# Wavelet-Based Approximations of Pointwise Bound Constraints in Lebesgue and Sobolev Spaces

S. Dahlke and T. M. Surowiec

March 18, 2020

**Abstract**

Many problems in optimal control, PDE-constrained optimization, and constrained variational problems include pointwise bound-constraints on the feasible controls and state variables. Most well-known approaches for treating such pointwise inequality constraints in numerical methods rely on finite element discretizations and interpolation arguments. We propose an alternative means of discretizing pointwise bound-constraints using a wavelet-based discretization. The main results show that the discrete, approximating sets converge in the sense of Mosco to the original sets. In situations of higher regularity, convergence rates follow immediately from the underlying wavelet theory. The approach exploits the fact that one can easily transform between a given multiscale wavelet representation and single-scale representation with linear complexity. This allows, for example, a direct treatment of variational problems involving fractional operators, without the need for lifting techniques. We demonstrate this fact with several numerical examples of fractional obstacle problems.

Keywords: Wavelets, PDE-Constrained Optimization, Bound Constraints, Fractional Obstacle Problem, Perpetual American Options
AMS MSC: 65T60, 65K10, 65K15, 49J53, 49J20, 49K20, 49M99 90C06

## 1 Introduction

Due to their many favorable properties, wavelets offer an attractive means of discretizing a wide array of variational problems and partial differential differential equations (PDEs). This is evidenced by the success of wavelet-based schemes for treating problems in signal and image processing [35], partial-differential equations [14, 15], and high-dimensional parametric or random partial differential equations [40], to name just a few. However, the literature is extremely scarce on wavelet-based approximations for bound-constrained variational problems such as obstacle problems or PDE-constrained optimization problems with control and/or state constraints. There is a good reason for this.

As their name suggests, the oscillatory properties of wavelets give rise to basis functions that have vanishing moments. This is in fact part of the reason for their success. However, from the perspective of approximation of inequality constraints and active-set-based optimization solvers, the undulatory behavior of wavelets appears to be highly problematic. Consider for example the set:

$$C := \left\{ v \in L^2(0,1) \,|\, -1 \le v(x) \le 1 \;\text{ a.e. } x \in (0,1) \right\}$$

and suppose we are given $v \in L^2(0,1)$ such that $v \notin C$. The $L^2(\Omega)$-projection $\mathrm{Proj}_C(v)$ of $v$ onto $C$ has a simple well-known formula in the continuous setting:

$$\mathrm{Proj}_C(v) = \inf_{u \in C} \|u - v\| = v - \max(0, v-1) + \max(0, -1-v). \tag{1.1}$$

This projection operator can then be used in function-space-based optimization algorithms such as projected gradient/projected Newton [8, 23, 31] or generalized (semismooth) Newton methods and active set strategies [6, 29, 43]. However, in practice (especially for first-order methods) we may need to calculate $\mathrm{Proj}_C(v)$ hundreds of times. For generalized Newton methods, we need to easily and accurately estimate the active and inactive sets in order to determine a generalized Hessian or Newton derivative.

Since $v \in L^2(0,1)$, it admits a multiscale wavelet expansion of the form

$$v = \sum_{j \in \mathbb{N}} \sum_{k \in \Lambda_j} \langle_{L^2(0,1)} v, \widetilde{\psi}_{j,k} \rangle \psi_{j,k},$$

where $j \in \mathbb{N}$ are the allowable scales and $\Lambda_j$ encode the locations of the wavelets for a given scale $j$. Without specifying further, we let $\{\psi_{j,k}\}_{j \in \mathbb{N}, k \in \Lambda_j}$, $\{\widetilde{\psi}_{j,k}\}_{j \in \mathbb{N}, k \in \Lambda_j}$ be a pair of primal and dual wavelets. It is only important for the current discussion to note that $\psi_{j,k}$ by nature oscillates between positive and negative values. We denote the usual inner product on $L^2(0,1)$ by $\langle \cdot, \cdot \rangle_{L^2(0,1)}$. Unless otherwise noted, we simply write $\langle \cdot, \cdot \rangle$. The positive part operator $\max(0, \cdot)$ is denoted by $(\cdot)_+$.

Fixing a maximum allowable scale $j_{\max}$ and returning to (1.1), one might then try to consider a finite-dimensional approximation of the $\mathrm{Proj}_C(c)$ suitable for numerical optimization methods by first projecting $v$ (denoted by $v_{j_{\max}}$) onto the finite-dimensional subspace (denoted by $V_{j_{\max}}$) and then by solving

$$\inf\{\|u - v_{j_{\max}}\| \text{ over } u \in V_{j_{\max}} \Big| \; u = \sum_{j=0}^{j_{\max}} \sum_{k \in \Lambda_j} d_{j,k} \psi_{j,k}, \; -1 \leq u \leq 1 \; \forall k \in \Lambda_j\}.$$

This (or any other related schemes directly involving the wavelet basis) will clearly not give good discrete estimates of the true active sets where $\mathrm{Proj}_C(v) = 1$ or $-1$; not to mention that it is difficult to see whether the finite dimensional problems converge to the original problem or that the inequality constraints are still technically understood for all $x \in (0,1)$.

However, $v_{j_{\max}}$ also admits an equivalent single-scale expansion using the generator (scaling) functions $\varphi_{j,k}$ of the form

$$v_{j_{\max}} = \sum_{k \in \Lambda_{j_{\max}}} \langle v, \widetilde{\varphi}_{j,k} \rangle \varphi_{j,k},$$

where $\widetilde{\varphi}_{j,k}$ is the associated dual generator function for the pair $(j,k)$ with $k \in \Lambda_j$. Since many generator functions $\varphi_{j,k}$ are non-negative, this opens up a wide variety of possibilities in which we first discretize using the single-scale expansions in nonnegative basis functions and whenever necessary apply the fast wavelet transform afterwards to return to the multiscale expansion if desired. This approach represents the basis for our study.

The desire to return to the multiscale basis mentioned above is again motivated by numerical methods. The two major advantages of using the multiscale expansion are in sparse representation of functions and preconditioning. For the former, we note that many functions, which are not sparse in the single scale (generator) basis, such as piecewise smooth functions, admit a very sparse representation in the multiscale (wavelet) basis. This is ultimately a direct consequence of the vanishing moments. Moreover, since wavelets characterize a number of function spaces, such as Sobolev and Besov spaces, the multiscale representation gives rise to very effiecient preconditioning strategies.

Our goal with this paper is twofold. First, we seek to demonstrate how wavelets admitting non-negative single-scale bases can be used to approximate bound constraints in certain Lebesgue and Sobolev spaces. Second, we will demonstrate these approximating sequences of sets converge with respect to an appropriate notion of set-convergence that is relevant to optimization, optimal control, and constrained variational problems. These include classical obstacle problems [32, 39], variational approximations of free-discontinuity problems [1], or constrained problems involving fractional derivatives [41]. In doing so, we hope to reach a broad set of researchers who either may not be familiar with wavelet-based approximation methods or, alternatively, for those who may not be familiar with variational convergence results for constrained minimization problems.

The rest of the paper is structured as follows. In Section 2, we introduce the necessary definitions arising in the theory of wavelets needed for our discussions as well as several concepts from set-valued and variational analysis. In Section 3, we propose a means of approximating bound constraints in $L^2$ using the classical Haar wavelet. This is particularly relevant for PDE-constrained optimization problems, where these types of inequalities often arise, see e.g., the well-known monographs [30, 34, 42] and the references therein. For variational problems in fractional Sobolev spaces, which do not admit embeddings into spaces of continuous functions, the Haar wavelet is also relevant. Afterwards, in Section 4, we consider a setting in which the constrained functions have higher regularity, i.e., in Sobolev spaces $H^s(\Omega)$ with $s > 1/2$. This is relevant for the variational problems mentioned above. In Section 5, we discuss an application of the results to a fractional obstacle problem and provide the results of several numerical experiments. We conclude with a summary and outlook on possible future directions.

## 2 Notation and Preliminary Results

In this section, we provide a quick overview on the construction of wavelets and some of their useful properties. The presentation closely follows a similar introduction in [18]. Moreover, recall the notion of set convergence relevant to our study.

### 2.1 A Brief Primer on Wavelets

Generally speaking, a wavelet basis comprises a collection of scaled, dilated and translated versions of a function $\psi \in L^2(\mathbb{R})$, i.e., the set

$$\Psi = \{\psi_{j,k}\}_{j,k\in\mathbb{Z}}, \qquad \psi_{j,k}(x) := 2^{j/2}\psi(2^j x - k) \tag{2.1}$$

forms a (stable) basis of $L^2(\mathbb{R})$. The *mother wavelet* $\psi$ may be chosen to be exponentially decaying or even compactly supported. Usually, such a wavelet basis is constructed by means of a *multiresolution analysis*

(MRA), which is a nested sequence $\{V_j\}_{j\in\mathbb{Z}}$ of closed linear subspaces of $L^2(\mathbb{R})$ whose union is dense while their intersection is zero. Moreover, the spaces $\{V_j\}_{j\in\mathbb{Z}}$ are related by dyadic dilation, i.e., $f \in V_j$ iff $f(2\cdot) \in V_{j+1}$. It is furthermore assumed that there exists a function $\varphi \in V_0$ with stable integer translates, the so–called *generator*, such that $V_0 := \overline{\text{span}\{\varphi(\cdot - k),\, k \in \mathbb{Z}\}}$. The nestedness and stability properties of an MRA imply that $\varphi$ is *refinable*, i.e., it satisfies a *two–scale relation*:

$$\varphi(x) = \sum_{k\in\mathbb{Z}} \alpha_k \varphi(2x - k), \tag{2.2}$$

with the *mask* $\mathbf{a} = \{\alpha_k\}_{k\in\mathbb{Z}} \in \ell_2(\mathbb{Z})$.

Because the union of the spaces $\{V_j\}_{j\in\mathbb{Z}}$ is dense in $L^2(\mathbb{R})$, it is easy to see that the construction of a wavelet basis reduces to finding a function whose translates span a complement space $W_0$ of $V_0$ in $V_1$, $V_1 = V_0 \oplus W_0,\ W_0 = \overline{\text{span}\{\psi(\cdot - k) \mid k \in \mathbb{Z}\}}$. Indeed, if we define $W_j := \{f(\cdot) \in L^2(\mathbb{R}) \mid f(2^{-j}\cdot) \in W_0\}$, it follows that $L^2(\mathbb{R}) = \bigoplus_{j=-\infty}^{\infty} W_j$, so that

$$\psi_{j,k}(\cdot) = 2^{j/2}\psi(2^j \cdot - k), \qquad j,k \in \mathbb{Z}, \tag{2.3}$$

forms a wavelet basis of $L^2(\mathbb{R})$. It is then perhaps clear that the wavelet $\psi$ can be found by means of a functional equation of the form

$$\psi(x) = \sum_{k\in\mathbb{Z}} b_k \varphi(2x - k), \tag{2.4}$$

where the sequence $\mathbf{b} := \{b_k\}_{k\in\mathbb{Z}}$ has to be judiciously chosen; see, e.g., [13], [21], [36] for details.

In practice, it is of course desirable to construct an *orthonormal* wavelet basis. This can be achieved if the integer translates of $\psi$ are pairwise orthogonal and span the orthogonal complement of $V_0$ in $V_1$. However, it has turned out that the orthonormality requirement is quite restrictive. One possible way out is to use the *biorthogonal* approach. Then, for a given wavelet basis $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ one is interested in finding a second system $\{\tilde{\psi}_{j,k}\}_{j,k\in\mathbb{Z}}$ satisfying

$$\langle \psi_{j,k}(\cdot), \tilde{\psi}_{j',k'}(\cdot)\rangle_{L^2(\mathbb{R})} = \delta_{j,j'}\delta_{k,k'}, \quad j,j',k,k' \in \mathbb{Z}. \tag{2.5}$$

Then all the computations are as simple as in the orthonormal case, i.e.,

$$v = \sum_{j,k\in\mathbb{Z}} \langle v, \tilde{\psi}_{j,k}\rangle_{L^2(\mathbb{R})}\psi_{j,k} = \sum_{j',k'\in\mathbb{Z}} \langle v, \psi_{j',k'}\rangle_{L^2(\mathbb{R})}\tilde{\psi}_{j',k'}.$$

To construct such a biorthogonal system, one needs *two* sequences of approximation spaces $\{V_j\}_{j\in\mathbb{Z}}$ and $\{\tilde{V}_j\}_{j\in\mathbb{Z}}$. Once again, one has to find bases for certain algebraic complement spaces $W_0$ and $\tilde{W}_0$ satisfying the biorthogonality conditions $V_0 \perp \tilde{W}_0,\ \tilde{V}_0 \perp W_0,\ V_0 \oplus W_0 = V_1,\ \tilde{V}_0 \oplus \tilde{W}_0 = \tilde{V}_1$. This is quite easy if the two generators $\varphi$ and $\tilde{\varphi}$ form a *dual pair*,

$$\langle \varphi(\cdot), \tilde{\varphi}(\cdot - k)\rangle_{L^2(\mathbb{R})} = \delta_{0,k}. \tag{2.6}$$

Indeed, then two biorthogonal wavelets $\psi$ and $\tilde{\psi}$ can be constructed as

$$\psi(x) = \sum_{k\in\mathbb{Z}} (-1)^k \beta_{1-k}\varphi(2x - k), \qquad \tilde{\psi}(x) = \sum_{k\in\mathbb{Z}} (-1)^k \alpha_{1-k}\tilde{\varphi}(2x - k) \tag{2.7}$$

where

$$\varphi(x) = \sum_{k\in\mathbb{Z}} \alpha_k \varphi(2x-k), \qquad \tilde{\varphi}(x) = \sum_{k\in\mathbb{Z}} \beta_k \tilde{\varphi}(2x-k). \tag{2.8}$$

Elegant constructions can be found, e.g., in [16]. In particular, the authors in [16] consider the important case in which the generator is a cardinal B-spline $\phi = N_m\ m \geq 1$.

The most important properties of wavelets can be summarized as follows.

- Polynomial exactness. If $\varphi$ is contained in $C_0^r(\mathbb{R}) := \{g \mid g \in C^r(\mathbb{R}) \text{ and } \operatorname{supp} g \text{ compact}\}$, then every monomial $x^\alpha$ has an expansion of the form

$$x^\alpha = \sum_{k\in\mathbb{Z}} c_k^\alpha \varphi(x-k), \quad \alpha \leq r. \tag{2.9}$$

- Oscillations. If the dual generator $\tilde{\varphi}$ is contained in $C_0^r(\mathbb{R})$, then the associated wavelet $\psi$ has vanishing moments up to order $r$, i.e.,

$$\int_{\mathbb{R}} x^\alpha \psi(x)dx = 0 \qquad \text{for all } 0 \leq \alpha \leq r. \tag{2.10}$$

- Approximation. For $f \in H^s(\mathbb{R}), 0 < s < r+1$, where $r$ is the order of polynomial exactness in $V_j$, the following Jackson-type inequality holds:

$$\|f - P_j f\|_{L^2(\mathbb{R})} \leq C 2^{-js} |f|_{H^s}, \quad P_j f := \sum_{k\in\mathbb{Z}} \langle f, \tilde{\varphi}_{j,k}\rangle_{L^2(\mathbb{R})} \varphi_{j,k}. \tag{2.11}$$

- Characterization of function spaces. Let $\varphi \in H^s$ and suppose that $V_j$ has polynomial exactness $r$. Then for all $t < \min\{r+1, s\}$

$$\|f\|_{H^t} \eqsim \|P_0 f\|_{L^2} + (\sum_{j\geq 0} \sum_{k\in\mathbb{Z}} 2^{2tj} |\langle f, \tilde{\psi}_{j,k}\rangle_{L^2(\mathbb{R})}|^2)^{1/2}. \tag{2.12}$$

Here, $\eqsim$, and later $\lesssim$, denote equality (inequality) up to constants.

**Remark 1.** 1. The construction outlined above can be generalized to $L^2(\mathbb{R}^n)$. The simplest way is to use tensor products, but also nonseparabel constructions including general scaling matrices are possible. Consequently, the number of mother wavelets that are needed depends on the determinant of the scaling matrix.

2. It is also possible to construct wavelet bases on nontrivial domains. In such situations, of course the boundary adaptation is a nontrivial task. We refer, e.g., to [19], [20] for details. All the important properties of wavelets carry over analogously.

3. Given a biorthogonal wavelet basis, any function $v \in V_j$ has two equivalent representations, the *single scale representation* with respect to the functions $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ and the *multiscale representation* which is based on the functions $\varphi_{0,k}, \psi_{l,m}, k, m \in \mathbb{Z}, 0 \leq l < j, \psi_{l,m} := 2^{j/2}\psi(2^j x - m)$.

From the coefficients of $v$ in the single scale representation, the coefficients in the multiscale representation can easily be obtained by filtering. Indeed, given $v = \sum_{k\in\mathbb{Z}} c_{j,k}\varphi_{j,k}$ and using the refinement equation (2.8) and the functional equation (2.7), it turns out that

$$v = \sum_{l\in\mathbb{Z}} 2^{-1/2}\Big(\sum_{k\in\mathbb{Z}} \overline{\beta}_{k-2l}c_{j,k}\Big)\varphi_{j-1,l} + \sum_{m\in\mathbb{Z}} 2^{-1/2}\Big(\sum_{k\in\mathbb{Z}}(-1)^k\overline{\alpha}_{1-k-2m}c_{j,k}\Big)\psi_{j-1,m}. \tag{2.13}$$

Here, $\overline{\beta}$ and $\overline{\alpha}$ denote the complex conjugates.

From (2.13) we observe that the information corresponding to $V_{j-1}$ and the one corresponding to the wavelet space $W_{j-1}$ can be obtained by applying a low–pass filter $H$ and a high–pass filter $D$, respectively,

$$\begin{aligned} c_{j-1} &= Hc_j, \quad c_{j-1,l} = \sum_{k\in\mathbb{Z}} 2^{-1/2}\overline{\beta}_{k-2l}c_{j,k}, \\ d_{j-1} &= Dc_j, \quad d_l^{j-1} = 2^{-1/2}\sum_{k\in\mathbb{Z}}(-1)^k\overline{\alpha}_{1-k-2l}c_{j,k}. \end{aligned} \tag{2.14}$$

By iterating this decomposition method, we obtain a pyramid algorithm, the so–called *fast wavelet transform*:

$$\begin{tikzcd} c_j \arrow[r, "H"] \arrow[dr, "D"'] & c_{j-1} \arrow[r, "H"] \arrow[dr, "D"'] & c_{j-2} \arrow[r, "H"] \arrow[dr, "D"'] & c_{j-3} \arrow[r, dotted] \arrow[dr, dotted] & {} \\ & d_{j-1} & d_{j-2} & d_{j-3} & {} \end{tikzcd}$$

A reconstruction algorithm can be obtained in a similar fashion. For numerical computations, the fast wavelet transform is especially attractive as it can be performed with linear complexity in the length of the initial single scale signal.

For further information on wavelet analysis, the reader is referred to one of the excellent textbooks on wavelets which have appeared quite recently [13], [21], [36], [44].

## 2.2 Set-Convergence

We will primarily work with the notion of set-convergence due to Mosco [37]. This can also be understood as Painlevé-Kuratowski set-convergence in which the upper/outer-limits are defined using the weak-topology and the lower/inner-limits are defined using the strong (norm)-topology, see e.g., [4] for details. Since this notion of convergence is equivalent to Mosco epi-convergence (weak-strong $\Gamma$-convergence) of the associated indicator functionals, it is best suited for approximating optimization and control problems, cf. [2, 3].

**Definition 2.** Let $X$ be a real reflexive Banach space and let $\{C_j\}$, $C_j \subset X$, be a sequence of nonempty, closed and convex subsets. Then $\{C_j\}$ is said to converge in the sense of Mosco to the closed convex set $C \subset X$ provided the following two conditions hold:

1. For every $x \in C$, there exists a sequence $\{x_j\}$ such that $x_j \in C_j$ and $x_j \to x$.

2. $C$ contains the set of all weak accumulation points of sequences $\{x_k\}$, where $x_k \in C_{j_k}$ and $\{C_{j_k}\}$ is a subsequence of $\{C_j\}$.

We use the standard notation $\to$ to denote norm convergence and $\rightharpoonup$ to denote weak convergence throughout the text. We occasionally specifically state the corresponding space for clarity.

# 3 Bilateral Constraints in $L^2(\Omega)$

In this section, we focus on bilateral constraints in the Lebesgue space $L^2(\Omega)$. This is a general technique that could also be employed for fractional Sobolev spaces $H^s(\Omega)$ when $s \in (0, 1/2]$ as the lack of higher regularity rules out any benefits obtained by employing wavelets with high regularity.

## 3.1 Approximation by Haar Wavelets

In the following discussion, we will assume that $n = 1$ and use Haar wavelets, cf. [27, 35], generated by the usual piecewise constant step functions, which for a given scale $j$ we denote by $\varphi_{j,k}$ for $k \in \Lambda_j$. Here, $\Lambda_j$ is the $j$-dependent set of translations. In other words, $\varphi_{j,k}$ is the (scaled) characteristic function of a subinterval of length $2^{-j}$, i.e.,

$$\varphi_{j,k}(x) = 2^{j/2}\chi_{[0,1]}(2^j x - k)$$

and $k \in \Lambda_j$ ensures that this subinterval is contained in $\Omega$. To some extent, this setting mirrors the usage of piecewise constant finite elements to approximate box constraints in the PDE-constrained optimization, see e.g., [24].

Given $\xi_0, \xi_1 \in L^2(\Omega)$ such that $\xi_0 < \xi_1$ a.e., we define the set $C \subset L^2(\Omega)$ by

$$C := \left\{ v \in L^2(\Omega) \,|\, \xi_0 \le v \le \xi_1 \right\}.$$

We first demonstrate that for each $j$, we can construct finite dimensional functions that approximately satisfy the bound constraints in $C$. Due to the stability of the basis, we will be able to show that these sequences of "inner approximations" of $C$ admit weakly convergent subsequences whose limit points are elements of $C$.

For a fixed scale $j \ge 0$, we let $V_j$ be the finite dimensional subspace of $L^2(\Omega)$ generated by the basis of generator functions $\varphi_{j,k}$, $k \in \Lambda_j$. We then define

$$C_j := \left\{ v_j := \sum_{k \in \Lambda_j} c_{j,k}\varphi_{j,k} \in V_j \,\middle|\, \langle \xi_0, \varphi_{j,k} \rangle \le c_{j,k} \le \langle \xi_1, \varphi_{j,k} \rangle \quad k \in \Lambda_j \right\}.$$

In the following discussion, we will prove the convergence of these sets $C_j$ to $C$ as $j \to +\infty$ in the sense of Mosco.

It is important to stress here that the functions $v_j \in C_j$ are most likely not what would be used in the discretization of the optimization problems. Instead, using a maximal scale $m \in \mathbb{N}$ as the desired accuracy of approximation, we would include the bound constraints in $C_j$ for a collection of scales $j \in \{0, \ldots, m\}$ and use the multiscale representation of $v_m$ given by

$$v_m = \sum_{j=0}^{m} \sum_{k \in \Lambda_j} d_{j,k}\psi_{j,k},$$

where for $j \in \{0, \ldots, m\}$ and $k \in \Lambda_j$, $\psi_{j,k}$ is the corresponding Haar wavelet and the coefficients $\{d_{j,k}\}$ are calculated using the associated $\{c_{j,k}\}$ the fast wavelet transform.

**Remark 3.** For the current setting, the smallest scale $j = 0$ is allowed. However, for some settings, e.g., with intervals of length less than one, it is necessary to choose $j$ large enough so that the scaled generated function is fully supported in the given interval.

This brings us to our first result.

**Lemma 4.** Every sequence $\{u_j\}$ such that $u_j \in C_j$ for all $j \in \mathbb{N}$ contains a weakly convergent subsequence $\{u_{j_\ell}\}$ such that $u_{j_\ell} \rightharpoonup \overline{u}$ as $\ell \to +\infty$. Moreover, we have $\overline{u} \in C$.

*Proof.* For a fixed scale $j \geq 0$, take $u_j \in C_j$ and let $c_j \in \mathbb{R}^{|\Lambda_j|}$ be the associated coefficients. By definition, we have

$$\langle \xi_0, \varphi_{j,k} \rangle \leq c_{j,k} \leq \langle \xi_1, \varphi_{j,k} \rangle \quad \text{for all } k \in \Lambda_j,$$

Since $\varphi_{j,k} \geq 0$ for every $k \in \Lambda_j$, it also holds for a.e. $x \in \Omega$ that

$$\langle \xi_0, \varphi_{j,k} \rangle \varphi_{j,k}(x) \leq c_{j,k}\varphi_{j,k}(x) \leq \langle \xi_1, \varphi_{j,k} \rangle \varphi_{j,k}(x) \quad \text{for all } k \in \Lambda_j,$$

Summing over $k \in \Lambda_j$ we have

$$\sum_{k \in \Lambda_j} \langle \xi_0, \varphi_{j,k} \rangle \varphi_{j,k} \leq \sum_{k \in \Lambda_j} c_{j,k}\varphi_{j,k} \leq \sum_{k \in \Lambda_j} \langle \xi_1, \varphi_{j,k} \rangle \varphi_{j,k}. \tag{3.1}$$

In fact, the upper and lower bounds are merely the projections $P_j\xi_0$ and $P_j\xi_1$ of $\xi_0$ and $\xi_1$ onto $V_j$, cf. (2.11), respectively. Whence we have

$$P_j\xi_0 \leq \sum_{k \in \Lambda_j} c_{j,k}\varphi_{j,k} \leq P_j\xi_1.$$

Next, consider that

$$\begin{aligned}
\|u_j\|^2_{L^2(\Omega)} = \Big\| \sum_{k \in \Lambda_j} c_{j,k}\varphi_{j,k} \Big\|^2_{L^2(\Omega)} &\lesssim \sum_{k \in \Lambda_j} |c_{j,k}|^2 \\
&\lesssim \sum_{k \in \Lambda_j} \max\{|\langle \xi_0, \varphi_{j,k} \rangle|^2, |\langle \xi_1, \varphi_{j,k} \rangle|^2\} \\
&\lesssim \sum_{k \in \Lambda_j} |\langle \xi_0, \varphi_{j,k} \rangle|^2 + |\langle \xi_1, \varphi_{j,k} \rangle|^2 \\
&\lesssim \|P_j\xi_0\|^2 + \|P_j\xi_1\|^2 \\
&\lesssim \|\xi_0\|^2_{L^2(\Omega)} + \|\xi_1\|^2_{L^2(\Omega)}.
\end{aligned}$$

Therefore, there exists a subsequence $\{j_\ell\}$ and some $\overline{u} \in L^2(\Omega)$ such that $u_{j_\ell} \rightharpoonup \overline{u}$.

Finally, since $P_{j_\ell}\xi_0 \to \xi_0$ and $P_{j_\ell}\xi_1 \to \xi_1$, we can show that $\overline{u} \in C$. To this end, let $\varphi \in L^2(\Omega)$ such that $\varphi \geq 0$ p.a.e. on $\Omega$. Then

$$0 \leq \langle P_{j_\ell}\xi_1 - u_{j_\ell}, \varphi \rangle$$

Since $P_{j_\ell}\xi_1 - u_{j_\ell} \rightharpoonup \xi_1 - \overline{u}$, we conclude that

$$0 \leq \langle \xi_1 - \overline{u}, \varphi \rangle \quad \text{for all } \varphi \in L^2(\Omega) : \varphi \geq 0.$$

It follows from the fundamental lemma of calculus of variations, that $\overline{u} \leq \xi_1$ p.a.e. on $\Omega$. An analogous argument shows that $\overline{u} \geq \xi_0$ p.a.e. on $\Omega$. This proves the assertion. $\square$

Returning to the preceding discussion on the multiscale representation, let $\{u_{j_\ell}\}$ be the weakly convergent sequence in the proof of Lemma 4. Then we may also write each $u_{j_\ell}$ via its multiscale representation using Haar wavelets

$$u_{j_\ell} = \sum_{j=0}^{j_\ell} \sum_{k \in \Lambda_j} d_{j,k}\psi_{j,k},$$

where, as mentioned above, the $d_{j,k}$ are calculated using the corresponding single scale coefficients $c_{j,k}$. As argued in the proof, $u_{j_\ell} \in C_{j_\ell}$ and $u_{j_\ell} \rightharpoonup \overline{u} \in C$ as $\ell \to +\infty$. Nevertheless, from the perspective of an active-set-based optimization algorithm, we would not determine, e.g., the upper active sets by checking where $\sum_{j=0}^{j_\ell} \sum_{k \in \Lambda_j} d_{j,k}\psi_{j,k}(x) = \xi_1(x)$. Instead, we use the inequalities on the single-scale coefficients $c_{j,k}$ for each $j \in \{0, \ldots, m\}$ and $k \in \Lambda_j$. If one were to use, e.g., piecewise constant finite elements to discretize the same set, e.g., as done in [24], then for a fixed mesh size $h$ we would obtain bounds on the coefficients as well similar to ours. However, the multiscale representation used here brings with it additional benefits such as compression, preconditioning, and a potentially sparse representation as mentioned in the introduction. Moreover, in Section 5, we can exploit the norm equivalence to easily solve variational problems in fractional Sobolev spaces.

We now show that any $u \in C$ can be strongly approximated by a sequence of elements $u_j \in C_j$ (without resorting to subsequences).

**Lemma 5.** For every $u \in C$ there exists a sequence $\{u_j\}$ such that $u_j \in C_j$ and $u_j \to u$ as $j \to +\infty$.

*Proof.* Fix an arbitrary $u \in C$. Then a.e. on $\Omega$, we have $\xi_0 \leq u \leq \xi_1$. Multiplying the inequality by an arbitrary generator function $\varphi_{j,k}$ and integrating over $\Omega$, we obtain for any $k \in \Lambda_j$

$$\langle \xi_0, \varphi_{j,k} \rangle \leq \langle u, \varphi_{j,k} \rangle \leq \langle \xi_1, \varphi_{j,k} \rangle.$$

Consequently, $u_j = P_j u = \sum_{k \in \Lambda_j} \langle u, \varphi_{j,k} \rangle \varphi_{j,k} \in C_j$ for every scale $j \geq 0$. Since $P_j u \to u$ as $j \to +\infty$, the assertion holds. $\square$

If we once again turn back to the multiscale discussion, then we showed that $u_j = P_j u$ in the single-scale sense is feasible for $C_j$. Alternatively, using the multiscale representation, $u_j$ would also be an element of the feasible set defined by the single scale bounds for the scales $0, \ldots, j$. Therefore, we can always be assured that an element $u \in C$ can be strongly approximated by a discrete counterpart; regardless of the representation. We now have our first main result.

**Theorem 6.** The sequence of convex sets $\{C_j\} \subset L^2(\Omega)$ converges in the sense of Mosco to $C$.

*Proof.* This is a direct consequence of Lemma 4 and Lemma 5 in light of the Definition 2. □

As mentioned in Section 2.2, the Mosco convergence of the sets $\{C_j\}$ provides us with a variational convergence result for the associated indicator functions $\{\chi_{C_j}\}$ to the indicator function $\chi_C$, where $\chi_{C_j}(u) = 0$ if $u \in C_j$ and $+\infty$, otherwise. Though this is a valuable result, it does not provide us with an a priori rate of convergence. Nevertheless, we can in fact obtain a rate of convergence once $u$ and the sequence $\{u_j\}$ in Lemma 5 exhibit even a minimal amount of Sobolev regularity.

**Corollary 7.** In the context of Lemma 5, suppose that $u \in C \cap H^t(\Omega)$ and $\{u_j\} \subset C_j \cap H^t(\Omega)$, then for all $s < t$ there exists a constant $c > 0$ such that the following error estimate holds:

$$\|u - u_j\|_{H^s(\Omega)} \leq c2^{-j(t-s)}\|u\|_{H^t(\Omega)} \tag{3.2}$$

In particular, if $s = 0 < t$, we have $\|u - u_j\|_{L^2(\Omega)} = O(2^{-jt})$.

*Proof.* This is merely an implication of Jackson's inequality, see e.g., [7], in light of the fact that $u_j$ in Lemma 5 is given by $u_j = P_j u$. □

**Remark 8.** Note that the error bound in (3.2) is taken from [7], where error bounds of this type are provided for both the projection operator $P_j$ and the interpolation operator $L_j$ used below.

### 3.2 Remarks and Extensions

A few remarks are in order. First, the arguments made above would work for any $L^2$-stable generator basis $\{\varphi_{j,k}\}_{k\in\Lambda_j}$ that contains only non-negative functions $\varphi_{j,k}$ which has the property $P_j u = \sum_{k\in\Lambda_j}\langle u, \varphi_{j,k}\rangle\varphi_{j,k}$ and $P_j u \to u$ strongly in $L^2(D)$. However, for the arguments used above we are basically limited to Haar wavelets since the primal and dual generating functions need to be the same. We remedy this with different arguments in Section 4.

Second, we never really use the fact that $D \subset \mathbb{R}^1$, so the results could easily be carried over to higher dimensions provided $\Omega$ can be split into a parametric image of hypercubes.

Third, there is a clear link between $C_j$ and the set $C \cap V_j$. Indeed, the latter is of the type:

$$C \cap V_j = \{v_j \in V_j \,|\xi_0 \leq v_j \leq \xi_1\},$$

where for $v_j \in C \cap V_j$ we have $\xi_0 \leq \sum_{k\in\Lambda_j} v_{j,k}\varphi_{j,k} \leq \xi_1$ a.e. $\Omega$. Since each pair $(\varphi_{j,\ell}, \varphi_{j,m}) \in \{\varphi_{j,k}\}_{k\in\Lambda_j}$ is pairwise orthogonal and $\|\varphi_{j,k}\|_{L^2(D)} = 1$, we have

$$\xi_0 \leq \sum_{k\in\Lambda_j} v_{j,k}\varphi_{j,k} \leq \xi_1,$$

and therefore,

$$\langle\xi_0, \varphi_{j,\ell}\rangle \leq \sum_{k\in\Lambda_j} v_{j,k}\langle\varphi_{j,k}, \varphi_{j,\ell}\rangle \leq \langle\xi_1, \varphi_{j,\ell}\rangle$$

and consequently,

$$\langle\xi_0, \varphi_{j,\ell}\rangle \leq v_{j,\ell} \leq \langle\xi_1, \varphi_{j,\ell}\rangle.$$

Therefore, $v_j \in C_j$, i.e., $C \cap V_j \subset C_j$. However, starting from $u_j \in C_j$ we initially only have (3.1), which in this context means

$$\sum_{k\in\Lambda_j} \langle \xi_0, \varphi_{j,k}\rangle \varphi_{j,k} \leq u_j \leq \sum_{k\in\Lambda_j} \langle \xi_1, \varphi_{j,k}\rangle \varphi_{j,k}. \tag{3.3}$$

Clearly, since $P_j\xi_i \neq \xi_i$ for $i = 1, 2$ and every $j$, we cannot conclude that $u_j \in C \cap V_j$. This would require

$$\xi_0 \leq \sum_{k\in\Lambda_j} \langle \xi_0, \varphi_{j,k}\rangle \varphi_{j,k} \quad \text{and} \quad \sum_{k\in\Lambda_j} \langle \xi_1, \varphi_{j,k}\rangle \varphi_{j,k} \leq \xi_1 \tag{3.4}$$

for all sufficiently large scales $j$, which cannot be expected even for simple piecewise linear $\xi_0, \xi_1$.

Finally, Corollary 7 provides us with some insight into the case in which we replace $L^2(\Omega)$ by $H^s(\Omega)$ with $s \in (0, 1/2)$. To this aim, let $C^s := C \cap H^s(\Omega)$ and assume that $\xi_0, \xi_1$ are such that (3.4) holds for sufficiently large scales $j$. As discussed above, this would mean that the finite dimensional counterparts of $C^s_j$, $j \in \mathbb{N}$, satisfy $C^s_j \subset C^s$. Since $C^s$ is a nonempty, closed, and convex set in $H^s(\Omega)$, this means that all weak accumulation points obtained from the sequence of sets $\{C^s_j\}$ are also contained in $C^s$. A stronger statement in the $L^2(\Omega)$ case is proven in Lemma 4. For the remaining inclusion, i.e., that the strong lower/inner limit of $\{C^s_j\}$ contains $C$, we note that by Corollary 7, the sequence $\{u_j\}$ as constructed in Lemma 5 converges strongly to $u$ in $H^{s'}(\Omega)$ with the rate $2^{-j(s-s')}$ for every $s' < s$. More importantly, we can also show that $u_j$ converges strongly in $H^s(\Omega)$ to $u$ (without a rate) as we do in the following section for the case when $s > 1/2$..

## 4 Bilateral Constraints in $H^s(\Omega)$ with $s > 1/2$

We now move to a setting in which the subsets satisfy $C \subset H^s(\Omega)$ for $s \in \mathbb{R}$, $s > 1/2$. For $s \leq 1/2$, it does not make sense to approximate the function spaces with wavelets generated by continuous bases, since (amongst other things) no increase of the approximation rate can be expected. For this case we refer to the preceding discussion at the end of Section 3.1. Moreover, we note that the interpolation arguments used in our proof of Mosco convergence require $H^s(\Omega)$ to be embedded into the space of continuous functions. This needs $s > 1/2$, cf. [22]. Our intention here is to demonstrate how wavelet bases with higher regularity may also be employed to approximate bound constraints, whenever the sets $C$ are in Sobolev spaces of an order high enough to guarantee continuity via the Sobolev embedding theorem.

For the approximation procedure here, it is necessary to restrict ourselves to certain classes of bounds $\xi_0, \xi_1$; as opposed to the $L^2(\Omega)$ in which rather general bounds were allowed. Analogous to the setting with lower regularity, we define our closed convex set $C$ by

$$C := \{v \in H^s(\Omega) \,|\, \xi_0 \leq v \leq \xi_1\}.$$

However, we assume here that $\xi_0, \xi_1 \in \mathbb{R}$ such that $\xi_0 < \xi_1$. We will comment on further generalizations below.

In the sequel, we will again assume that $n = 1$, only now we use biorthogonal wavelets generated by piecewise linear B-splines, which we denote by $N_{j,k}$ (primal) and $\widetilde{N}_{j,k}$ (dual) for $k \in \Lambda_j$ and $j \geq 0$ instead of the piecewise constant scaling functions used to define Haar wavelets. For any scale $j \geq 0$, we consider the following sets:

$$C_j := \left\{ u = \sum_{k \in \Lambda_j} c_{j,k} N_{j,k} \;\middle|\; \langle \xi_0, \widetilde{N}_{j,k} \rangle \leq c_{j,k} \leq \langle \xi_1, \widetilde{N}_{j,k} \rangle \;\forall k \in \Lambda_j \right\}.$$

Furthermore, we have

$$\langle \xi_0, \widetilde{N}_{j,k} \rangle = \xi_0 \int_\Omega \widetilde{N}_{j,k} \,\mathrm{d}x = \xi_0 \int_\Omega \left( \sum_{\ell \in \Lambda_j} 2^{-j/2} N_{j,\ell} \right) \widetilde{N}_{j,k} \,\mathrm{d}x = \xi_0 2^{-j/2} \sum_{\ell \in \Lambda_j} \langle N_{j,\ell}, \widetilde{N}_{j,k} \rangle = 2^{-j/2} \xi_0.$$

Therefore, $C_j$ can also be written in the case of constant bounds as

$$C_j := \left\{ u = 2^{-j/2} \sum_{k \in \Lambda_j} c_{j,k} N_{j,k} \;\middle|\; \xi_0 \leq c_{j,k} \leq \xi_1 \quad \forall k \in \Lambda_j \right\}.$$

For general non-constant continuous bounds $\xi_0, \xi_1$ this is obviously not the case unless the $\xi_0, \xi_1$ are given by polynomials that can be reproduced by the generator basis or if a condition such as (3.4) can be guaranteed for all scales.

Finally, for any function $v \in H^s(\Omega)$ and $s > 1/2$, it follows from the Sobolev embedding theorem, that $v$ admits a continuous representative and can therefore be evaluated at each point $x_{j,k} := 2^{-j}k$ with $k \in \Lambda_j$. We may then define the following interpolation operator $L_j$ by

$$L_j v := 2^{-j/2} \sum_{k \in \Lambda_j} v_{j,k} N_{j,k}, \tag{4.1}$$

where $v_{j,k} := v(2^{-j}k)$. We first prove the following technical lemma for $L_j$.

**Lemma 9.** Under the standing assumptions, $L_j v \to v$ strongly in $H^s(\Omega)$ for every $v \in H^s(\Omega)$.

**Remark 10.** In the following proof, we make use of the Jackson inequality for fractional Sobolev spaces (3.2) (cf. Remark 8) in the context of the interpolation operator $L_j$ (4.1). We refer the reader to [7, Eq. (2.14)] and the surrounding discussion. In particular, the upper bound of 2 on $t$ arises there.

*Proof.* Fix $u \in H^s(\Omega)$ and some real $t$ such that $2 \geq t > s$. Then $L_j u := 2^{-j/2} \sum_{k \in \Lambda_j} u_{j,k} N_{j,k}$, where $u_{j,k} := u(2^{-j}k)$. Moreover, $H^t(\Omega)$ is continuously embedded into $H^s(\Omega)$. Furthermore, by construction, one readily shows that $H^t(\Omega)$ is dense in $H^s(\Omega)$. Hence, we have for every $\varepsilon > 0$ a function $u_\varepsilon \in H^t(\Omega)$ such that $\|u - u_\varepsilon\|_{H^s(\Omega)} < \varepsilon$. Next, for each scale $j \geq 0$, we have by Jackson's inequality that there exists $\widehat{C} > 0$ such that

$$\|u_\varepsilon - L_j u_\varepsilon\|_{H^s(\Omega)} \leq \widehat{C} 2^{-j(t-s)} \|u_\varepsilon\|_{H^t(\Omega)}.$$

In particular, we can choose $j_\varepsilon \geq 0$ such that $2^{-j_\varepsilon(t-s)} \|u_\varepsilon\|_{H^t(\Omega)} < \varepsilon$. Finally, it follows from [7, Eq. (2.14)] that

$$\|u - L_j u\|_{H^s(\Omega)} \leq \widetilde{C} \|u\|_{H^s(\Omega)},$$

where $\widetilde{C}$ is independent of $j$ and $u$. Consequently, we can deduce the inequality

$$\|L_j(u_\varepsilon - u)\|_{H^s(\Omega)} \leq (1 + \widetilde{C})\|u_\varepsilon - u\|_{H^s(\Omega)}.$$

It then follows from the triangle inequality that for all $\varepsilon > 0$ there exists $j_\varepsilon \geq 0$ such that for every $j \geq j_\varepsilon$ we have

$$\|u - L_j u\|_{H^s(\Omega)} \leq (2 + \max\{\widetilde{C}, \widehat{C}\})\varepsilon;$$

whence we have the assertion. □

We immediately obtain the following.

**Lemma 11.** The set of all weak accumulation points of sequences $\{u_j\}$, $u_j \in C_j$, is contained in $C$.

*Proof.* Since each $N_{j,k}$ is continuous and piecewise linear, $u_j \in C_j$ is both Lipschitz and an element of $H^s(\Omega)$ with $s < 3/2$. As argued at the end of Section 3.2, $C$ is a nonempty, closed, and convex subset of $H^s(\Omega)$ and therefore weakly closed.

Suppose now that $u_j \in C_j$. It follows by definition that

$$\sum_{k\in\Lambda_j} \langle \xi_0, \widetilde{N}_{j,k}\rangle N_{j,k} \leq u_j \leq \sum_{k\in\Lambda_j} \langle \xi_1, \widetilde{N}_{j,k}\rangle N_{j,k}.$$

In addition, we note that

$$\sum_{k\in\Lambda_j} \langle \xi_0, \widetilde{N}_{j,k}\rangle N_{j,k} = \xi_0 \sum_{k\in\Lambda_j} 2^{-j/2} N_{j,k} = L_j\xi_0 = \xi_0$$

As a result of this observation, we have for any $u_j \in C_j$ that

$$\xi_0 \leq u_j \leq \xi_1. \tag{4.2}$$

It follows then that $u_j \in C$ for each $j$. Then since $C$ is weakly closed, any weak accumulation points of the sequence $\{u_j\}$ with $u_j \in C_j$ are contained in $C$, as was to be shown.

□

**Remark 12.** We can easily generalize the previous result for more complex upper and lower bounds $\xi_1, \xi_0$, provided $\xi_1, \xi_0 \in L^2(\Omega)$. In particular, we have $P_j\xi_0 \to \xi_0$ strongly in $L^2(\Omega)$. This holds analogously for the upper bound. Now, for any $u_j \in C_j$ we obtain

$$P_j\xi_0 \leq u_j \leq P_j\xi_1. \tag{4.3}$$

This does not guarantee that $u_j \in C$. Nevertheless, we define the sequence $\{u_j\}$ such that $u_j \in C_j$ and $u_j \rightharpoonup u$ (weakly in $H^s(\Omega)$). Clearly, for each $j \geq 0$, $u_j - P_j\xi_0$ and $P_j\xi_1 - u_j$ define bounded linear functionals on $H^s(\Omega)$ via the usual $L^2(\Omega)$ inner product. Therefore, for any $\varphi \in H^s(\Omega)$ such that $\varphi \geq 0$ (pointwise almost everywhere on $\Omega$), we have the inequalities

$$0 \leq \langle u_j - P_j\xi_0, \varphi\rangle \text{ and } 0 \leq \langle P_j\xi_1 - u_j, \varphi\rangle.$$

Passing to the limit in $j$, we obtain

$$0 \leq \langle u - \xi_0, \varphi \rangle \text{ and } 0 \leq \langle \xi_1 - u, \varphi \rangle.$$

Consequently, $u \in C$.

Compared to Lemma 4, the statement of Lemma 11 is strictly weaker. However, it suffices to prove Mosco convergence of the sets $\{C_j\}$ to $C$. In order to see the difference in the two statements, consider that

$$C = \bigcup_{m \in \mathbb{N}} \left[ C \cap \mathbb{B}_m(0) \right],$$

where $\mathbb{B}_m(0)$ is the closed ball of radius $m$ in $H^s(\Omega)$. Indeed, let $u \in C \cap \mathbb{B}_m(0)$ for some $m$, then $u \in H^s(\Omega)$ and $\xi_0 \leq u \leq \xi_1$ (pointwise almost everywhere), i.e., $u \in C$. The union of all such sets is in $C$. On the other hand, if $u \in C$, then $\xi_0 \leq u \leq \xi_1$ and there exists $m \in \mathbb{N}$ such that $\|u\|_{H^s(\Omega)} \leq m$. It follows that $u \in C \cap \mathbb{B}_m(0)$. Following this line of thought, let

$$C_{j,m} := \left\{ u = \sum_{k \in \Lambda_j} c_{j,k} N_{j,k} \;\middle|\; \langle \xi_0, \widetilde{N}_{j,k} \rangle \leq c_{j,k} \leq \langle \xi_1, \widetilde{N}_{j,k} \rangle \; \forall k \in \Lambda_j \text{ and } \|u\|_{H^s(\Omega)} \leq m \right\} \quad m \in \mathbb{N}.$$

Then, fixing $m \in \mathbb{N}$, we have an analogous result to Lemma 4.

**Proposition 13.** Fix $m \in \mathbb{N}$. Then any sequence $\{u_j\}$ such that $u_j \in C_{j,m}$ for all $j \in \mathbb{N}$ contains a weakly convergent subsequence $\{u_{j_\ell}\}$ such that $u_{j_\ell} \rightharpoonup \overline{u}$ as $\ell \to +\infty$ and $\overline{u} \in C \cap \mathbb{B}_m(0)$, i.e., $\overline{u} \in C$.

*Proof.* Any sequence as defined in the hypothesis is automatically bounded uniformly in $H^s(\Omega)$. Therefore, there exists a subsequence of $\{u_j\}$, denoted still by the index $j$, that converges weakly in $H^s(\Omega)$. Denote the limit point by $\overline{u}$. Due to the weak lower semicontinuity of the $H^s(\Omega)$-norm, we have

$$\|\overline{u}\|_{H^s(\Omega)} \leq \liminf_j \|u_j\|_{H^s(\Omega)} \leq m.$$

Furthermore, since $N_{j,k} \geq 0$, we may proceed as in the proof of Lemma 4 to argue that $\xi_0 \leq u_j \leq \xi_1$. The rest follows as in the proof of Lemma 11. $\square$

Proposition 13 is obviously not the same result as Lemma 4. However, it is impossible to ensure that the weak derivatives of the $u_j$ remain bounded using the bound constraints alone, as there is nothing to ensure that the $u_j$ do not rapidly oscillate with increasing frequency as $j \to +\infty$. In the context of optimal control, optimization or constrained variational problems, it is often possible to obtain an implicit bound on the correct norm due to the properties of the objective function. We refer the reader to our example in Section 5.

Our next result takes advantage of the higher regularity properties of $u$ in the current setting along with Jackson's inequality.

**Lemma 14.** For every $u \in C$ there exists a sequence $\{u_j\}$ such that $u_j \in C_j$ and $u_j \to u$ as $j \to +\infty$.

*Proof.* Let $u \in C$. As we discussed above, since $n = 1$ and $s > 1/2$, it follows from the Sobolev embedding theorem that $u$ is continuous up to the boundary on $\Omega$. Therefore, by taking the sample points $2^{-j}k$ for $k \in \Lambda_j$ we then have

$$\xi_0(2^{-j}k) = \xi_0 \leq u(2^{-j}k) \leq \xi_1 = \xi_1(2^{-j}k).$$

As a result, $L_j u \in C_j$ for each $j$. Then by Lemma 9, $L_j u \to u$ and the assertion follows. □

**Theorem 15.** The sequence of convex sets $\{C_j\} \subset H^s(\Omega)$ converges in the sense of Mosco to $C$.

*Proof.* This is a direct consequence of Lemmas 11 and 14. □

# 5 Applications and Experiments

## 5.1 An Application to a Fractional Obstacle Problem

In this final section, we consider an application of the approximation results to the discretization of a fractional obstacle problem, cf. [41]. As discussed in [41] and outlined in more detail in [10, 17], the highest expected value $u^\star$ of a perpetual American put option, i.e., an open-ended contract that allows you to sell a given underlying asset at an agreed upon strike price $K > 0$ from the time of signing until the option is exercised, can be modeled by the solution of a fractional obstacle problem.

More specifically, we assume that the value of the underlying asset $X_t$ at time $t \geq 0$ is real-valued and driven by an $\alpha$-stable symmetric Lévy distribution with initial value $X_0 = x \in \mathbb{R}$. We furthermore assume that the expected payoff function is given by $\mathbb{E}[e^{-t}\eta(X_t)]$ with $\eta(x) := \max(0, K - \exp(x))$. Then $u^\star$ is the unique solution of the variational problem

$$\min\left\{\frac{1}{2}\|u\|^2_{H^s(\mathbb{R})} \text{ over } u \in H^s(\mathbb{R}) \mid u(x) \geq \eta(x) \text{ a.e. } x \in \mathbb{R}\right\},$$

assuming $s = \alpha/2$ and that the feasible set is augmented by the condition that $\lim_{|x|\to+\infty} u(x) = 0$. If we were then to modify this problem by making the not too unrealistic assumption that the values of the underlying are in a bounded interval $\Omega \subset \mathbb{R}$, then we could consider as an approximation the problem:

$$\min\left\{\frac{1}{2}\|u\|^2_{H^s(\Omega)} \text{ over } u \in H^s(\Omega) \mid u(x) \geq \eta(x) \text{ a.e. } x \in \Omega\right\}. \tag{5.1}$$

Even with this modification, the first-order optimality conditions for (5.1) contain the non-local operator $(-\Delta)^s + I$. Here, $(-\Delta)^s$ is the (regional) fractional Laplace operator given as the variational derivative of the (squared) Gagliardo seminorm associated with $H^s(\Omega)$.

There are a number of ways to define $(-\Delta)^s$. One particularly popular way of treating problems in which $(-\Delta)^s$ arises is to lift the domain $\Omega$ and consequently, the entire problem, one dimension higher, see [12]. This removes the non-locality of the differential operator at the price of increasing the underlying dimension and losing the boundedness of the domain. The new differential operator is now only uniformly elliptic if we define the problem on a corresponding Muckenhoupt-weighted Sobolev space on the infinite half-cylinder $\Omega \times \mathbb{R}_+ \setminus \{0\}$. In the context of fractional obstacle problems such as (5.1), this is the approach taken for the numerical approximation by finite elements, see e.g., [38].

In contrast to the lifting approach, our results from Sections 3 and 4 make it possible to attack (5.1) directly using wavelet-based discretizations. For the sake of discussion, we let $\alpha = 1/2$ (and $\beta = 1$). These are specific to the input of the Levy process, and not to be confused with the wavelet discussion in Section 2.1.These constants correspond to the underlying $X_t$ following a standard Lévy distribution. In this case, (5.1) is to be considered in the space $H^{1/4}(\Omega)$. This allows us to characterize any feasible function in the continuous setting directly via its wavelet transform using Haar wavelets. For $\alpha > 1$, we could use the biorthogonal B-spline approach. Furthermore, if we fix a maximum scale $j \in \mathbb{N}$, then we first can approximate (5.1) in the single-scale basis using the generator functions $\{\varphi_{j,k}\}_{k\in\Lambda_j}$. In particular, we approximate the feasible set in the single-scale $j$ by the constraints:

$$c_{j,k} \geq \langle \eta, \varphi_{j,k} \rangle \quad k \in \Lambda_j.$$

Using these coefficient vectors $c_j$, we could write (5.1) in the space $V_j$. However, this would be equivalent to a finite element discretization using piecewise constant finite elements on a uniform (dyadic) grid on $\Omega$, which would not provide any computational benefits. On the other hand, there exists a linear transformation $W_j$, i.e., the fast wavelet transform, that transforms $c_j$ into a set of wavelet, i.e., multiscale coefficients $d_\ell \in \mathbb{R}^{|\Lambda_\ell|}$ for each $\ell \in \mathbb{N}$, $0 \leq \ell \leq j-1$. This essential fact provides us with an equivalent objective function via

$$\frac{1}{2}\|u_j\|^2_{H^s(\Omega)} \eqsim \frac{1}{2} \sum_{0\leq\ell\leq j-1} \sum_{k\in\Lambda_\ell} 2^{2s\ell}|d_{\ell,k}|^2 = \frac{1}{2} \sum_{0\leq\ell\leq j-1} \sum_{k\in\Lambda_\ell} 2^{2s\ell}|W_j c_j|^2_{\ell,k}$$

Consequently, we obtain a family of approximations of (5.1) indexed by the maximum scale $j$:

$$\min\left\{ \frac{1}{2} \sum_{0\leq\ell\leq j-1} \sum_{k\in\Lambda_\ell} 2^{\ell/2}|W_j c_j|^2_{\ell,k} \text{ over } c_j \in \mathbb{R}^{|\Lambda_j|} \;\middle|\; c_{j,k} \geq \langle \eta, \varphi_{j,k} \rangle \text{ for } k \in \Lambda_j \right\}. \tag{5.2}$$

Combining our findings in Sections 3 and 4, we can argue that feasible sets $C_j$ converge in the sense of Mosco to the original feasible set. Then by classical results on the approximation of elliptic variational inequalities, see e.g., [25, Theorem 5.2], the unique optimal solution to (5.2) converges strongly in $H^{1/4}(\Omega)$ to the optimal solution of (5.1).

**Example 16** (Conic Constraint and Forcing Term)**.** In this first example, we consider a simple modification of (5.1):

$$\min\left\{ \frac{1}{2}\|u\|^2_{H^s(\Omega)} - \langle f, u \rangle_{-s,s} \text{ over } u \in H^s(\Omega) \mid u(x) \geq 0 \text{ a.e. } x \in \Omega \right\}. \tag{5.3}$$

Here, we set $f := [(-\Delta)^s + I](\varphi)]$. For our numerical experiments, we choose two instances of $\varphi$

$$\varphi^1(x) := \max\{0, -3(x-0.5)+0.25\} \text{ and } \varphi^2(x) := \begin{cases} \max\{0, 1.0-(0.5-x)^(1/9)\}, & x \in [0.25, 0.5) \\ \max\{0, 1.0-(x-0.5)^(1/9)\}, & x \in [0.5, 0.75) \\ 0.0, & \text{else.} \end{cases}$$

Notice that $\varphi^1$ is globally Lipschitz and piecewise smooth except for two kinks and $\varphi^2$ is piecewise smooth with two discontinuities and a third non-differentiability at 0.5. Both $\varphi^i$, $i = 1, 2$, are feasible for (5.3) and both solve the unconstrained problem. Hence, they are exact solutions of (5.3). This allows us to investigate the rate of convergence.

**Example 17** (Pricing a Perpetual American Put Option)**.** In this example, we consider (5.2) using the payoff $\eta(x) = \max\{0, K - \exp(x)\}$ and strike $K = \exp(0.5)$.

### 5.2 Numerical Experiments

In this final section, we discuss the the performance of the proposed discretization and optimization algorithm using Examples 16 and 17 for illustration. We highlight here a few points: The evaluation of the gradient and discrete projection onto the feasible set has the same (linear) complexity O($|\Lambda_j|$) as the wavelet transform. The evaluation of the objective has quadratic complexity O($|\Lambda_j|^2$). Therefore, fast first-order methods of convex optimization are a viable option to solve (5.2), cf. [5] and the many references therein. As an alternative, one could solve (5.2) with a semismooth Newton approach as in [29] or [43]. From an implementation standpoint, this would require matrix slicing and the solution of (sparse) linear systems at each step. For our numerical experiments, we use a standard projected gradient method with backtracking line search allowing for step size increases as outlined in [8]. In addition, we use a nested-scale approach by first solving (5.2) for a given scale $j \in \mathbb{N}$ and then (via nearest-neighbor interpolation) using the interpolated solution (scaled by $2^{-1/2}$) as a starting point for the subsequent scale $j+1$. This has proven to be very effective and requires no linear system solves.

All experiments were implemented in the Julia language [9] version 1.2.0-1 on a 2016 MackBook Pro with 3,1 GHz Intel Core i5 processor, 16 GB 2133 MHz LPDDR3 memory, running macOS Version 10.12.6. All graphics were generated using the GR() backend. In both examples we set $\Omega = (0,1)$ for simplicity.

For ease of discussion, we denote the objective function by $f(c_j)$ and the optimal solution $c_j^\star$. We set $m_j := |\Lambda_j|$ for readability and $c_j^k(\alpha) := \mathrm{Proj}_{C_j}(c_j^k - \alpha \nabla f(c_j^k))$. We recall the sufficient decrease condition used in the line search: Given the current iterate $c_j^k \in \mathbb{R}^{m_j}$, we set $\alpha_l^k = m^l \zeta$, $l = 1, 2, ...$, until

$$f(c_j^k(\alpha_l^k)) \leq f(c_j^k) - \sigma(a_l^k)^{-1} \|c_j^k(\alpha_l^k) - c_j^k\|^2_{\mathbb{R}^{m_j}},$$

at which point we set $c_j^{k+1} := c_j^k(\alpha_l^k)$. In our implementation, we always set: $\sigma = 0.1$ and $m = 0.5$. At the start of a new scale $j$, we set $\zeta = 1.0$ otherwise we use the previously accepted stepsize $\alpha_l^k$. Our implementation also allows for lengthening the step sizes by using the rule $\zeta/m^l$. The line search updates stop and accept the current step if $l = 30$ and the overall algorithm was set to stop if $k = 5000$. We did not observe either of these behaviors.

The results of our experiments can be seen in Tables 1 and 2 as well as Figures 1, 2, and 3. Starting with Example 16, we see that the algorithm converges in both cases in less than 200 iterations with the number of iterations on the highest scale less than 25 iterations in both cases. We calculated the function space norms by first projecting the solution $c_j^\star$ onto $V_{14}$, i.e., the finite-dimensional function space spanned by the translated and dilated generator functions for the scale $j = 14$. For both the $L^2$- and $H^s$-norms we exploited the norm equivalence used in our discretization scheme. This required the calculation of the single-scale coefficients for the true solutions $\varphi^i$. Although the potentially scale-dependent residual $\|c_j - \mathrm{Proj}_{\mathbb{R}^n_+}(c_j - \nabla f(c_j))\|_{\mathbb{R}^n}$ was used as a stopping criterion, we note that the relative rate of change of the objective functions:

$$f^k_{\mathrm{rate}} := |f(c_j^{k+1}) - f(c_j^k)| / |f(c_j^1) - f(c_j^0)|$$

appeared independent of scale in the sense that the smallest index $k$ for which this $f^k_{\mathrm{rate}}$ <1e-3 was only very mildly dependent on $j$; increasing by 1-2 iterations on average with each new $j$. This was consistent in both Examples 16 and 17 and is one way of viewing scale (mesh) dependence in optimal control problems with bound constraints since $f$ is strongly convex, cf. [31]. As expected by the theory, the smoother solution associated with $\varphi^1$ exhibits a faster rate of convergence than the discontinuous solution for $\varphi^2$, see Figures 1 and 2. This is consistent with the theoretical estimates based on Jackson's inequality given above.

For Example 17 no explicit solution was available for comparison to obtain a rate of convergence. In order to remedy this, we used the solution $u^\star_{14}$ obtained by projecting $c^\star_{14}$ onto $V_{14}$ and compared the rate of convergence in the $L^2$-norm. This was done exactly using the properties of the single-scale expansions and the $L^2$-inner product. See Figure 3. Here, it appears that the rate of convergence is also on the order of $2^{-j}$. The algorithm converged in 264 iterations, but only required 20 iterations on the finest scale. For $j = 1, \dots, 14$, $f^k_{\text{rate}} <$ 1e-3 was reached in no more than 13 iterations for each scale with no indication of scale dependence. Finally, in contrast to Example 16 it is more difficult to see the true active sets. In order to remedy this, we included Figure 4, which shows $\log_{10}|u^\star_{14}(x) - \eta(x)|$ for sample points $x_p \in [0, 1]$ such that $x_p = p*$1e-5, $p = 1, \dots, 1e5$. We conclude that any function values less than 1e-4 are most likely active in the true continuous solution.

| | total iter | $j_{\max}$ iter | $\|c_j - \text{Proj}_{\mathbb{R}^n_+}(c_j - \nabla f(c_j))\|_{\mathbb{R}^n}$ | $\|u^\star_j - \varphi^i\|_{L^2}$ | $\|u^\star_j - \varphi^i\|_{H^{1/4}}$ |
|---|---|---|---|---|---|
| $\varphi^1$ | 111 | 7 | 8.8982e-5 | 7.8025e-5 | 9.5287e-3 |
| $\varphi^2$ | 182 | 23 | 9.3283e-5 | 1.2206e-3 | 5.4127e-2 |

Table 1: Results for Example 16. "total iter" is the sum total of all iteration counts over the scales from $j = 1, \dots, 15$. "$j_{\max}$ iter" is the total number of iterations needed for $j = 14$. $\|c_j - \text{Proj}_{\mathbb{R}^n_+}(c_j - \nabla f(c_j))\|_{\mathbb{R}^n}$ is the Euclidean norm of the residual at the optimal solution $c_j$. $\|u^\star_j - \varphi^i\|_{L^2}$ and $\|u^\star_j - \varphi^i\|_{H^{1/4}}$ are the errors of the projected optimal solution $u^\star_j$ to the true solution $\varphi^i$ for $j = 14$. The algorithm was stopped once $\|c^k_j - \text{Proj}_{\mathbb{R}^n_+}(c^k_j - \nabla f(c^k_j))\|_{\mathbb{R}^n} <$ 1e-4 for the current iterate $c^k_j$.

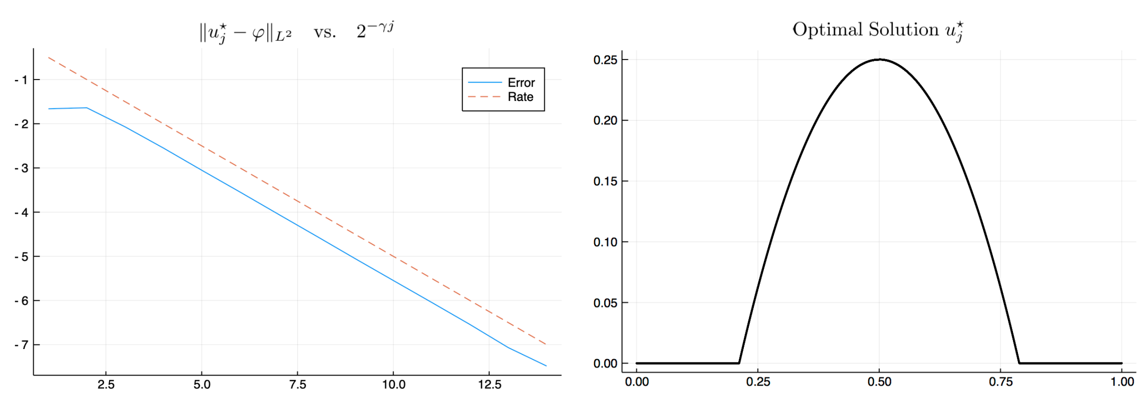


Figure 1: Plots for Example 16 using $\varphi^1$: (l.) The error $\|u^\star_j - \varphi^1\|_{L^2}$ as a function of scale $j$. The $x$-axis corresponds to scale $j$, the $y$-axis is $\log_2$-scale. We use $\gamma = 1$ for comparison (dashed line, "Rate" in plot). (r.) The estimated continuous solution $u^\star_j$ obtained by projecting onto the finite dimensional function space using the optimal coefficients $c^\star_j$. We use here $j = 14$.

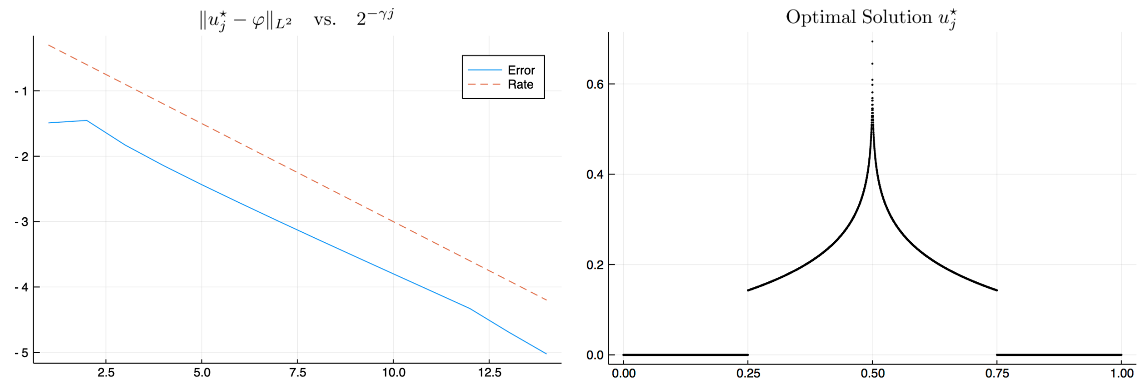


Figure 2: Plots for Example 16 using $\varphi^2$: (l.) The error $\|u_j^\star - \varphi^2\|_{L^2}$ as a function of scale $j$. The $x$-axis corresponds to scale $j$, the $y$-axis is $\log_2$-scale. We use $\gamma = 0.3$ for comparison (dashed line, "Rate" in plot). (r.) The estimated continuous solution $u_j^\star$ obtained by projecting onto the finite dimensional function space using the optimal coefficients $c_j^\star$. We use here $j = 14$.

| total iter | $j_{\max}$ iter | $\|c_j - \mathrm{Proj}_{\mathbb{R}^n_+}(c_j - \nabla f(c_j))\|_{\mathbb{R}^n}$ |
|---|---|---|
| 264 | 20 | 8.5174e-5 |

Table 2: Results for Example 17. "total iter" is the sum total of all iteration counts over the scales from $j = 1, \ldots, 14$. "$j_{\max}$ iter" is the total number of iterations needed for $j = 14$. $\|c_j - \mathrm{Proj}_{\mathbb{R}^n_+}(c_j - \nabla f(c_j))\|_{\mathbb{R}^n}$ is the Euclidean norm of the residual at the optimal solution $c_j$. The algorithm was stopped once $\|c_j^k - \mathrm{Proj}_{\mathbb{R}^n_+}(c_j^k - \nabla f(c_j^k))\|_{\mathbb{R}^n} <$ 1e-4 for the current iterate $c_j^k$.

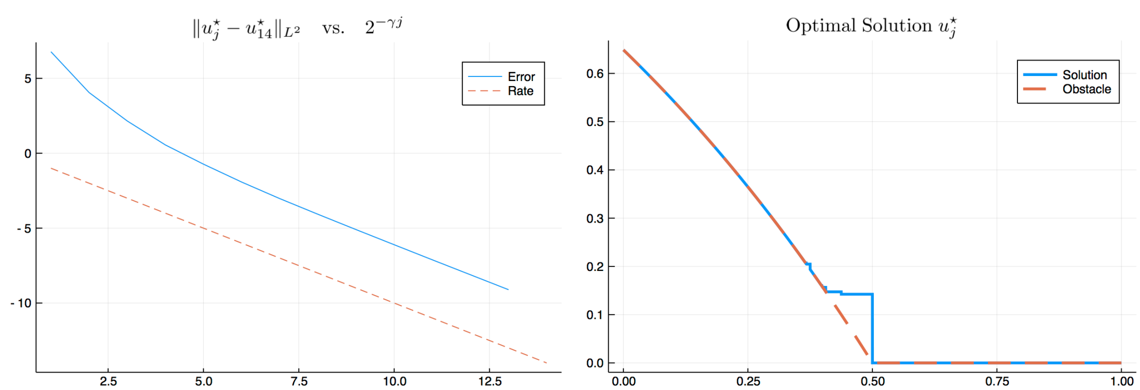


Figure 3: Plots for Example 17: (l.) The error $\|u_j^\star - u_{14}^\star\|_{L^2}$ as a function of scale $j$. The $x$-axis corresponds to scale $j$, the $y$-axis is $\log_2$-scale. We use $\gamma = 1.0$ for comparison (dashed line, "Rate" in plot). The true continuous solution $u_{14}^\star$ is estimated by projecting onto the finite dimensional function space using the optimal coefficients $c_{14}^\star$. (r.) Optimal solution $u_{14}^\star$ (bold line, "Solution" in plot) versus the payoff obstacle $\eta$ (dashed line, "Obstacle" in plot).

# 6 Conclusion

We have shown that it is indeed possible to discretize bound constraints relevant to optimal control, PDE-constrained optimization, and variational inequalities using wavelet-based methods. Though wavelets have been used in the context of optimal control before, see e.g., [11, 26, 33], we are only aware of the paper [28], in which inequality constraints appear in a slightly different, but clearly related context, for a variational prob-

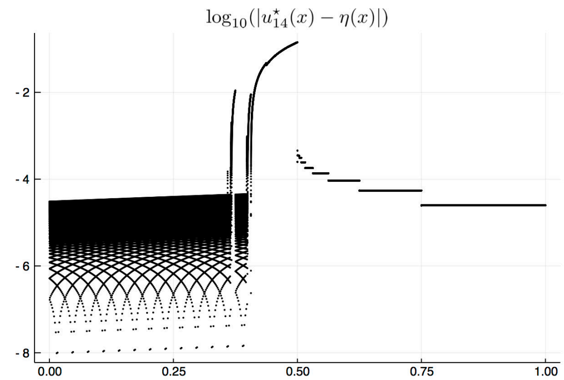


Figure 4: Estimation of the true active set in Example 17. All values below 1e-4 are most likely active in the true solution. The $y$-axis is $\log_{10}$.

lem in negative-order fractional Sobolev spaces. In the theoretical sections, we demonstrated the crucial set convergence results and provided rates of convergence for situations involving functions with higher regularity. As noted in the remarks above, these arguments can be easily extended to related function spaces, higher dimensions, and more complex domains, which makes the approach very attractive for practical problems. In addition, the final section of our paper demonstrated clear computational benefits for using our proposed scheme to solve variational problems involving non-local operators.

# References

[1] L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via $\Gamma$-convergence. *Comm. Pure Appl. Math.*, 43(8):999–1036, 1990.

[2] H. Attouch. *Variational convergence for functions and operators*. Applicable Mathematics Series. Pitman Advanced Publishing Program, 1984.

[3] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces*, volume 6 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006.

[4] J.-P. Aubin and H. Frankowska. *Set-valued analysis*. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA, 2009. Reprint of the 1990 edition [MR1048347].

[5] A. Beck. *First-order methods in optimization*, volume 25 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2017.

[6] S. J. Benson and T. S. Munson. Flexible complementarity solvers for large-scale applications. *Optim. Methods Softw.*, 21(1):155–168, 2006.

[7] S. Bertoluzza. An adaptive collocation method based on interpolating wavelets. In *Multiscale wavelet methods for partial differential equations*, volume 6 of *Wavelet Anal. Appl.*, pages 109–135. Academic Press, San Diego, CA, 1997.

[8] D. P. Bertsekas. On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Trans. Automatic Control*, AC-21(2):174–184, 1976.

[9] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.

[10] S. I. Boyarchenko and S. Z. Levendorskii. Perpetual American options under Lévy processes. *SIAM J. Control Optim.*, 40(6):1663–1696, 2002.

[11] C. Burstedde and A. Kunoth. Fast iterative solution of elliptic control problems in wavelet discretization. *J. Comput. Appl. Math.*, 196(1):299–319, 2006.

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

[13] Charles K. Chui. *An introduction to wavelets*, volume 1 of *Wavelet Analysis and its Applications*. Academic Press, Inc., Boston, MA, 1992.

[14] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70(233):27–75, 2001.

[15] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods. II. Beyond the elliptic case. *Found. Comput. Math.*, 2(3):203–245, 2002.

[16] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45(5):485–560, 1992.

[17] R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.

[18] S. Dahlke. Multiscale approximation. In Willi Freeden, M. Zuhair Nashed, and Thomas Sonar, editors, *Handbook of Geomathematics*, pages 1221–1241. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[19] W. Dahmen and R. Schneider. Composite wavelet bases for operator equations. *Math. Comp.*, 68(228):1533–1567, 1999.

[20] W. Dahmen and R. Schneider. Wavelets on manifolds. I. Construction and domain decomposition. *SIAM J. Math. Anal.*, 31(1):184–230, 1999.

[21] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[22] E. Di Nezza, G. Palatucci, and E. Valdinoci. Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.*, 136(5):521–573, 2012.

[23] J. C. Dunn. Global and asymptotic convergence rate estimates for a class of projected gradient processes. *SIAM J. Control Optim.*, 19(3):368–400, 1981.

[24] R. S. Falk. Approximation of a class of optimal control problems with order of convergence estimates. *J. Math. Anal. Appl.*, 44:28–47, 1973.

[25] R. Glowinski. *Lectures on numerical methods for nonlinear variational problems*, volume 65 of *Tata Institute of Fundamental Research Lectures on Mathematics and Physics*. Tata Institute of Fundamental Research, Bombay; sh Springer-Verlag, Berlin-New York, 1980. Notes by M. G. Vijayasundaram and M. Adimurthi.

[26] M. D. Gunzburger and A. Kunoth. Space-time adaptive wavelet methods for optimal control problems constrained by parabolic evolution equations. *SIAM J. Control Optim.*, 49(3):1150–1170, 2011.

[27] A. Haar. Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.*, 69(3):331–371, 1910.

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

[29] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13(3):865–888 (2003), 2002.

[30] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York, 2009.

[31] C. T. Kelley and E. W. Sachs. Mesh independence of the gradient projection method for optimal control problems. *SIAM J. Control Optim.*, 30(2):477–493, 1992.

[32] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*, volume 31 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1980 original.

[33] A. Kunoth. Adaptive wavelet schemes for an elliptic control problem with Dirichlet boundary control. *Numer. Algorithms*, 39(1-3):199–220, 2005.

[34] J.-L. Lions. *Optimal control of systems governed by partial differential equations*. Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin, 1971.

[35] S. Mallat. *A wavelet tour of signal processing*. Academic Press, Inc., San Diego, CA, 1998.

[36] Y. Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. Translated from the 1990 French original by D. H. Salinger.

[37] U. Mosco. Convergence of convex sets and of solutions of variational inequalities. *Advances in Math.*, 3:510–585, 1969.

[38] R. H. Nochetto, E. Otárola, and A. J. Salgado. Convergence rates for the classical, thin and fractional elliptic obstacle problems. *Philos. Trans. Roy. Soc. A*, 373(2050):20140449, 14, 2015.

[39] J.-F. Rodrigues. *Obstacle problems in mathematical physics*, volume 134 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1987. Notas de Matemática [Mathematical Notes], 114.

[40] C. Schwab and C. J. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.

[41] L. Silvestre. Regularity of the obstacle problem for a fractional power of the Laplace operator. *Comm. Pure Appl. Math.*, 60(1):67–112, 2007.

[42] F. Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.

[43] M. Ulbrich. Semismooth Newton methods for operator equations in function spaces. *SIAM J. Optim.*, 13(3):805–842 (2003), 2002.

[44] P. Wojtaszczyk. *A mathematical introduction to wavelets*, volume 37 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1997.