(\bar{u}^\gamma - u_b)_+\|_{L^2(Q)}$ as γ increases.

In Figure 6.5 we show some snapshots of a numerical experiment in which we keep $s = 0.8$, but now use

$$u_d(x,y,t) := 10(1+t)\widetilde{u},$$

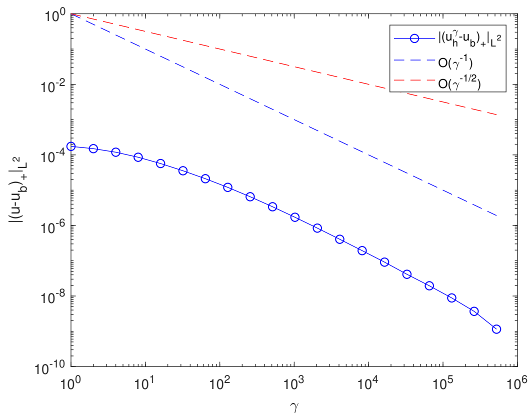

Figure 6.4: Convergence of $\|(\bar{u}^\gamma - u_b)_+\|_{L^2(Q)}$ as γ increases for $s = 0.8$.

and $u_b = 1$, constant in space and time. For $\gamma = 2,097,152$, we show the desired state optimal state, control, adjoint, and Lagrange multiplier at $t = 0.75$ in Figure 6.5. Note the similarity between the desired state and the optimal state, but due to the state constraints the values of the optimal state remain below 1. We also clearly notice that the Lagrange multiplier is a measure. In addition, the adjoint variable is non-smooth at the bottom of its inverted dome, as expected.

6.5 Conclusion and Open Questions

In this chapter, we have presented and analyzed optimal control problems with fractional PDEs as constraints as well as control and state constraints. We introduced the reduced form of the control problem and derived the first order optimality conditions. Notice that the elliptic case was considered in the previous chapter. With a perspective towards implementation, we have introduced the Moreau-Yosida regularization for the (elliptic and parabolic) control problems and have shown convergence of these regularized approximations to the solution of the original problems. We have also derived finite element error estimates in a specific elliptic setting.

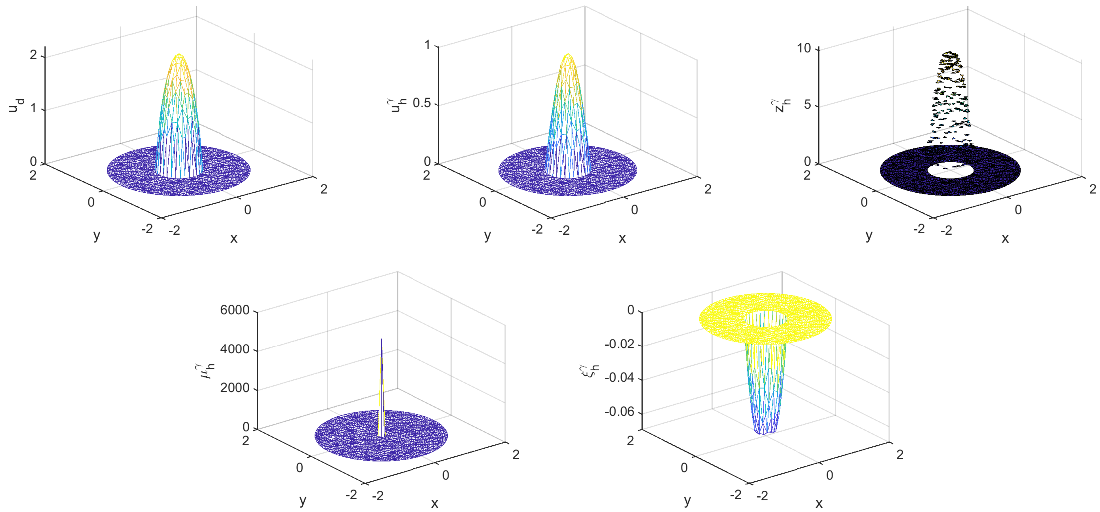

Figure 6.5: The desired state u_d (upper left), optimal state (upper middle), optimal control (upper right), Lagrange multiplier (lower left), and optimal adjoint (lower right) for $s = 0.8$ and $\gamma = 2,097,153$ at time $t = 0.75$.

The focus of the chapter is on the analysis and algorithm for the fractional optimal control problems with state constraints. But there remain several open questions, especially on the regularity of the original state equation. It remains open to establish the continuity of solutions to the fractional parabolic problem when the right-hand-side belongs to a dual space. Notice that this result is known in the elliptic case, see Chapter 5. We have established finite element error estimates for the elliptic problem, but the error bounds depends on the Moreau-Yosida regularization parameter γ. It remains open to show error estimates which are independent of γ. The discretization error estimates for the fully discrete problem are completely open.

Chapter 7: Novel Deep Neural Networks for solving Bayesian Statistical Inverse Problems

7.1 Introduction

Large-scale statistical inverse problems governed by partial differential equations (PDEs) are increasingly found in different areas of computational science and engineering [23, 33, 37, 63, 93]. The basic use of this class of problems is to recover certain physical quantities from limited and noisy observations. They are generally computationally demanding and can be solved using the Bayesian framework [33, 58, 63]. In the Bayesian approach to statistical inverse problems, one models the solution as a posterior distribution of the unknown parameters conditioned on the observations. Such mathematical formulations are often ill-posed and regularization is introduced in the model via prior information. The posterior distribution completely characterizes the uncertainty in the model. Once it has been computed, statistical quantities of interest can then be obtained from the posterior distribution. Unfortunately, many problems of practical relevance do not admit analytic representation of the posterior distribution. Thus, the posterior distributions are generally sampled using Markov chain Monte Carlo (MCMC)-type schemes.

We point out here that a large number of samples are generally needed by MCMC methods to obtain convergence. This makes MCMC methods computationally intensive to simulate Bayesian inverse problems. Moreover, for statistical inverse problems governed by PDEs, one needs to solve the forward problem corresponding to each MCMC sample. This task further increases the computational complexity of the problem, especially if the governing PDE is a large-scale nonlinear PDE [58]. Hence, it is reasonable to construct a computationally cheap surrogate to replace the forward model solver [101]. Problems that involve large numbers of input parameters often lead to the so-called *curse of dimensionality*. Projection-based reduced-order models e.g., reduced basis methods and the discrete empirical interpolation method (DEIM) are typical examples of dimensionality reduction methods for tackling parameterized PDEs [5, 24, 43, 57, 64, 84, 110]. However, they are intrusive by design in the sense that they do not allow reuse of existing codes in the forward solves, especially for nonlinear forward models. To mitigate this computational issue, the goal of this work is to demonstrate the use of deep neural networks (DNN) to construct surrogate models specifically for nonlinear parameterized PDEs governing statistical Bayesian inverse problems. In particular, we will focus on fractional Deep Neural Networks (fDNN) which have been recently introduced and applied to classification problems [17].

The use of DNNs for surrogate modeling in the framework of PDEs has received increasing attention in recent years, see e.g., [44,76,80,102,109,111,124,137,139] and the references therein. Some of these references also cover the type of Bayesian inverse problems considered in this chapter. To accelerate or replace computationally-expensive PDE solves, a DNN is trained to approximate the mapping from parameters in the PDE to observations obtained from its solution. In a supervised learning setting, training data consisting of inputs and outputs are available and the learning problem aims at tuning the weights of the DNN. To obtain an effective surrogate model, it is desirable to train the DNN to a high accuracy.

We consider a different approach than the aforementioned works. We propose novel dynamical systems based neural networks which allow connectivity among all the network layers. Specifically, we consider a fractional DNN technique recently proposed in [17] in the framework of classification problems. We note that [17] is motivated by [77]. Both papers are in the spirit of the push to develop rigorous mathematical models for the analysis and understanding of DNNs. The idea is to consider DNNs as dynamical systems. More precisely, in [25, 77, 79], DNN is thought of as an optimization problem constrained by a discrete ordinary differential equation (ODE). As pointed out in [17], designing the DNN solution algorithms at the continuous level has the appealing advantage of architecture independence; in other words, the number of optimization iterations remains the same even if the number of layers is increased. Moreover, [17] specifically considers continuous fractional ODE constraint. Unlike standard DNNs, the resulting fractional DNN allows the network to access historic information of input and gradients across all subsequent layers since all the layers are connected to one another.

We demonstrate that our fractional DNN leads to a significant reduction in the overall computational time used by MCMC algorithms to solve inverse problems, while maintaining accurate evaluation of the relevant statistical quantities. We note here that, although the papers [102,124,139] consider DNNs for inverse problems governed by PDEs, the DNNs used in these studies do not, in general, have the optimization-based formulation considered here.

The remainder of the chapter is organized as follows: In section 7.2 we state the generic parameterized PDEs under consideration, as well as discuss the well-known surrogate modeling approaches. Our main DNN algorithm to approximate parameterized PDEs is provided in section 7.3. We first discuss a ResNet architecture to approximate parameterized PDEs, which is followed by our fractional ResNet (or fractional DNN) approach to carry out the same task. A brief discussion on error estimates has been provided. Next, we discuss the application of fractional DNN to statistical Bayesian inverse problems in section 7.4. We conclude the chapter with several illustrative numerical examples in section 7.5.

7.2 Parameterized PDEs

In this section, we present an abstract formulation of the forward problem as a discrete partial differential equation (PDE) depending on parameters. We are interested in the following model which represents (finite element, finite difference or finite volume) discretization of a (possibly nonlinear) parameterized PDE

$$F(\mathbf{u}(\boldsymbol{\xi}); \boldsymbol{\xi}) = 0, \tag{7.1}$$

where $\mathbf{u}(\boldsymbol{\xi}) \in \mathcal{U} \subset \mathbb{R}^{N_x}$ and $\boldsymbol{\xi} \in \mathcal{P} \subset \mathbb{R}^{N_{\boldsymbol{\xi}}}$ denote the solution of the PDE and the parameter in the model, respectively. Here, $\mathcal{P}$ denotes the parameter domain and $\mathcal{U}$ the solution manifold. For a fixed parameter $\boldsymbol{\xi} \in \mathcal{P}$, we seek the solution $\mathbf{u}(\boldsymbol{\xi}) \in \mathcal{U}$. In other words, we have the functional relation given by the parameter-to-solution map

$$\boldsymbol{\xi} \mapsto \Phi(\boldsymbol{\xi}) \equiv \mathbf{u}(\boldsymbol{\xi}). \tag{7.2}$$

Several processes in computational sciences and engineering are modelled via parameter-dependent PDEs, which when discretized are of the form (7.1), for instance, in viscous flows governed by Navier-Stokes equations, which is parameterized by the Reynolds number [59]. Similarly, in unsteady natural convection problems modelled via Boussinesq equations, the Grashof or Prandtl numbers are important parameters [110], etc. Approximating the high-fidelity solution of (7.1) can be done with nonlinear solvers, such as Newton or Picard methods combined with Krylov solvers [59]. However, computing solutions to (7.1) can become prohibitive, especially when they are required for many parameter values and $N_{\boldsymbol{\xi}}$ is large. Besides, a relatively large N_x (resulting from a really fine mesh in the discretization of the PDE) yields large (nonlinear) algebraic systems which are computationally expensive to solve and may also lead to huge storage requirements. This is, for instance, the case in Bayesian inference problems governed by PDEs where several forward solves are required to adequately sample posterior distributions through MCMC-type schemes [26, 104]. As a result of the above computational challenges, it is reasonable to replace the high-fidelity model by a surrogate model which is relatively easy to evaluate.

In what follows, we discuss two classes of surrogate models, namely: reduced-order models and deep learning models.

7.2.1 Surrogate Modeling

Surrogate models are cheap-to-evaluate models designed to replace computationally costly models. The major advantage of surrogate models is that approximate solution of the models can be easily evaluated at any new parameter instance with minimal loss of accuracy, at a cost independent of the dimension of the original high-fidelity problem.

A popular class of surrogate models are the reduced-order models (ROMs), see e.g., [5,24, 42, 56, 57, 64, 84, 110]. Typical examples of ROMs include the reduced basis method [110], proper orthogonal decomposition [42, 72, 84, 110] and the discrete empirical interpolation method (DEIM) and its variants [5, 24, 42, 57, 64]. A key feature of the ROMs is that they use the so-called *offline-online paradigm*. The offline step essentially constructs the low-dimensional approximation to the solution space; this approximation is generally known as the *reduced basis*. The online step uses the reduced basis to solve a smaller reduced problem. The resulting reduced solution accurately approximates the solution of the original problem.

Deep neural network (DNN) models constitute another class of surrogate models which are well-known for their high approximation capabilities. The basic idea of DNNs is to approximate multivariate functions through a set of layers of increasing complexity [71]. Examples of DNNs for surrogate modeling include Residual Neural Network (ResNet) [79, 82], physics-informed neural network (PINNs) [111] and fractional DNN [17].

Note that the ROMs require system solves and they are highly intrusive especially for nonlinear problems [5, 42, 57, 58]. In contrast, the DNN approach is fully non-intrusive, which is essential for legacy codes. Although rigorous error estimates for ROMs under various assumptions have been well studied [110], we like the advantage of DNN being nonintrusive, but recognize that error analysis is not yet as strong.

In this work, we propose a surrogate model based on the combination of POD and fractional DNN. Before we discuss our proposed model, we first review POD.

7.2.2 Proper Orthogonal Decomposition

For sufficiently large $N_s \in \mathbb{N}$, suppose that $\mathbb{E} := \{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots, \boldsymbol{\xi}_{N_s}\}$ is a set of parameter samples with $\boldsymbol{\xi}_i \in \mathbb{R}^{N_{\boldsymbol{\xi}}}$, and $\{\mathbf{u}(\boldsymbol{\xi}_1), \mathbf{u}(\boldsymbol{\xi}_2), \cdots, \mathbf{u}(\boldsymbol{\xi}_{N_s})\}$ the corresponding snapshots (solutions of the model (7.1), with $\mathbf{u}_i \in \mathbb{R}^{N_x}$). Here, we assume that $\text{span}\{\mathbf{u}(\boldsymbol{\xi}_1), \mathbf{u}(\boldsymbol{\xi}_2), \cdots, \mathbf{u}(\boldsymbol{\xi}_{N_s})\}$ sufficiently approximates the space of all possible solutions of (7.1). Next, we denote by

$$\mathbb{S} = [\mathbf{u}_1 | \mathbf{u}_2 | \cdots | \mathbf{u}_{N_s}] \in \mathbb{R}^{N_x \times N_s} \tag{7.3}$$

the matrix whose columns are the solution snapshots. Then, the singular value decomposition (SVD) of $\mathbb{S}$ is given by

$$\mathbb{S} = \widetilde{\mathbb{V}}\Sigma\mathbb{W}^T, \tag{7.4}$$

where $\widetilde{\mathbb{V}} \in \mathbb{R}^{N_x \times r}$ and $\mathbb{W} \in \mathbb{R}^{N_s \times r}$ are orthogonal matrices called the left and right singular vectors, and $r \leq \min\{N_s, N_x\}$ is the rank of $\mathbb{S}$. Thus, $\widetilde{\mathbb{V}}^T\widetilde{\mathbb{V}} = I_r, \quad \mathbb{W}^T\mathbb{W} = I_r$, and $\Sigma = \mathrm{diag}(\rho_1, \rho_2, \cdots, \rho_r) \in \mathbb{R}^{r \times r}$, where $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_r \geq 0$ are the singular values of $\mathbb{S}$.

Now, denote by $\mathbb{V} \subset \widetilde{\mathbb{V}}$ the first $k \leq r$ left singular vectors of $\mathbb{S}$. Then, the columns of $\mathbb{V} \in \mathbb{R}^{N_x \times k}$ form a POD basis of dimension k. According to the Schmidt-Eckart-Young theorem [54, 69], the POD basis $\mathbb{V}$ minimizes, over all possible k-dimensional orthonormal bases $\mathbb{Z} \in \mathbb{R}^{N_x \times k}$, the sum of the squares of the errors between each snapshot vector $\mathbf{u}_i$ and its projection onto the subspace spanned by $\mathbb{Z}$. More precisely,

$$\sum_{i=1}^{N_s} ||\mathbf{u}_i - \mathbb{V}\mathbb{V}^T\mathbf{u}_i||_2^2 = \min_{\mathbb{Z}\in\mathcal{Z}} \sum_{i=1}^{N_s} ||\mathbf{u}_i - \mathbb{Z}\mathbb{Z}^T\mathbf{u}_i||_2^2 = \sum_{i=k+1}^{r} \rho_i, \tag{7.5}$$

where $\mathcal{Z} := \{\mathbb{Z} \in \mathbb{R}^{N_x \times k} : \mathbb{Z}^T\mathbb{Z} = I_k\}$. Note from (7.5) that the POD basis $\mathbb{V}$ solves a least squares minimization problem, which guarantees that the approximation error is controlled by the singular values.

For every $\boldsymbol{\xi}$, we then approximate the continuous solution $\mathbf{u}(\boldsymbol{\xi})$ as $\mathbf{u}(\boldsymbol{\xi}) \approx \mathbb{V}\widehat{\mathbf{u}}(\boldsymbol{\xi})$, where $\widehat{\mathbf{u}}(\boldsymbol{\xi})$ solves the reduced problem

$$\mathbb{V}^T F(\mathbb{V}\widehat{\mathbf{u}}(\boldsymbol{\xi}); \boldsymbol{\xi}) = 0. \tag{7.6}$$

Notice that, in some reduced-modeling techniques such as DEIM, additional steps are needed to fully reduce the dimensionality of the problem (7.6). Nevertheless, one still needs to solve a nonlinear (reduced) system like (7.6) to evaluate $\widehat{\mathbf{u}}$.

We conclude this section by emphasizing that the above approach is "linear" because

$\mathbf{u}(\boldsymbol{\xi}) \approx \widehat{\Phi}(\boldsymbol{\xi})$, where $\widehat{\Phi}(\boldsymbol{\xi}) = \mathbb{V}\widehat{\mathbf{u}}(\boldsymbol{\xi})$ is an approximation of the map $\Phi(\boldsymbol{\xi})$ given in (7.2).

7.3 Deep Neural Network

The DNN approach to modeling surrogates produces a nonlinear approximation $\widehat{\Phi}$ of the input-output map $\Phi : \mathbb{R}^{N_{\boldsymbol{\xi}}} \to \mathbb{R}^{N_x}$ given in (7.2), where $\widehat{\Phi}$ depends implicitly on a set of parameters $\boldsymbol{\theta} \in \mathbb{R}^{N_\theta}$ configured as a layered set of latent variables that must be trained. We represent this dependence using the notation $\widehat{\Phi}(\boldsymbol{\xi}; \boldsymbol{\theta})$. In the context of PDE surrogate modeling, training a DNN requires a data set $(\mathbb{E}, \mathbb{S})$, where the parameter samples $\boldsymbol{\xi}_j \in \mathbb{E}$ are the inputs and the corresponding snapshots $\mathbf{u}_j \in \mathbb{S}$ are targets; training then consists of constructing $\boldsymbol{\theta}$ so that the DNN $\widehat{\Phi}(\boldsymbol{\xi}_j; \boldsymbol{\theta})$ matches $\mathbf{u}_j$ for each $\boldsymbol{\xi}_j$. This matching is determined by a *loss functional.* The learning problem therefore involves computing the optimal parameter $\boldsymbol{\theta}$ that minimizes the loss functional and satisfies $\widehat{\Phi}(\boldsymbol{\xi}_j; \boldsymbol{\theta}) \approx \mathbf{u}_j$. The ultimate goal is that this approximation also holds with the same optimal parameter $\boldsymbol{\theta}$ for a different data set; in other words, for $\boldsymbol{\xi} \notin \mathbb{E}$, we take $\widehat{\Phi}(\boldsymbol{\xi}; \boldsymbol{\theta})$ to represent a good approximation to $\Phi(\boldsymbol{\xi})$.

Deep learning can be either supervised or unsupervised depending on the data set used in training. In the supervised learning technique for DNN, all the input samples $\boldsymbol{\xi}_j$ are available for all the corresponding samples of the targets $\mathbf{u}_j$. In contrast, the unsupervised learning framework does not require all the outputs to accomplish the training phase. We adopt the supervised learning approach to model our surrogate and apply it to Bayesian inverse problems. In particular, we shall discuss Residual neural network (ResNet)[79, 82] and Fractional DNN [17] in the context of PDE surrogate modeling.

7.3.1 Residual Neural Network

The residual neural network (ResNet) model was originally proposed in [82]. For a given input datum $\boldsymbol{\xi}$, ResNet approximates Φ through the following recursive expression

$$\begin{aligned} \phi_1 &= \sigma(W_0 \boldsymbol{\xi} + \mathbf{b}_0), \\ \phi_j &= \phi_{j-1} + h\sigma(W_{j-1}\phi_{j-1} + \mathbf{b}_{j-1}), \quad 2 \le j \le L-1, \\ \phi_L &= W_{L-1}\phi_{L-1}, \end{aligned} \tag{7.7}$$

where $\{W_j, \mathbf{b}_j\}$ are the weights and the biases, $h > 0$ is the stepsize and L is the number of layers and σ is an activation function which is applied element-wise on its arguments. Typical examples of σ include the hyperbolic tangent function, the logistic function or the rectified linear unit (or ReLU) function. ReLU is a nonlinear function given by $\sigma(x) = \max\{x, 0\}$. In this work, we use a smoothed ReLU function defined, for $\varepsilon > 0$, as

$$\sigma(x) := \begin{cases} x, & x > \varepsilon \\ 0, & x < -\varepsilon \\ \frac{1}{4\varepsilon}x^2 + \frac{1}{2}x + \frac{\varepsilon}{4}, & -\varepsilon \le x \le \varepsilon. \end{cases} \tag{7.8}$$

Note that as $\varepsilon \to 0$, smooth ReLU approaches ReLU, see Figure 7.1.

It follows from (7.7) that

$$\widehat{\Phi}(\boldsymbol{\xi}; \boldsymbol{\theta}) = \phi_L(\boldsymbol{\xi}) = W_{L-1}\Big(\big(I + h(\sigma \circ \mathcal{K}_{L-1})\big) \circ \cdots \circ \big(I + h(\sigma \circ \mathcal{K}_1)\big) \circ \mathcal{K}_0\Big)(\boldsymbol{\xi}), \tag{7.9}$$

where $\mathcal{K}_j(\mathbf{y}) = W_j\,\mathbf{y} + \mathbf{b}_j$, for all $j = 0, \ldots, L-1$ and for any $\mathbf{y}$.

To this end, the following two critical questions naturally come to mind:

(a) How well does $\widehat{\Phi}(\boldsymbol{\xi}; \boldsymbol{\theta})$ approximate $\Phi(\boldsymbol{\xi})$?

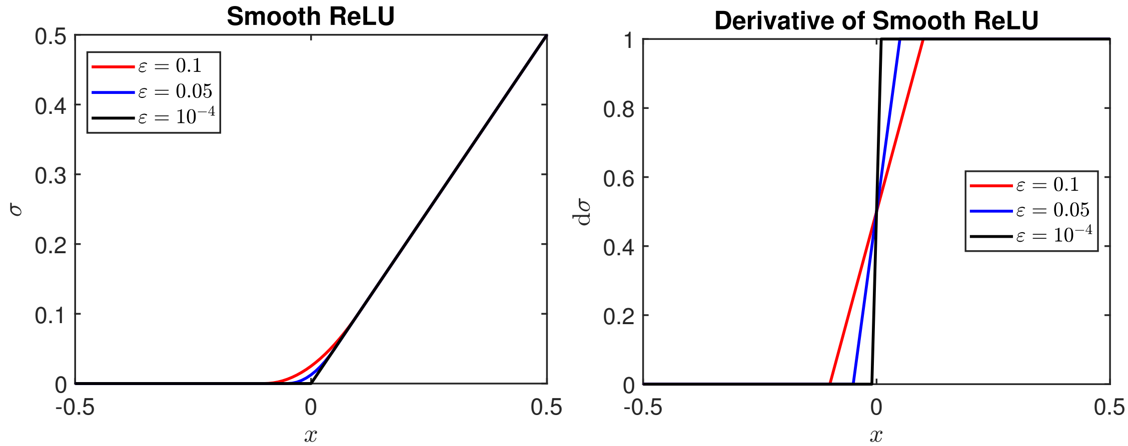

Figure 7.1: Smooth ReLU and derivative for different values of ε.

(b) How do we determine $\boldsymbol{\theta}$?

We will address (b) first and leave (a) for a discussion provided in section 7.3.3. Now, setting $\boldsymbol{\theta} = \{W_j, \mathbf{b}_j\}$, it follows that the problem of approximating Φ via ResNet is essentially the problem of learning the unknown parameter $\boldsymbol{\theta}$. More specifically, the learning problem is the solution of the minimization problem [79]:

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}) \tag{7.10}$$

subject to constraints (7.7), where $\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{u})$ is suitable loss function.

In this work, we consider the mean squared error, together with a regularization term, as our loss functional:

$$\mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}) = \frac{1}{2N_s} \sum_{j=1}^{N_s} \|\hat{\Phi}(\boldsymbol{\xi}_j; \boldsymbol{\theta}) - \mathbf{u}_j\|_2^2 + \frac{\lambda}{2} ||\boldsymbol{\theta}||_2^2, \tag{7.11}$$

where λ is the regularization parameter. Due to its highly non-convex nature, this is a very difficult optimization problem. Indeed, a search over a high dimensional parameter space for the global minimum of a non-convex function can be intractable. The current state-of-the-art approaches to solve these optimization problems are based on the stochastic gradient descent method [102, 124, 139].

As pointed out in, for instance, [79], the middle equation in expression (7.7) mimics the forward Euler discretization of a nonlinear differential equation:

$$d_t\phi(t) = \sigma(W(t)\phi(t) + b(t)), \quad t \in (0, T],$$

$$\phi(0) = \phi_0.$$

It is known that standard DNNs are prone to vanishing gradient problem [128], leading to loss of information during the training phase. ResNets do help, but more helpful is the so-called DenseNet [90]. Notice that in a typical DenseNet, each layer takes into account all the previous layers. However, this is an ad hoc approach with no rigorous mathematical framework. Recently in [17], a mathematically rigorous approach based on fractional derivatives has been introduced. This ResNet is called *Fractional DNN.* In [17], the authors numerically establish that the fractional derivative based ResNet outperforms ResNet in overcoming the vanishing gradient problem. This is not surprising because fractional derivatives allow connectivity among all the layers. Building rigorously on the idea of [17], we proceed to present the fractional DNN surrogate model for the discrete PDE in (7.1).

7.3.2 Fractional Deep Neural Network

As pointed out in the previous section, the fractional DNN approach is based on the replacement of the standard ODE constraint in the learning problem by a fractional differential equation. To this end, we first introduce the definitions and concepts on which we shall rely to discuss fractional DNN.

Let $u : [0, T] \to \mathbb{R}$ be an absolutely continuous function and assume $\gamma \in (0, 1)$. Next, consider the following fractional differential equations

$$d_t^\gamma u(t) = f(u(t)), \quad u(0) = u_0, \tag{7.12}$$

and

$$d_{T-t}^{\gamma} u(t) = f(u(t)), \quad u(T) = u_T. \tag{7.13}$$

Here, d_t^{γ} and d_{T-t}^{γ} denote the left and right Caputo derivatives, respectively [17].

Then, setting $u(t_j) = u_j$, and using the L^1-scheme (see e.g., [17]) for the discretization of (7.12) and (7.13) yields, respectively,

$$u_{j+1} = u_j - \sum_{k=0}^{j-1} a_{j-k} \ (u_{k+1} - u_k) + h^{\gamma}\Gamma(2-\gamma) f(u_j), \quad j = 0, ..., L-1, \tag{7.14}$$

and

$$u_{j-1} = u_j + \sum_{k=j}^{L-1} a_{k-j} \ (u_{k+1} - u_k) - h^{\gamma}\Gamma(2-\gamma) f(u_j), \quad j = L, ..., 1, \tag{7.15}$$

where $h > 0$ is the step size, $\Gamma(\cdot)$ is the Euler-Gamma function, and

$$a_{j-k} = (j+1-k)^{1-\gamma} - (j-k)^{1-\gamma}. \tag{7.16}$$

After this brief overview of the fractional derivatives, we are ready to introduce our fractional DNN (cf. (7.7))

$$\begin{aligned}
&\phi_1 = \sigma(W_0 \phi_0 + \mathbf{b}_0); \quad \phi_0 = \boldsymbol{\xi}, \\
&\phi_j = \phi_{j-1} - \sum_{k=0}^{j-2} a_{j-1-k} \ (\phi_{k+1} - \phi_k) + h^{\gamma}\Gamma(2-\gamma)\sigma(W_{j-1}\phi_{j-1} + \mathbf{b}_{j-1}), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 2 \le j \le L-1, \\
&\phi_L = W_{L-1}\phi_{L-1}.
\end{aligned} \tag{7.17}$$

Our learning problem then amounts to

$$\min_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}) \tag{7.18}$$

subject to constraints (7.17). Notice that the middle equation in (7.17) mimics the L^1-in time discretization of the following nonlinear fractional differential equation

$$d_t^\gamma \phi(t) = \sigma(W(t)\phi(t) + b(t)), \quad t \in (0, T],$$

$$\phi(0) = \phi_0.$$

There are two ways to approach the constrained optimization problem (7.18). The first approach is the so-called reduced approach, where we eliminate the constraints (7.17) and consider the minimization problem (7.18) only in terms of $\boldsymbol{\theta}$. The resulting problem can be solved by using a gradient based method such as BFGS, see e.g., [96, Chapter 4]. During every step of the gradient method, one needs to solve the state equation (7.17) and an adjoint equation. These two solves enable us to derive an expression of the gradient of the reduced loss function with respect to $\boldsymbol{\theta}$. Alternatively, one can derive the gradient and adjoint equations by using the Lagrangian approach. It is well-known that the gradient with respect to $\boldsymbol{\theta}$ for both approaches coincides, see [8, pg. 14] for instance. We next illustrate how to evaluate this gradient using the Lagrangian approach.

The Lagrangian functional associated with the discrete constrained optimization problem (7.18) is given by

$$\mathcal{L}(\mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\psi}) := \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}) + \langle \phi_1 - \sigma(W_0\phi_0 + \mathbf{b}_0), \psi_1 \rangle$$

$$+ \sum_{j=2}^{L-1} \left\langle \phi_j - \phi_{j-1} + \sum_{k=0}^{j-2} a_{j-1-k} \ (\phi_{k+1} - \phi_k) + h^\gamma \Gamma(2-\gamma)\sigma(W_{j-1}\phi_{j-1} + \mathbf{b}_{j-1}), \psi_j \right\rangle$$

$$+ \langle \phi_L - W_{L-1}\phi_{L-1}, \psi_L \rangle, \quad (7.19)$$

where ψ_j's are the Lagrange multi1pliers, also called adjoint variables, corresponding to (7.17) and $\langle \cdot, \cdot \rangle$ is the inner product on $\mathbb{R}^{N_x}$.

Next we write the state and adjoint equations fulfilled at a stationary point of the Lagrangian $\mathcal{L}$. In addition, we state the derivative of $\mathcal{L}$ with respect to the design variable $\boldsymbol{\theta}$.

(i) State Equation.

$$\begin{aligned} \phi_1 &= \sigma(W_0\boldsymbol{\xi} + \mathbf{b}_0), \\ \phi_j &= \phi_{j-1} - \sum_{k=0}^{j-2} a_{j-1-k} \ (\phi_{k+1} - \phi_k) + h^\gamma \Gamma(2-\gamma)\sigma(W_{j-1}\phi_{j-1} + \mathbf{b}_{j-1}), \\ &\qquad\qquad 2 \le j \le L-1, \\ \phi_L &= W_{L-1}\phi_{L-1}. \end{aligned} \quad (7.20a)$$

(ii) Adjoint Equation.

$$\psi_j = \psi_{j+1} + \sum_{k=j+1}^{L-2} a_{k-j} \ (\psi_{k+1} - \psi_k) - \qquad j = L-2, ..., 1$$

$$h^\gamma \Gamma(2-\gamma) \left[-W_j^T \left(\psi_{j+1} \odot \sigma' \left(W_j \phi_{j+1} + \mathbf{b}_j \right) \right) \right], \tag{7.20b}$$

$$\psi_{L-1} = -W_{L-1}^T \psi_L,$$

$$\psi_L = \partial_{\phi_L} \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}).$$

(iii) Derivative with respect to $\boldsymbol{\theta}$.

$$\partial_{W_{L-1}} \mathcal{L} = \psi_L \ \phi_{L-1}^T = \partial_{\phi_L} \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}) \ \phi_{L-1}^T + \partial_{W_{L-1}} \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}),$$

$$\partial_{W_j} \mathcal{L} = - \ \phi_j \ \left(\psi_{j+1} \odot \sigma'(W_j \phi_j + \mathbf{b}_j) \right)^T + \partial_{W_j} \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}),$$

$$j = 0, ..., L-2, \tag{7.20c}$$

$$\partial_{\mathbf{b}_j} \mathcal{L} = - \ \ \psi_{j+1}^T \ \sigma'(W_j \phi_j + \mathbf{b}_j) + \partial_{\mathbf{b}_j} \mathcal{J}(\boldsymbol{\theta}; \boldsymbol{\xi}, \mathbf{u}).$$

$$j = 0, ..., L-2 \, .$$

The right-hand-side of (7.20c) represents the gradient of $\mathcal{L}$ with respect to $\boldsymbol{\theta}$. We then use a gradient-based method (BFGS in our case) to identify $\boldsymbol{\theta}$.

7.3.3 Error Analysis

In this section we briefly address the question of how well the deep neural approximation approximates the PDE solution map $\Phi : \mathbb{R}^{N_{\boldsymbol{\xi}}} \to \mathbb{R}^{N_x}$. The approximation capabilities of neural networks has received a lot of attention recently in the literature, see e.g., [55, 89, 117, 138] and the references therein. The papers [89, 117] obtain results based on general activation functions for which the approximation rate is $\mathcal{O}(n^{-1/2})$, where n is the total number of hidden neurons. This implies that neural networks can overcome the curse of dimensionality, as the approximation rate is independent of dimension N_x.

We begin by making the observation that the fractional DNN can be written as a linear combination of activation functions evaluated at different layers. Indeed, observe from (7.17) that if $\Phi(\boldsymbol{\xi})$ is approximated by a one-hidden layer network $\hat{\Phi}(\boldsymbol{\xi}, \boldsymbol{\theta}) := \phi_2$ (that is, $L = 2$) then it can be expressed as:

$$\phi_2 = W_1 \sigma(W_0 \boldsymbol{\xi} + \mathbf{b}_0). \tag{7.21}$$

By a one-hidden layer network, we mean a network with the input layer, one hidden layer and the output layer [55, 89, 100, 117]. Next, for $L = 3$, we obtain

$$\phi_3 = W_2 \phi_2 = W_2 \left[\alpha_0 \sigma(W_0 \phi_0 + \mathbf{b}_0) + \alpha_1 \sigma(W_1 \phi_1 + \mathbf{b}_1) + \alpha_2 \phi_0 \right], \tag{7.22}$$

where $\alpha_0 = 1 - a_1,\ \alpha_1 = \tau := h^\gamma \Gamma(2-\gamma)$ and $\alpha_2 = a_1$. Similarly, if we set $L = 4$, then we get

$$\begin{aligned} \phi_4 &= W_3 \phi_3 \\ &= W_3 \left[\alpha_0 \sigma(W_0 \phi_0 + \mathbf{b}_0) + \alpha_1 \sigma(W_1 \phi_1 + \mathbf{b}_1) + \alpha_2 \sigma(W_2 \phi_2 + \mathbf{b}_2) + \alpha_3 \phi_0 \right], \end{aligned} \tag{7.23}$$

where $\alpha_0 = (1 - a_1 + a_1^2 - a_2),\ \alpha_1 = (1 - a_1)\tau,\ \alpha_2 = \tau$, and $\alpha_3 = 2a_2 - a_1 - a_1^2$.

Proceeding in a similar fashion yields the following result regarding multilayer fractional DNN.

Proposition 7.1 (Representation of multilayer fractional DNN)**.** For $L \geq 3$, the fractional DNN given by (7.17) fulfills

$$\phi_L = W_{L-1} \left[\alpha_{L-1} \phi_0 + \sum_{i=0}^{L-2} \alpha_i \sigma(W_i \phi_i + \mathbf{b}_i) \right], \tag{7.24}$$

where α_i are constants depending on τ and a_i, as given by (7.16). Observe that a one-hidden layer fDNN (that is, $L = 2$) coincides with a one-hidden layer standard DNN.

Error analysis for multilayer networks is generally challenging. Some papers that study this include [81,138] and the references therein. The results of these two papers focus mainly on the ReLU activation function. In particular, [81] discusses the approximation of linear finite elements by ReLU deep and shallow networks.

In this work, we restrict the analysis discussion to one-hidden layer fDNN i.e., $L = 2$. In particular, we consider a one-hidden layer network with finitely many neurons. Notice that for $L = 2$, fDNN coincides with standard DNN according to Proposition 7.1. Therefore, it is possible to extend the result from [117] to our case. To state the result from [117], we first introduce some notation.

To this end, we make the assumption that the function $\Phi(\boldsymbol{\xi})$ is defined on a bounded domain $\mathcal{P} \subset \mathbb{R}^{N_{\boldsymbol{\xi}}}$ and has a bounded Barron norm

$$||\Phi||_{\mathcal{B}^m} = \int_{\mathcal{P}} (1+\omega)^m |\hat{\Phi}(\omega)| \, d\omega, \;\; m \geq 0. \tag{7.25}$$

Now, let $m \in \mathbb{N} \cup \{0\}$ and $p \in [1, \infty]$. Recall that the Sobolev space $\mathcal{W}^{m,p}(\mathcal{P})$ is the space of functions in $L^p(\mathcal{P})$ whose weak derivatives of order m are also in $L^p(\mathcal{P})$:

$$\mathcal{W}^{m,p}(\mathcal{P}) := \{f \in L^p(\mathcal{P}) : D^{\mathbf{m}} f \in L^p(\mathcal{P}), \; \forall \mathbf{m} \text{ with } |\mathbf{m}| \leq m\}, \tag{7.26}$$

where $\mathbf{m} = (m_1, \ldots, m_{N_\xi}) \in \{0, 1, \ldots\}^{N_\xi}$, $|\mathbf{m}| = m_1 + m_2 \ldots + m_{N_\xi}$ and $D^{\mathbf{m}} f$ is the weak derivative. In particular, the space $\mathcal{H}^m(\mathcal{P}) := \mathcal{W}^{m,2}(\mathcal{P})$ is a Hilbert space with inner product and norm given respectively by

$$(f, g)_{\mathcal{H}^m(\mathcal{P})} = \sum_{|\mathbf{m}| \leq m} \int_{\mathcal{P}} D^{\mathbf{m}} f(\mathbf{x}) D^{\mathbf{m}} g(\mathbf{x}) \, d\mathbf{x},$$

and $||f||_{\mathcal{H}^m(\mathcal{P})} = (f,f)^{1/2}_{\mathcal{H}^m(\mathcal{P})}$. For $p = \infty$, we note that the Sobolev space $\mathcal{W}^{m,\infty}(\mathcal{P})$ is equipped with the norm

$$||f||_{\mathcal{W}^{m,\infty}(\mathcal{P})} = \max_{\mathbf{m}:|\mathbf{m}|\leq m} \operatorname*{ess\,sup}_{\mathbf{x}\in\mathcal{P}} |D^{\mathbf{m}} f(\mathbf{x})|. \tag{7.27}$$

Also, we say that $f \in \mathcal{W}^{m,p}_{loc}(\mathcal{P})$ if $f \in \mathcal{W}^{m,p}(\mathcal{P}')$, $\forall$ compact $\mathcal{P}' \subset \mathcal{P}$.

Next, since the mapping $\Phi : \mathbb{R}^{N_{\boldsymbol{\xi}}} \to \mathbb{R}^{N_x}$ can be computed using N_x mappings $\Phi_j : \mathbb{R}^{N_{\boldsymbol{\xi}}} \to \mathbb{R}$, it suffices to focus on networks with one output unit. Observe from (7.21) that

$$\hat{\Phi}(\boldsymbol{\xi},\boldsymbol{\theta}) := \phi_2 = \sum_{i=1}^{n} c_i \cdot \sigma(\omega_i \cdot \xi + b_i), \tag{7.28}$$

where n is the total number of neurons in the hidden layer, b_i an element of the bias vector, c_i and ω_i are rows of W_1 and W_0 respectively. Now, for a given activation function σ, define the set

$$\mathcal{F}^n_{N_{\boldsymbol{\xi}}}(\sigma) = \left\{ \sum_{i=1}^{n} c_i \cdot \sigma(\omega_i \cdot \xi + b_i), \quad \omega_i \in \mathbb{R}^{N_\xi},\ b_i \in \mathbb{R} \right\}. \tag{7.29}$$

We can now state the following result for any non-zero activation functions satisfying some regularity conditions [117, Corollary 1]. After this result, we will show that our activation function in (7.8) fulfills the required assumptions.

Theorem 7.1. Let $\mathcal{P}$ be a bounded domain and $\sigma \in \mathcal{W}^{m,\infty}_{loc}(\mathbb{R})$ be a non-zero activation function. Suppose there exists a function $\nu \in \mathcal{F}^q_1(\sigma)$ satisfying

$$|\nu^{(k)}(t)| \leq C_p(1+|t|)^{-p}, \ \ 0 \leq k \leq m, \ \ p > 1. \tag{7.30}$$

Then, for any $\Phi \in \mathcal{B}^m$, we have

$$\inf_{\hat{\Phi}_n \in \mathcal{F}^n_{N_\xi}(\sigma)} ||\Phi - \hat{\Phi}_n||_{\mathcal{H}^m(\mathcal{P})} \leq C(\sigma, m, p, \beta)|\mathcal{P}|^{1/2} q^{1/2} n^{-1/2} ||\Phi||_{\mathcal{B}^{m+1}}, \tag{7.31}$$

where $\beta = \operatorname{diam}(\mathcal{P})$ and C is a constant depending on σ, m, p and β.

Note that the bound in (7.31) is $\mathcal{O}(n^{-1/2})$. Here, the function $\nu(x)$ is a linear combination of the shifts and dilations of the activation $\sigma(x)$. Moreover, the activation function itself need not satisfy the decay (7.30). The theorem says that it suffices to find some function $\nu \in \mathcal{F}^q_1(\sigma)$ (cf. (7.29) with $n = q,\ N_\xi = 1$) for which (7.30) holds.

In this work, we are working with smooth ReLU (see (7.8)) as the activation function $\sigma(x)$ for which $m = 2$. One choice of ν for which (7.30) holds is $\nu \in \mathcal{F}^3_1(\sigma)$ given by

$$\nu(t) = \sigma(t+1) + \sigma(t-1) - 2\sigma(t). \tag{7.32}$$

It can be shown that, ν satisfies (7.30) with $C_p = 20$ and $p = 1.5$, that is,

$$|\nu^{(k)}(t)| \leq 20(1 + |t|)^{-1.5}, \ \ 0 \leq k \leq 2. \tag{7.33}$$

This can be verified from Figure 7.2.

In what follows, we discuss the application of our proposed fractional DNN to the solution of two statistical Bayesian inverse problems.

7.4 Application to Bayesian Inverse Problems

The inverse problem associated with the forward problem (7.1) essentially requires estimating a parameter vector $\boldsymbol{\xi} \in \mathbb{R}^{N_{\boldsymbol{\xi}}}$ given some observed noisy limited data $d \in \mathbb{R}^{N_x}$. In the Bayesian inference framework, the posterior probability density, $\pi(\boldsymbol{\xi}|d) : \mathbb{R}^{N_{\boldsymbol{\xi}}} \to \mathbb{R}$, solves the statistical inverse problem. In principle, $\pi(\boldsymbol{\xi}|d)$ encodes the uncertainty from the set of

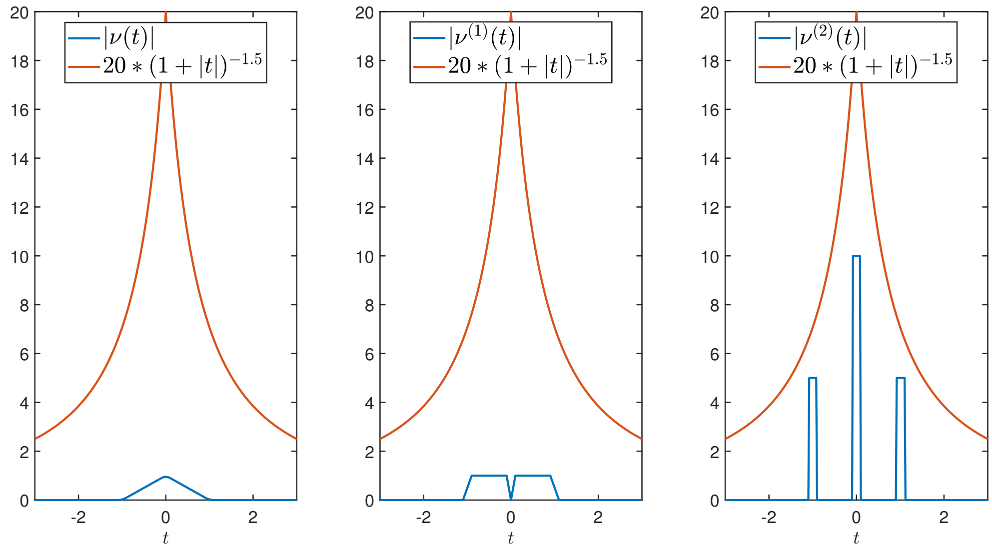

Figure 7.2: Verification of the decay condition eq. (7.30) for ν in eq. (7.32), which has been defined using the smooth ReLU in eq. (7.8).

observed data and the sought parameter vector. More formally, the posterior probability density is given by Bayes' rule as

$$\pi_{pos}(\boldsymbol{\xi}) := \pi(\boldsymbol{\xi}|d) \propto \pi(\boldsymbol{\xi})\pi(d|\boldsymbol{\xi}), \tag{7.34}$$

where $\pi(\cdot) : \mathbb{R}^{N_{\boldsymbol{\xi}}} \to \mathbb{R}$ is the prior and $\pi(d|\boldsymbol{\xi})$ is the likelihood function.

The standard approach to Bayesian inference uses the assumption that the observed data are of the form (see e.g., [37, 63])

$$d = \Phi(\boldsymbol{\xi}) + \eta, \tag{7.35}$$

where Φ is the *parameter-to-observable* map and $\eta \sim \mathcal{N}(0, \Sigma_\eta)$. In our numerical experiments, we assume that $\Sigma_\eta = \kappa^2 I$, where I is the identity matrix with appropriate dimensions and κ denotes the variance. Now, let the log-likelihood be given by

$$L(\boldsymbol{\xi}) = \frac{1}{2}\left|\left|\Sigma_\eta^{1/2}(d - \Phi(\boldsymbol{\xi}))\right|\right|^2; \tag{7.36}$$

then Bayes' rule in (7.34) becomes

$$\pi_{pos}(\boldsymbol{\xi}) := \pi(\boldsymbol{\xi}|d) \propto \pi(\boldsymbol{\xi})\pi(d|\boldsymbol{\xi}) = \frac{1}{Z}\exp\left(-L(\boldsymbol{\xi})\right)\pi(\boldsymbol{\xi}), \tag{7.37}$$

where $Z = \int_{\mathcal{P}} \exp(-L(\boldsymbol{\xi}))\pi(\boldsymbol{\xi})\, d\boldsymbol{\xi}$ is a normalizing constant.

We note here that, in practice, the posterior density rarely admits analytic expression. Indeed, it is generally evaluated using Markov chain Monte Carlo (MCMC) sampling techniques. In MCMC schemes one has to perform several thousands of forward model simulations. This is a significant computational challenge, especially if the forward model represents large-scale high fidelity discretization of a nonlinear PDE. It is therefore reasonable to construct a cheap-to-evaluate surrogate for the forward model in the offline stage for use in online computations in MCMC algorithms.

In what follows, we discuss MCMC schemes for sampling posterior probability distributions, as well as how our fractional DNN surrogate modeling strategy can be incorporated in them.

7.4.1 Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are very powerful and flexible numerical sampling approaches for approximating posterior distributions, see e.g., [26,37,120]. Prominent MCMC methods include Metropolis-Hastings algorithm (MH) [120], Gibbs algorithm [23] and Hamiltonian Monte Carlo algorithm (HMC) [26,33], Metropolis-adjusted Langevin algorithm (MALA) [26], and preconditioned Crank Nicolson algorithm (pCN) [23], and their variants. In our numerical experiments, we will use MH, pCN, HMC, and MALA algorithms. We will describe only MH and refer the reader to [26] for the details of the derivation and properties of the many variants of the other three algorithms.

To this end, consider a given parameter sample $\boldsymbol{\xi} = \boldsymbol{\xi}^i$. The MH algorithm generates a new sample $\boldsymbol{\xi}^{i+1}$ as follows:

(1) Generate a proposed sample $\boldsymbol{\xi}^*$ from a proposal density $q(\boldsymbol{\xi}^*|\boldsymbol{\xi})$, and compute $q(\boldsymbol{\xi}|\boldsymbol{\xi}^*)$.

(2) Compute the acceptance probability

$$\alpha(\boldsymbol{\xi}^*|\boldsymbol{\xi}) = \min\left\{1, \frac{\pi(\boldsymbol{\xi}^*|d)q(\boldsymbol{\xi}^*|\boldsymbol{\xi})}{\pi(\boldsymbol{\xi}|d)q(\boldsymbol{\xi}|\boldsymbol{\xi}^*)}\right\}. \tag{7.38}$$

(3) If Uniform(0; 1] $< \alpha(\boldsymbol{\xi}^*|\boldsymbol{\xi})$, then $\boldsymbol{\xi}^{i+1} = \boldsymbol{\xi}^*$. Else, set $\boldsymbol{\xi}^{i+1} = \boldsymbol{\xi}$.

Observe from (7.36) and (7.37) that each evaluation of the likelihood function, and hence, the acceptance probability (7.38) requires the evaluation of the forward model to compute $\pi(\boldsymbol{\xi}^*|d)$. In practice, one has to do tens of thousands of forward solves for the HM algorithm to converge. We propose to replace the forward solves by the fractional DNN surrogate model. In our numerical experiments, we follow [78] in which an adaptive Gaussian proposal is used:

$$q(\boldsymbol{\xi}^*|\boldsymbol{\xi}^i) = \exp\left(-\frac{1}{2}(\boldsymbol{\xi}^* - \boldsymbol{\xi}^{i-1})^T C_{i-1}(\boldsymbol{\xi}^* - \boldsymbol{\xi}^{i-1})\right), \tag{7.39}$$

where $C_0 = I$,

$$C_{i-1} = \frac{1}{i}\sum_{j=0}^{i-1}(\boldsymbol{\xi}^j - \bar{\boldsymbol{\xi}})(\boldsymbol{\xi}^j - \bar{\boldsymbol{\xi}})^T + \vartheta I,$$

$\bar{\boldsymbol{\xi}} = i^{-1}\sum_{j=0}^{i-1}\boldsymbol{\xi}^j$, and $\vartheta \approx 10^{-8}$. When the proposal is chosen adaptively as specified above, the HM method is referred to as an Adaptive Metropolis (AM) method [22, 78].

7.5 Numerical Experiments

In this section, we consider two statistical inverse problems. The first one is a diffusion-reaction problem in which two parameters need to be inferred [58]. The second one is a more challenging problem – a thermal fin problem from [26], which involves one hundred parameters to be identified. All experiments were performed using MATLAB R2020b on a

BFGS iterations	Error	Time (in sec)
400	2.26×10^{-2}	7.99
800	1.10×10^{-2}	15.10
1600	4.09×10^{-3}	29.45
3200	3.43×10^{-3}	56.14
6400	2.56×10^{-3}	105.66

Table 7.1: Diffusion-reaction problem: Number of BFGS iterations, relative errors and times for training the fractional DNN.

Mac desktop with RAM 32GB and CPU 3.6 GHz Quad-Core Intel Core i7.

In both of these experiments we train using a 3-hidden layer network with 15 neurons in each hidden layer (that is, $L = 4,\ n = 45$) and $k = 400$ neurons in the output layer, where k matches the dimension of POD basis as described in section 7.5.1 below. We chose $\varepsilon = 0.1$ in the smooth ReLU activation function, final time $T = 1$ and step-size $h = \frac{1}{3}$. We set the fractional exponent $\gamma = 0.5$ in the Caputo Fractional derivative and the regularization parameter $\lambda = 10^{-6}$. To train the network, we use the BFGS optimization method [96, Chapter 4].

Tables 7.1 and 7.2 show the number of BFGS iterations and the CPU times required to train the data from the diffusion-reaction problem and the thermal fin problem, respectively. Also shown in these tables are the relative errors obtained by evaluating the trained network at a parameter $\boldsymbol{\xi}^e$ not in the training set $\mathbb{E}$; here,

$$Error = \frac{||u(\boldsymbol{\xi}^e) - \hat{\Phi}(\boldsymbol{\xi}^e)||_\infty}{||u(\boldsymbol{\xi}^e)||_\infty},$$

where $u(\boldsymbol{\xi}^e)$ and $\hat{\Phi}(\boldsymbol{\xi}^e)$ are the true and surrogate solutions at $\boldsymbol{\xi}^e$, respectively. Note from both tables that after 1600 BFGS iterations, the decrease in errors is not significant. Hence, for all the experiments discussed below, we use 1600 BFGS iterations.

BFGS iterations	Error	Time (in sec)
400	1.87×10^{-2}	10.69
800	1.15×10^{-2}	18.51
1600	7.39×10^{-3}	33.52
3200	4.80×10^{-3}	62.90
6400	3.73×10^{-3}	117.55

Table 7.2: Thermal fin problem: Number of BFGS iterations, relative errors and times for training the fractional DNN.

7.5.1 Diffusion-Reaction Example

We consider the following nonlinear diffusion-reaction problem posed in a two-dimensional spatial domain [42, 72]

$$\begin{aligned} -\Delta u + g(u;\boldsymbol{\xi}) &= f, \quad \text{in } \Omega = (0,1)^2, \\ u &= 0, \quad \text{on } \partial\Omega, \end{aligned} \tag{7.40}$$

where $g(u;\boldsymbol{\xi}) = \frac{\boldsymbol{\xi}_2}{\boldsymbol{\xi}_1}\left[\exp(\boldsymbol{\xi}_1 u) - 1\right]$ and $f = 100\sin(2\pi x_1)\sin(2\pi x_2)$. Moreover, the parameters are $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \in [0.01, 10]^2 \subset \mathbb{R}^2$.

Equation (7.40) is discretized on a uniform mesh in Ω with 64 grid points in each direction using centered differences resulting in 4096 degrees of freedom. We obtained the solution of the resulting system of nonlinear equations using an inexact Newton-GMRES method as described in [94]. The stopping tolerance was 10^{-6}.

To train the network, we first computed $N_s = 900$ solution snapshots $\mathbb{S}$ corresponding to a set $\mathbb{E}$ of 900 parameters $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ drawn from the parameter space $[0.01, 10]^2$. These parameters were chosen using Latin hypercube sampling. Each solution snapshot is of dimension $N_x = 4096$. Next, we computed the SVD of the matrix of solution snapshots $\mathbb{S} = \widetilde{\mathbb{V}}\Sigma\mathbb{W}^T$, and set our POD basis $\mathbb{V} = \widetilde{\mathbb{V}}(:, 1 : k)$, where $k = 400$. Our training set then consisted of $\mathbb{E}$ as inputs and $\mathbb{V}^T\mathbb{S} \in \mathbb{R}^{k \times N_s}$ as our targets. As reported by Hesthaven and Ubbiali in [83] in the context of solving parameterized PDEs, the POD-DNN approach accelerates the online computations for surrogate models. Thus, we follow this approach in our numerical experiments; in particular, using k-dimensional data (where $k \ll N_x$) confirms an overall speed up in the solution of the statistical inverse problems.

Next, in both numerical examples considered here, we solved the inverse problems using $M = 20,000$ MCMC samples. In each case, the first $10,000$ samples were discarded for "burn-in" effects, and the remaining $10,000$ samples were used to obtain the reported statistics. Here, we used an initial Gaussian proposal (7.39) with covariance $C_0 = I$ and updated the covariance after every 100th sample. As in [58], we assume for this problem, that $\boldsymbol{\xi}$ is uniformly distributed over the parameter space $[0.01, 10]^2$. Hence, (7.37) becomes

$$\pi_{pos}(\boldsymbol{\xi}) \propto \begin{cases} \exp\left(-\frac{1}{2\kappa^2}(d-\Phi(\boldsymbol{\xi}))^T(d-\Phi(\boldsymbol{\xi}))\right) & \text{if } \boldsymbol{\xi} \in [0.01, 10]^2. \\ 0 & \text{otherwise.} \end{cases} \tag{7.41}$$

We generated the observations d by using (7.35) the true parameter to be identified $\boldsymbol{\xi}^e = (1, 0.1)$ and a Gaussian noise vector η with $\kappa = 10^{-2}$.

Figures 7.3, 7.4 and 7.5, represent, respectively, the histogram (which depicts the posterior distribution), the Markov chains and the autocorrelation functions corresponding to the parameters using the high fidelity model (Full) and the deep neural network surrogate (DNN) models in the MCMC algorithm. Recall that, for a Markov chain $\{\delta_j\}_{j=1}^{J}$ generated by the Metropolis-Hastings algorithm, with variance κ^2, the autocorrelation function (ACF) ϱ of the δ-chain is given by

$$\varrho(j) = \mathrm{cov}(\delta_1, \delta_{1+|j|})/\kappa^2.$$

and the integrated autocorrelation time, τ_{int}, (IACT) of the chain is defined as

$$\tau_{int}(\delta) := \sum_{j=-J+1}^{J-1} \varrho(j) \approx 1 + 2\sum_{j=1}^{J-1}\left(1 - \frac{j}{J}\right)\mathrm{cov}(\delta_1, \delta_{1+j})/\kappa^2. \tag{7.42}$$

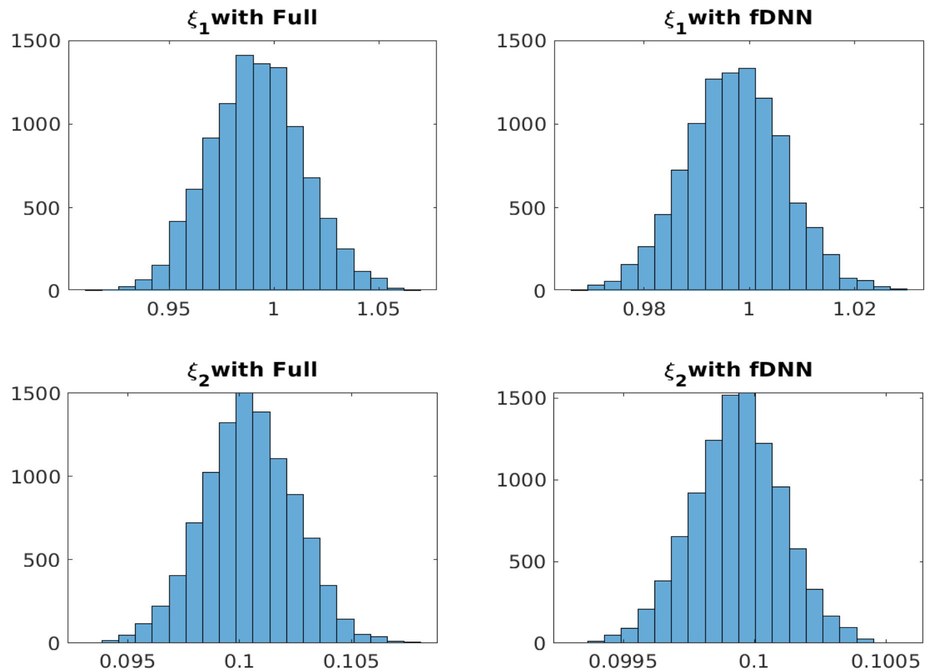

Figure 7.3: Histograms of the posterior distributions for the parameters $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$. They have been obtained from Full (left) and fDNN (right) models with $M = 10,000$ MCMC samples.

ACF decays very fast to zero as $J \to \infty$. In practice, J is often taken to be $\lfloor 10 \log_{10} M \rfloor$, [23]. If IACT $= K$, this means that the roughly every Kth sample of the δ chain is independent. We have used the following estimators to approximate $\varrho(j)$ and τ_{int} in our computations [22, 121]:

$$\varrho(j) := B(j)/B(0), \quad \text{and} \quad \tau_{int} := \sum_{j=-\widehat{J}}^{\widehat{J}} \varrho(j),$$

where $B(j) = \frac{1}{J-j} \sum_{k=1}^{J-j} (\delta_k - \bar{\delta})(\delta_{k+|j|} - \bar{\delta})$, $\bar{\delta}$ is the mean of the δ-chain, and $\widehat{J}$ is chosen be the smallest integer such that $\widehat{J} \geq 3\tau_{int}$.

Observe from Figs. 7.3 and 7.4 that, for both Full and fDNN models, the respective histograms and Markov chains are centered around the parameters of interest $(1, 0.1)$. In fact, for the full model, the 95% confidence intervals (CIs) for ξ_1 and ξ_2 are $[0.9496, 1.0492]$

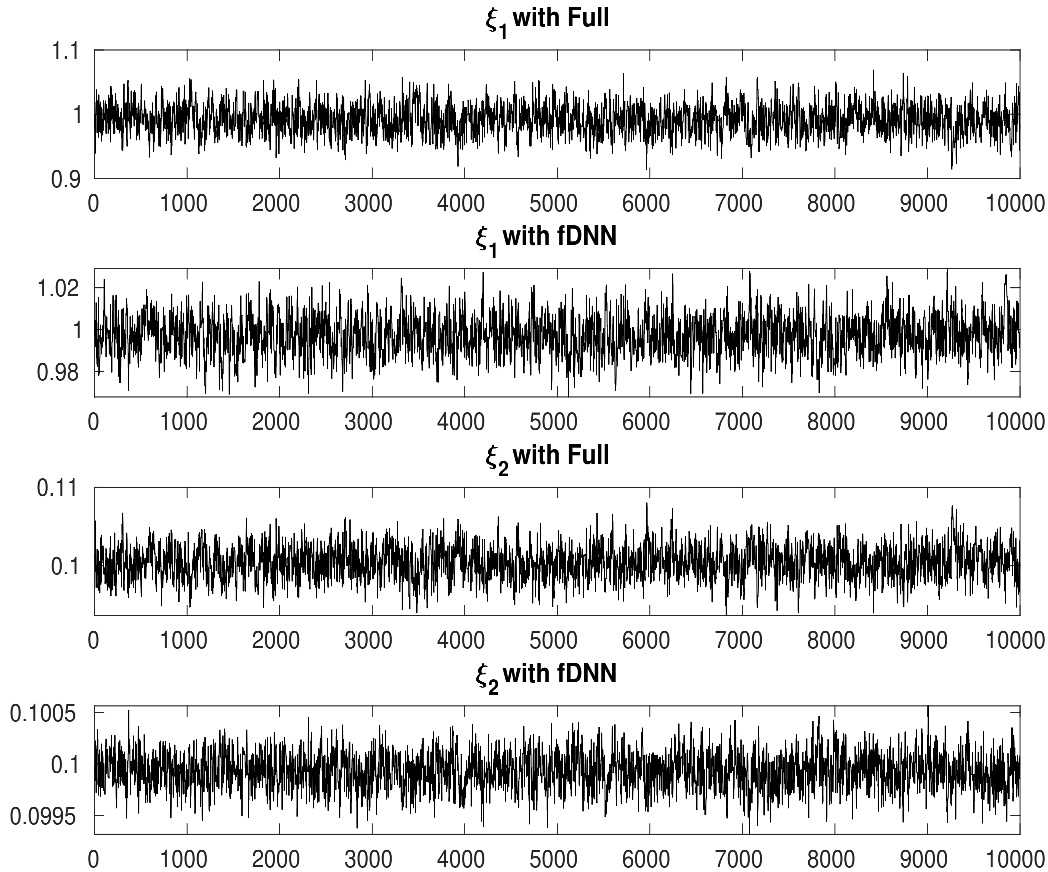

Figure 7.4: MCMC samples for the parameters $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ using Full (first and third) and fDNN (second and fourth) models. For the full model, the 95% confidence interval for ξ_1 and ξ_2 are $[0.9496, 1.0492]$ and $[0.0954, 0.1048]$. For the fDNN model, the 95% confidence interval for ξ_1 and ξ_2 are $[0.9818, 1.0196]$ and $[0.0995, 0.1005]$.

and $[0.0954, 0.1048]$. For the fDNN model, the CIs are $[0.9818, 1.0196]$ and $[0.0995, 0.1005]$. Thus, fDNN identifies $(1, 0.1)$ appropriately.

We also examine the impact of training accuracy in Fig 7.5. The figure shows that the ACFs decay very fast to zero, as expected. The value of IACT is roughly 10 for all the MCMC chains generated by the full model (left). The right of the figure shows the IACT for fDNN model, where the red-colored curves show the results obtained with 1600 training steps and the blue curves show those obtained with 6400 steps. In general, the IACT values imply that roughly every $10th$ sample of the MCMC chains generated by both the full and the fDNN models are independent.

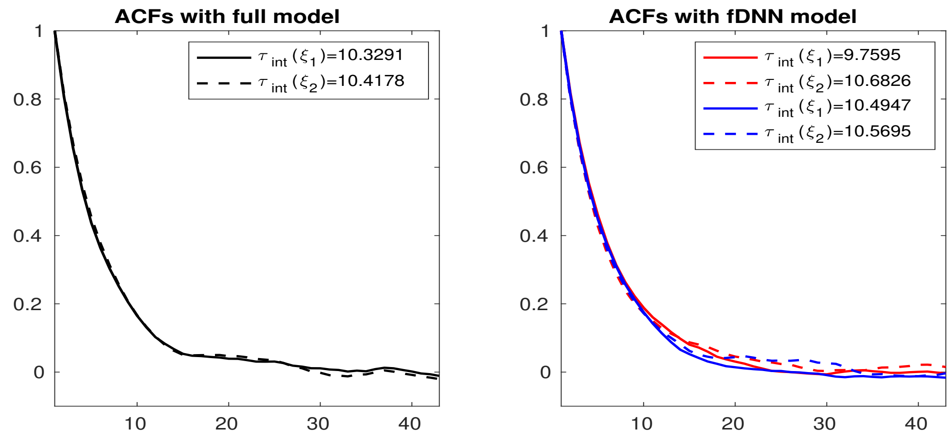

Figure 7.5: Autocorrelation functions (ACFs) for $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ chains computed with Full (left) and DNN (right) models. The red and blue lines are ACFs obtained with 1600 and 6400 BFGS iterations, respectively.

These results indicate that the fDNN surrogate model produces results with accuracy comparable to those obtained using the full model. The advantage of the surrogate lies in its reduced costs. In this example, using the HM algorithm, the full model required 2347.7 seconds of CPU time to solve the inverse problem, whereas the fractional DNN model required 29.3 seconds (the online costs for fDNN), a reduction in CPU time of a factor of 80. As always for surrogate approximations, there is overhead, the offline computations, associated with construction of the surrogate, i.e., identification of the parameter set $\boldsymbol{\theta}$ that defines the fDNN. For this example, this entailed construction of the 900 snapshot solutions used to produce targets for the training process, computation of the SVD of the matrix $\mathbb{S}$ of snapshots giving the targets, together with the training of the network (using BFGS). The times required for these computations were 127.8 seconds to construct the inputs/targets, 0.52 second to compute the SVD and 29.5 seconds to train the network. Thus, the offline and online times sum to 187.1 seconds, which is a lot smaller than the time needed for the full solution (that is, 2347.7 seconds). Of course, the offline computations represent a one-time expense which need not be repeated if more samples are used to perform an

MCMC simulation or if different MCMC algorithms are used.[1] This is, for instance, the case with the Differential Evolution Adaptive Metropolis (DREAM) method which runs multiple different chains simultaneously when used to sample the posterior distributions [131].

7.5.2 Thermal Fin Example

Next, we consider the following thermal fin problem from [26]:

$$
\begin{aligned}
-\mathrm{div}\,(e^{\boldsymbol{\xi}(\mathbf{x})}\nabla u) &= 0, && \text{in } \Omega, \\
(e^{\boldsymbol{\xi}(\mathbf{x})}\nabla u)\cdot\mathbf{n} + 0.1u &= 0, && \text{on } \partial\Omega\setminus\Gamma, \\
(e^{\boldsymbol{\xi}(\mathbf{x})}\nabla u(\mathbf{x}))\cdot\mathbf{n} &= 1, && \text{on } \Gamma = (-0.5, 0.5)\times\{0\}.
\end{aligned}
\tag{7.43}
$$

These equations (7.43) represent a forward model for heat conduction over the non-convex domain Ω as shown in Fig. 7.6. Given the heat conductivity function $e^{\boldsymbol{\xi}(\mathbf{x})}$, the forward problem (7.43) is used to compute the temperature u. The goal of this inverse problem is to infer 100 unknown parameters $\boldsymbol{\xi}$ from 262 noisy observations of u. Fig. 7.6 shows the location of the observations on the boundary $\partial\Omega\setminus\Gamma$, as well as the forward PDE solution u at the true parameter $\boldsymbol{\xi}$.

To train the network, we first computed, as in the diffusion-reaction case, $N_s = 900$ solution snapshots $\mathbb{S}$ corresponding to 900 parameters $\boldsymbol{\xi}\in\mathbb{R}^{100}$ drawn using Latin hypercube sampling. This problem is a lot more difficult than the previous problem in the sense that the dimension of the parameter space in this case is 100 rather than 2. Next, we compute the SVD of $\mathbb{S}$ and proceed as before.

In what follows, the infinite variants of Riemannian manifold Metropolis-adjusted

[1] We did not try to optimize the number of samples used for training and in particular $N_s = 900$ was an essentially arbitrary choice. The results obtained here with $N_s = 900$ were virtually the same as those found using 10,000 training samples, with the smaller number of samples incurring dramatically lower offline costs. Further reduction in computational time could be achieved using a smaller training data set. In [83], the authors used $\mathcal{O}(400)$ snapshots to train the feedforward network and still achieved good results.

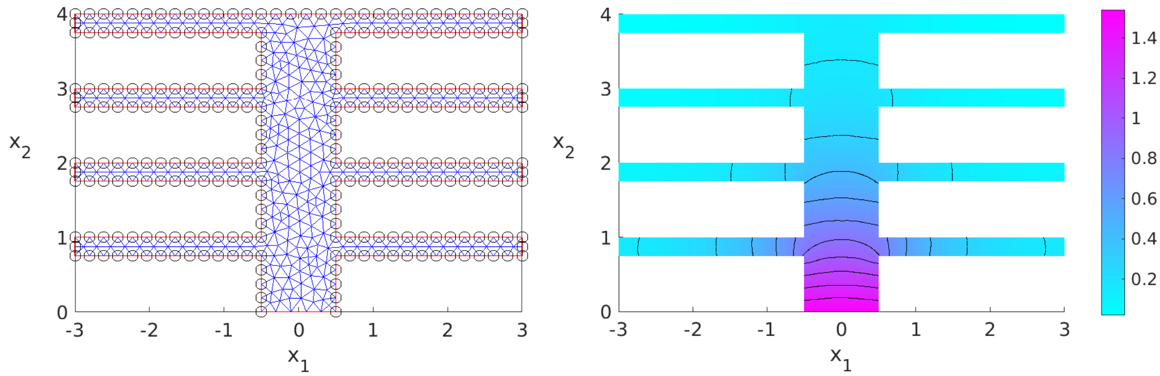

Figure 7.6: The location of observations (circles) (left) and the forward PDE solution u under the true parameter $\boldsymbol{\xi}$ (right).

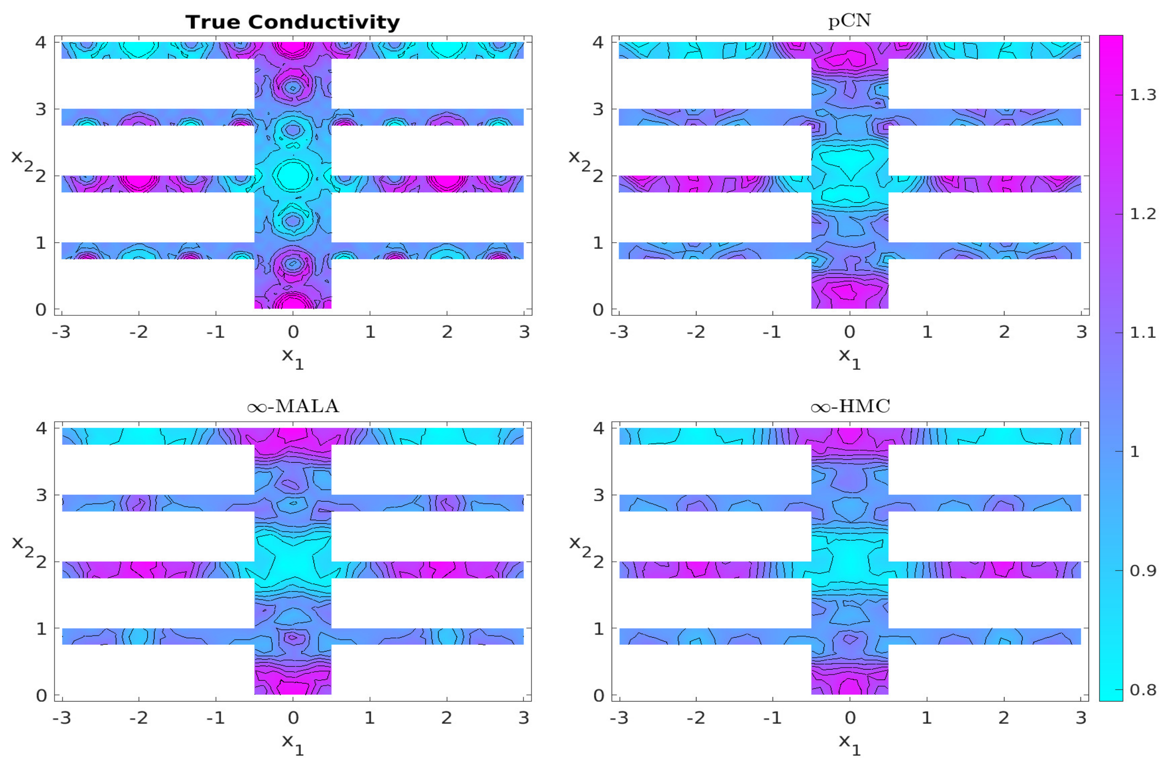

Figure 7.7: The true heat conductivity field $e^{\boldsymbol{\xi}(\mathbf{x})}$ (upper left) and the mean estimates of the posterior obtained by the different MCMC methods using fractional DNN as a surrogate model.

Langevin algorithm (MALA) and Hamiltonian Monte-Carlo (HMC) algorithms as presented in [26] are denoted, respectively, by ∞-MALA and ∞-HMC. In particular, Fig. 7.7 shows the true conductivity, as well as the posterior mean estimates obtained by three MCMC methods – pCN, ∞-MALA and ∞-HMC – using the fractional DNN as a surrogate model.

As in [26], for this problem we used a Gaussian prior defined on the domain $\mathcal{D} := [-3,3] \times [0,4]$ with covariance $\mathcal{C}$ of eigen-structure given by

$$\mathcal{C} = \sum_{i \in I} \mu_i^2 \{\varphi_i \otimes \varphi_i\},$$

where $\mu_i^2 = \{\pi^2((i_1 + 1/2)^2 + (i_2 + 1/2)^2)\}^{-1.2}$,

$$\varphi_i(\mathbf{x}) = 2|\mathcal{D}|^{-1/2} \cos(\pi(i_1 + 1/2)x_1) \cos(\pi(i_2 + 1/2)x_2),$$

and $I = \{i = (i_1, i_2), i_1 \geq 0,\ i_2 \geq 0\}$. Moreover, the log-conductivity $\boldsymbol{\xi}$ was the true parameter to be inferred with coordinates $\boldsymbol{\xi}_i = \mu_i \sin((i_1 - 1/2)^2 + (i_2 - 1/2)^2), \quad i_1 \leq 10,\ i_2 \leq 10$.

Table 7.3 shows the average acceptance rates for these models and the computational times required to perform the MCMC simulation for different variants of the MCMC algorithm, using both fDNN surrogate computations and full-order discrete PDE solution. These results indicate that the acceptance rates are comparable for the fDNN surrogate model and the full order forward discrete PDE solver. Moreover, there is about a 90% reduction in the computational times when fDNN surrogate solution is used, a clear advantage of the fDNN approach. The costs of the offline computations for fDNN were 63.2 seconds to generate the data used for the fractional DNN models and 33.5 seconds to train the network (with 1600 BFGS iterations) a total of 99.7 seconds. As in the case of the diffusion-reaction problem, the overhead required for the one-time offline computations is offset by the short CPU time required for the online phase with the MCMC algorithms. Note that the same offline computations were used for each of the three MCMC simulations tested.

Model	pCN	∞-MALA	∞-HMC
Acc. Rate (fDNN)	0.67	0.67	0.79
Acc. Rate (Full)	0.66	0.70	0.75
CPU time (fDNN)	16.46	98.0	228.4
CPU time (Full)	157.8	958.9	2585.3

Table 7.3: Acceptance rates and computational times needed to solve the inverse problem by pCN, ∞-MALA and ∞-HMC algorithms together with fDNN and full forward models.

7.6 Conclusions

This chapter has introduced a novel deep neural network (DNN) based approach to approximate nonlinear parameterized partial differential equations (PDEs). The proposed DNN helps learn the parameter to PDE solution operator. We have used this learnt solution operator to solve two challenging Bayesian inverse problems. The proposed approach is highly efficient without compromising on accuracy. We emphasize, that the proposed approach shows several advantages over the traditional surrogate approaches for parameterized PDEs such as reduced basis methods. For instance the proposed approach is fully non-intrusive and therefore it can be directly used in legacy codes, unlike the reduced basis method for nonlinear PDEs which can be highly intrusive.

Bibliography

[1] G. Acosta, F.M. Bersetche, and J.P. Borthagaray, *A short fe implementation for a 2d homogeneous dirichlet problem of a fractional laplacian*, Computers & Mathematics with Applications **74** (2017), no. 4, 784–816.

[2] Hans Wilhelm Alt, *Linear functional analysis*, Universitext, Springer-Verlag London, Ltd., London, 2016, An application-oriented introduction, Translated from the German edition by Robert Nürnberg. MR 3497775

[3] H. Antil and S. Bartels, *Spectral Approximation of Fractional PDEs in Image Processing and Phase Field Modeling*, Comput. Methods Appl. Math. **17** (2017), no. 4, 661–678. MR 3709055

[4] H. Antil, A. Drăgănescu, and K. Green, *A note on multigrid preconditioning for fractional pde-constrained optimization problems*, Results in Applied Mathematics **9** (2021), 100133.

[5] H. Antil, M. Heinkenschloss, and D. C. Sorensen, *Application of the discrete empirical interpolation method to reduced order modeling of nonlinear and parametric systems*, Reduced order methods for modeling and computational reduction, MS&A. Model. Simul. Appl., vol. 9, Springer, Cham, 2014, pp. 101–136. MR 3241209

[6] H. Antil, R. Khatri, and M. Warma, *External optimal control of nonlocal PDEs*, Inverse Problems **35** (2019), no. 8, 084003, 35. MR 3988258

[7] H. Antil, D.P. Kouri, M.-D. Lacasse, and D. Ridzal (eds.), *Frontiers in PDE-constrained optimization*, The IMA Volumes in Mathematics and its Applications, vol. 163, Springer, New York, 2018. MR 3839310

[8] H. Antil and D. Leykekhman, *A brief introduction to PDE-constrained optimization*, Frontiers in PDE-constrained optimization, IMA Vol. Math. Appl., vol. 163, Springer, New York, 2018, pp. 3–40. MR 3839730

[9] H. Antil, R.H. Nochetto, and P. Venegas, *Controlling the Kelvin force: basic strategies and applications to magnetic drug targeting*, Optim. Eng. **19** (2018), no. 3, 559–589. MR 3843442

[10] ———, *Optimizing the Kelvin force in a moving target subdomain*, Math. Models Methods Appl. Sci. **28** (2018), no. 1, 95–130. MR 3737079

[11] H. Antil, J. Pfefferer, and S. Rogovs, *Fractional operators with inhomogeneous boundary conditions: analysis, control, and discretization*, Commun. Math. Sci. **16** (2018), no. 5, 1395–1426. MR 3900347

[12] H. Antil, J. Pfefferer, and M. Warma, *A note on semilinear fractional elliptic equation: analysis and discretization*, ESAIM: M2AN **51** (2017), no. 6, 2049–2067.

[13] H. Antil and C.N. Rautenberg, *Sobolev spaces with non-Muckenhoupt weights, fractional elliptic operators, and applications*, SIAM J. Math. Anal. **51** (2019), no. 3, 2479–2503. MR 3961985

[14] H. Antil and M. Warma, *Optimal control of the coefficient for the regional fractional p-Laplace equation: approximation and convergence*, Math. Control Relat. Fields **9** (2019), no. 1, 1–38. MR 3924856

[15] Harbir Antil, Tyrus Berry, and John Harlim, *Fractional diffusion maps*, Applied and Computational Harmonic Analysis **54** (2021), 145–175.

[16] Harbir Antil, Zichao Wendy Di, and Ratna Khatri, *Bilevel optimization, deep learning and fractional Laplacian regularization with applications in tomography*, Inverse Problems **36** (2020), no. 6, 064001, 22. MR 4105331

[17] Harbir Antil, Ratna Khatri, Rainald Löhner, and Deepanshu Verma, *Fractional deep neural network via constrained optimization*, Machine Learning: Science and Technology **2** (2020), no. 1, 015003.

[18] Harbir Antil and Mahamadi Warma, *Optimal control of fractional semilinear PDEs*, ESAIM Control Optim. Calc. Var. **26** (2020), Paper No. 5, 30. MR 4055455

[19] W. Arendt, C. J. K. Batty, M. Hieber, and F. Neubrander, *Vector-valued Laplace transforms and Cauchy problems*, second ed., Monographs in Mathematics, vol. 96, Birkhäuser/Springer Basel AG, Basel, 2011. MR 2798103

[20] W. Arendt and R. Nittka, *Equivalent complete norms and positivity*, Arch. Math. (Basel) **92** (2009), no. 5, 414–427. MR 2506943

[21] H. Attouch, G. Buttazzo, and G. Michaille, *Variational analysis in Sobolev and BV spaces*, second ed., MOS-SIAM Series on Optimization, vol. 17, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2014, Applications to PDEs and optimization. MR 3288271

[22] J. M. Bardsley, *Computational Uncertainty Quantification for Inverse Problems*, SIAM, 2018.

[23] J. M. Bardsley, A. Solonen, H. Haario, and M. Laine, *Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems*, SIAM Journal on Scientific Computing **36(4)** (2014), A1895 – A1910.

[24] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, *An 'empirical interpolation' method: Application to efficient reduced-basis discretization of partial differential equations*, Comptes Rendus Mathematique **339** (2004), no. 9, 667 – 672.

[25] M. Benning, E. Celledoni, M. Ehrhardt, B. Owren, and C.-B. Schnlieb, *Deep learning as optimal control problems: Models and numerical methods*, J. of Comp. Dynamics **6** (2019), no. 1, 171 – 198.

[26] A. Beskos, M. Girolami, S. Lan, P. E. Farrell, and A. M. Stuart, *Geometric MCMC for infinite-dimensional inverse problems*, J. of Comp. Physics **335** (2017), 327 – 351.

[27] U. Biccari, M. Warma, and E. Zuazua, *Local regularity for fractional heat equations*, Recent Advances in PDEs: Analysis, Numerics and Control, Springer, 2018, pp. 233–249.

[28] C. Bjorland, L. Caffarelli, and A. Figalli, *Nonlocal tug-of-war and the infinity fractional Laplacian*, Comm. Pure Appl. Math. **65** (2012), no. 3, 337–380. MR 2868849

[29] K. Bogdan, K. Burdzy, and Z.-Q. Chen, *Censored stable processes*, Probab. Theory Related Fields **127** (2003), no. 1, 89–152. MR 2006232

[30] J. P. Borthagaray, D. Leykekhman, and R. H. Nochetto, *Local energy estimates for the fractional laplacian*, arXiv preprint arXiv:2005.03786 (2020), arXiv:2005.03786.

[31] J.P. Borthagaray and P. Ciarlet, Jr., *On the convergence in H^1-norm for the fractional Laplacian*, SIAM J. Numer. Anal. **57** (2019), no. 4, 1723–1743. MR 3984304

[32] L. Brasco, E. Parini, and M. Squassina, *Stability of variational eigenvalues for the fractional p-Laplacian*, Discrete Contin. Dyn. Syst. **36** (2016), no. 4, 1813–1845. MR 3411543

[33] T. Bui-Thanh and M. Girolami, *Solving large-scale PDE-constrained Bayesian inverse problems with Riemann manifold Hamiltonian Monte Carlo*, Inverse Problems **2014** (30), 114014.

[34] L. Caffarelli and L. Silvestre, *An extension problem related to the fractional Laplacian*, Comm. Partial Differential Equations **32** (2007), no. 7-9, 1245–1260. MR 2354493

[35] L.A. Caffarelli, J.-M. Roquejoffre, and Y. Sire, *Variational problems for free boundaries for the fractional Laplacian*, J. Eur. Math. Soc. (JEMS) **12** (2010), no. 5, 1151–1179. MR 2677613

[36] L.A. Caffarelli, S. Salsa, and L. Silvestre, *Regularity estimates for the solution and the free boundary of the obstacle problem for the fractional Laplacian*, Invent. Math. **171** (2008), no. 2, 425–461. MR 2367025

[37] D. Calvetti and E. Somersalo, *Introduction to Bayesian Scientific Computing*, Springer, 2007.

[38] E. Casas, *Control of an elliptic problem with pointwise state constraints*, SIAM J. Control Optim. **24** (1986), no. 6, 1309–1318. MR 861100

[39] ———, *Boundary control of semilinear elliptic equations with pointwise state constraints*, SIAM J. Control Optim. **31** (1993), no. 4, 993–1006. MR 1227543

[40] ______, *Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations*, SIAM J. Control Optim. **35** (1997), no. 4, 1297–1327. MR 1453300

[41] E. Casas, M. Mateos, and B. Vexler, *New regularity results and improved error estimates for optimal control problems with state constraints*, ESAIM Control Optim. Calc. Var. **20** (2014), no. 3, 803–822. MR 3264224

[42] S. Chaturantabut, *Nonlinear Model Reduction via Discrete Empirical Interpolation*, Ph.D. thesis, Rice University, Houston, 2011.

[43] S. Chaturantabut and D. C. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM Journal on Scientific Computing **32** (2010), 2737 – 2764.

[44] Y. Chen, L. Lu, G. E. Karniadakis, and L. D. Negro, *Physics-informed neural networks for inverse problems in nano-optics and metamaterials*, Optics Express **28** (2020), no. 8, 11618–11633.

[45] Burkhard Claus and Mahamadi Warma, *Realization of the fractional Laplacian with nonlocal exterior conditions via forms method*, J. Evol. Equ. **20** (2020), no. 4, 1597–1631. MR 4181960

[46] Rama Cont and Peter Tankov, *Financial modelling with jump processes*, Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2004. MR 2042661

[47] E.B. Davies, *Heat kernels and spectral theory*, Cambridge Tracts in Mathematics, vol. 92, Cambridge University Press, Cambridge, 1990. MR 1103113

[48] M. D'Elia and M. Gunzburger, *Optimal distributed control of nonlocal steady diffusion problems*, SIAM J. Control Optim. **52** (2014), no. 1, 243–273. MR 3158780

[49] E. Di Nezza, G. Palatucci, and E. Valdinoci, *Hitchhiker's to the fractional Sobolev spaces*, Bull. Sci. Math. **136** (2012), no. 5, 521–573. MR 2944369

[50] S. Dipierro, X. Ros-Oton, and E. Valdinoci, *Nonlocal problems with Neumann boundary conditions*, Rev. Mat. Iberoam. **33** (2017), no. 2, 377–416. MR 3651008

[51] Serena Dipierro, Giampiero Palatucci, and Enrico Valdinoci, *Dislocation dynamics in crystals: a macroscopic theory in a fractional Laplace setting*, Comm. Math. Phys. **333** (2015), no. 2, 1061–1105. MR 3296170

[52] G. Dore and A. Venni, *On the closedness of the sum of two closed operators*, Math. Z. **196** (1987), no. 2, 189–201. MR 910825

[53] Q. Du, M. Gunzburger, R.B. Lehoucq, and K. Zhou, *A nonlocal vector calculus, nonlocal volume-constrained problems, and nonlocal balance laws*, Math. Models Methods Appl. Sci. **23** (2013), no. 3, 493–540. MR 3010838

[54] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika **1** (1936), 211 – 218.

[55] S. Ellacott, *Aspects of the numerical analysis of neural networks*, Acta Numerica **3** (1994), 145 – 202.

[56] H. C. Elman and V. Forstall, *Preconditioning techniques for reduced basis methods for parameterized elliptic partial differential equations*, SIAM Journal on Scientific Computing **37(5)** (2015), S177 – S194.

[57] ———, *Numerical solution of the steady-state Navier-Stokes equations using empirical interpolation methods*, CMAME **317** (2017), 380 – 399.

[58] H. C. Elman and A. Onwunta, *Reduced-order modeling for nonlinear bayesian statistical inverse problems*, hhttps://arxiv.org/abs/1909.02539 (2019), 1 – 20.

[59] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite Elements and Fast Iterative Solvers*, vol. Second Edition, Oxford University Press, 2014.

[60] A Ern and J.-L. Guermond, *Theory and practice of finite elements*, Applied Mathematical Sciences, vol. 159, Springer-Verlag, New York, 2004. MR 2050138

[61] Gaetano Fichera, *Elastostatics problems with unilateral constraints: The Signorini problem with ambiguous boundary conditions*, Seminari 1962/63 Anal. Alg. Geom. e Topol., Vol. 2, Ist. Naz. Alta Mat, Ediz. Cremonese, Rome, 1965, pp. 613–679. MR 0191160

[62] A. Fiscella, R. Servadei, and E. Valdinoci, *Density properties for fractional Sobolev spaces*, Ann. Acad. Sci. Fenn. Math. **40** (2015), no. 1, 235–253. MR 3310082

[63] H. P. Flath, L. C. Wilcox, V. Akcelik, J. Hill, B. van Bloemen Waanders, and O. Ghattas, *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*, SIAM Journal on Scientific Computing **33** (2011), 407 – 432.

[64] V. Forstall, *Iterative Solution Methods for Reduced-order Models of Parameterized Partial Differential Equations*, Ph.D. thesis, University of Maryland, College Park, 2015.

[65] C. G. Gal and M. Warma, *Nonlocal transmission problems with fractional diffusion and boundary conditions on non-smooth interfaces*, Comm. Partial Differential Equations **42** (2017), no. 4, 579–625. MR 3642095

[66] Paolo Gatto and Jan S. Hesthaven, *Numerical approximation of the fractional Laplacian via hp-finite elements, with an application to image denoising*, J. Sci. Comput. **65** (2015), no. 1, 249–270. MR 3394445

[67] C. Geuzaine and J.-F. Remacle, *Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities*, International journal for numerical methods in engineering **79** (2009), no. 11, 1309–1331.

[68] Tuhin Ghosh, Mikko Salo, and Gunther Uhlmann, *The Calderón problem for the fractional Schrödinger equation*, Anal. PDE **13** (2020), no. 2, 455–475. MR 4078233

[69] G. H. Golub and C. H. Van Loan, *Matrix Computations*, JHU press, 1996.

[70] W. Gong, M. Hinze, and Z. Zhou, *Finite element method and a priori error estimates for Dirichlet boundary control problems governed by parabolic PDEs*, J. Sci. Comput. **66** (2016), no. 3, 941–967. MR 3456959

[71] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT Press, 2016.

[72] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, Mathematical Modelling and Numerical Analysis **41(3)** (2007), 575 – 605.

[73] P. Grisvard, *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics, vol. 24, Pitman (Advanced Publishing Program), Boston, MA, 1985. MR 775683

[74] G. Grubb, *Local and nonlocal boundary conditions for μ-transmission and fractional elliptic pseudodifferential operators*, Anal. PDE **7** (2014), no. 7, 1649–1682. MR 3293447

[75] ———, *Fractional Laplacians on domains, a development of Hörmander's theory of μ-transmission pseudodifferential operators*, Adv. Math. **268** (2015), 478–528. MR 3276603

[76] M. Gulian, M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Machine learning of space-fractional differential equations*, SIAM Journal on Scientific Computing **41** (2019), A2485 – A2509.

[77] S. Günther, L. Ruthotto, J. B. Schroder, E. C. Cyr, and N. R. Gauger, *Layer-parallel training of deep residual neural networks*, SIAM J. on Math. of Data Science **2** (2020), no. 1, 1 – 23.

[78] H. Haario, E. Saksman, and J. Tamminen, *An adaptive Metropolis algorithms*, Bernoulli **7** (2001), no. 2, 223 – 242.

[79] E. Haber and L. Ruthotto, *Stable architectures for deep neural networks*, Inverse Problems **34** (2017), no. 1, 014004.

[80] J. Han, A. Jentzen, and W. E., *Solving high-dimensional partial differential equations using deep learning*, PNAS **115** (2018), no. 34, 8505 – 8510.

[81] J. He, L. Li, J. Xu, and C. Zheng, *ReLU deep neural newtworks and linear finite elements*, Journal of Computational Mathematics **38** (2020), no. 3, 502.

[82] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[83] J. Hesthaven and S. Ubbiali, *Non-intrusive reduced order modeling of nonlinear problems using neural networks*, Journal of Computational Physics **363** (2018), 55 –78.

[84] J. S. Hesthaven, G. Rozza, and B. Stamm, *Certified reduced basis methods for parametrized partial differential equations*, SpringerBriefs in Mathematics, Springer, Bilbao, 2016. MR 3408061

[85] M. Hintermüller and M. Hinze, *Moreau-Yosida regularization in state constrained elliptic control problems: error estimates and parameter adjustment*, SIAM J. Numer. Anal. **47** (2009), no. 3, 1666–1683. MR 2505869

[86] M. Hintermüller and K. Kunisch, *Feasible and noninterior path-following in constrained minimization with low multiplier regularity*, SIAM J. Control Optim. **45** (2006), no. 4, 1198–1221. MR 2257219

[87] M. Hinze, *A variational discretization concept in control constrained optimization: the linear-quadratic case*, Comput. Optim. Appl. **30** (2005), no. 1, 45–61. MR 2122182

[88] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE constraints*, Mathematical Modelling: Theory and Applications, vol. 23, Springer, New York, 2009. MR 2516528

[89] K. Hornik, M. Stinchcombe, H. White, and P. Auer, *Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives*, Neural Computation **6** (1994), no. 6, 1262 – 1275.

[90] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, *Densely connected convolutional networks*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[91] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach spaces. Vol. I. Martingales and Littlewood-Paley theory*, Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics, vol. 63, Springer, Cham, 2016. MR 3617205

[92] K. Ito and K. Kunisch, *Lagrange multiplier approach to variational problems and applications*, Advances in Design and Control, vol. 15, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008. MR 2441683

[93] J. Kaipio and E. Somersalo, *Statistical and Computional Inverse Problems*, Springer, 2005.

[94] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, 1995.

[95] C. T. Kelley, *Iterative methods for optimization*, Frontiers in Applied Mathematics, vol. 18, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999. MR 1678201

[96] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, 1999.

[97] D. Kinderlehrer and G. Stampacchia, *An introduction to variational inequalities and their applications*, Academic Press, New York, 1980.

[98] Mateusz Kwaśnicki, *Ten equivalent definitions of the fractional Laplace operator*, Fract. Calc. Appl. Anal. **20** (2017), no. 1, 7–51. MR 3613319

[99] T. Leonori, I. Peral, A. Primo, and F. Soria, *Basic estimates for solutions of a class of nonlocal elliptic and parabolic equations*, Discrete Contin. Dyn. Syst. **35** (2015), no. 12, 6031–6068. MR 3393266

[100] M. Leshno, V.Y. Lin, A. Pinkus, and S. Schocken, *Multilayer feedforward networks with a nonpolynomial activation function can approximate any function*, Neural networks **6** (1993), no. 6, 861–867.

[101] J. Li and Y. M Marzouk, *Adaptive construction of surrogates for the bayesian solution of inverse problems*, SIAM Journal on Scientific Computing **36** (2014), no. 3, A1163 – A1186.

[102] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, *Fourier neural operator for parametric partial differential equations*, arXiv preprint arXiv:2010.08895 (2020), arXiv:2010.08895.

[103] C. Louis-Rose and M. Warma, *Approximate controllability from the exterior of space-time fractional wave equations*, Applied Mathematics & Optimization (2018), 1–44.

[104] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, *A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion*, SIAM Journal on Scientific Computing **34** (2012), A1460 – A1487.

[105] I. Neitzel and F. Tröltzsch, *On regularization methods for the numerical solution of parabolic control problems with pointwise state constraints*, ESAIM Control Optim. Calc. Var. **15** (2009), no. 2, 426–453. MR 2513093

[106] R. Nittka, *Inhomogeneous parabolic Neumann problems*, Czechoslovak Math. J. **64(139)** (2014), no. 3, 703–742. MR 3298555

[107] Xavier Ros Oton, *Integro-differential equations : regularity theory and pohozaev identities*, Ph.D. thesis, Universitat Politécnica de Catalunya, 2014.

[108] El-M. Ouhabaz, *Analysis of heat equations on domains*, London Mathematical Society Monographs Series, vol. 31, Princeton University Press, Princeton, NJ, 2005. MR 2124040

[109] G. Pang, L. Lu, and G. E. Karniadakis, *Fpinns: Fractional physics-informed neural networks*, SIAM Journal on Scientific Computing **41** (2019), A2603 – A2626.

[110] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced basis methods for partial differential equations: an introduction*, vol. 92, Springer, 2015.

[111] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, Journal of Compuational Physics **378** (2019), 686 – 707.

[112] X. Ros-Oton and J. Serra, *The Dirichlet problem for the fractional Laplacian: regularity up to the boundary*, J. Math. Pures Appl. (9) **101** (2014), no. 3, 275–302. MR 3168912

[113] ———, *The extremal solution for the fractional Laplacian*, Calc. Var. Partial Differential Equations **50** (2014), no. 3-4, 723–750. MR 3216831

[114] A. Rösch, *Error estimates for linear-quadratic control problems with control constraints*, Optim. Methods Softw. **21** (2006), no. 1, 121–134. MR 2192592

[115] A. Rösch and R. Simon, *Superconvergence properties for optimal control problems discretized by piecewise linear and discontinuous functions*, Numer. Funct. Anal. Optim. **28** (2007), no. 3-4, 425–443. MR 2311374

[116] R. Servadei and E. Valdinoci, *On the spectrum of two different fractional operators*, Proc. Roy. Soc. Edinburgh Sect. A **144** (2014), no. 4, 831–855. MR 3233760

[117] J. W. Siegel and Jinchao Xu, *Approximation rates for neural networks with general activation functions*, Neural Networks (2020), 313 – 321.

[118] Luis Silvestre, *Regularity of the obstacle problem for a fractional power of the Laplace operator*, Comm. Pure Appl. Math. **60** (2007), no. 1, 67–112. MR 2270163

[119] Luis Enrique Silvestre, *Regularity of the obstacle problem for a fractional power of the Laplace operator*, ProQuest LLC, Ann Arbor, MI, 2005, Thesis (Ph.D.)–The University of Texas at Austin. MR 2707618

[120] R. C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

[121] A. Sokal, *Monte Carlo methods in statistical mechanics: Foundations and new algorithms*, Functional Integration (C. DeWitt-Morette, P. Cartier, and A. Folacci, eds.), Springer, 1997, pp. 131 – 192.

[122] Renming Song and Zoran Vondraček, *Potential theory of subordinate killed brownian motion in a domain*, Probability Theory and Related Fields **125** (2003), no. 4, 578–592.

[123] Pablo Raúl Stinga, *User's guide to the fractional Laplacian and the method of semigroups*, Handbook of fractional calculus with applications. Vol. 2, De Gruyter, Berlin, 2019, pp. 235–265. MR 3965397

[124] R. K. Tripathy and I. Bilionis, *Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification*, JCP **375** (2018), 565 – 588.

[125] F. Tröltzsch, *Optimal control of partial differential equations*, Graduate Studies in Mathematics, vol. 112, American Mathematical Society, Providence, RI, 2010, Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels. MR 2583281

[126] E. Valdinoci, *From the long jump random walk to the fractional Laplacian*, Bol. Soc. Esp. Mat. Apl. SeMA (2009), no. 49, 33–44. MR 2584076

[127] Juan Luis Vázquez, *Nonlinear diffusion with fractional Laplacian operators*, Nonlinear partial differential equations, Abel Symp., vol. 7, Springer, Heidelberg, 2012, pp. 271–298. MR 3289370

[128] A. Veit, M. J. Wilber, and S. Belongie, *Residual networks behave like ensembles of relatively shallow networks*, Advances in neural information processing systems, 2016, pp. 550–558.

[129] M. I. Višik and G. I. Èskin, *Convolution equations in a bounded region*, Uspehi Mat. Nauk **20** (1965), no. 3 (123), 89–152. MR 0185273

[130] M.I. Višik and G.I. Èskin, *Convolution equations in a bounded region*, Uspehi Mat. Nauk **20** (1965), no. 3 (123), 89–152. MR 0185273

[131] J. A. Vrugt, C. J. F. Ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon, *Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling*, International Journal of Nonlinear Sciences and Numerical Simulation **10** (2009), no. 3, 273 – 290.

[132] M. Warma, *A fractional Dirichlet-to-Neumann operator on bounded Lipschitz domains*, Commun. Pure Appl. Anal. **14** (2015), no. 5, 2043–2067. MR 3359558

[133] ———, *The fractional relative capacity and the fractional Laplacian with Neumann and Robin boundary conditions on open sets*, Potential Anal. **42** (2015), no. 2, 499–547. MR 3306694

[134] ———, *Approximate controllability from the exterior of space-time fractional diffusive equations*, SIAM J. Control Optim. **57** (2019), no. 3, 2037–2063. MR 3968237

[135] C. J. Weiss, B.G. van Bloemen Waanders, and H. Antil, *Fractional operators applied to geophysical electromagnetics*, Geophysical Journal International **220** (2020), no. 2, 1242–1259.

[136] Wojbor A. Woyczyński, *Lévy processes in the physical sciences*, Lévy processes, Birkhäuser Boston, Boston, MA, 2001, pp. 241–266. MR 1833700

[137] L. Yang, X. Meng, and G. E. Karniadakis, *B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data*, Journal of Computational Physics **425** (2021), 109913.

[138] D. Yarotsky, *Error bounds for approximations with deep ReLU networks*, Neural Networks **94** (2017), 103 – 114.

[139] Y. Zhu and N. Zabaras, *Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification*, J. of Comp. Physics **366** (2018), 415 – 447.

Curriculum Vitae

Deepanshu Verma received his Bachelor of Science (Hons.) in Mathematics from S.G.T.B. Khalsa College, Delhi University in 2015. He received his Master of Science in Mathematics from Indian Institute of Technology Bombay in 2018.