A Wavelet-Based Approach for the Optimal Control of Non-Local Operator Equations

S. Dahlke, H. Harbrecht, T. M. Surowiec

July 6, 2020

Abstract

The optimal control of partial differential equations (PDEs) driven by non-local operators presents many numerical challenges. In contrast to the existing methods available in the literature, we propose a wavelet-based approach. This allows us to directly treat the non-local operators without the need to extend the underlying PDE into a higher spatial dimension. Due to their possessing vanishing moments, wavelets offer efficient compression strategies that lead to O(N)-algorithms for the forward equation, where N is the number of degrees of freedom. These computational advantages carry over to the solution of the class of control problems under consideration. The latter are equivalent to a coupled system of nonsmooth operator equations with non-local operators.

Keywords: Wavelets, PDE-Constrained Optimization, Bound Constraints, Fractional Laplacian, Non-local Operators, Semismooth Newton
AMS MSC: 65T60, 65K10, 65K15, 49J20, 65.49, 90C06

1 Introduction

The purpose of this study is to propose a wavelet-based approach for the optimal control of a class of non-local equations. In doing so, we do not strive for the highest form of abstraction, rather we seek to demonstrate the viability of the approach as an alternative to what can be found in the literature.

We start by briefly introducing the non-local operator equation by following the notation and definitions employed in [13, 14]. Note also that we employ the terminology of optimal control (control, state, forward problem, etc.) as is often done in the literature, despite the fact that the non-local equation under consideration is time-independent. To this aim, let $\Omega \subset \mathbb{R}^n$, $n \geq 1$, be an open and bounded domain and define the operator $\mathcal{L}$ for some $u : \Omega \to \mathbb{R}$ by

$$(\mathcal{L}u)(x) := 2 \int_{\mathbb{R}^n} \big(u(y) - u(x)\big)\kappa(x, y)\mathrm{d}y, \quad x \in \Omega.$$

Here, the kernel $\kappa : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a non-negative, symmetric mapping, i.e., $\kappa \geq 0$ and $\kappa(x, y) = \kappa(y, x)$. As noted in [13], a special case of $\mathcal{L}$ is the fractional Laplacian operator $(-\Delta)^s$. The non-local equation of interest in our study is given by

$$\begin{aligned} -\mathcal{L}u &= Bz + f \qquad && \text{on } \Omega, \\ u &= 0 && \text{on } \Omega_{\mathcal{I}}, \end{aligned} \tag{1.1}$$

where $f : \Omega \to \mathbb{R}$ is a fixed forcing term, $z \in Z$ is a decision variable, e.g., control, and B is a bounded linear operator that maps z into the non-local equation. The set $\Omega_{\mathcal{I}} \subset \mathbb{R}^n$ is known as the "interaction volume" and is assumed to be disjoint from Ω. In particular, it might be also a closed set. We will specify more details about the underlying spaces and necessary assumptions on κ below.

One concrete message of the paper is that wavelet methods provide important computational advantages compared to more classical discretization schemes such as finite elements or spectral methods for (1.1). Indeed, in the finite element setting, the non-locality gives rise to densely populated stiffness matrices unless the underlying domain is extended into a higher dimension on an infinite cylinder; see the seminal papers on the FEM for fractional diffusion [7, 25], which build on the lifting approach suggested in [6]. Spectral methods can be theoretically used to treat the fractional Laplacian, but require full knowledge of the spectrum of the Laplacian and thus rule out nontrivial domains. Furthermore, they are inadequate in a control setting whenever bound constraints are present. In contrast, wavelets are oscillating functions with local supports and vanishing moments. They allow not only a direct treatment of the non-local operators without increasing the underlying dimension, but also, the vanishing moment property can be used to devise very efficient compression and preconditioning strategies. This leads to sparse well-conditioned stiffness matrices. Recently, it has also been demonstrated that certain wavelet bases can be used for pointwise inequality constraints, see [8, 17], thus rendering them applicable to bound constrained optimal control problems.

Treating (1.1) as the forward problem, we consider the following optimization problem:

$$\begin{aligned} \inf \ & \frac{1}{2}\|Cu - u_d\|_H^2 + \frac{\nu}{2}\|z\|_Z^2 \text{ over } (z, u) \in Z_{\text{ad}} \times V \\ \text{s.t.} \quad & -\mathcal{L}u = Bz + f \quad \text{on } \Omega, \\ & \quad\ u = 0 \qquad\qquad \text{on } \Omega_{\mathcal{I}}. \end{aligned} \tag{1.2}$$

Here, H and Z are a real Hilbert spaces, $Z_{\text{ad}} \subset Z$ is a nonempty, closed, and convex set, $\nu > 0$, and C is a bounded linear operator whose image represents the observation of the state u. The state space V will be introduced below.

Despite exhibiting a familiar structure to classical optimal control of elliptic PDEs, the presence of the nonlocal operator is a major source of computational challenges. While theoretical issues such as existence and uniqueness of solutions as well as optimality conditions follow directly from the standard theory, the numerical solution of (1.2) is nontrivial. We refer the reader to [2, 3] for an approach using FEM.

It is worth noting here that we could easily consider more general objective functionals without major difficulties. On the other hand, introducing state constraints of the type

$$u(x) \geq \psi(x), \quad x \in \Omega, \tag{1.3}$$

for some sufficiently regular $\psi : \Omega \to \mathbb{R}$, would represent a major challenge. This is due to the fact that the solutions u will rarely satisfy the necessary regularity requirements needed to prove the existence of Lagrange multipliers. This lack of regularity would also significantly complicate standard relaxation-based approaches, where (1.3) is removed as a constraint and the term

$$\frac{\alpha}{2} \int_{\Omega} \max\{0, \psi(x) - u(x)\}_+^2 \mathrm{d}x$$

is added to the objective. Indeed, the convergence of these schemes (as $\alpha \uparrow \infty$) requires a uniform bound on the approximating adjoint states. However, the latter also requires a regularity condition. Despite this, there has been some major progress in this area when $\mathcal{L}$ is the fractional Laplacian [1,4].

Returning to the non-local equation, we now specify the assumptions on the kernel and define the state space V. Let $s \in (0,1)$ be a fixed real number and $\sigma : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ a function for which there exist constants $\gamma_1, \gamma_2 > 0$ such that

$$\gamma_1 \leq \sigma(x,y) \leq \gamma_2 \text{ for all } (x,y) \in \mathbb{R}^n \times \mathbb{R}^n$$

and

$$\kappa(x,y) := \frac{\sigma(x,y)}{|y-x|^{n+2s}} \tag{1.4}$$

is symmetric. Note that the symmetry properties imply that $\mathcal{L}$ itself is a symmetric operator. A much broader class of kernels is possible, cf. [13, 14], however, this will suffice for our discussion. Following [13, 14], we assume that $\Omega, \Omega_{\mathcal{I}}$, and $\Omega \cup \Omega_{\mathcal{I}}$ are bounded with piecewise smooth boundaries and satisfy the interior cone condition. Next, we define the energy norm

$$|||v||| := \left(\frac{1}{2} \int_{\Omega \cup \Omega_{\mathcal{I}}} \int_{\Omega \cup \Omega_{\mathcal{I}}} \big(u(y) - u(x)\big) \kappa(x,y) \big(u(y) - u(x)\big) \mathrm{d}y \mathrm{d}x \right)^{1/2}.$$

As demonstrated in [14], this is an equivalent norm on the Sobolev-Slobodeckij space $H^s(\Omega \cup \Omega_{\mathcal{I}})$. Consequently, the space

$$V := \{v \in H^s(\Omega \cup \Omega_{\mathcal{I}}) : v = 0 \text{ on } \Omega_{\mathcal{I}}\}$$

is a Hilbert space when equipped with the energy norm $|||\cdot|||$. In order to highlight the dependence on the kernel κ, we denote the inner product on V by:

$$(u,v)_\kappa := \frac{1}{2} \int_{\Omega \cup \Omega_{\mathcal{I}}} \int_{\Omega \cup \Omega_{\mathcal{I}}} \big(u(y) - u(x)\big) \kappa(x,y) \big(v(y) - v(x)\big) \mathrm{d}y \mathrm{d}x$$

and set

$$\|u\|_\kappa := |||u||| = \sqrt{(u,u)_\kappa}.$$

These facts allow us to more concretely define what is meant by a solution to the forward problem. In fact, the proofs of existence and uniqueness of optimal controls $z^\star$ as well as the derivation of optimality conditions reduces to the classical approach as detailed in the seminal work [24]. We collect these results in Theorem 1 below. For readability, we use a slight abuse of notation and let $\mathcal{L}_\kappa$ be the uniformly V-elliptic bounded linear operator defined by the inner product $(\cdot,\cdot)_\kappa$, i.e., we let $\mathcal{L}_\kappa : V \to V^*$ (the topological dual for V) be defined by

$$\langle -\mathcal{L}_\kappa u, v\rangle = (u,v)_\kappa.$$

It then follows from the Lax-Milgram lemma that the solution operator for (1.1) is given by

$$u = S(z) := (-\mathcal{L}_\kappa)^{-1}(Bz + f), \quad z \in Z.$$

Clearly, $S : Z \to V$ is continuously Fréchet differentiable and affine in z.

Theorem 1. Under the standing assumptions, the optimal control problem (1.2) admits a unique solution $z^\star \in Z_{\mathrm{ad}}$. Furthermore, there exists an adjoint state $\lambda^\star \in V$ such that

$$-\mathcal{L}_\kappa u^\star = B\mathfrak{P}\Big(-\frac{1}{\nu}B^*\lambda^\star\Big), \tag{1.5a}$$
$$-\mathcal{L}_\kappa \lambda^\star = C^*(u_d - Cu^\star). \tag{1.5b}$$

Here, $\mathfrak{P} : Z \to Z_{\mathrm{ad}}$ is the usual metric projection onto the closed convex set Z_{ad}.

Proof. Cf. [24, Chap. 2]. □

Note that in some cases there are potentially nontrivial Riesz mappings involved with the H- and Z-norms, e.g., if H involves either a trace or differential operator. If this were the case, then these must be included in the optimality system (1.5). In particular, if Λ_H is the canonical isomorphism of H into H^* and and Λ_Z is the canonical isomorphism of Z into Z^*, then (1.5) becomes

$$-\mathcal{L}_\kappa u^\star = B\mathfrak{P}\Big(-\frac{1}{\nu}\Lambda_Z^{-1}B^*\lambda^\star\Big), \tag{1.6a}$$
$$-\mathcal{L}_\kappa \lambda^\star = C^*\Lambda_H(u_d - Cu^\star). \tag{1.6b}$$

Taking these arguments into consideration, we see that the solution of (1.5) reduces to the solution of a coupled set of smooth and nonsmooth non-local equations. We thus turn our focus to the development of function-space algorithms, e.g., a generalized Newton solver in this non-local setting, and an efficient numerical discretization. To the best of our knowledge our study is the first of its kind for a system of nonsmooth, nonlocal equations.

We will assume in the sequel that the interaction volume $\Omega_{\mathcal{I}}$ encloses the domain Ω such that $\Omega_{\mathcal{I}} \cup \Omega$ is a simply connected domain and $\Omega \Subset \Omega \cup \Omega_{\mathcal{I}}$, see also Figure 1 for an illustration. Consequently, we deduce that

$$V \cong H^s(\Omega)/\mathbb{R} \quad \text{for } 0 < s < \frac{1}{2}$$

and

$$V \cong H_0^s(\Omega) \text{ for } \quad \frac{1}{2} < s < 1.$$

In the limit case $s = 1/2$, it holds $V \cong H_{00}^{1/2}(\Omega)$, see [6].

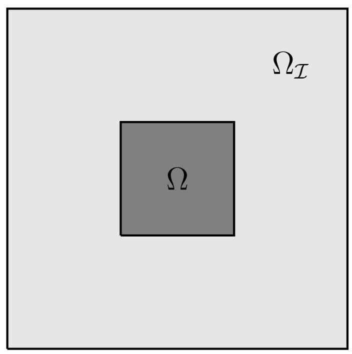

Figure 1: The domains Ω and $\Omega_{\mathcal{I}}$.

The rest of the paper is organized as follows. In Section 2 we briefly recall the basic concepts from wavelet analysis as far as it is needed for our purpose. Following this, we discuss the crucial components of wavelet matrix compression in Section 3, which are exploited in our numerical method. Section 4 contains the numerical results of the proposed method for an unconstrained ($Z_{\text{ad}} = L^2(\Omega \cup \Omega_{\mathcal{I}})$ and pointwise bilateral constraints ($Z_{\text{ad}} \subsetneq L^2(\Omega \cup \Omega_{\mathcal{I}})$). These numerical experiments clearly demonstrate the computational advantages and applicability of the approach.

2 Wavelets and Multiresolution Analysis

The construction of wavelet bases starts with a so-called multiresolution analysis. In general, a multiresolution analysis is a sequence of hierarchical trial spaces

$$\{0\} = V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset V_j \subset L^2(\Omega \cup \Omega_{\mathcal{I}}), \tag{2.1}$$

each spanned by trial functions:

$$V_j = \operatorname{span} \Phi_j, \quad \text{where} \quad \Phi_j = \{\varphi_{j,\mathbf{k}} : \mathbf{k} \in \Delta_j\}.$$

Here, Δ_j is an index set for the single-scale basis of the space V_j with cardinality $|\Delta_j| \sim 2^{jn}$. A final requirement is that the bases Φ_j are uniformly stable, i.e., the function $v_j = \sum_{\mathbf{k}\in\Delta_j} v_{\mathbf{k}}\varphi_{j,\mathbf{k}} \in V_j$ satisfies

$$\|v_j\|^2_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \sim \sum_{\mathbf{k}\in\Delta_j} |v_{\mathbf{k}}|^2.$$

Notice that the single-scale bases have local supports, which satisfy

$$\operatorname{diam}\operatorname{supp}\varphi_{j,\mathbf{k}} \sim 2^{-j}.$$

Additional properties of the spaces V_j are required for using them as trial spaces in a Galerkin scheme. More specifically, the trial spaces are required to have *approximation order* $d \in \mathbb{N}$ and *regularity* $\gamma > 0$, where

$$\begin{aligned} \gamma &= \sup\{t \in \mathbb{R} : V_j \subset H^t(\Omega\cup\Omega_{\mathcal{I}})\}, \\ d &= \sup\left\{t \in \mathbb{R} : \inf_{v_j\in V_j} \|v - v_j\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \lesssim 2^{-jt}\|v\|_{H^t(\Omega\cup\Omega_{\mathcal{I}})}\right\}. \end{aligned} \tag{2.2}$$

The parameter d corresponds to the order of polynomials that can be represented locally. For ansatz spaces based on smoothest splines, we have $\gamma = d - 1/2$, since they are globally C^{d-1}-smooth, whereas using Lagrangian finite element shape functions leads to $\gamma = 3/2$, since these are globally continuous. Note especially that conformity of the Galerkin scheme induces $\gamma > s$, which means that the wavelets are required to be globally continuous if $s \geq 1/2$.

Instead of using only a single-scale j, the idea of wavelet concepts is to keep track of the increment of information between two adjacent scales $j-1$ and j. Since $V_{j-1} \subset V_j$, one decomposes $V_j = V_{j-1} \oplus W_j$ with some complementary space W_j, not necessarily orthogonal to V_{j-1}. Recursive splitting of the trial spaces leads to the wavelet decomposition $V_J = \bigoplus_{j=0}^J W_j$.

Of practical interest are the bases of the complementary spaces W_j in V_j

$$\Psi_j = \{\psi_{j,\mathbf{k}} : \mathbf{k} \in \nabla_j := \Delta_j \setminus \Delta_{j-1}\}.$$

It is supposed that the collections $\Phi_{j-1} \cup \Psi_j$ are also uniformly stable bases of V_j. If $\Psi = \bigcup_{j\geq 0} \Psi_j$, where $\Psi_0 := \Phi_0$, is a Riesz-basis of $L^2(\Omega\cup\Omega_{\mathcal{I}})$, then it is called a wavelet basis. For any function $v \in L^2(\Omega\cup\Omega_{\mathcal{I}})$, we have the mutliscale wavelet expansion

$$v = \sum_{j=0}^{\infty} \sum_{\mathbf{k}\in\Delta_j} v_{j,\mathbf{k}}\psi_{j,\mathbf{k}}$$

and hence

$$\|v\|^2_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \sim \sum_{j=0}^{\infty} \sum_{\mathbf{k}\in\nabla_j} |v_{j,\mathbf{k}}|^2. \tag{2.3}$$

In particular, this estimate implies the wavelets are normalized

$$\|\psi_{j,\mathbf{k}}\|_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \sim 1.$$

We shall further assume that the functions $\psi_{j,k}$ are also local with respect to the corresponding scale,

$$\operatorname{diam}\operatorname{supp}\psi_{j,\mathbf{k}} \sim 2^{-j}.$$

This ensures that the fast wavelet transform. i.e., the change between the single-scale basis and the wavelet basis can be performed in linear complexity.

At first glance it would be very convenient to deal with a single orthonormal system of wavelets. However, it was shown in [10, 12, 28] that orthogonal wavelets are not completely appropriate for the efficient compression of non-local operators. For this reason, we employ biorthogonal wavelet bases. More specifically, we have a biorthogonal, or dual, multiresolution analysis, i.e., dual single-scale bases $\widetilde{\Phi}_j = \{\widetilde{\varphi}_{j,\mathbf{k}} : \mathbf{k} \in \Delta_j\}$ and wavelets $\widetilde{\Psi}_j = \{\widetilde{\psi}_{j,\mathbf{k}} : \mathbf{k} \in \nabla_j\}$ that are coupled to the primal ones via

$$(\Phi_j, \widetilde{\Phi}_j)_{L^2(\Omega\cup\Omega_{\mathcal{I}})} = \mathbf{I} \quad \text{and} \quad (\Psi_j, \widetilde{\Psi}_j)_{L^2(\Omega\cup\Omega_{\mathcal{I}})} = \mathbf{I}.$$

Therefore, the associated spaces $\widetilde{V}_j := \operatorname{span}\widetilde{\Phi}_j$ and $\widetilde{W}_j := \operatorname{span}\widetilde{\Psi}_j$ satisfy

$$V_{j-1} \perp \widetilde{W}_j \quad \text{and} \quad \widetilde{V}_{j-1} \perp W_j. \tag{2.4}$$

Analogously to (2.2), the dual spaces provide some approximation order $\widetilde{d} \in \mathbb{N}$ and regularity $\widetilde{\gamma} > 0$, as well. The relation (2.4) implies that the wavelets provide *vanishing moments* of order $\widetilde{d}$

$$\left|(v, \psi_{j,\mathbf{k}})_{L^2(\Omega\cup\Omega_{\mathcal{I}})}\right| \lesssim 2^{-j(n/2+\widetilde{d})}|v|_{W^{\widetilde{d},\infty}(\operatorname{supp}\psi_{j,k})}. \tag{2.5}$$

Here $|v|_{W^{\widetilde{d},\infty}(\Omega)} := \sup_{|\boldsymbol{\alpha}|=\widetilde{d}} \|\partial^{\boldsymbol{\alpha}} v\|_{L^\infty(\Omega)}$ denotes the semi-norm in $W^{\widetilde{d},\infty}(\Omega)$. We refer to [9] for further details.

Finally, it turns out that properly scaled versions of these wavelets constitute Riesz bases for a whole scale of Sobolev spaces. In fact, in accordance with [9, 16], we have the well known *norm equivalences*

$$\|v\|^2_{H^t(\Omega\cup\Omega_{\mathcal{I}})} \sim \sum_{j=0}^{\infty} 2^{2jt} \sum_{\mathbf{k}\in\nabla_j} |v_{j,\mathbf{k}}|^2, \quad t \in (-\widetilde{\gamma}, \gamma). \tag{2.6}$$

These are essential to develop optimal compression and preconditioning strategies, see [10]. As a last remark, we note that piecewise constant and bilinear wavelets on arbitrary domains or surfaces, which provide the above properties, have been constructed in [18, 20].

3 Wavelet matrix compression

The main component needed for wavelet matrix compression relies on the smoothness of the kernel κ. To this end, we will henceforth assume that the function $\sigma(x, y)$ in (1.4) is (sufficiently) smooth. As a result, the kernel $\kappa(x, y)$ is smooth except for the singularity along the diagonal $x = y$. In addition, we obtain the following crucial estimate of decay based on mixed partial derivatives of κ:

$$\left|\partial_x^{\boldsymbol{\alpha}}\partial_y^{\boldsymbol{\beta}}\kappa(x, y)\right| \le \frac{c_{\boldsymbol{\alpha},\boldsymbol{\beta}}}{|x-y|^{n+2s+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}, \quad x \ne y, \tag{3.1}$$

provided we have

$$|\boldsymbol{\alpha}| + |\boldsymbol{\beta}| > n + 2s.$$

The estimate (3.1) is the key building block for the wavelet matrix compression.

In view of the wavelets' vanishing moments (2.5), the decay estimate (3.1) implies the following estimate for the matrix coefficients of the discrete non-local operator:

$$(\mathcal{L}\psi_{j,\mathbf{k}}, \psi_{j',\mathbf{k}'})_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \lesssim \frac{2^{-(j+j')(n/2+\widetilde{d})}}{\operatorname{dist}(\Omega_{j,\mathbf{k}}, \Omega_{j',\mathbf{k}'})^{n+2s+2\widetilde{d}}}. \tag{3.2}$$

Here, we use the abbreviation $\Omega_{j,\mathbf{k}} := \operatorname{supp}\psi_{j,\mathbf{k}}$ and $\Omega_{j',\mathbf{k}'} := \operatorname{supp}\psi_{j',\mathbf{k}'}$ for the support of the wavelets $\psi_{j,\mathbf{k}}$ and $\psi_{j',\mathbf{k}'}$, respectively.

Estimate (3.2) implies that the discrete operator becomes quasi-sparse in wavelet coordinates and is the main foundation of the compression estimates derived in [10]. Based on (3.2), we can set all matrix entries to zero, for which the distance of the supports between the associated trial and test functions is larger than a level dependent cut-off parameter $\mathcal{B}_{j,j'}$.

Further compression, reflected by a cut-off parameter $\mathcal{B}^{\mathcal{S}}_{j,j'}$, is achieved by neglecting some of those matrix entries, for which the corresponding trial and test functions have overlapping supports. We refer to compression with respect to the $\mathcal{B}_{j,j'}, \mathcal{B}^{\mathcal{S}}_{j,j'}$ parameters as "a priori compression."

To formulate this result, we introduce the abbreviation $\Omega^{\mathcal{S}}_{j,\mathbf{k}} := \operatorname{sing\,supp}\psi_{j,\mathbf{k}}$, which denotes the *singular support* of the wavelet $\psi_{j,\mathbf{k}}$, i.e., that subset of $\Omega_{j,\mathbf{k}}$ where the wavelet $\psi_{j,\mathbf{k}}$ is not smooth.

Theorem 2 (A-priori compression [10])**.** Let $\Omega_{j,\mathbf{k}}$ and $\Omega^{\mathcal{S}}_{j,\mathbf{k}}$ be given as above and define the compressed system matrix $\mathbf{L}_J$, corresponding to the non-local operator $\mathcal{L}$, by

$$[\mathbf{L}^{\psi}_J]_{(j,\mathbf{k}),(j',\mathbf{k}')} := \begin{cases} 0, & \operatorname{dist}(\Omega_{j,\mathbf{k}}, \Omega_{j',\mathbf{k}'}) > \mathcal{B}_{j,j'} \text{ and } j, j' > 0, \\ 0, & \operatorname{dist}(\Omega_{j,\mathbf{k}}, \Omega_{j',\mathbf{k}'}) \leq 2^{-\min\{j,j'\}} \text{ and} \\ & \operatorname{dist}(\Omega^{\mathcal{S}}_{j,\mathbf{k}}, \Omega_{j',\mathbf{k}'}) > \mathcal{B}^{\mathcal{S}}_{j,j'} \text{ if } j' > j \geq 0, \\ & \operatorname{dist}(\Omega_{j,\mathbf{k}}, \Omega^{\mathcal{S}}_{j',\mathbf{k}'}) > \mathcal{B}^{\mathcal{S}}_{j,j'} \text{ if } j > j' \geq 0, \\ (\mathcal{L}\psi_{j',\mathbf{k}'}, \psi_{j,\mathbf{k}})_{L^2(\Omega\cup\Omega_{\mathcal{I}})}, & \text{otherwise.} \end{cases} \tag{3.3}$$

Fixing

$$a > 1, \qquad d < \delta < \widetilde{d} + 2s, \tag{3.4}$$

the cut-off parameters $\mathcal{B}_{j,j'}$ and $\mathcal{B}^{\mathcal{S}}_{j,j'}$ are set as follows

$$\begin{aligned} \mathcal{B}_{j,j'} &= a\ \max\left\{2^{-\min\{j,j'\}}, 2^{\frac{2J(\delta-s)-(j+j')(\delta+\widetilde{d})}{2(\widetilde{d}+s)}}\right\}, \\ \mathcal{B}^{\mathcal{S}}_{j,j'} &= a \max\left\{2^{-\max\{j,j'\}}, 2^{\frac{2J(\delta-s)-(j+j')\delta-\max\{j,j'\}\widetilde{d}}{\widetilde{d}+2s}}\right\}. \end{aligned} \tag{3.5}$$

As a consequence, the system matrix $\mathbf{L}^{\psi}_J$ has only $\mathcal{O}(N_J)$ nonzero coefficients, where $N_J = 2^{Jn}$ denotes the degrees of freedom in the space V_J. Moreover, the error estimate

$$\|u - u_J\|_{H^{2s-d}(\Omega\cup\Omega_{\mathcal{I}})} \lesssim 2^{-2J(d-s)}\|u\|_{H^d(\Omega\cup\Omega_{\mathcal{I}})} \tag{3.6}$$

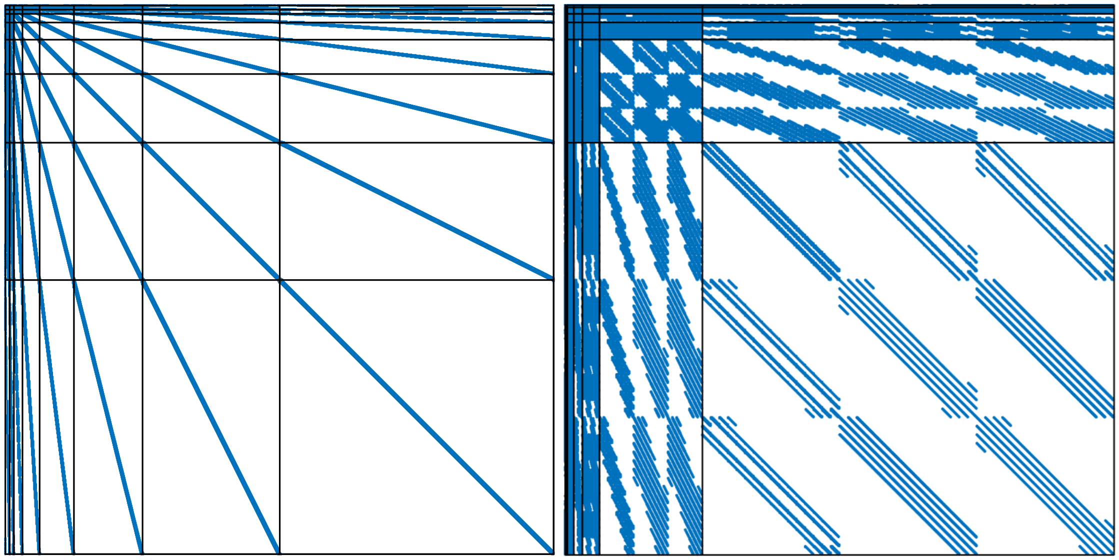

Figure 2: Compression pattern in case of an interval (left) and a square (right).

holds for the solution u_J of the compressed Galerkin system provided that u, Ω, and $\Omega_{\mathcal{I}}$ are sufficiently regular.

The compressed system matrix can be assembled in linear complexity if one employs the exponentially convergent hp–quadrature method proposed in [19]. Moreover, for performing faster matrix-vector multiplications, an additional a-posteriori compression might be applied which reduces again the number of nonzero coefficients by a factor 2–5 [10]. The pattern of the compressed system matrix exhibit the typical *finger structure*, compare Figure 2.

Since the boundary integral operator $\mathcal{L}$ has an order s different from 0, the compressed system matrix $\mathbf{L}_J$ becomes more and more ill-conditioned when the level J increases. More precisely, the condition number of the system matrix will asymptotically grow like $2^{2J|s|}$ as the level J increases. However, as an immediate consequence of the norm equivalences (2.6) of wavelet bases, normalizing the wavelets relative to the energy norm leads to uniformly bounded condition numbers.

Theorem 3 (Preconditioning [11, 28])**.** Let the diagonal matrix $\mathbf{D}_J^r$ be defined by

$$\big[\mathbf{D}_J^r\big]_{(j,\mathbf{k}),(j',\mathbf{k}')} = 2^{rj}\delta_{(j,\mathbf{k}),(j',\mathbf{k}')} \text{ for all } k \in \nabla_j,\ \mathbf{k}' \in \nabla_{j'},\ 0 \le j, j' \le J.$$

Then, if the regularity $\widetilde{\gamma}$ of the dual wavelets satisfies $\widetilde{\gamma} > -s$, the diagonal matrix $\mathbf{D}_J^{2s}$ defines an asymptotically optimal preconditioner to $\mathbf{L}_J$, i.e.,

$$\operatorname{cond}_{\ell^2}(\mathbf{D}_J^{-s}\mathbf{L}_J^{\psi}\mathbf{D}_J^{-s}) \sim 1.$$

Remark 4. The entries on the main diagonal of $\mathbf{L}_J$ satisfy

$$(\mathcal{L}\psi_{j,\mathbf{k}}, \psi_{j,\mathbf{k}})_{L^2(\Omega\cup\Omega_{\mathcal{I}})} \sim 2^{2|s|j}.$$

Therefore, the above preconditioning can be replaced by a diagonal scaling. Indeed, the diagonal scaling improves and even simplifies the standard wavelet preconditioning.

4 Numerical results

4.1 Set-up and implementation

Our realization is based on piecewise constant wavelets in two spatial dimensions. We are hence able to cover the range of parameters $0 < s < 1/2$. Since the fractional Laplacian is of positive order, we can simply choose Haar wavelets for the discretization, which implies $d = \widetilde{d} = 1$ in (3.4). In particular, Haar wavelets are orthonormal, which implies that the primal wavelets $\{\psi_{j,\mathbf{k}}\}$ and the dual wavelets $\{\widetilde{\psi}_{j,\mathbf{k}}\}$ coincide.

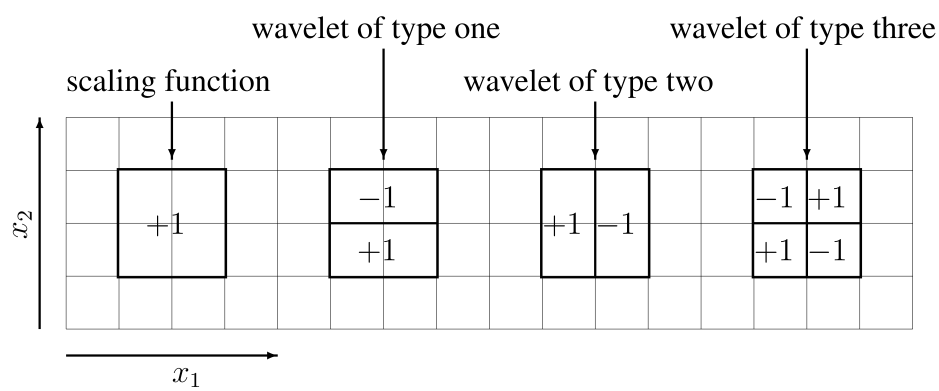

Figure 3: Illustration of the Haar basis in two spatial dimensions.

Haar wavelets on the unit square can easily constructed recursively by starting on the fine grid j. Having there $2^j \times 2^j$ piecewise constant ansatz functions $\{\varphi_{j,\mathbf{k}}\}$, one gets the basis functions for level $j-1$ as the piecewise constant ansatz functions $\{\varphi_{j-1,\mathbf{k}}\}$ and the corresponding wavelets $\{\psi_{j,\mathbf{k}}\}$ on level j by simple agglomeration of four fine grid functions each, see Figure 3. Having wavelets on the unit square, one can define wavelets on arbitrary domains via parametrization by quadrangular patches, see [18] for the details. For our numerical experiments, we consider the unit circle which is parametrized by five patches as seen in Figure 4.

The implementation of the wavelet matrix compression basically follows [19]. Singular integrals are treated by the Duffy trick, cf. [15,27], but we can only get analytical integrands in case of kernel singularities with integer exponent. Therefore, we always refine quasi-singular integrals until we are on the finest level J. This ensures an accurate quadrature while the over-all complexity of the wavelet matrix compression still scales linearly in the number of degrees of freedom.

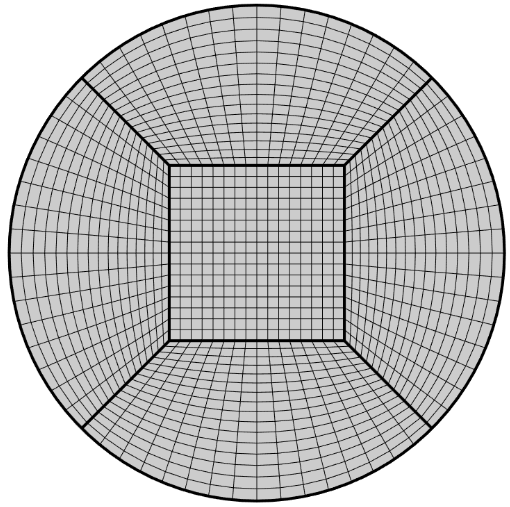

Figure 4: The parametrization of the unit circle by five patches and mesh on level $j = 4$.

4.2 Compression rates

Our first numerical tests are concerned with the compression rates of the wavelet matrix compression. We compute for different values of the parameter s the compression rates by the ratio of the number of nonzero matrix coefficients and the square of the number of degrees of freedom N_J, corresponding to the number of coefficients in the uncompressed system matrix. These numbers, specified in percent, are tabulated in Table 1 for both, the a-priori compression (the respective columns are entitled "a-priori") and for the a-posteriori compression (the respective columns are entitled "a-posteriori").

degrees of freedom		compression rates					
		$s = 1/8$		$s = 1/4$		$s = 3/8$	
J	N_J	a-priori	a-posteriori	a-priori	a-posteriori	a-priori	a-posteriori
1	20	100	58.5	100	54.5	100	57.5
2	80	55.4	40.2	54.3	36.6	52.8	33.4
3	320	22.5	12.9	21.1	14.5	19.9	13.3
4	1280	7.92	4.63	7.12	3.81	6.55	4.44
5	5120	2.61	1.37	2.24	1.19	1.93	0.98
6	20480	0.82	0.40	0.66	0.31	0.54	0.28
7	81920	0.25	0.12	0.19	0.09	0.15	0.08
8	327680	0.07	0.03	0.05	0.02	0.04	0.02

Table 1: A-priori and a-posteriori compression rates for different choices of s on the unit circle.

We observe that the a-priori compression becomes better as s increases. This results from the fact that the best possible convergence rate decreases as s increases and, hence, the accuracy requirements are reduced. We also deduce from Table 1 that the a-posteriori compression improves the compression rate at least by a factor 2.

4.3 Preconditioning

We briefly discuss the wavelet preconditioning approach used in our optimization solver here. For different values of the parameter s and discretization levels J, we compute the ℓ^2-condition numbers of the system matrix with and without diagonal scaling. Since the constant functions lie in the operator's kernel and, hence, the smallest eigenvalue is zero, we compute the ℓ^2-condition numbers of the modified system matrix $\mathbf{L}_J^\psi + \mathbb{1}\mathbb{1}^\intercal$, where $\mathbb{1}$ corresponds to the discretization of the function $f(\mathbf{x}) \equiv 1$.

degrees of freedom		condition numbers					
		$s = 1/8$		$s = 1/4$		$s = 3/8$	
J	N_J	unscaled	scaled	unscaled	scaled	unscaled	scaled
1	20	10.4	6.74	15.4	9.44	24.0	14.0
2	80	16.5	7.31	27.4	10.8	48.6	17.1
3	320	28.4	7.73	52.0	11.8	104	19.9
4	1280	42.9	8.02	93.6	12.9	203	22.7
5	5120	60.2	8.25	149	13.6	377	25.3
6	20480	80.5	8.42	234	14.3	751	29.0

Table 2: Condition numbers of the system matrix with respect to the fractional Laplacian with and without preconditioning.

As can be seen from Table 2, the condition of the unscaled system matrix (the respective columns are entitled "unscaled") indeed grows approximately like 2^{2sJ} when increasing the discretization level J. In contrast, we observe that the condition numbers of the diagonally scaled system matrices (the respective columns are entitled "scaled") seem to stay bounded by a constant; though the latter does in fact depend on the order $2s$ of the fractional Laplacian.

4.4 An unconstrained optimal control problem: $Z_{\text{ad}} = L^2(\Omega \cup \Omega_{\mathcal{I}})$

We solve the optimal control problem (1.2) subject to (1.1) for different values of the parameter s and Ω being the unit circle. The observation operator C under consideration is the projection of a given function $f \in L^2(\Omega)$ onto the square $\square := (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^2$, which coincides with the central patch of our parametrization, see Figure 4. The operator B is chosen as the identity operator and thus preserves the regularity of λ^*.

We directly exploit the solution formulae (1.5a) and (1.5b) to compute the optimal solution of the problem under consideration. The desired function u_d is chosen to be $u_d(x_1, x_2) = x_1 x_2$ and ν is set to 10^{-3}. We discretize u and z by Haar wavelets and exploit the fact that $\lambda^\star = -\nu z^\star$. The

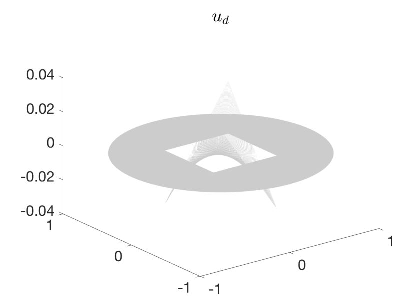

Figure 5: The desired state $u_d(x,y) = xy$ on the $\square := (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^2$.

discretized optimality system now becomes the following linear system of equations after applying the compression strategies:

$$\begin{bmatrix} \mathbf{L}_J^{\psi} & -\mathbf{M}_J^{\psi} \\ \mathbf{N}_J^{\psi} & \nu \mathbf{L}_J^{\psi} \end{bmatrix} \begin{bmatrix} \mathbf{u}_J^{\psi} \\ \mathbf{z}_J^{\psi} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{u}_{J,d}^{\psi} \end{bmatrix}. \tag{4.1}$$

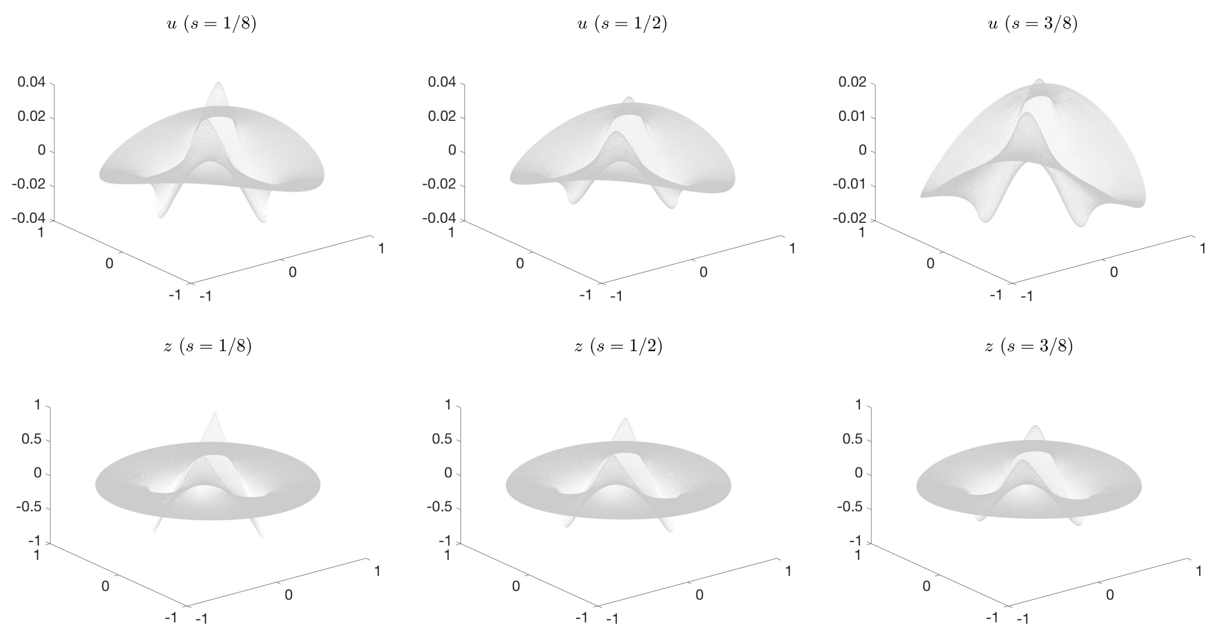

Figure 6: The optimal states u (first row) and controls z (second row) in case of $s = 1/8$ (left column), $s = 1/4$ (middle column), and $s = 3/8$ (right column).

In (4.1), $\mathbf{L}_J^{\psi}$ denotes the compressed system matrix with respect to the fractional Laplacian while

$$\mathbf{M}_J^{\psi} = \left[\int_{\Omega} \psi_{j,\mathbf{k}}(\mathbf{x}) \psi_{j',\mathbf{k}'}(\mathbf{x}) \,\mathrm{d}\mathbf{x} \right]_{(j,\mathbf{k}),(j',\mathbf{k}')}, \quad \mathbf{N}_J^{\psi} = \left[\int_{\square} \psi_{j,\mathbf{k}}(\mathbf{x}) \psi_{j',\mathbf{k}'}(\mathbf{x}) \,\mathrm{d}\mathbf{x} \right]_{(j,\mathbf{k}),(j',\mathbf{k}')},$$

are the mass matrix and the truncated mass matrix, respectively. Note that we do not compute these mass matrices in practice since they are more dense than the corresponding matrices with respect to single-scale ansatz functions. Therefore, by using the corresponding matrices with respect to classical piecewise constant ansatz functions

$$\mathbf{M}_J^{\varphi} = \left[\int_{\Omega} \varphi_{J,\mathbf{k}}(\mathbf{x}) \varphi_{J,\mathbf{k}'}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right]_{(J,\mathbf{k}),(J,\mathbf{k}')}, \quad \mathbf{N}_J^{\varphi} = \left[\int_{\square} \varphi_{J,\mathbf{k}}(\mathbf{x}) \varphi_{J,\mathbf{k}'}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right]_{(J,\mathbf{k}),(J,\mathbf{k}')},$$

we find

$$\mathbf{M}_J^{\psi} = \mathbf{T}_J^{\varphi \to \psi} \mathbf{M}_J^{\varphi} \mathbf{T}_J^{\psi \to \varphi}, \quad \mathbf{N}_J^{\psi} = \mathbf{T}_J^{\varphi \to \psi} \mathbf{N}_J^{\varphi} \mathbf{T}_J^{\psi \to \varphi}$$

with $\mathbf{T}_J^{\varphi \to \psi}$ and $\mathbf{T}_J^{\psi \to \varphi}$ being the fast wavelet transform and its inverse. The wavelet transform has linear complexity and, hence, the matrix-vector product $\mathbf{M}_J^{\psi}\mathbf{x}$ is computable also in linear complexity. In the current setting, we have the advantageous fact that $(\mathbf{T}_J^{\varphi \to \psi})^{\intercal} = \mathbf{T}_J^{\psi \to \varphi}$, which is a consequence of the fact that Haar wavelets constitute an orthonormal basis.

Likewise, we derive the right hand side $\mathbf{u}_d^{\psi}$. Instead of computing

$$\mathbf{v}_J^{\psi} = \left[\int_{\square} u_d(\mathbf{x}) \psi_{j,\mathbf{k}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right]_{(j,\mathbf{k})}$$

and solving

$$\mathbf{M}_J^{\psi} \mathbf{u}_{J,d}^{\psi} = \mathbf{v}_J^{\psi},$$

we directly compute the data vector with respect to the classical single-scale basis

$$\mathbf{v}_J^{\varphi} = \left[\int_{\square} u_d(\mathbf{x}) \varphi_{J,\mathbf{k}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right]_{(J,\mathbf{k})}$$

to get

$$\mathbf{u}_{J,d}^{\psi} = \mathbf{T}_J^{\varphi \to \psi} \big(\mathbf{M}_J^{\varphi} \big)^{-1} \mathbf{v}_J^{\varphi}.$$

Here, we exploited that the mass matrix $\mathbf{M}_J^{\varphi}$ is a diagonal matrix and hence easily invertible.

We use a diagonal scaling as preconditioner of the linear system of equations (4.1). It greatly reduces the number of iterations required by the GMRES-method, compare [26]. Indeed, we only need about 40 iterations of GMRES to reduce the ℓ^2-norm of the residuum to being smaller than 10^{-8}, basically independent of the discretization level J.

In Figure 6, we visualized the optimal states u and optimal controls z for $s = 1/8$, $s = 1/4$, and $s = 3/8$ for discretization level $J = 7$, which corresponds to $N_J = 81920$ degrees of freedom for both, the state and the control. One observes that u corresponds well with the desired function u_d on the square $\square$. Outside this square, it rapidly tends to zero, where we find that the decay becomes slower as s grows.

4.5 Pointwise Bilateral Constraints: $Z_{\mathrm{ad}} \subsetneq L^2(\Omega \cup \Omega_{\mathcal{I}})$

In contrast to the previous section, we consider pointwise bound constraints on the control of the type:

$$z_{\min} \le z \le z_{\max} \quad \text{on } \Omega. \tag{4.2}$$

In order to solve the respective constraint optimization problem, we apply the primal-dual active set strategy as introduced in [5, 22, 23]. The essential idea of this iterative strategy is to replace successively the inequality constraints by the related equality constraints for all the indices where the constraint becomes active. This strategy has the advantage that, although initially introduced as an active set strategy, it can be reinterpreted as a semi-smooth Newton method and converges thus superlinearly, see [21].

An important observation for realizing the active set strategy in the context of a wavelet discretization is that the wavelets are oscillating functions. Hence, we have to switch to the single-scale basis to compare two functions. We first introduce the discrete bounds

$$\begin{aligned}\mathbf{z}_{J,\min}^{\varphi} &:= \big(\mathbf{M}_J^{\varphi}\big)^{-1}\left[\int_{\square} z_{\min}(\mathbf{x})\varphi_{J,\mathbf{k}}(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]_{(J,\mathbf{k})},\\ \mathbf{z}_{J,\max}^{\varphi} &:= \big(\mathbf{M}_J^{\varphi}\big)^{-1}\left[\int_{\square} z_{\max}(\mathbf{x})\varphi_{J,\mathbf{k}}(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]_{(J,\mathbf{k})}.\end{aligned}$$

This leads to the discrete analogue of the box constraint (4.2) for the discrete control $\mathbf{z}_J^{\varphi}$:

$$\mathbf{z}_{J,\min}^{\varphi} \le \mathbf{z}_J^{\varphi} \le \mathbf{z}_{J,\max}^{\varphi},$$

where the inequality relation for vectors has to be understood componentwise.

In the ℓ-th iteration step, given the iterate $(\mathbf{u}_J^{\varphi,(\ell)}, \mathbf{z}_J^{\varphi,(\ell)})$ and the Lagrange multipliers $(\boldsymbol{\mu}_{J,\min}^{\varphi,(\ell)}, \boldsymbol{\mu}_{J,\max}^{\varphi,(\ell)})$, we compute the active sets

$$\begin{aligned}\mathbb{I}^{(\ell)} &:= \big\{\mathbf{k}\in\Delta_J : c\big(z_{J,\mathbf{k}}^{\varphi,(\ell)} - z_{\min,(J,\mathbf{k})}^{\varphi,(\ell)}\big) + \mu_{\min,(J,\mathbf{k})}^{\varphi,(\ell)} < 0\big\},\\ \mathbb{J}^{(\ell)} &:= \big\{\mathbf{k}\in\Delta_J : c\big(z_{J,\mathbf{k}}^{\varphi,(\ell)} - z_{\max,(J,\mathbf{k})}^{\varphi,(\ell)}\big) + \mu_{\max,(J,\mathbf{k})}^{\varphi,(\ell)} > 0\big\},\end{aligned}$$

where $c > 0$ is an appropriately chosen parameter (we choose $c = 10^{-4}$ in our experiments). The sets $\mathbb{I}^{(\ell)}$ and $\mathbb{J}^{(\ell)}$ contain the indices of all coefficients for which the lower and upper box constraints become active, respectively. Therefore, we have to solve the saddle point problem

$$\left[\begin{array}{cc|cc} \mathbf{L}_J^{\psi} & -\mathbf{M}_J^{\psi} & & \\ \mathbf{N}_J^{\psi} & \nu\mathbf{L}_J^{\psi} & \mathbf{T}_J^{\varphi\to\psi}\mathbf{I}_{\mathbb{I}^{(\ell)}} & \mathbf{T}_J^{\varphi\to\psi}\mathbf{I}_{\mathbb{J}^{(\ell)}} \\ \hline & \mathbf{I}_{\mathbb{I}^{(\ell)}}^{\intercal}\mathbf{T}_J^{\psi\to\varphi} & & \\ & \mathbf{I}_{\mathbb{J}^{(\ell)}}^{\intercal}\mathbf{T}_J^{\psi\to\varphi} & & \end{array}\right] \left[\begin{array}{c} \mathbf{u}_J^{\psi,(\ell+1)} \\ \mathbf{z}_J^{\psi,(\ell+1)} \\ \hline \boldsymbol{\mu}_{J,\min,\mathbb{I}^{(\ell)}}^{\varphi,(\ell+1)} \\ \boldsymbol{\mu}_{J,\max,\mathbb{J}^{(\ell)}}^{\varphi,(\ell+1)} \end{array}\right] = \left[\begin{array}{c} \mathbf{0} \\ \mathbf{u}_{J,d}^{\psi} \\ \hline \mathbf{z}_{J,\min,\mathbb{I}^{(\ell)}}^{\varphi} \\ \mathbf{z}_{J,\max,\mathbb{J}^{(\ell)}}^{\varphi} \end{array}\right]. \tag{4.3}$$

Here, the matrices $\mathbf{I}_{\mathbb{I}^{(\ell)}} \in \mathbb{R}^{N_J\times|\mathbb{I}^{(\ell)}|}$ and $\mathbf{I}_{\mathbb{J}^{(\ell)}} \in \mathbb{R}^{N_J\times|\mathbb{J}^{(\ell)}|}$ are obtained from the identity matrix in $\mathbb{R}^{N_J}$ by removing those columns whose indices are not contained in the index sets $\mathbb{I}^{(\ell)}$ and $\mathbb{J}^{(\ell)}$, respectively. Likewise, the vectors $\boldsymbol{\mu}_{J,\min,\mathbb{I}^{(\ell)}}^{\varphi,(\ell+1)} \in \mathbb{R}^{|\mathbb{I}^{(\ell)}|}$ and $\boldsymbol{\mu}_{J,\max,\mathbb{J}^{(\ell)}}^{\varphi,(\ell+1)} \in \mathbb{R}^{|\mathbb{J}^{(\ell)}|}$ consist only of those components of $\boldsymbol{\mu}_{J,\min}^{\varphi,(\ell+1)}$ and $\boldsymbol{\mu}_{J,\max}^{\varphi,(\ell+1)}$ which are contained in the index sets $\mathbb{I}^{(\ell)}$ and $\mathbb{J}^{(\ell)}$, respectively. The same holds true for $\mathbf{z}_{J,\min,\mathbb{I}^{(\ell)}}^{\varphi}$ and $\mathbf{z}_{J,\max,\mathbb{I}^{(\ell)}}^{\varphi}$.

For all inactive indices, the box constraints will be ignored and the associated components of the Lagrange multipliers are set to 0:

$$\mu^{\varphi,(\ell+1)}_{\min,(J,\mathbf{k})} := 0 \text{ for all } \mathbf{k} \in \Delta_J \setminus \mathbb{I}^{(\ell)},$$
$$\mu^{\varphi,(\ell+1)}_{\max,(J,\mathbf{k})} := 0 \text{ for all } \mathbf{k} \in \Delta_J \setminus \mathbb{J}^{(\ell)}.$$

Finally, the iteration index is increased $\ell \mapsto \ell + 1$ and the loop restarted.

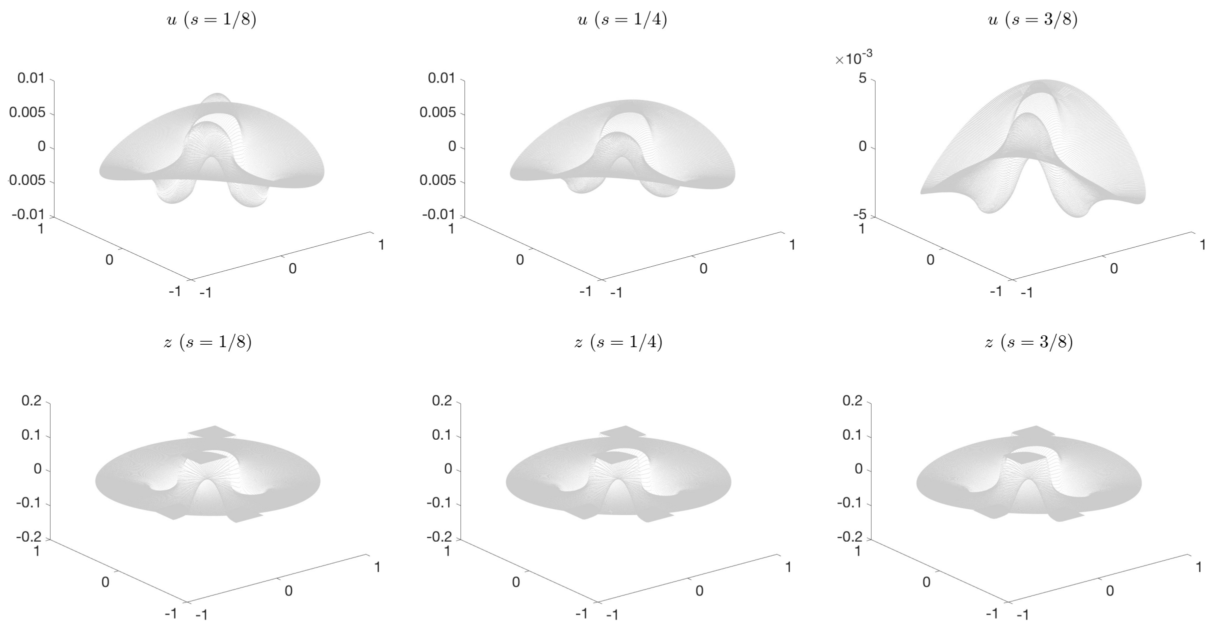

Figure 7: The optimal states u (first row) and controls z (second row) in case of $s = 1/8$ (left column), $s = 1/4$ (middle column), and $s = 3/8$ (right column).

We consider again the optimal control problem from Subsection 4.4, but impose the box constraints (4.2) with $z_{\min} \equiv -0.1$ and $z_{\max} \equiv 0.1$ to the control. The discretization level is $J = 7$, corresponding to $N_J = 81920$ piecewise constant ansatz functions each to discretize the state and the control. The active set strategy is started with $\mathbb{I}^{(0)} = \mathbb{J}^{(0)} = \emptyset$. In each iteration of the active set strategy, the resulting saddle point problem (4.3) is solved by the GMRES-method, where we use again the diagonal as preconditioner for the blocks corresponding to the primal variables while for the dual variables no preconditioner is applied. Note that the active set strategy never needs more than five iterations. The computed controls and states for $s = 1/8$, $s = 1/4$, and $s = 3/8$ are found in Figure 7, where we observe in any case that the box constraints become active.

5 Conclusion and Outlook

Our study is the first of its kind to present a wavelet-based approach for the solution of a semismooth system of equations with non-local operators. In addition, it provides a roadmap for the application of wavelets for other bound-constrained optimal control problems with local partial differential operators, as well. The approach has the advantage that the non-local operators can be treated directly, without the need to transform the forward problem. In fact, the wavelet characterization of Sobolev spaces allows us to apply the method for $\mathcal{L}_\kappa = (-\Delta)^s$ for any $s \in (0,1]$, even below the critical threshold of $1/2$. As expected, the solver behaved scale-independently (cf. mesh-independence when using FEM), and a full theoretical proof will be the subject of future research.

References

[1] H. Antil, T. S. Brown, and D. Verma. Moreau-Yosida regularization for optimal control of fractional elliptic problems with state and control constraints. *arXiv preprint arXiv:1912.05033*, 2019.

[2] H. Antil, and E. Otárola. A FEM for an optimal control problem of fractional powers of elliptic operators. *SIAM J. Control Optim.*, 53(6), 3432-3456, 2015.

[3] H. Antil, and E. Otárola. An a posteriori error analysis for an optimal control problem involving the fractional Laplacian. *IMA J. Numer. Anal.*, 38 (1), 198–226, 2017.

[4] 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.*, https://doi.org/10.1007/s10957-020-01684-z, 2020

[5] M. Bergounioux, K. Ito, and K. Kunisch. Primal-dual strategy for constrained optimal control problems. *SIAM J. Control Optim.*, 37:1176–1194, 1999.

[6] L. Caffarelli and L. Silvestre. An extension problem related to the fractional Laplacian. *Comm. Partial Differential Equations*, 32(7–9):1245–1260, 2007.

[7] L. Chen, R. H. Nochetto, E. Otárola, and A. J. Salgado. A PDE approach to fractional diffusion: a posteriori error analysis. *J. Comput. Phys.*, 293:339–358, 2015.

[8] S. Dahlke, T. M. Surowiec. Wavelet-Based Approximations of Pointwise Bound Constraints in Lebesgue and Sobolev Spaces. *Bericht Mathematik* Nr. 2020-02 des Fachbereichs Mathematik und Informatik, Philipps-Universität Marburg, 2020.

[9] W. Dahmen. Wavelet and multiscale methods for operator equations. *Acta Numer.*, 6:55–228, 1997.

[10] W. Dahmen, H. Harbrecht, and R. Schneider. Compression techniques for boundary integral equations. Optimal complexity estimates. *SIAM J. Numer. Anal.*, 43:2251–2271, 2006.

[11] W. Dahmen and A. Kunoth. Multilevel preconditioning. *Numer. Math.*, 63:315–344, 1992.

[12] W. Dahmen, S. Prößdorf, and R. Schneider. Multiscale methods for pseudo-differential equations on smooth closed manifolds. In C.K. Chui, L. Montefusco, and L. Puccio, editors, *Proceedings of the International Conference on Wavelets: Theory, Algorithms, and Applications*, pages 385–424, 1994.

[13] M. D'Elia and M. Gunzburger. The fractional Laplacian operator on bounded domains as a special case of the nonlocal diffusion operator. *Comput. Math. Appl.*, 66(7):1245–1260, 2013.

[14] Q. Du, M. Gunzburger, R.B. Lehoucq, and K. Zhou. Analysis and approximation of nonlocal diffusion problems with volume constraints. *SIAM Rev.*, 54(4):667–696, 2012.

[15] Michael G. Duffy. Quadrature over a pyramid or cube of integrands with a singularity at a vertex. *SIAM J. Numer. Anal.*, 19(6):1260–1262, 1982.

[16] S. Jaffard. Wavelet methods for fast resolution of elliptic equations. *SIAM J. Numer. Anal.*, 29:965–986, 1992.

[17] H. Harbrecht, W.L. Wendland, and N. Zorii. On Riesz minimal energy problems. *J. Math. Anal. Appl.*, 393(2):397–412, 2012.

[18] H. Harbrecht and R. Schneider. Biorthogonal wavelet bases for the boundary element method. *Math. Nachr.*, 269–270:167–188, 2004.

[19] H. Harbrecht and R. Schneider. Wavelet Galerkin schemes for boundary integral equations. Implementation and quadrature. *SIAM J. Sci. Comput.*, 27(4):1347–1370, 2006.

[20] H. Harbrecht and R. Stevenson. Wavelets with patchwise cancellation properties. *Math. Comput.*, 75(256):1871–1889, 2006.

[21] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13:865–888, 2003.

[22] K. Ito and K. Kunisch. The primal-dual active set method for nonlinear optimal control problems with bilateral constraints. *SIAM J. Contr. Optim.*, 43:357–376, 2004.

[23] K. Kunisch and A. Rösch. Primal-dual active set strategy for a general class of constrained optimal control problems. *SIAM J. Optim.*, 13:321–334, 2002.

[24] J.L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*, volume 170 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin–Heidelberg, 1971.

[25] R. H. Nochetto, E. Otárola, and A. J. Salgado. A PDE approach to fractional diffusion in general domains: a priori error analysis. *Found. Comput. Math.*, 15(3):733–791, 2015.

[26] Y. Saad and M.H. Schulz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. and Stat. Comput.*, 7(3):856–869, 1986.

[27] S.A. Sauter and C. Schwab. Quadrature for hp-Galerkin BEM in $\mathbb{R}^3$. *Numer. Math.*, 78(2):211–258, 1997.

[28] R. Schneider. *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur Lösung großer vollbesetzter Gleichungssysteme*. Teubner, Stuttgart, 1998.