Preprint No. 1239, Department of Mathematics, University of Utrecht, May 2002. Submitted to SIAM J. Numer. Anal.

ADAPTIVE SOLUTION OF OPERATOR EQUATIONS USING WAVELET FRAMES

ROB STEVENSON

Abstract. In "Adaptive Wavelet Methods II - Beyond the Elliptic case" of Cohen, Dahmen and DeVore ([CDD00]), an adaptive method has been developed for solving general operator equations. Using a *Riesz basis* of wavelet type for the energy space, the operator equation is transformed into an equivalent matrix-vector system. This system is solved iteratively, where the application of the infinite stiffness matrix is replaced by an adaptive approximation. Assuming that the stiffness matrix is sufficiently compressible, i.e., that it can be sufficiently well appproximated by sparse matrices, it was proved that the adaptive method has optimal computational complexity in the sense that it converges with the same rate as the best N-term approximation for the solution assuming it would be explicitly available. The condition concerning compressibility requires that, dependent on their order, the wavelets have sufficiently many vanishing moments, and that they are sufficiently smooth. Yet, except on tensor product domains, wavelets that satisfy this smoothness requirement are difficult to construct.

In this paper we write the domain or manifold on which the operator equation is posed as an *overlapping* union of subdomains, each of them being the image under a smooth parametrization of the hypercube. By lifting wavelets on the hypercube to the the subdomains we obtain a *frame* for the energy space. With this frame the operator equation is transformed into a matrix-vector system, after which this system is solved iteratively by an adaptive method similar to the one from [CDD00]. With this approach, frame elements that have sufficiently many vanishing moments and are sufficiently smooth, which is needed for the compressibility, are easily constructed. By handling additional difficulties due to the fact that a frame gives rise to an underdetermined matrix-vector system, we prove that this adaptive method has optimal computational complexity.

1. Introduction

For some boundedly invertible $L : H \to H'$, where H is some Hilbert space with dual H', and some $g \in H'$, we consider the problem of finding $u \in H$ such that

$$Lu = g.$$

As typical examples we think of linear differential- or integral equations in variational form. Although also systems of such equations fit into the framework, for ease of exposition in this introduction let us consider scalar equations so that H is typically a Sobolev space H^t of some order $t \in \mathbb{R}$.

Date: May 21, 2002.

2000 *Mathematics Subject Classification.* 41A25, 41A46, 65F10, 65N12, 65N55, 65T60.

Key words and phrases. Operator equations, adaptive methods, wavelets, frames, optimal computational complexity, best N-term approximation.

Assuming that we have a Riesz *basis* Ψ for H^t available, which we formally view as a column vector, by writing $u = \mathbf{u}^T\Psi$ above problem is equivalent to finding $\mathbf{u} \in \ell_2$ satisfying the infinite matrix-vector system

$$\mathbf{M}\mathbf{u} = \mathbf{g},$$

where $\mathbf{M} := \langle \Psi, L\Psi\rangle : \ell_2 \to \ell_2$ is boundedly invertible, and $\mathbf{g} := \langle \Psi, g\rangle \in \ell_2$. Here $\langle\, ,\, \rangle$ denotes the duality product on (H^t, H^{-t}).

Let us denote by $\mathbf{u}_N$ a *best N-term approximation* for $\mathbf{u}$, i.e., a vector with at most N non-zero coefficients that has distance to $\mathbf{u}$ less or equal to that of any vector with a support of that size. Note that $\|u - \mathbf{u}_N^T\Psi\|_{H^t} \eqsim \|\mathbf{u} - \mathbf{u}_N\|_{\ell_2}$. Considering bases Ψ of sufficiently smooth *wavelet* type, the theory of nonlinear approximation learns us ([DeV98, Coh00]) that if both

$$0 < s < \tfrac{d-t}{n},$$

where d is the *order* of the wavelets and n is the space dimension, and u is in the *Besov space* $B^{sn+t}_\tau(L_\tau)$ with $\tau = (\frac{1}{2} + s)^{-1}$, then

$$\sup_{N\in\mathbb{N}} N^s \|\mathbf{u} - \mathbf{u}_N\|_{\ell_2} < \infty.$$

The condition here involving Besov regularity is much milder that the corresponding condition $u \in H^{sn+t}$ involving Sobolev regularity that would be needed to guarantee the same rate of convergence with linear approximation in the span of N wavelets corresponding to the 'coarsest levels'. Indeed, assuming a sufficiently smooth right-hand side, for several boundary value problems it has been proved that the solution has a much higher Besov- than Sobolev regularity ([DD97, Dah99a]). Note that a rate higher than $\frac{d-t}{n}$ can never be expected with wavelets of order d, except when the solution u happens to be a finite linear combination of wavelets.

So far we discussed the approximation of $\mathbf{u}$, which however is only implicitly given as the solution of $\mathbf{M}\mathbf{u} = \mathbf{g}$. In [CDD01, CDD00], an iterative adaptive method for solving this system has been developed that given a tolerance $\varepsilon > 0$ yields an approximate solution $\mathbf{u}_\varepsilon$ with $\|\mathbf{u} - \mathbf{u}_\varepsilon\| \le \varepsilon$, where the number of operations and storage locations it requires is of the same order as the lenght of the smallest best N-term approximation for $\mathbf{u}$ on distance ε, meaning that the method has *optimal computational complexity.*

When L and thus $\mathbf{M}$ are symmetric and positive definite, the method consists of the application of the simple damped Richardson iteration onto the infinite system, where the multiplication of $\mathbf{M}$ with the current, finitely supported approximation vector for $\mathbf{u}$ is replaced by an adaptive approximation. In each iteration, each column of $\mathbf{M}$ is replaced by a finitely supported approximation with a tolerance that decreases as function of the modulus of the corresponding entry in the vector. Note that even for a differential operator the matrix $\mathbf{M}$ is not sparse due to the interaction between wavelets from different levels. A second ingredient of the method is the application after each K steps, with K being some fixed number, of a clean-up or coarsening procedure that removes the smallest entries from the current approximation in order to ensure an optimal work-accuracy balance.

For nonsymmetric or indefinite $\mathbf{M}$, one can simply apply the adaptive method to the normal equations, or alternatively one can apply more advanced iterations which may lead to quantitatively better results ([CDD00, DDU01, DUV02]).

The proof of the optimality of the method requires that $\mathbf{M}$ is sufficient *compressible*, meaning that given some tolerance $\delta > 0$, there exists another infinite matrix on distance less than δ, which in each row and column has only a finite, and sufficiently small number of non-zero entries. For large classes of differential- and integral operators this property can indeed be verified when, dependent on the order d, the wavelets have *sufficiently many vanishing moments* and are *sufficiently smooth* (see also [Ste02]).

The bottleneck for the application of this adaptive wavelet method is the availability of suitable wavelet bases on general, non-rectangular domains or manifolds. An approach to construct wavelet bases is to write the domain as a *non-overlapping* union of subdomains, that are the images of the hypercube under smooth parametrizations. Wavelets, or 'initial stable completions', living on the hypercube are lifted to the subdomains. Since in general more than one subdomain is needed, there is the problem of 'stitching' functions over the interfaces.

The approach from [DS99b] yields wavelet bases that in principal satisfy all requirements. Yet, since suitable extension operators from one subdomain into neighbouring subdomains enter the construction, it seems not easy to implement. The approaches from [DS99a, CTU99, CM00] yield wavelets which over the interfaces between subdomains are only continuous. For example thinking of a differential equation of order 2 on a two-dimensional domain, with this restricted smoothness only for orders $d \leq 2$ sufficient compressibility of the matrix $\mathbf{M}$ can be shown. Yet, with these low orders an adaptive method can at most give a small improvement in the order of convergence compared to non-adaptive methods, which in practice might not compensate for the overhead it requires.

Again because of their lack of smoothness beyond continuity, also finite element wavelets as constructed in [DS99c, CES00, Ste00] seem not very suited for the adaptive method.

The approach followed in this paper is to apply an *overlapping* decomposition of the domain or manifold into subdomains. By lifting wavelets on the hypercube to those subdomains, and by multiplying them by smooth weight functions that vanish at the internal boundaries of these subdomains, a countable set of functions is obtained, that we again denote by Ψ, which is dense in H^t and which for each $u \in H^t$ yields *some* representation $u = \mathbf{u}^T\Psi$ with $\|u\|_{H^t} \eqsim \|\mathbf{u}\|_{\ell_2}$. Such a set Ψ is called a *frame* for H^t. By writing $u = \mathbf{u}^T\Psi$, solving $Lu = g$ is again equivalent to solving $\mathbf{Mu} = \mathbf{g}$, where $\mathbf{M} = \langle \Psi, L\Psi\rangle$ and $\mathbf{g} = \langle \Psi, g\rangle$. Yet, due to the overlapping decomposition the representation $u = \mathbf{u}^T\Psi$ will not be unique, and so the system $\mathbf{Mu} = \mathbf{g}$ will have more solutions, that however all correspond to the unique solution of $Lu = g$.

When L is symmetric and positive definite, $\mathbf{M}$ is symmetric and *semi*-positive definite, and $\mathbf{Mu} = \mathbf{g}$ can be solved by the damped Richardson iteration. In each iteration the norm of the defect is reduced by a constant factor less than one. For nonsymmetric or indefinite L, the iteration can be applied to the normal equations.

Following [CDD00] for the Riesz basis case, in the practical algorithm the application of $\mathbf{M}$ will be replaced by the adaptive approximation. To be able to prove that the method has optimal computational complexity, again it is needed that $\mathbf{M}$ is sufficiently compressible, i.e., that dependent on the order d, the wavelets have sufficiently many vanishing moments and are sufficiently smooth. Since its construction does not involve stitching of functions over interfaces, the advantage of this frame approach is that these conditions concerning vanishing moments and smoothness are easily satisfied.

Furthermore, because of the multiplication with the weight functions, boundary conditions at the internal boundaries of the subdomains can be chosen at ones convenience. In particular, in case of a closed manifold, this gives the additional advantage that all wavelet bases on the hypercube can be chosen to satisfy *periodic boundary conditions.* Such bases are the most easy to implement, and they have much better quantitative properties than available wavelet bases satisfying other boundary conditions.

A final advantage is that generally an overlapping domain decomposition is much easier to construct with simpler parametrizations than a non-overlapping one, which might also give a favourable quantitative effect.

The use of a frame instead of a Riesz basis gives also rise to a problem: Since in the adaptive method the matrix-vector product is replaced by an adaptive approximation, each time it is invoked it gives an error that might have a component in the, non-trivial, kernel of $\mathbf{M}$. Also the clean-up or coarsening step may introduce such components. Just because these components are in the kernel of $\mathbf{M}$, they will not be affected by subsequent Richardson steps, meaning that in the cause of the iteration the component of the current approximation in the kernel of $\mathbf{M}$ may increase. Although this component has no influence on the obtained approximation for the solution of $Lu = g$, that is, after forming the series with the frame elements, it might be responsible for the major part of the computational costs of each iteration. Indeed, recall that in the adaptive approximation of the matrix-vector product the accuracy with which the columns of $\mathbf{M}$ are approximated is determined by the moduli of the corresponding entries in the vector.

Under some technical assumption on the frame, specifically on the projector, called $\mathbf{Q}$, onto the complement of the kernel of $\mathbf{M}$ in ℓ_2, we will prove that above effect will not occur or only to such an extent that also in the frame case the adaptive method has optimal computational complexity. Unfortunately, although we expect it to hold more generally, for our frame construction based on overlapping decompositions, so far we could give a complete proof that this technical assumption holds in only one situation that $t = 0$ and that the wavelet bases on the hypercube are L_2-orthogonal.

Above problem lead us to introduce a modified adaptive algorithm to which a projection step is added that is applied before each coarsening step. This projector only affects the redundant representation in the overlap regions in a way that the component of the current approximation in the kernel of $\mathbf{M}$ is controlled. The projector itself, called $\mathbf{P}$, is given by an infinite matrix, and in the algorithm, as $\mathbf{M}$, it is only applied approximately using the adaptive matrix-vector product. We show that $\mathbf{P}$ is sufficiently compressible, and prove that this modified algorithm has optimal computational complexity in the general case.

This paper is organized as follows: In §2, we recall the concept of a frame, and show how it can be used to transform an operator equation into an infinite, underdetermined matrix-vector equation. We discuss iterative schemes to solve such equations. Next, we replace those ingredients of such schemes that involve infinite vectors or matrices by practical realizable approximations, and show that they, together with a coarsening routine, give rise to a convergent algorithm **SOLVE**. In addition, we introduce a convergent modified algorithm **modSOLVE** that contains the inexact application of a projector $\mathbf{P}$ that explicitly controls size of the component of the current approximation in the kernel of $\mathbf{M}$.

In §3 we study the *rate* of convergence and the computational costs of both algorithms. First we recall some theory dealing with best N-term approximation. We formulate a condition on the compressibility of the stiffness matrix $\mathbf{M}$, and for **modSOLVE** also of $\mathbf{P}$, that for $\mathbf{M}$ is known to be satisfied for wavelets that, dependent on their order, have sufficiently many vanishing moments and are sufficiently smooth. Under these conditions, it is proved that both **SOLVE** and **modSOLVE** have optimal computational complexity, where for **SOLVE** we need the aforementioned assumption on the projector $\mathbf{Q}$.

In §4 we outline the construction of suitable frames using overlapping domain decompositions. Having specified the construction of a frame, we now discuss the condition on $\mathbf{Q}$. Furthermore, we define a suitable $\mathbf{P}$ and show that it is sufficiently compressible.

In order to avoid the repeated use of generic but unspecified constants, in this paper by $C \lesssim D$ we mean that C can be bounded by a multiple of D, independently of parameters which C and D may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

2. The basic concept

2.1. **Frames.** Let H be a separable real Hilbert space. A countable collection $\Psi \subset H$ is called a *frame* for H when there exist two positive constants A_Ψ, B_Ψ such that

$$A_\Psi \|f\|_{H'}^2 \leq \|f(\Psi)\|^2 \leq B_\Psi \|f\|_{H'}^2, \qquad (f \in H'). \tag{2.1}$$

Here with $f(\Psi)$ we mean the sequence $(f(\psi))_{\psi\in\Psi}$, with $\|f(\Psi)\|$ denoting its ℓ_2-norm. We adapted the definition of a frame given in [Dau92, §3] by identifying H with its dual H' via the Riesz mapping. As a consequence of (2.1), the frame operators

$$F : H' \to \ell_2 : f \mapsto f(\Psi),$$

and its dual

$$F' : \ell_2 \to H : \mathbf{c} \mapsto \mathbf{c}^T\Psi$$

are bounded with norm less or equal to $B_\Psi^{\frac{1}{2}}$. Here we used $\mathbf{c}^T\Psi$ as shorthand notation for $\sum_{\psi\in\Psi} c_\psi \psi$. The composition $F'F : H' \to H$ is boundedly invertible with $\|(F'F)^{-1}\|_{H'\leftarrow H} \leq A_\Psi^{-1}$. The collection $\tilde{\Psi} := (F'F)^{-1}\Psi$ is a frame for H' (the "canonical" dual frame) with frame operators

$$\tilde{F} := F(F'F)^{-1}, \;\; \tilde{F}' = (F'F)^{-1}F'$$

and frame constants B_Ψ^{-1}, A_Ψ^{-1}.

The property of Ψ being a frame for H with constants A_Ψ, B_Ψ can be shown to be equivalent to $\operatorname{clos}\operatorname{span}\Psi = H$ and

$$B_\Psi^{-1}\|u\|_H^2 \leq \inf_{\mathbf{c}\in\ell_2,\, F'\mathbf{c}=u} \|\mathbf{c}\|^2 \leq A_\Psi^{-1}\|u\|_H^2, \qquad (u\in H). \tag{2.2}$$

We have

$$\ell_2 = \operatorname{Ran} F \oplus^\perp \operatorname{Ker} F',$$

and

$$\mathbf{Q} := F(F'F)^{-1}F' : \ell_2 \to \ell_2 \tag{2.3}$$

is the orthogonal projector onto $\operatorname{Ran} F$. The frame Ψ is a Riesz basis for H iff $\operatorname{Ker} F' = 0$.

2.2. **Transformation of an operator equation to an ℓ_2-problem.** Let $L : H \to H'$ be a boundedly invertible linear operator, and let Ψ be a frame for H. Given a $g \in H'$, we consider the problem of finding $u \in H$ such that

$$Lu = g. \tag{2.4}$$

As examples, one may think of L as being a linear differential- or integral operator in variational form that defines a homeomorphism between a relevant Sobolev space, or a closed subspace of that, and its dual. A possible construction of a frame will be discussed in §4.1.

Apart from scalar equations also systems of differential- and/or integral equations fit into this framework. Examples can be found e.g. in [CDD00, §3]. In this case, H is a product of relevant Sobolev spaces, and it can be equipped with a frame defined as the product of frames for the coordinate spaces.

Writing $u = F'\mathbf{u}$ for some $\mathbf{u} \in \ell_2$, $\mathbf{u}$ satisfies

$$\mathbf{M}\mathbf{u} = \mathbf{g},$$

where

$$\mathbf{M} := FLF' \quad \text{and} \quad \mathbf{g} := Fg.$$

From

$$\left.\begin{aligned} \tilde{F}L^{-1}\tilde{F}'FLF' &= \tilde{F}F' \\ FLF'\tilde{F}L^{-1}\tilde{F}' &= F\tilde{F}' \end{aligned}\right\} = \mathbf{Q} = \mathbf{id} \quad \text{on } \operatorname{Ran} F,$$

we conclude that $\mathbf{M}|_{\operatorname{Ran} F} : \operatorname{Ran} F \to \operatorname{Ran} F$ is boundedly invertible, with $\|\mathbf{M}\| \leq B_\Psi\|L\|_{H'\leftarrow H}$ and $\|\mathbf{M}|_{\operatorname{Ran} F}^{-1}\| \leq A_\Psi^{-1}\|L^{-1}\|_{H\leftarrow H'}$, whereas $\operatorname{Ker}\mathbf{M} = \operatorname{Ker} F'$.

2.3. **Iterative schemes to solve the infinite dimensional system $\mathbf{Mu} = \mathbf{g}$.** In case L is *symmetric* and *positive definite*, i.e., $L' = L$ and $\inf_{0\neq v\in H}(Lv)(v)/\|v\|^2 > 0$, then $\mathbf{M} = \mathbf{M}^* \geq 0$. With $\lambda_{\max} := \lambda_{\max}(\mathbf{M}) = \|\mathbf{M}\|$ and $\lambda_{\min}^+ := \lambda_{\min}(\mathbf{M}|_{\operatorname{Ran} F}) = \|\mathbf{M}|_{\operatorname{Ran} F}^{-1}\|^{-1}$, for $0 < \alpha < 2/\lambda_{\max}$, we consider the *damped Richardson iteration*

$$\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} - \alpha(\mathbf{M}\mathbf{u}^{(i)} - \mathbf{g}). \tag{2.5}$$

From $\mathbf{u} - \mathbf{u}^{(i+1)} = (\mathbf{id} - \alpha\mathbf{M})(\mathbf{u} - \mathbf{u}^{(i)})$ and $\mathbf{QM} = \mathbf{MQ}$, we infer that

$$\|\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i+1)})\| \leq \rho\|\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i)})\|, \tag{2.6}$$

where $\rho := \|(\mathbf{id} - \alpha\mathbf{M})|_{\operatorname{Ran} F}\| = \max\{\alpha\lambda_{\max} - 1, 1 - \alpha\lambda^+_{\min}\} < 1$, with minimum $\frac{\kappa-1}{\kappa+1}$ when $\alpha = 2/(\lambda_{\max} + \lambda^+_{\min})$, where $\kappa = \lambda_{\max}/\lambda^+_{\min}$. Note that $u - F'\mathbf{u}^{(i)} = F'\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i)})$.

We will study an inexact version of (2.5) in which the application of the infinite matrix $\mathbf{M}$ is approximated. A difficulty will be that errors made in $\ker F'$ are not reduced in subsequent iterations. Although obviously these errors do not affect $F'\mathbf{u}^{(i)}$, they might hamper a cheap, but sufficiently accurate matrix-vector multiplication. Under some condition on $\mathbf{Q}$, i.e., on the frame, we will prove that this is not the case, in the sense that these errors do not pile up too much.

For handling cases where $\mathbf{Q}$ might not satisfy this condition, we consider a modified algorithm that contains the explicit application of a projector to reduce error components in $\operatorname{Ker} F'$: Let $\mathbf{P} : \ell_2 \to \ell_2$ be some bounded projector with

$$\operatorname{Ker} \mathbf{P} = \operatorname{Ker} F',$$

so that $\ell_2 = \operatorname{Ran} \mathbf{P} \oplus \operatorname{Ker} F'$ is a 'stable' splitting. Let $\mathbf{u}^{(i+1)}$ denote the result of applying $\mathbf{P}$ to the result of K damped Richardson iterations starting with $\mathbf{u}^{(i)}$. Using $\mathbf{MPu} = \mathbf{g}$ and $\mathbf{P}(\mathbf{id} - \mathbf{Q}) = 0$, we arrive at

$$\mathbf{Pu} - \mathbf{u}^{(i+1)} = \mathbf{P}(\mathbf{id} - \alpha\mathbf{M})^K(\mathbf{Pu} - \mathbf{u}^{(i)}) = \mathbf{P}(\mathbf{id} - \alpha\mathbf{M})^K\mathbf{Q}(\mathbf{Pu} - \mathbf{u}^{(i)}), \tag{2.7}$$

and so

$$\|\mathbf{Pu} - \mathbf{u}^{(i+1)}\| \le \|\mathbf{P}\|\rho^K\|\mathbf{Pu} - \mathbf{u}^{(i)}\|,$$

showing convergence when K is chosen such that $\|\mathbf{P}\|\rho^K < 1$. Note that $u - F'\mathbf{u}^{(i)} = F'(\mathbf{Pu} - \mathbf{u}^{(i)})$.

Except when the condition number κ is close to one, the Richardson iteration is known to converge relatively slow, and quantitatively better results can be expected by more advanced iterations. Yet, for simplicity we confine the analysis to the easiest algorithm.

The case of L being non-symmetric or indefinite can be treated by considering the *normal equations*

$$\mathbf{M}^*\mathbf{Mu} = \mathbf{M}^*\mathbf{g}. \tag{2.8}$$

Both $\mathbf{M}|_{\operatorname{Ran} F}$ and $\mathbf{M}^*|_{\operatorname{Ran} F}$ are boundedly invertible on $\operatorname{Ran} F$ and so is $\mathbf{M}^*\mathbf{M}|_{\operatorname{Ran} F}$, whereas $\mathbf{M}^*\mathbf{M}(\operatorname{Ker} F') = 0$. By redefining $\lambda_{\max} = \lambda_{\max}(\mathbf{M}^*\mathbf{M}) = \|\mathbf{M}\|^2$ and $\lambda^+_{\min} = \lambda_{\min}(\mathbf{M}^*\mathbf{M}|_{\operatorname{Ran} F}) = \|\mathbf{M}|^{-1}_{\operatorname{Ran} F}\|^{-2}$, the damped Richardson iteration, possibly alternated with the projection step, can now be applied to solve (2.8).

In this paper, we confine the analysis to the symmetric positive definite case. Yet, following the lines of [CDD00, §7], everything that will be said about the SPD case can be easily generalized to the iteration applied to (2.8).

As an alternative for saddle-point problems, in [CDD00, DDU01, DUV02] the Uzawa algorithm or a reformulation as a positive definite system are studied, with the aim to obtain quantitatively better algorithms by avoiding the squaring of the condition number κ. It can be expected that also these methods can be based on frames.

2.4. **Approximate iterations.** Obviously, since in actual computations neither we can handle the generally infinite vector $\mathbf{g}$, nor we can apply the infinite matrix $\mathbf{M}$, the damped Richardson iteration, possibly alternated with the projection, is not a practical algorithm. In this section, we study convergence of the iterations in which these ingredients are approximated. Following [CDD00], we assume that we have the following routines at our disposal:

$\mathbf{RHS}[\varepsilon, \mathbf{g}] \to \mathbf{g}_\varepsilon$
determines a finitely supported $\mathbf{g}_\varepsilon \in \ell_2$ *satisfying*

$$\|\mathbf{g} - \mathbf{g}_\varepsilon\| \leq \varepsilon.$$

$\mathbf{APPLY}[\varepsilon, \mathbf{N}, \mathbf{v}] \to \mathbf{w}_\varepsilon$
determines for a finitely supported $\mathbf{v} \in \ell_2$*, and for* $\mathbf{N} = \mathbf{M}$ *(or* $\mathbf{P}$ *or* $\mathbf{M}^*$*), a finitely supported* $\mathbf{w}_\varepsilon$ *satisfying*

$$\|\mathbf{N}\mathbf{v} - \mathbf{w}_\varepsilon\| \leq \varepsilon.$$

$\mathbf{COARSE}[\varepsilon, \mathbf{v}] \to \mathbf{v}_\varepsilon$
creates, for a finitely supported $\mathbf{v} \in \ell_2$*, a vector* $\mathbf{v}_\varepsilon$ *by replacing all but* N *coefficients of* $\mathbf{v}$ *by zeros, such that*

$$\|\mathbf{v} - \mathbf{v}_\varepsilon\| \leq \varepsilon, \tag{2.9}$$

whereas N *is at most a constant multiple of the minimal value of* N *for which* (2.9) *is valid.*

In §3.1 and §3.2, we will discuss suitable realizations of **COARSE** and **APPLY** respectively. The routine **COARSE** will be necessary to obtain an optimal work/accuracy balance. The realization of **RHS** depends on the problem at hand.

Based on above routines, we consider the following inexact version of the damped Richardson iteration:

$\mathbf{SOLVE}[\varepsilon, \mathbf{M}, \mathbf{g}] \to \mathbf{u}_\varepsilon$:
Let $\theta < 1/3$ *and* $K \in \mathbb{N}$ *be fixed such that* $3\rho^K < \theta$
$i := 0$, $\mathbf{u}^{(0)} := 0$, $\varepsilon_0 := \|\mathbf{M}|^{-1}_{\mathrm{Ran}\, F}\|\, \|\mathbf{g}\|$
While $\varepsilon_i > \varepsilon$ do
    $i := i + 1$
    $\varepsilon_i := 3\rho^K \varepsilon_{i-1}/\theta$
    $\mathbf{g}^{(i)} := \mathbf{RHS}[\frac{\theta\varepsilon_i}{6\alpha K}, \mathbf{g}]$
    $\mathbf{v}^{(i,0)} := \mathbf{u}^{(i-1)}$
    For $j = 1, \ldots, K$ do
        $\mathbf{v}^{(i,j)} := \mathbf{v}^{(i,j-1)} - \alpha\left(\mathbf{APPLY}[\frac{\theta\varepsilon_i}{6\alpha K}, \mathbf{M}, \mathbf{v}^{(i,j-1)}] - \mathbf{g}^{(i)}\right)$
    od
    $\mathbf{u}^{(i)} := \mathbf{COARSE}[(1-\theta)\varepsilon_i, \mathbf{v}^{(i,K)}]$
od

$\mathbf{u}_\varepsilon := \mathbf{u}^{(i)}$

Proposition 2.1. *Let $\mathbf{u} \in \ell_2$ be some solution of $\mathbf{Mu} = \mathbf{g}$. Then the vectors $\mathbf{u}^{(i)}$, $\mathbf{v}^{(i,K)}$ produced in* **SOLVE**$[\varepsilon, \mathbf{M}, \mathbf{g}]$ *satisfy*

$$\|\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i)})\| \le \varepsilon_i, \qquad (i \ge 0), \tag{2.10}$$

and so in particular $\|\mathbf{Q}(\mathbf{u} - \mathbf{u}_\varepsilon)\| \le \varepsilon$. Furthermore,

$$\|\mathbf{Qu} + (\mathbf{id} - \mathbf{Q})\mathbf{u}^{(i-1)} - \mathbf{v}^{(i,K)}\| \le \tfrac{2}{3}\theta\varepsilon_i, \qquad (i \ge 1), \tag{2.11}$$

which will be used in §3.3.

Proof. For $i = 0$, (2.10) follows from $\mathbf{Qu} = \mathbf{M}|^{-1}_{\operatorname{Ran} F}\mathbf{g}$.

Now for an $i \ge 1$, let $\|\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i-1)})\| \le \varepsilon_{i-1}$. Since $\mathbf{MQu} = \mathbf{g}$ and $\|\mathbf{id} - \alpha\mathbf{M}\| \le 1$, we have

$$\|\mathbf{Qu} - \mathbf{v}^{(i,K)} - (\mathbf{id} - \alpha\mathbf{M})^K(\mathbf{Qu} - \mathbf{u}^{(i-1)})\| \le K(\alpha\tfrac{\theta\varepsilon_i}{6\alpha K} + \alpha\tfrac{\theta\varepsilon_i}{6\alpha K}) = \tfrac{\theta\varepsilon_i}{3}.$$

From

$$(\mathbf{id} - \alpha\mathbf{M})^K(\mathbf{Qu} - \mathbf{u}^{(i-1)}) = (\mathbf{id} - \alpha\mathbf{M})^K\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i-1)}) - (\mathbf{id} - \mathbf{Q})\mathbf{u}^{(i-1)}$$

and $\|(\mathbf{id} - \alpha\mathbf{M})^K\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i-1)})\| \le \rho^K\|\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i-1)})\| \le \rho^K\varepsilon_{i-1} = \frac{\theta\varepsilon_i}{3}$, we conclude (2.11). The definition of $\mathbf{u}^{(i)}$ now shows that $\|\mathbf{Qu} + (\mathbf{id} - \mathbf{Q})\mathbf{u}^{(i-1)} - \mathbf{u}^{(i)}\| \le (\frac{2\theta}{3} + (1 - \theta))\varepsilon_i = (1 - \frac{\theta}{3})\varepsilon_i$, and so $\|\mathbf{Q}(\mathbf{u} - \mathbf{u}^{(i)})\| \le (1 - \frac{\theta}{3})\varepsilon_i \le \varepsilon_i$. □

Remark 2.2. Compared to the Riesz basis setting discussed in [CDD00], we see that in **SOLVE** we have to pay for working with a frame. When going from $\mathbf{u}^{(i)}$ to $\mathbf{u}^{(i+1)}$, each of the "evaluation" errors made in the K intermediate steps, and in fact even the sum, should be less than the error that is allowed in $\mathbf{u}^{(i+1)}$. Indeed, since only $\|\mathbf{id} - \alpha\mathbf{M}\| \le 1$, these errors might not be reduced by the iteration.

The inexact version of the damped Richardson iteration alternated with the inexact application of the projector $\mathbf{P}$ is given by

modSOLVE$[\varepsilon, \mathbf{M}, \mathbf{g}] \to \mathbf{u}_\varepsilon$:
Let $\theta < 1/3$ and $K \in \mathbb{N}$ be fixed such that $3\rho^K\|\mathbf{P}\| < \theta$
$i := 0$, $\mathbf{u}^{(0)} := 0$, $\varepsilon_0 := \|\mathbf{P}\|\,\|\mathbf{M}|^{-1}_{\operatorname{Ran} F}\|\,\|\mathbf{g}\|$
While $\varepsilon_i > \varepsilon$ do
  $i := i + 1$
  $\varepsilon_i := 3\rho^K\|\mathbf{P}\|\varepsilon_{i-1}/\theta$
  $\mathbf{g}^{(i)} := \mathbf{RHS}[\frac{\theta\varepsilon_i}{6\alpha K\|\mathbf{P}\|}, \mathbf{g}]$
  $\mathbf{v}^{(i,0)} := \mathbf{u}^{(i-1)}$
  For $j = 1, \dots, K$ do
    $\mathbf{v}^{(i,j)} := \mathbf{v}^{(i,j-1)} - \alpha(\mathbf{APPLY}[\frac{\theta\varepsilon_i}{6\alpha K\|\mathbf{P}\|}, \mathbf{M}, \mathbf{v}^{(i,j-1)}] - \mathbf{g}^{(i)})$
  od
  $\mathbf{z}^{(i)} := \mathbf{APPLY}[\frac{\theta\varepsilon_i}{3}, \mathbf{P}, \mathbf{v}^{(i,K)}]$
  $\mathbf{u}^{(i)} := \mathbf{COARSE}[(1 - \theta)\varepsilon_i, \mathbf{z}^{(i)}]$
od

$\mathbf{u}_\varepsilon := \mathbf{u}^{(i)}$

Proposition 2.3. *Let $\mathbf{u} \in \ell_2$ be some solution of $\mathbf{Mu} = \mathbf{g}$. Then the vectors $\mathbf{u}^{(i)}$, $\mathbf{z}^{(i)}$ produced in* **modSOLVE**$[\varepsilon, \mathbf{M}, \mathbf{g}]$ *satisfy*

$$\|\mathbf{Pu} - \mathbf{u}^{(i)}\| \leq \varepsilon_i, \qquad (i \geq 0), \tag{2.12}$$

and so in particular $\|\mathbf{Pu} - \mathbf{u}_\varepsilon\| \leq \varepsilon$. Furthermore,

$$\|\mathbf{Pu} - \mathbf{z}^{(i)}\| \leq \theta\varepsilon_i, \qquad (i \geq 1), \tag{2.13}$$

which will be used in §3.3.

Proof. For $i = 0$, (2.12) follows from $\mathbf{Pu} = \mathbf{PM}|^{-1}_{\operatorname{Ran} F}\mathbf{g}$.

Now for an $i \geq 1$, let $\|\mathbf{Pu} - \mathbf{u}^{(i-1)}\| \leq \varepsilon_{i-1}$. Since $\mathbf{MPu} = \mathbf{g}$ and $\|\mathbf{id} - \alpha\mathbf{M}\| \leq 1$, we have

$$\|\mathbf{Pu} - \mathbf{v}^{(i,K)} - (\mathbf{id} - \alpha\mathbf{M})^K(\mathbf{Pu} - \mathbf{u}^{(i-1)})\| \leq K(\alpha\tfrac{\theta\varepsilon_i}{6\alpha K\|\mathbf{P}\|} + \alpha\tfrac{\theta\varepsilon_i}{6\alpha K\|\mathbf{P}\|}) = \tfrac{\theta\varepsilon_i}{3\|\mathbf{P}\|}.$$

From $(\mathbf{id} - \alpha\mathbf{M})^K = (\mathbf{id} - \alpha\mathbf{M})^K\mathbf{Q} + \mathbf{id} - \mathbf{Q}$ and $\mathbf{P}(\mathbf{id} - \mathbf{Q}) = 0$, we have

$$\|\mathbf{Pu} - \mathbf{Pv}^{(i,K)} - \mathbf{P}(\mathbf{id} - \alpha\mathbf{M})^K\mathbf{Q}(\mathbf{Pu} - \mathbf{u}^{(i-1)})\| \leq \tfrac{\theta\varepsilon_i}{3},$$

and so

$$\|\mathbf{Pu} - \mathbf{z}^{(i)} - \mathbf{P}(\mathbf{id} - \alpha\mathbf{M})^K\mathbf{Q}(\mathbf{Pu} - \mathbf{u}^{(i-1)})\| \leq \tfrac{2\theta\varepsilon_i}{3}.$$

Using $\|\mathbf{P}(\mathbf{id} - \alpha\mathbf{M})^K\mathbf{Q}\| \leq \|\mathbf{P}\|\rho^K$, we conclude that

$$\|\mathbf{Pu} - \mathbf{z}^{(i)}\| \leq \tfrac{2\theta\varepsilon_i}{3} + \|\mathbf{P}\|\rho^K\varepsilon_{i-1} = \theta\varepsilon_i,$$

and so $\|\mathbf{Pu} - \mathbf{u}^{(i)}\| \leq \theta\varepsilon_i + (1-\theta)\varepsilon_i = \varepsilon_i$. □

3. Convergence rates and computational costs

3.1. **Best N-term approximation and COARSE.** To assess the efficiency of **SOLVE** or **modSOLVE**, following [CDD00] we will consider the following benchmark: Suppose that for some solution $\mathbf{u} \in \ell_2$ of $\mathbf{Mu} = \mathbf{g}$ we would have *all* coefficients available. Then the most economical approximation for $\mathbf{u}$ on distance less than ε would be $\mathbf{u}_N$, defined by replacing all but the N largest coefficients in modulus of $\mathbf{u}$ by zeros, with $N = N(\varepsilon, \mathbf{u})$ being the smallest integer such that

$$\|\mathbf{u} - \mathbf{u}_N\| \leq \varepsilon. \tag{3.1}$$

For $N \in \mathbb{N}$, the vector $\mathbf{u}_N$ is called the *best N-term approximation* for $\mathbf{u}$. If for some $s > 0$,

$$\|\mathbf{u} - \mathbf{u}_N\| \lesssim N^{-s}, \qquad (N \in \mathbb{N}), \tag{3.2}$$

then $N = N(\varepsilon, \mathbf{u})$ from (3.1) would satisfy $N(\varepsilon, \mathbf{u}) \lesssim \varepsilon^{-1/s}$.

Assuming (3.2), in §3.3 we will prove that for the vector $\mathbf{u}_\varepsilon$ produced by **(mod)SOLVE**, it holds that $\#\operatorname{supp}\mathbf{u} \lesssim \varepsilon^{-1/s}$, whereas the number of floating point operations to compute it is of the same order. In view of (3.2), we may conclude that **(mod)SOLVE** is of optimal computational complexity.

The question *whether*, and if so, *for which* s (3.2) is valid, is related to properties of the frame and the (Besov-) regularity of the solution $u \in H$ of the operator equation (2.4). It will be discussed in §4.2.

Vectors $\mathbf{u} \in \ell_2$ that satisfy (3.2) can be characterized as follows ([DeV98]): Let $\gamma_n(\mathbf{u})$ denote the nth largest coefficient in modulus of $\mathbf{u}$. For $0 < \tau < 2$, the space ℓ_τ^w is defined by

$$\ell_\tau^w = \{\mathbf{u} \in \ell_2 : |\mathbf{u}|_{\ell_\tau^w} := \sup_n n^{1/\tau}|\gamma_n(\mathbf{u})| < \infty\}.$$

It is easily verified that $\ell_\tau \hookrightarrow \ell_\tau^w \hookrightarrow \ell_{\tau+\delta}$ for any $\delta \in (0, 2-\tau]$, justifying why ℓ_τ^w is called *weak* ℓ_τ. The expression $|\mathbf{u}|_{\ell_\tau^w}$ only defines a quasi-norm since it does not necessarily satisfy the triangle inequality. Yet, for each $0 < \tau < 2$, there exists a $C_1(\tau) > 0$ with

$$|\mathbf{v} + \mathbf{w}|_{\ell_\tau^w} \le C_1(\tau)\,(|\mathbf{v}|_{\ell_\tau^w} + |\mathbf{w}|_{\ell_\tau^w}), \qquad (\mathbf{v}, \mathbf{w} \in \ell_\tau^w), \tag{3.3}$$

or equivalently ([BL76, Lemma 3.10.1]), for $\mu = \mu(\tau) > 0$ sufficiently small it holds that

$$|\mathbf{v} + \mathbf{w}|_{\ell_\tau^w}^\mu \le |\mathbf{v}|_{\ell_\tau^w}^\mu + |\mathbf{w}|_{\ell_\tau^w}^\mu, \qquad (\mathbf{v}, \mathbf{w} \in \ell_\tau^w).$$

With these ℓ_τ^w-spaces at hand, it can be shown that the property (3.2) is equivalent to $\mathbf{u} \in \ell_\tau^w$, with τ related to s according to $\tau = (\frac{1}{2} + s)^{-1}$. In particular for each $\tau \in (0, 2)$,

$$\sup_N N^s \|\mathbf{u} - \mathbf{u}_N\| \eqsim |\mathbf{u}|_{\ell_\tau^w}. \tag{3.4}$$

The routine $\mathbf{v}_\varepsilon = \mathbf{COARSE}[\varepsilon, \mathbf{v}]$ might be defined by taking $\mathbf{v}_\varepsilon = \mathbf{v}_N$ with N being the smallest integer such that $\|\mathbf{v} - \mathbf{v}_N\| \le \varepsilon$. Yet, since the determination of the best N-term approximation requires sorting all elements of $\mathbf{v}$ by their modulus, this algorithm cannot be implemented in linear time. It requires the order of $(\#\operatorname{supp}\mathbf{v}) \cdot \log(\#\operatorname{supp}\mathbf{v})$ operations, with $\#\operatorname{supp}\mathbf{v}$ denoting the number of non-zero coefficients of $\mathbf{v}$.

Following ideas from [Bar02, Met02] we use a routine **COARSE** with which this log-*factor* is avoided:

COARSE$[\varepsilon, \mathbf{v}] \to \mathbf{v}_\varepsilon$:

- $q := \lceil \log((\#\operatorname{supp}\mathbf{v})^{1/2}\|\mathbf{v}\|/\varepsilon) \rceil$.
- *Devide the elements of* $\mathbf{v}$ *into sets* $V_0, \dots, V_q$*, where for* $0 \le i \le q-1$*,* V_i *contains the elements with modulus in* $(2^{-i-1}\|\mathbf{v}\|, 2^{-i}\|\mathbf{v}\|]$*, and possible remaining elements are put into* V_q*.*
- *Create* $\mathbf{v}_\varepsilon$ *by extracting elements first from* V_0 *and when it is empty from* V_1 *and so forth, until* $\|\mathbf{v} - \mathbf{v}_\varepsilon\| \le \varepsilon$*.*

The value of q is chosen such that the sum of squares of the elements in V_q is less or equal to ε^2, meaning that the last element added to $\mathbf{v}_\varepsilon$ (assuming that this vector is non-zero) originates from V_i for some $i < q$. Since then also $\mathbf{v}_N$ with $\|\mathbf{v} - \mathbf{v}_N\| \le \varepsilon$ must contain

elements from this V_i, and since within each V_i the squared values of the elements differ at most a factor 4, we obtain the following result:

Proposition 3.1. *For $\mathbf{v}_\varepsilon$ yielded by above routine, it holds that $\|\mathbf{v}-\mathbf{v}_\varepsilon\|\le\varepsilon$ and*

$$\#\operatorname{supp}\mathbf{v}_\varepsilon \le 4\min\{N : \|\mathbf{v}-\mathbf{v}_N\|\le\varepsilon\} \tag{3.5}$$

meaning that it defines a valid procedure **COARSE**. *The number of operations needed for this routine is of the order*

$$\#\operatorname{supp}\mathbf{v}+q \lesssim \#\operatorname{supp}\mathbf{v}+\log(\varepsilon^{-1}\|\mathbf{v}\|). \tag{3.6}$$

Later, it will appear that the latter log-*term* is harmless.

Below in Proposition 3.2, we recall a crucial result proven in [CDD01]. It shows that for any fixed $\theta<1/3$, a finitely supported approximation of a target vector in ℓ_τ^w can always be coarsened such that the resulting approximation has an error that is at most $1/\theta$ times the original error, whereas the size of its support is at most some fixed multiple of that of the best N-term approximation with that error. Although this result was proven for best N-term approximations, from (3.5) it is obvious that it is also valid for the current routine **COARSE**.

Proposition 3.2 ([CDD01, Corollary 5.2])**.** *Let $\theta<1/3$, $\tau\in(0,2)$ and $\tau=(\frac12+s)^{-1}$. Then for any $\varepsilon>0$, $\mathbf{v}\in\ell_\tau^w$, and finitely supported $\mathbf{w}\in\ell_2$ with*

$$\|\mathbf{v}-\mathbf{w}\|\le\theta\varepsilon,$$

for $\overline{\mathbf{w}}=\mathbf{COARSE}[(1-\theta)\varepsilon,\mathbf{w}]$ it holds that

$$\#\operatorname{supp}\overline{\mathbf{w}}\lesssim\varepsilon^{-1/s}|\mathbf{v}|_{\ell_\tau^w}^{1/s},$$

and obviously $\|\mathbf{v}-\overline{\mathbf{w}}\|\le\varepsilon$.

Remark 3.3. In [CDD01, Corollary 5.2] this result was formulated for $\theta=1/5$. However an inspection of the proof, and an easy generalization of [CDD01, (5.4)] concerning thresholding, shows the result for any $\theta<1/3$. Applying **COARSE** with a θ larger than $1/5$ might give a quantitative improvement of **(mod)SOLVE**, since then it increases the error with a smaller factor. It is easily seen that in any case Proposition 3.2 can not be valid for $\theta>1/2$.

Controlling the sizes of the supports of approximations of an ℓ_τ^w-function relative to their errors, implies controlling their ℓ_τ^w-(quasi-)norms. Indeed, an easy application of the next proposition shows that in the situation of Proposition 3.2, in addition we have that

$$|\overline{\mathbf{w}}|_{\ell_\tau^w}\le C_2(\tau)|\mathbf{v}|_{\ell_\tau^w}, \tag{3.7}$$

for some constant $C_2(\tau)$ independent of ε.

Proposition 3.4 ([CDD01, Lemma 4.11])**.** *Let $\tau\in(0,2)$ and $\tau=(\frac12+s)^{-1}$. Then for any $\mathbf{v}\in\ell_\tau^w$, and finitely supported $\mathbf{z}\in\ell_2$, we have*

$$|\mathbf{z}|_{\ell_\tau^w}\lesssim|\mathbf{v}|_{\ell_\tau^w}+(\#\operatorname{supp}\mathbf{z})^s\|\mathbf{v}-\mathbf{z}\|$$

Proof. For convenience we recall the short proof. Let $N = \#\mathrm{supp}\,\mathbf{z}$, then

$$|\mathbf{z}|_{\ell_\tau^w} \lesssim |\mathbf{z}-\mathbf{v}_N|_{\ell_\tau^w} + |\mathbf{v}_N|_{\ell_\tau^w} \lesssim (2N)^s\|\mathbf{z}-\mathbf{v}_N\| + |\mathbf{v}|_{\ell_\tau^w},$$

where we used $\#\mathrm{supp}(\mathbf{z}-\mathbf{v}_N) \le 2N$ and (3.4). The proof is completed by

$$\|\mathbf{z}-\mathbf{v}_N\| \le \|\mathbf{z}-\mathbf{v}\| + \|\mathbf{v}-\mathbf{v}_N\| \le 2\|\mathbf{z}-\mathbf{v}\|.$$

□

3.2. **Requirements on the infinite dimensional system.** In order to be able to show that **(mod)SOLVE** has optimal computational complexity, we will have to impose some conditions on the matrix $\mathbf{M}$, and for **modSOLVE** also on $\mathbf{P}$, as well as on the right-hand side $\mathbf{g}$. Our treatment closely follows [CDD01, CDD00], except that, following ideas from [Bar02, Met02], we avoid some log-factors in the operations count due to sorting.

Definition 3.5. Let $s^* > 0$. A bounded $\mathbf{N} : \ell_2 \to \ell_2$ is called s^*-*admissable*, when for a suitable routine **APPLY**, for each $s \in (0, s^*)$, for all $\varepsilon > 0$ and finitely supported vectors $\mathbf{v}$, with $\mathbf{w}_\varepsilon = \mathbf{APPLY}[\varepsilon, \mathbf{N}, \mathbf{v}]$ the following is valid:

(I) $\#\mathrm{supp}\,\mathbf{w}_\varepsilon \lesssim \varepsilon^{-1/s}|\mathbf{v}|_{\ell_\tau^w}^{1/s}$,
(II) the number of arithmetic operations used to compute it is at most a fixed multiple of $\varepsilon^{-1/s}|\mathbf{v}|_{\ell_\tau^w}^{1/s} + \#\mathrm{supp}\,\mathbf{v}$,

Remark 3.6. Let $\mathbf{N}$ be s^*-admissable. Then for any $s \in (0, s^*)$, with $\tau = (\frac{1}{2}+s)^{-1}$, $\mathbf{N} : \ell_\tau^w \to \ell_\tau^w$ is bounded. Indeed, let $\mathbf{v} \in \ell_\tau^w$. Part (I) from Definition 3.5 can be written as $\#\mathrm{supp}\,\mathbf{w}_\varepsilon \le C\varepsilon^{-1/s}|\mathbf{v}|_{\ell_\tau^w}^{1/s}$ for some constant C. For any $N \in \mathbb{N}$, take $\varepsilon = C^s|\mathbf{v}|_{\ell_\tau^w}N^{-s}$, or equivalently, $N = C\varepsilon^{-1/s}|\mathbf{v}|_{\ell_\tau^w}^{1/s}$. Let $(\mathbf{Nv})_N$ denote the best N-term approximation for $\mathbf{Nv}$. Then

$$N^s\|\mathbf{Nv} - (\mathbf{Nv})_N\| \le N^s\|\mathbf{Nv}-\mathbf{w}_\varepsilon\| \le N^s\varepsilon = C^s|\mathbf{v}|_{\ell_\tau^w},$$

showing $|\mathbf{Nv}|_{\ell_\tau^w} \lesssim |\mathbf{v}|_{\ell_\tau^w}$ by (3.4).

Secondly, for any $s \in (0, s^*)$, and $\tau = (\frac{1}{2}+s)^{-1}$, the mapping $\mathbf{v} \mapsto \mathbf{w}_\varepsilon := \mathbf{APPLY}[\varepsilon, \mathbf{N}, \mathbf{v}]$ is bounded on ℓ_τ^w uniformly in $\varepsilon > 0$. Indeed Proposition 3.4, Part (I) of Definition 3.5 and the boundedness of $\mathbf{N}$ demonstrated above show that

$$|\mathbf{w}_\varepsilon|_{\ell_\tau^w} \lesssim |\mathbf{Nv}|_{\ell_\tau^w} + (\#\mathrm{supp}\,\mathbf{w}_\varepsilon)^s\|\mathbf{NV}-\mathbf{w}_\varepsilon\| \lesssim |\mathbf{Nv}|_{\ell_\tau^w} + |\mathbf{v}|_{\ell_\tau^w} \lesssim |\mathbf{v}|_{\ell_\tau^w}.$$

It will turn out that a matrix is s^*-admissable when it is s^*-compressible, a property that can be verified for the matrices at hand.

Definition 3.7. Let $s^* > 0$. A bounded $\mathbf{N} : \ell_2 \to \ell_2$ is called s^*-*compressible*, when for each $j \in \mathbb{N}$ there exist constants α_j and C_j, and an infinite matrix $\mathbf{N}_j$ having at most $\alpha_j 2^j$ non-zero entries in each column, such that

$$\|\mathbf{N}-\mathbf{N}_j\| \le C_j, \tag{3.8}$$

$(\alpha_j)_{j\in\mathbb{N}}$ is summable, and for any $s < s^*$, $(C_j 2^{sj})_{j\in\mathbb{N}}$ is summable.

For s^*-compressible $\mathbf{N}$, we will make use of the following routine **APPLY**:

APPLY$[\varepsilon, \mathbf{N}, \mathbf{v}] \to \mathbf{w}_\varepsilon$:

- $q := \lceil \log((\#\operatorname{supp} \mathbf{v})^{1/2}\|\mathbf{v}\|\|\mathbf{N}\|2/\varepsilon)\rceil$.
- *Devide the elements of* $\mathbf{v}$ *into sets* $V_0, \ldots, V_q$*, where for* $0 \le i \le q-1$*,* V_i *contains the elements with modulus in* $(2^{-i-1}\|\mathbf{v}\|, 2^{-i}\|\mathbf{v}\|]$*, and possible remaining elements are put into* V_q*.*
- *For* $k = 0, 1, \ldots$*, generate vectors* $\mathbf{v}_{[k]}$ *by subsequently extracting* $2^k - \lfloor 2^{k-1}\rfloor$ *elements from* $\cup_i V_i$*, starting from* V_0 *and when it is empty continuing with* V_1 *and so forth, until for some* $k = \ell$ *either* $\cup_i V_i$ *becomes empty or*

$$\|\mathbf{N}\|\|\mathbf{v} - \sum_{k=0}^{\ell} \mathbf{v}_{[k]}\| \le \varepsilon/2. \tag{3.9}$$

  In both cases $\mathbf{v}_{[\ell]}$ *may contain less than* $2^\ell - \lfloor 2^{\ell-1}\rfloor$ *elements.*
- *Compute the smallest* $j \ge \ell$ *such that*

$$\sum_{k=0}^{\ell} C_{j-k}\|\mathbf{v}_{[k]}\| \le \varepsilon/2. \tag{3.10}$$

- *For* $0 \le k \le \ell$*, compute the non-zero entries in the matrices* $\mathbf{N}_{j-k}$ *which have a column index in common with one of the entries of* $\mathbf{v}_{[k]}$*, and compute*

$$\mathbf{w}_\varepsilon := \mathbf{N}_j\mathbf{v}_{[0]} + \mathbf{N}_{j-1}\mathbf{v}_{[1]} + \ldots + \mathbf{N}_{j-\ell}\mathbf{v}_{[\ell]}. \tag{3.11}$$

The sizes of the entries of $\mathbf{v}$ determine the accuracy with which the corresponding columns of $\mathbf{N}$ are approximated, which justifies why we speak about an *adaptive* solution method.

Proposition 3.8. *For* $\mathbf{w}_\varepsilon$ *yielded by above routine, indeed we have*

$$\|\mathbf{N}\mathbf{v} - \mathbf{w}_\varepsilon\| \le \varepsilon.$$

Moreover, when $\mathbf{N}$ *is* s^**-compressible, this* **APPLY** *realizes* (I), (II) *of Definition 3.5, and so* $\mathbf{N}$ *is* s^**-admissable.*

Proof. From (3.9), (3.8) and (3.10), we have

$$\|\mathbf{N}\mathbf{v} - \mathbf{w}_\varepsilon\| \le \varepsilon/2 + \sum_{k=0}^{\ell} C_{j-k}\|\mathbf{v}_{[k]}\| \le \varepsilon.$$

Let $s \in (0, s^*)$ be given. The number of operations needed for generating the vectors $\mathbf{v}_{[k]}$ is of the order

$$\#\operatorname{supp} \mathbf{v} + q \lesssim \#\operatorname{supp} \mathbf{v} + \log(\varepsilon^{-1}\|\mathbf{v}\|) \lesssim \#\operatorname{supp} \mathbf{v} + \varepsilon^{-1/s}|\mathbf{v}|_{\ell^w_\tau}^{1/s}.$$

The value of q was chosen such that the sum of squares of elements in V_q is less or equal to $(\varepsilon/(2\|\mathbf{N}\|))^2$, meaning that for all $k<\ell$, $\mathbf{v}_{[k]}$ only contains elements from V_i for $i<q$. Since within each of these V_i the squared values of the elements differ at most a factor 4, for $k\le\ell$ we obtain that

$$\|\mathbf{v}_{[k]}\|\le\|\mathbf{v}-\sum_{j=0}^{k-1}\mathbf{v}_{[j]}\|\le\|\mathbf{v}-\mathbf{v}_{\lceil 2^{k-1}/4\rceil}\|,$$

with $\mathbf{v}_N$ denoting the best N-term approximation of $\mathbf{v}$. Using (3.4), with $\tau=(\frac{1}{2}+s)^{-1}$ we infer that

$$\sum_{k=0}^{\ell}C_{j-k}\|\mathbf{v}_{[k]}\|\lesssim\sum_{k=0}^{\ell}C_{j-k}2^{-ks}|\mathbf{v}|_{\ell_\tau^w}\lesssim 2^{-js}|\mathbf{v}|_{\ell_\tau^w},$$

and so $2^j\lesssim\varepsilon^{-1/s}|\mathbf{v}|_{\ell_\tau^w}^{1/s}$ by definition of j. Now $\#\operatorname{supp}\mathbf{w}_\varepsilon$ and the number of operations needed for the evaluation of (3.11) can be bounded by $\sum_{k=0}^{\ell}\alpha_{j-k}2^{j-k}2^k\lesssim 2^j\lesssim\varepsilon^{-1/s}|\mathbf{v}|_{\ell_\tau^w}^{1/s}$. □

We will consider right-hand sides $\mathbf{g}$ that satisfy the following definition.

Definition 3.9. A vector $\mathbf{g}\in\ell_2$ is called *s^*-optimal*, when for a suitable routine **RHS**, for each $s\in(0,s^*)$ and all $\varepsilon>0$, with $\mathbf{g}_\varepsilon=\mathbf{RHS}[\varepsilon,\mathbf{g}]$ the following is valid:

(I) $\#\operatorname{supp}\mathbf{g}_\varepsilon\lesssim\varepsilon^{-1/s}|\mathbf{g}|_{\ell_\tau^w}^{1/s}$,
(II) the number of arithmetic operations used to compute it is at most a multiple of $\varepsilon^{-1/s}|\mathbf{g}|_{\ell_\tau^w}^{1/s}$,

Remark 3.10. A direct consequence of Proposition 3.4 and Part (I) of Definition 3.9 is that

$$|\mathbf{g}_\varepsilon|_{\ell_\tau^w}\lesssim|\mathbf{g}|_{\ell_\tau^w}. \tag{3.12}$$

Implicitly, in the proof of Proposition 3.8 we assumed that each element of $\mathbf{N}_j$ can be computed at unit costs. For a discussion under which circumstances this as well as $\mathbf{g}$ being s^*-optimal can be expected we refer to [CDD00, §6.2].

3.3. **The complexity of (mod)SOLVE.** We show that **SOLVE** and **modSOLVE** are of optimal computational complexity. We start with **modSOLVE**, since for this routine the proof follows closely the one given in [CDD00] for the Riesz basis case.

Theorem 3.11. *For some $s^*>0$, assume that $\mathbf{M}$ and $\mathbf{P}$ are s^*-admissible, $\mathbf{g}$ is s^*-optimal, and that for some $s\in(0,s^*)$, with $\tau=(\frac{1}{2}+s)^{-1}$, $\mathbf{Mu}=\mathbf{g}$ has a solution $\mathbf{u}\in\ell_\tau^w$.*

Then for all $\varepsilon>0$, $\mathbf{u}_\varepsilon=\mathbf{modSOLVE}[\varepsilon,\mathbf{M},\mathbf{g}]$ satisfies

(I) $\#\operatorname{supp}\mathbf{u}_\varepsilon\lesssim\varepsilon^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$,
(II) *the number of arithmetic operations used to compute it is at most a multiple of* $\varepsilon^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$,

Furthermore, as was shown in Proposition 2.3, $\|\mathbf{Pu}-\mathbf{u}_\varepsilon\|\le\varepsilon$, and so $\|u-F'\mathbf{u}_\varepsilon\|_H\le B_\Psi^{\frac{1}{2}}\varepsilon$.

Proof. It suffices to prove the statements for any $\varepsilon = \varepsilon_i$ with $\varepsilon_i = (3\rho^K\|\mathbf{P}\|/\theta)^i\varepsilon_0$ as in the algorithm **modSOLVE**.

As noted in Remark 3.6, the fact that $\mathbf{P}$ is s^*-admissible implies that it bounded on ℓ_τ^w. For any $i \geq 1$, from $\|\mathbf{P}\mathbf{u} - \mathbf{z}^{(i)}\| \leq \theta\varepsilon_i$ proven in Proposition 2.3, Proposition 3.2 and the assumption $\mathbf{u} \in \ell_\tau^w$ show that $\mathbf{u}^{(i)} := \mathbf{COARSE}[(1-\theta)\varepsilon_i, \mathbf{z}^{(i)}]$ satisfies

$$\#\mathrm{supp}\,\mathbf{u}^{(i)} \lesssim \varepsilon_i^{-1/s}|\mathbf{P}\mathbf{u}|_{\ell_\tau^w}^{1/s} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}, \tag{3.13}$$

i.e. (I), and by (3.7), also

$$|\mathbf{u}^{(i)}|_{\ell_\tau^w} \lesssim |\mathbf{P}\mathbf{u}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}. \tag{3.14}$$

Here we emphasize that both these results are valid uniformly in i.

To compute $\mathbf{u}^{(i)}$ from $\mathbf{u}^{(i-1)}$, **modSOLVE** uses one application of **RHS**, K applications of **APPLY** involving $\mathbf{M}$, $2K$ vector updates, one application of **APPLY** involving $\mathbf{P}$, and finally an application of **COARSE**. From the key estimates (3.13), (3.14), and the fact that K is some *fixed constant*, in the following three paragraphs we show that these computations take not more than a multiple of $\varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ operations. Since $(\varepsilon_i)_i$ is a geometrically decreasing sequence, we may therefore conclude (II).

Since $\mathbf{M}$ is s^*-admissible, it is bounded on ℓ_τ^w. As a consequence, $|\mathbf{g}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$, and so the assumption of $\mathbf{g}$ being s^*-optimal gives $\#\mathrm{supp}\,\mathbf{g}^{(i)} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ and $|\mathbf{g}^{(i)}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$ by (3.12), whereas the number of operations used to compute it is at most a multiple of $\varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$.

Because of $|\mathbf{g}^{(i)}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$ and $|\mathbf{u}^{(i-1)}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$, from the assumption that $\mathbf{M}$ is s^*-admissable it follows that $|\mathbf{v}^{(i,j)}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$ by Remark 3.6. Again since $\mathbf{M}$ is s^*-admissable, the latter result shows that $\#\mathrm{supp}\,\mathbf{v}^{(i,j)} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ for $1 \leq j \leq K$, whereas by $\#\mathrm{supp}\,\mathbf{v}^{(i,0)} \lesssim \varepsilon_{i-1}^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ ((3.13)) and $\#\mathrm{supp}\,\mathbf{g}^{(i)} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$, its computation takes not more than a multiple of $\varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ operations.

Since $|\mathbf{v}^{(i,K)}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$, $\#\mathrm{supp}\,\mathbf{v}^{(i,K)} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ and $\mathbf{P}$ is s^*-admissible, the computation of $\mathbf{z}^{(i)} := \mathbf{APPLY}[\theta\varepsilon_i/3, \mathbf{P}, \mathbf{v}^{(i,K)}]$ takes a number of operations that is at most a multiple of $\varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$, $\#\mathrm{supp}\,\mathbf{z}^{(i)} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$, and $|\mathbf{z}^{(i)}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$. Finally, by (3.6), the latter result implies that also $\mathbf{COARSE}[(1-\theta)\varepsilon_i, \mathbf{z}^{(i)}]$ needs at most a multiple of $\#\mathrm{supp}\,\mathbf{z}^{(i)} + \log(\varepsilon_i^{-1}\|\mathbf{z}^{(i)}\|) \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ operations. □

The key to the proof of Theorem 3.11 is the fact that the iterands produced by **modSOLVE** are uniformly bounded in ℓ_τ^w. Unfortunately, generally this will not be the case with **SOLVE**. Since **SOLVE** does not contain a projection onto a complement space of $\ker F'$, it is not capable to reduce errors once made in $\ker F'$. Recall that such errors are not reduced by the Richardson steps since they are in the kernel of $\mathbf{M}$. Although, because of the geometric decrease of the tolerances, these errors are summable in ℓ_2, we cannot show this in ℓ_τ^w, and so boundedness of the iterands in ℓ_τ^w is *not* guaranteed. For example, thinking of **RHS** and **APPLY** as being performed exactly, i.e., with zero tolerances, each

time **COARSE**$[(1-\theta)\varepsilon_i, \mathbf{v}^{(i,K)}]$ is invoked it gives an error, which might be completely contained in $\ker F'$, for which we can say not more than that its ℓ_τ^w-norm is less or equal to $|\mathbf{v}^{(i,K)}|_{\ell_\tau^w}$, i.e., that it is bounded.

For $\mathbf{u}^{(i)}$, ε_i as in **SOLVE**, for some $\check{s} \in (s, s^*)$, and with $\check{\tau} = (\frac{1}{2}+\check{s})^{-1}$, in the proof of Theorem 3.12 given below we will show that

$$\varepsilon_i^{(\check{s}/s)-1}|\mathbf{u}^{(i)}|_{\ell_{\check{\tau}}^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}^{\check{s}/s}, \quad \#\mathrm{supp}\,\mathbf{u}^{(i)} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s},$$

and, and as a consequence of these results, that the method is of optimal complexity. Note that for any $\mathbf{v}$ with finite support,

$$|\mathbf{v}|_{\ell_{\check{\tau}}^w} \leq (\#\mathrm{supp}\,\mathbf{v})^{\check{s}-s}|\mathbf{v}|_{\ell_\tau^w}. \tag{3.15}$$

So $|\mathbf{u}^{(i)}|_{\ell_\tau^w} \lesssim 1$ and $\#\mathrm{supp}\,\mathbf{u}^{(i)} \lesssim \varepsilon_i^{-1/s}$ would give $\varepsilon_i^{(\check{s}/s)-1}|\mathbf{u}^{(i)}|_{\ell_{\check{\tau}}^w} \lesssim |\mathbf{u}^{(i)}|_{\ell_\tau^w}^{\check{s}/s} \lesssim 1$. Yet, conversely, under no condition on $\#\mathrm{supp}\,\mathbf{u}^{(i)}$, uniform boundedness of $\varepsilon_i^{(\check{s}/s)-1}|\mathbf{u}^{(i)}|_{\ell_{\check{\tau}}^w}$ implies that of $|\mathbf{u}^{(i)}|_{\ell_\tau^w}$. In other words, it will turn out that uniform boundedness in ℓ_τ^w of the iterands is not a necessary condition for obtaining an optimal complexity result.

Theorem 3.12. *For some $s^* > 0$, assume that $\mathbf{M}$ is s^*-admissible, $\mathbf{g}$ is s^*-optimal, and that for some $s \in (0, s^*)$, with $\tau = (\frac{1}{2}+s)^{-1}$, $\mathbf{M}\mathbf{u} = \mathbf{g}$ has a solution $\mathbf{u} \in \ell_\tau^w$. In addition, assume that there exists an $\check{s} \in (s, s^*)$ such that with $\check{\tau} = (\frac{1}{2}+\check{s})^{-1}$,*

$$\mathbf{Q} \text{ is bounded on } \ell_{\check{\tau}}^w. \tag{3.16}$$

Then, if the parameter K in **SOLVE** *is sufficiently large; sufficient is*

$$3\rho^K < \theta \min\left\{1, \left[C_1(\check{\tau})C_2(\check{\tau})|(\mathbf{id}-\mathbf{Q})|_{\ell_{\check{\tau}}^w \leftarrow \ell_{\check{\tau}}^w}\right]^{s/(\check{s}-s)}\right\}, \tag{3.17}$$

where $C_1(\check{\tau})$, $C_2(\check{\tau})$ are the constants from (3.3), (3.7) *respectively; then for all $\varepsilon > 0$, $\mathbf{u}_\varepsilon = \mathbf{SOLVE}[\varepsilon, \mathbf{M}, \mathbf{g}]$ satisfies*

(I) $\#\mathrm{supp}\,\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$,

(II) *the number of arithmetic operations used to compute it is at most a multiple of $\varepsilon^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$,*

Furthermore, as shown in Proposition 2.1, $\|\mathbf{Q}(\mathbf{u}-\mathbf{u}_\varepsilon)\| \leq \varepsilon$, and so $\|u - F'\mathbf{u}_\varepsilon\|_H \leq B_\Psi^{\frac{1}{2}}\varepsilon$.

Proof. It suffices to prove the statements for any $\varepsilon = \varepsilon_i$ with $\varepsilon_i = (3\rho^K/\theta)^i\varepsilon_0$ as in the algorithm **SOLVE**.

Since $\mathbf{Q}$ is bounded on ℓ_2, and by assumption it is bounded on $\ell_{\check{\tau}}^w$, an interpolation argument (cf. [DeV98, (4.24)]) shows that it is bounded on ℓ_τ^w as well. Let N_i be the smallest integer such that $\|\mathbf{Q}\mathbf{u} - (\mathbf{Q}\mathbf{u})_{N_i}\| \leq \theta\varepsilon_i/3$, where $(\mathbf{Q}\mathbf{u})_N$ is the best N-term approximation for $\mathbf{Q}\mathbf{u}$. Then using the assumption $\mathbf{u} \in \ell_\tau^w$, (3.4) shows that

$$N_i \lesssim \varepsilon_i^{-1/s}|\mathbf{Q}\mathbf{u}|_{\ell_\tau^w}^{1/s} \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s},$$

and so by (3.15),

$$\varepsilon_i^{(\check{s}/s)-1}|(\mathbf{Q}\mathbf{u})_{N_i}|_{\ell_{\check{\tau}}^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}^{(\check{s}/s)-1}|(\mathbf{Q}\mathbf{u})_{N_i}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}^{(\check{s}/s)-1}|\mathbf{Q}\mathbf{u}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}^{\check{s}/s}. \tag{3.18}$$

From $\|\mathbf{Q}\mathbf{u}+(\mathbf{id}-\mathbf{Q})\mathbf{u}^{(i-1)}-\mathbf{v}^{(i,K)}\|\leq 2\theta\varepsilon_i/3$ proven in Proposition 2.1, we get

$$\|(\mathbf{Q}\mathbf{u})_{N_i}+(\mathbf{id}-\mathbf{Q})\mathbf{u}^{(i-1)}-\mathbf{v}^{(i,K)}\|\leq\theta\varepsilon_i.$$

From (3.7) and then (3.3), it follows that $\mathbf{u}^{(i)}:=\mathbf{COARSE}[(1-\theta)\varepsilon_i,\mathbf{v}^{(i,K)}]$ satisfies

$$\begin{aligned}|\mathbf{u}^{(i)}|_{\ell^w_{\check\tau}}&\leq C_2(\check\tau)|(\mathbf{Q}\mathbf{u})_{N_i}+(\mathbf{id}-\mathbf{Q})\mathbf{u}^{(i-1)}|_{\ell^w_{\check\tau}}\\&\leq C_1(\check\tau)C_2(\check\tau)|(\mathbf{Q}\mathbf{u})_{N_i}|_{\ell^w_{\check\tau}}+C_1(\check\tau)C_2(\check\tau)|(\mathbf{id}-\mathbf{Q})|_{\ell^w_{\check\tau}\leftarrow\ell^w_{\check\tau}}|\mathbf{u}^{(i-1)}|_{\ell^w_{\check\tau}},\end{aligned}$$

and so by (3.18) and $\varepsilon_i=3\rho^K\varepsilon_{i-1}/\theta$,

$$\left(\varepsilon_i^{(\check s/s)-1}|\mathbf{u}^{(i)}|_{\ell^w_{\check\tau}}\right)\leq C|\mathbf{u}|^{\check s/s}_{\ell^w_\tau}+C_1(\check\tau)C_2(\check\tau)|(\mathbf{id}-\mathbf{Q})|_{\ell^w_{\check\tau}\leftarrow\ell^w_{\check\tau}}(3\rho^K/\theta)^{(\check s/s)-1}\left(\varepsilon_{i-1}^{(\check s/s)-1}|\mathbf{u}^{(i-1)}|_{\ell^w_{\check\tau}}\right)$$

for some constant $C>0$. We may conclude that if K satisfies (3.17), then solutions of the homogeneous part of this recursion convergence to zero, and so

$$\varepsilon_i^{(\check s/s)-1}|\mathbf{u}^{(i)}|_{\ell^w_{\check\tau}}\lesssim|\mathbf{u}|^{\check s/s}_{\ell^w_\tau},\tag{3.19}$$

which, as we emphasize here, holds uniformly in i.

Knowing (3.19), Proposition 3.2 and (3.18) show that

$$\begin{aligned}\#\operatorname{supp}\mathbf{u}^{(i)}&\lesssim\varepsilon_i^{-1/\check s}|(\mathbf{Q}\mathbf{u})_{N_i}+(\mathbf{id}-\mathbf{Q})\mathbf{u}^{(i-1)}|^{1/\check s}_{\ell^w_{\check\tau}}\\&\lesssim\varepsilon_i^{-1/s}\left(\varepsilon_i^{(\check s/s)-1}\left[|(\mathbf{Q}\mathbf{u})_{N_i}|_{\ell^w_{\check\tau}}+|\mathbf{id}-\mathbf{Q}|_{\ell^w_{\check\tau}\leftarrow\ell^w_{\check\tau}}|\mathbf{u}^{(i-1)}|_{\ell^w_{\check\tau}}\right]\right)^{1/\check s}\\&\lesssim\varepsilon_i^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau},\end{aligned}$$

i.e., (I) is valid.

The remainder of the proof resembles the one of Theorem 3.11. We have to show that the K intermediate steps that transfer $\mathbf{u}^{(i-1)}$ to $\mathbf{u}^{(i)}$ take a number of operations that is bounded by some multiple of $\varepsilon_i^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau}$.

As in the proof of Theorem 3.11, since $\mathbf{M}$ is s^*-admissible and $\mathbf{g}$ is s^*-optimal, we have $\#\operatorname{supp}\mathbf{g}^{(i)}\lesssim\varepsilon_i^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau}$ and $|\mathbf{g}^{(i)}|_{\ell^w_\tau}\lesssim|\mathbf{u}|_{\ell^w_\tau}$, and so in addition, $\varepsilon_i^{(\check s/s)-1}|\mathbf{g}^{(i)}|_{\ell^w_{\check\tau}}\lesssim|\mathbf{u}|^{\check s/s}_{\ell^w_\tau}$.

Since $\mathbf{M}$ is s^*-admissible, this last result together with $\varepsilon_{i-1}^{(\check s/s)-1}|\mathbf{u}^{(i-1)}|_{\ell^w_{\check\tau}}\lesssim|\mathbf{u}|^{\check s/s}_{\ell^w_\tau}$ show that

$$\varepsilon_i^{(\check s/s)-1}|\mathbf{v}^{(i,j)}|_{\ell^w_{\check\tau}}\lesssim|\mathbf{u}|^{\check s/s}_{\ell^w_\tau},\qquad(0\leq j\leq K),\tag{3.20}$$

by Remark 3.6 (use $\check s<s^*$).

A new element in this proof is the observation that, instead of uniform boundedness in ℓ^w_τ, (3.20) is already sufficient to guarantee that the supports have the appropriate sizes. Indeed, again since $\mathbf{M}$ is s^*-admissible (use $\check s<s^*$), it follows that

$$\#\operatorname{supp}\mathbf{v}^{(i,j)}\lesssim\varepsilon_i^{-1/\check s}|\mathbf{v}^{(i,j-1)}|^{1/\check s}_{\ell^w_{\check\tau}}\lesssim\varepsilon_i^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau},\qquad(1\leq j\leq K),\tag{3.21}$$

whereas by $\#\operatorname{supp}\mathbf{v}^{(i,0)}\lesssim\varepsilon_{i-1}^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau}$ by (I), $\#\operatorname{supp}\mathbf{g}^{(i)}\lesssim\varepsilon_i^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau}$, and the second inequality in (3.21), its computation takes not more a multiple of $\varepsilon_i^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau}$ operations.

Finally, by (3.6), the application of $\mathbf{COARSE}[(1-\theta)\varepsilon_i, \mathbf{v}^{(i,K)}]$ needs at most a multiple of $\#\mathrm{supp}\,\mathbf{v}^{(i,K)} + \log(\varepsilon_i^{-1}\|\mathbf{v}^{(i,K)}\|) \lesssim \varepsilon_i^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ operations. □

Remark 3.13. The conditions imposed in Theorem 3.12 do not exclude the possibility that $\mathbf{u}_\varepsilon \in \mathrm{Ran}\,F$, and so $\mathbf{Q}\mathbf{u}_\varepsilon = \mathbf{u}_\varepsilon$. Then analogously to Remark 3.6, the estimates $\|\mathbf{Q}(\mathbf{u}-\mathbf{u}_\varepsilon)\| \le \varepsilon$ and $\#\mathrm{supp}\,\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ *imply* that $\mathbf{Q} : \ell_\tau^w \to \ell_\tau^w$ is bounded. In this sense, the condition imposed in Theorem 3.12 that for some $\check\tau < \tau$, $\mathbf{Q} : \ell_{\check\tau}^w \to \ell_{\check\tau}^w$ is bounded, is an almost necessary one.

4. Construction of frames

Recall that $L : H \to H'$ was assumed to be a boundedly invertible operator, where we have in mind a linear differential- or integral operator. When L is an operator of order $2t$, typically H is a Sobolev space of order t. We also briefly discussed the case of having systems of such equations, which however poses no principal additional difficulties. So here we restrict ourselves to scalar equations, where in addition we assume that the equation is imposed on a *domain* $\Omega \subset \mathbb{R}^n$. In particular in connection with integral equations, it is also relevant to study the case of the equation being formulated on a manifold. Yet, in that case the construction of a frame may follow same principles as in the domain case that we outline here.

4.1. **Overlapping domain decompositions.**

Theorem 4.1. *For some $\Gamma^D \subset \partial\Omega$, possibly $\Gamma^D = \emptyset$, and $t \in \mathbb{R}$, let*

$$\mathcal{H}^t = \begin{cases} H^t_{0,\Gamma^D}(\Omega) & \text{when } t \ge 0, \\ (H^{-t}_{0,\Gamma^D}(\Omega))' & \text{when } t < 0, \end{cases}$$

where for $t \ge 0$,

$$H^t_{0,\Gamma^D}(\Omega) = \mathrm{clos}_{H^t(\Omega)}\{u \in H^t(\Omega) \cap C^\infty(\Omega) : \mathrm{supp}\,u \cap \Gamma^D = \emptyset\}.$$

Let $\Omega = \cup_{i=1}^m \Omega_i$ be an open covering, with which we mean that the sets Ω_i are open, and that there exists a partition of unity $\{\chi_i\}$ relative to $\{\Omega_i\}$, i.e., $\chi_i \in C^\infty(\Omega)$, $0 \le \chi_i \le 1$, χ_i vanishes outside Ω_i and $\sum_i \chi_i = 1$.

With $\Gamma_i^D = \begin{cases} \partial\Omega_i \cap (\Omega \cup \Gamma^D) & \text{when } t \ge 0 \\ \partial\Omega_i \cap \Gamma^D & \text{when } t < 0 \end{cases}$, *let* $\mathcal{H}_i^t = \begin{cases} H^t_{0,\Gamma_i^D}(\Omega_i) & \text{when } t \ge 0, \\ (H^{-t}_{0,\Gamma_i^D}(\Omega_i))' & \text{when } t < 0, \end{cases}$

and let $\Psi^{(i)}$ be a Riesz basis, or more generally, a frame for $\mathcal{H}_i^t$.

Let $\{\omega_i\}_{1\le i\le m}$ be a collection of non-negative functions on Ω, with ω_i smooth on Ω_i and zero outside Ω_i, such that there exists an open covering $\Omega = \cup_{i=1}^m \hat\Omega_i$ with $\hat\Omega_i \subset \Omega_i$ and $\omega_i \eqsim 1$ on $\hat\Omega_i$.

Then

$$\cup_i \omega_i \Psi^{(i)} \text{ is a frame for } \mathcal{H}^t.$$

Proof. First we demonstrate that for $u \in \mathcal{H}^t$,

$$\|u\|_{\mathcal{H}^t}^2 \eqsim \inf_{\omega_i^{-1}u_i \in \mathcal{H}_i^t,\, \sum_i u_i = u} \sum_i \|\omega_i^{-1}u_i\|_{\mathcal{H}_i^t}^2. \tag{4.1}$$

Here by writing $\omega_i^{-1}u_i \in \mathcal{H}_i^t$, in particular we implicitly state that u_i vanishes outside $\operatorname{supp}\omega_i$.

Let $\omega_i^{-1}u_i \in \mathcal{H}_i^t$. Then since ω_i is smooth on Ω_i, $u_i \in \mathcal{H}_i^t$. Furthermore, the spaces $\mathcal{H}_i^t$ are selected in such a way that the trivial extension with zero of a function on Ω_i extends to a bounded mapping from $\mathcal{H}_i^t \to \mathcal{H}^t$. To see this for $t<0$, note that the restriction of a function on Ω to Ω_i, which is the adjoint of the zero extension, is a bounded mapping from $H^{-t}_{0,\Gamma_D}(\Omega)$ to $H^{-t}_{0,\partial\Omega_i\cap\Gamma^D}(\Omega_i)$. We conclude that for any $\omega_i^{-1}u_i \in \mathcal{H}_i^t$, the function $u=\sum_i u_i \in \mathcal{H}^t$, with $\|u\|^2_{\mathcal{H}^t} \lesssim \sum_i \|u_i\|^2_{\mathcal{H}_i^t} \lesssim \sum_i \|\omega_i^{-1}u_i\|^2_{\mathcal{H}_i^t}$.

Conversely, let $\{\hat\chi_i\}$ be a partition of unity relative to $\{\hat\Omega_i\}$. Then any $u\in\mathcal{H}^t$ can be written as $u=\sum_i \hat\chi_i u$, where, because of $\omega_i \eqsim 1$ on $\hat\Omega_i$, $\omega_i^{-1}\hat\chi_i u \in \mathcal{H}_i^t$ and $\|\omega_i^{-1}\hat\chi_i u\|_{\mathcal{H}_i^t} \lesssim \|\hat\chi_i u\|_{\mathcal{H}_i^t} \lesssim \|u\|_{\mathcal{H}^t}$ completing the proof of (4.1).

Since $\Psi^{(i)}$ is a frame for $\mathcal{H}_i^t$, for $v_i\in\mathcal{H}_i^t$ we have $\|v_i\|^2_{\mathcal{H}_i^t} \eqsim \inf_{\mathbf{c}_i\in\ell_2,\,\mathbf{c}_i^T\Psi^{(i)}=v_i}\|\mathbf{c}_i\|^2_{\ell_2}$. With v_i of the form $\omega_i^{-1}u_i$, $\mathbf{c}_i^T\Psi^{(i)}=v_i$ is equivalent to $\mathbf{c}_i^T\omega_i\Psi^{(i)}=u_i$. Now from (4.1), we conclude that for $u\in\mathcal{H}^t$,

$$\begin{aligned}\|u\|^2_{\mathcal{H}^t} &\eqsim \inf_{\omega_i^{-1}u_i\in\mathcal{H}_i^t,\,\sum_i u_i=u}\ \sum_i\ \inf_{\mathbf{c}_i\in\ell_2,\,\mathbf{c}_i^T\omega_i\Psi^{(i)}=u_i}\|\mathbf{c}_i\|^2\\ &= \inf_{(\mathbf{c}_1^T,\dots,\mathbf{c}_m^T)^T\in\ell_2,\,\sum_i \mathbf{c}_i^T\omega_i\Psi^{(i)}=u}\ \sum_i\|\mathbf{c}_i\|^2\end{aligned}$$

meaning that $\cup_i\omega_i\Psi^{(i)}$ is a frame for $\mathcal{H}^t$. □

Remarks 4.2. If Theorem 4.1 is applied with ω_i being the characteristic function of $\Omega_i=\hat\Omega_i$, then it shows that $\cup_i\Psi^{(i)}$ is a frame for $\mathcal{H}^t$.

If each ω_i is selected such that it *vanishes at* the internal boundary $\partial\Omega_i\cap\Omega$, then above proof shows that boundary conditions on that part of $\partial\Omega_i$ can actually be chosen at ones convenience, i.e., any $\partial\Omega_i\cap\Gamma^D\subset\Gamma_i^D\subset\partial\Omega_i\cap(\Omega\cup\Gamma^D)$ will do.

To construct collections $\Psi^{(i)}$ that serve as ingredients in Theorem 4.1, we may proceed as follows: Suppose that for each $1\le i\le m$, we have a sufficiently smooth regular parametrization κ_i between $(0,1)^n$, or another reference domain, and Ω_i (see Figure 1). With $\Gamma^D_{i,\square}=\kappa_i^{-1}(\Gamma_i^D)$, let $\Psi^{(i)}_\square$ be a Riesz basis for $H^t_{0,\Gamma^D_{i,\square}}(0,1)^n$ when $t\ge 0$, or for $(H^{-t}_{0,\Gamma^D_{i,\square}}(0,1)^n)'$ otherwise. Then we may conclude that $\Psi^{(i)}=\Psi^{(i)}_\square\circ\kappa_i^{-1}$ is a Riesz basis for $\mathcal{H}_i^t$.

At least if the parametrizations are constructed such that the image of a face of $[0,1]^n$ has either empty intersection with Γ_D or that it is fully contained in Γ_D, then $\Psi^{(i)}_\square$ of wavelet type can easily be constructed by taking tensor products of wavelet bases on the interval with appropriate boundary conditions.

With the construction of spline wavelets on the interval from [DKU99], only wavelets with supports near the endpoints depend on the boundary condition. This means that if the weights ω_i in Theorem 4.1 vanish in a sufficiently large *neighbourhood of* the internal

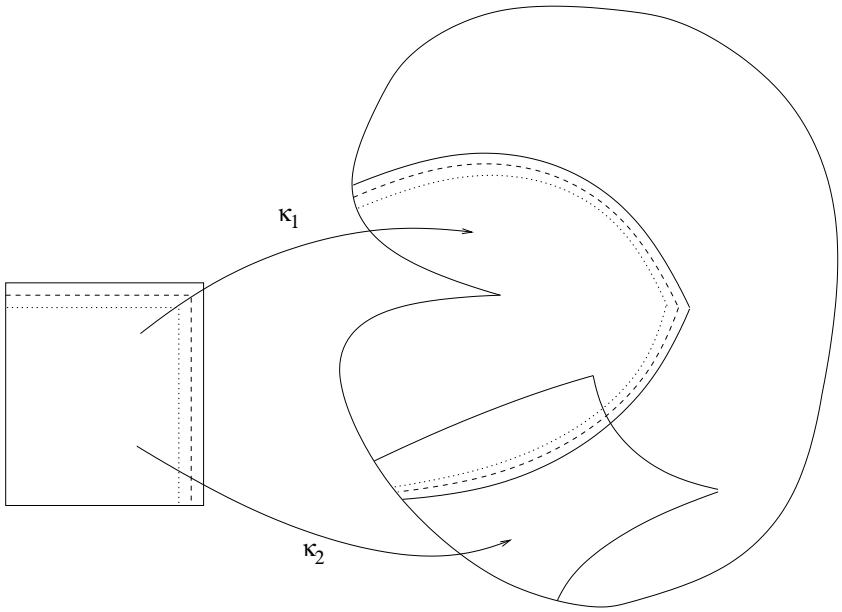

FIGURE 1. Overlapping domain decomposition (the dashed and dotted lines will get their meaning in §4.4).

boundaries $\partial\Omega_i \cap \Omega$, then boundary conditions at these internal boundaries are irrelevant since they have no influence on the constructed frame.

Compared to the construction of wavelet *bases* for $\mathcal{H}^t$ based on a *non-overlapping* decomposition of the domain, the frame approach seems to have the following advantages:

• It is easier to construct parametrizations corresponding to an overlapping domain decomposition; only local parametrizations of $\partial\Omega$ are needed. Having less complicated κ_i may also have a favourable quantitative effect on the frame constants A_Ψ and B_Ψ.

• Constructions of wavelet bases based on non-overlapping domain decompositions all involve a kind of 'stitching' of wavelets from different subdomains at the interfaces. The construction from [DS99b] yields wavelets with all desired theoretical properties, but it seems difficult to implement. The constructions proposed in [DS99a, CTU99, CM00] yield near the interfaces wavelets which are only continuous, which has the following disadvantages: As we will see in the next section (cf. (4.3)), this limited smoothness might restrict the range of s for which convergence of order N^{-s} can be expected. Secondly, it restricts the value of s^* for which $\mathbf{M}$ (and $\mathbf{P}$) are s^*-compressible (see [CDD01, Proposition 6.2.2], [Ste02] and §4.5).

• When a frame construction similar as in Theorem 4.1 is applied on a *closed manifold* with weights ω_i that vanish at the internal boundaries, then the wavelet bases on $(0,1)^n$ that serve as ingredients may satisfy *periodic boundary conditions*. Not only the implementation of such bases is much easier, they are also much better conditioned than available wavelet bases that satisfy Dirichlet or Neumann boundary conditions.

4.2. **Regularity.** As we have seen, under some conditions the routine **SOLVE** or **mod-SOLVE** exhibits an error decay of order N^{-s}, with N being the number of operations spent and coefficients stored, *in case* $\mathbf{u} \in \ell_\tau^w$ with $\tau = (\frac{1}{2}+s)^{-1}$. Recall that $\mathbf{u}$ is some solution of $\mathbf{Mu} = \mathbf{g}$, that is, $u = \mathbf{u}^T\Psi$ is the solution of $Lu = g$.

In case Ψ is a *Riesz basis* for $\mathcal{H}^t$ of biorthogonal wavelet type of *order* d, meaning that $d-1$ is the order of local polynomial reproduction, then it is known that for

$$0 < s < (d-t)/n,$$

it holds that

$$\mathbf{u} \in \ell_\tau \quad \text{iff} \quad u \in B_\tau^{sn+t}(L_\tau(\Omega)), \tag{4.2}$$

at least when the wavelets are contained in $B_\tau^{sn+t}(L_\tau(\Omega))$ and $s \le 1/2$ if $t < -n/2$ (see Figure 2). Recall that $\mathbf{u} \in \ell_\tau$ implies $\mathbf{u} \in \ell_\tau^w$. Here for $\nu \ge 0$, $B_p^\nu(L_p(\Omega))$ is the usual

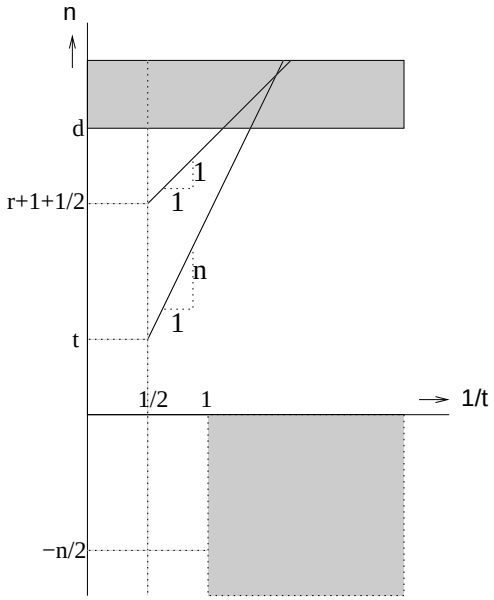

FIGURE 2. $B_\tau^\nu(L_\tau(\Omega))$ with the line $\nu = sn + t$, where $\tau = (\frac{1}{2}+s)^{-1}$, and the line $\nu = r + 1 + 1/\tau$.

Besov space, in which possible boundary conditions are incorporated, measuring "ν orders of smoothness in L_p", and for $\nu < 0$, $B_p^\nu(L_p(\Omega)) := (B_{p'}^{-\nu}(L_{p'}(\Omega))'$ with $1/p + 1/p' = 1$, and so necessarily $p \ge 1$. This latter restriction induces the afore-mentioned condition $s \le 1/2$ if $t < -n/2$. For details about Besov spaces and proofs of (4.2) in various circumstances we refer to [Coh00].

Remark 4.3. If the wavelets are piecewise smooth, globally C^r-functions for some $r \in \mathbb{N} \cup \{-1\}$, with $r = -1$ meaning that they satisfy no global smoothness requirements, then it is known that they are contained in $B_\tau^\nu(L_\tau(\Omega))$ when $\nu < r + 1 + 1/\tau$, whereas they

are not contained in this space when $\nu > r + 1 + 1/\tau$. using result So if for $s = (d-t)/n$ with $\tau = (\frac{1}{2} + s)^{-1}$, it holds that $r + 1 + 1/\tau \geq sn + t$, i.e.,

$$r \geq -\tfrac{3}{2} + d + (t-d)/n. \tag{4.3}$$

then smoothness of the wavelets does not limit the range for which (4.2) is valid.

With spline wavelets we have $r = d-2$, meaning that (4.3) reads as the mild requirement $(d-t)/n \geq 1/2$.

Remark 4.4. Note that to obtain a rate N^{-s} with a linear, non-adaptive, method it is needed that the solution u is in $H^{sn+t}(\Omega)$ which is a much smaller space than $B_\tau^{sn+t}(L_\tau(\Omega))$ Recently, a number of regularity proofs appeared showing that various operators have indeed a much larger regularity in above Besov scale than in the Sobolev scale (e.g. see [DD97, Dah99b]). A particular example is the operator corresponding to Poisson's equation on a two-dimensional polygonal domain which has been shown to have infinity regularity in the Besov scale (see [Dah99a]). This means that the Besov regularity of the solution is only limited by smoothness of the right-hand side, and so that it can be arbitrarily large.

Let now $\Psi = \cup_i \omega_i \Psi^{(i)}$ be a *frame* for $\mathcal{H}^t$ as constructed in Theorem 4.1, where the $\Psi^{(i)}$ are sufficiently smooth, biorthogonal wavelet bases of order d for $\mathcal{H}_i^t$. Let $\{\hat{\chi}_i\}$ be a partition of unity relative to $\{\hat{\Omega}_i\}$. Then $u \in B_\tau^{sn+t}(L_\tau(\Omega))$ implies $\hat{\chi}_i u \in B_\tau^{sn+t}(L_\tau(\Omega_i))$ and also $\omega_i^{-1}\hat{\chi}_i u \in B_\tau^{sn+t}(L_\tau(\Omega_i))$. So if $0 < s < (d-t)/n$, and $s \leq 1/2$ if $t < -n/2$, then (4.2) learns that each $\omega_i^{-1}\hat{\chi}_i u$ has a unique expansion $\mathbf{u}_i^T \Psi^{(i)}$, and so $\hat{\chi}_i u = \mathbf{u}_i^T \omega_i \Psi^{(i)}$, where $\mathbf{u}_i \in \ell_\tau$. We conclude that u has a representation $\sum_i \mathbf{u}_i^T \omega_i \Psi^{(i)}$ with $(\mathbf{u}_1^T, \ldots, \mathbf{u}_m^T)^T \in \ell_\tau \subset \ell_\tau^w$ under the same condition on u as is needed in the Riesz basis case.

4.3. **Boundedness of Q, i.e. condition** (3.16)**.** Assuming $\mathbf{Mu} = \mathbf{g}$ has a solution $\mathbf{u} \in \ell_\tau^w$, in Theorem 3.12 optimal computational complexity of **SOLVE** was proved under the condition that for some $\check{\tau} < \tau$, $\mathbf{Q}$ is bounded on $\ell_{\check{\tau}}^w$. Recall that $\mathbf{Q} = F(F'F)^{-1}F'$, where with the frame construction from Theorem 4.1 and biorthogonal wavelet bases $\Psi^{(i)}$ on the subdomains, $F' : \ell_2 \to \mathcal{H}^t : \mathbf{c} = (\mathbf{c}_1^T, \ldots, \mathbf{c}_m^T)^T \mapsto \sum_i \omega_i \mathbf{c}_i^T \Psi^{(i)}$, and so $F : (\mathcal{H}^t)' \to \ell_2 : u \mapsto ((\langle u, \omega_i \Psi^{(i)} \rangle_{L_2(\Omega)})_i)^T = ((\langle \omega_i u, \Psi^{(i)} \rangle_{L_2(\Omega_i)})_i)^T$.

For wavelets that are sufficiently smooth, from (4.2) we know that $\mathbf{c}_i \mapsto \mathbf{c}_i^T \Psi_i$ is bounded from $\ell_{\check{\tau}} \to B_{\check{\tau}}^{\check{s}n+t}(L_{\check{\tau}}(\Omega_i))$ when $0 < \check{s} < (d-t)/n$ and $\check{s} \leq 1/2$ if $t < -n/2$.

The mapping $v_i \mapsto \langle v_i, \Psi^{(i)} \rangle_{L_2(\Omega_i)}^T$ is the inverse of $\mathbf{d}_i \mapsto \mathbf{d}_i^T \tilde{\Psi}^{(i)}$, where $\tilde{\Psi}^{(i)}$ is the dual wavelet basis. When the dual wavelets are sufficiently smooth, the latter mapping is boundedly invertible from $\ell_{\check{\tau}}$ to $B_{\check{\tau}}^{\check{s}n-t}(L_{\check{\tau}}(\Omega_i))$ when $0 < \check{s} < (\tilde{d}+t)/n$, where $\tilde{d}$ is the order of the dual multi-resolution analysis, and $\check{s} \leq 1/2$ if $-t < -n/2$.

If we now in addition assume that the ω_i vanish at the internal boundaries $\partial\Omega_i \cap \Omega$, and so that these weights are globally smooth on Ω, then we may conclude that $F' : \ell_{\check{\tau}} \to B_{\check{\tau}}^{\check{s}n+t}(L_{\check{\tau}}(\Omega))$ and $F : B_{\check{\tau}}^{\check{s}n-t}(L_{\check{\tau}}(\Omega)) \to \ell_{\check{\tau}}$ are bounded when $0 < \check{s} < \min\{(d-t)/n, (\tilde{d}+t)/n\}$ and $\check{s} \leq 1/2$ if $|t| > n/2$.

Unfortunately, so far we are able to verify boundedness of $(F'F)^{-1} : B_{\check{\tau}}^{\check{s}n+t}(L_{\check{\tau}}(\Omega)) \to B_{\check{\tau}}^{\check{s}n-t}(L_{\check{\tau}}(\Omega))$, that in combination with above boundedness of F' and F would show the

desired property of $\mathbf{Q}$, in only one particular situation that $t=0$ and the $\Psi^{(i)}$ are $L_2(\Omega_i)$-orthonormal bases. In that case, $FF'u=\sum_i \omega_i\langle\omega_i u,\Psi^{(i)}\rangle_{L_2(\Omega_i)}\Psi^{(i)}=(\sum_i\omega_i^2)u$, and so $(FF')^{-1}u=(\sum_i\omega_i^2)^{-1}u$ meaning that by the global smoothness of the weights, $(FF')^{-1}$ clearly has above property for any s. Note that on the other hand if we do not damp the wavelets near the internal boundaries, i.e., if ω_i is just the characteristic function of Ω_i, then $(F'F)^{-1}$ will be bounded on $B^{\check{s}n}_{\check{\tau}}(L_{\check{\tau}}(\Omega))$ for $\check{s}$ in a limited range only.

So at least for $t=0$ and with sufficiently smooth orthonormal $\Psi^{(i)}$ (which implies $\tilde{d}=d$) and weights that vanish at internal boundaries $\partial\Omega_i\cap\Omega$, for any τ with $0<s=1/\tau-1/2<d/n$ (which is the full range for which one may expect that $\mathbf{u}\in\ell^w_\tau$), there exists a $\check{\tau}<\tau$ such that $\mathbf{Q}$ is bounded on $\ell_{\check{\tau}}$.

4.4. **Construction of P.** The fact that we have no general answer whether $\mathbf{Q}$ satisfies (3.16) was the motivation to introduce the routine **modSOLVE** that contains the inexact application of a suitable projector $\mathbf{P}$.

In the situation of Theorem 4.1, so with $\{\hat{\chi}_i\}$ a partition of unity relative to $\{\hat{\Omega}_i\}$ and ω_i the weights, and where $\Psi^{(i)}$ are biorthogonal wavelet bases for $\mathcal{H}^t_i$ with duals $\tilde{\Psi}^{(i)}$, let us define $Z:u\mapsto((\langle\hat{\chi}_i\omega_i^{-1}u,\tilde{\Psi}^{(i)}\rangle_{L_2(\Omega_i)})_i)^T$ which is a bounded mapping from $\mathcal{H}^t\to\ell_2$. It holds that $F'Zu=\sum_i\omega_i\langle\hat{\chi}_i\omega_i^{-1}u,\tilde{\Psi}^{(i)}\rangle_{L_2(\Omega_i)}\Psi^{(i)}=u$. So defining

$$\mathbf{P}=ZF':(\mathbf{c}_1^T,\dots,\mathbf{c}_m^T)^T\mapsto((\langle\hat{\chi}_i\omega_i^{-1}\textstyle\sum_{\check{\imath}}\omega_{\check{\imath}}\mathbf{c}_{\check{\imath}}^T\Psi^{(\check{\imath})},\tilde{\Psi}^{(i)}\rangle_{L_2(\Omega_i)})_i^T)^T,$$

we infer that $\mathbf{P}:\ell_2\to\ell_2$ is a bounded projector with $\operatorname{Ker}\mathbf{P}=\operatorname{Ker}F'$, which are the basic requirements on $\mathbf{P}$ imposed in §2.3, and which guarantee that **modSOLVE** is convergent, i.e., Proposition 2.3 is valid. Note that the application of $\mathbf{P}$ may only change coefficients corresponding to wavelets with supports or that of the corresponding dual wavelets intersect more than one Ω_i.

To apply above $\mathbf{P}$ we need a practical construction of the partition of unity $\{\hat{\chi}_i\}$. Apart from this, we discuss here the construction of weights in Theorem 4.1 that vanish at, or even in a neighbourhood of the internal boundaries $\partial\Omega_i\cap\Omega$. As we have seen, the application of weights that vanish at the internal boundaries seems necessary for $\mathbf{Q}$ satisfying (3.16), whereas even when the application of $\mathbf{P}$ is necessary such weights may have a favourable quantitative effect. Furthermore, weights that vanish even in a neighbourhood of the internal boundaries allow us to ignore boundary conditions at these boundaries with the construction of wavelets on the subdomains.

Let $\Omega=\cup_i\Omega_i$ be an open covering, and $\kappa_i:(0,1)^n\to\Omega_i$ smooth regular parametrizations, where we assume that the image of a face of $[0,1]^n$ under κ_i has either empty intersection with $\partial\Omega$ or that it is contained in $\partial\Omega$. Then there exist $0\le\hat{a}_j^{(i)}\le\check{a}_j^{(i)}<\check{b}_j^{(i)}\le\hat{b}_j^{(i)}\le 1$ such that $\cup_{i=1}^m\kappa_i\left(\prod_{j=1}^n(\check{a}_j^{(i)},\check{b}_j^{(i)})\right)=\Omega$, whereas strict inequalities $0<\hat{a}_j^{(i)}<\check{a}_j^{(i)}$ or $\check{b}_j^{(i)}<\hat{b}_j^{(i)}<1$ hold if (and only if) the face corresponds to an internal boundary (cf. dashed and dotted boundaries in Figure 1).

Now let $\eta_i,\phi_i\in C^\infty((0,1)^n)$ with $0\le\eta_i,\phi_i\le 1$, such that $\eta_i\eqsim 1$ on $\prod_j(\hat{a}_j^{(i)},\hat{b}_j^{(i)})$, whereas it vanishes at, or even in a neighbourhood of faces of $[0,1]^n$ that correspond to internal

boundaries, and $\phi_i = 1$ on $\prod_j(\check{a}_j^{(i)}, \check{b}_j^{(i)})$, whereas it vanishes at faces of $\prod_{j=1}^n(\hat{a}_j^{(i)}, \hat{b}_j^{(i)})$ that correspond to internal boundaries.

Defining $\hat{\Omega}_i = \kappa_i(\prod_j(\hat{a}_j^{(i)}, \hat{b}_j^{(i)}))$ and $\omega_i = \begin{cases} \eta_i \circ \kappa_i^{-1} & \text{on } \Omega_i \\ 0 & \text{on } \Omega\backslash\Omega_i \end{cases}$, as desired we have that $\omega_i \in C^\infty(\Omega)$, $0 \le \omega_i \le 1$, $\omega_i \eqq 1$ on $\hat{\Omega}_i$, and ω_i vanishes at or even in a neighbourhood of $\partial\Omega_i \cap \Omega$.

Defining $\check{\Omega}_i = \kappa_i(\prod_j(\check{a}_j^{(i)}, \check{b}_j^{(i)}))$ and $\hat{\chi}_i^{(i)} = \begin{cases} \phi_i \circ \kappa_i^{-1} & \text{on } \Omega_i \\ 0 & \text{on } \Omega\backslash\Omega_i \end{cases}$, we have $0 \le \hat{\chi}_i^{(i)} \le 1$, $\hat{\chi}_i^{(i)} = 1$ on $\check{\Omega}_i$, and $\hat{\chi}_i^{(i)}$ vanishes outside $\hat{\Omega}_i$. A partition of unity relative to $\{\hat{\Omega}_i\}$ is now given by $\{\hat{\chi}_i^{(m)} : 1 \le i \le m\}$, where for $2 \le k \le m$, $\{\hat{\chi}_i^{(k)} : 1 \le i \le k-1\}$ is defined by $\hat{\chi}_i^{(k)} := \hat{\chi}_i^{(k-1)}(1 - \hat{\chi}_k^{(k)})$. Indeed, an induction argument shows that $\sum_{i=1}^k \hat{\chi}_i^{(k)} = 1$ on $\cup_{i=1}^k \check{\Omega}_i$, and so in particular $\sum_{i=1}^m \hat{\chi}_i^{(m)} = 1$ on $\cup_{i=1}^m \check{\Omega}_i = \Omega$. Furthermore, $\hat{\chi}_i^{(m)} \in C^\infty(\Omega)$, $0 \le \hat{\chi}_i^{(m)} \le 1$, and, since $\operatorname{supp} \hat{\chi}_i^{(m)} \subset \operatorname{supp} \hat{\chi}_i^{(i)}$, $\hat{\chi}_i^{(m)}$ vanishes outside $\hat{\Omega}_i$.

4.5. **Compressibility, i.e., the value of s^*.** Let $\Psi = \cup_{i=1}^m \omega_i \Psi^{(i)}$ be a frame for $\mathcal{H}^t$ as constructed in Theorem 4.1, where the $\Psi^{(i)}$ are biorthogonal wavelet bases for $\mathcal{H}_i^t$ of order d, with dual bases $\tilde{\Psi}^{(i)}$ of order $\tilde{d}$.

We write $\Psi^{(i)} = \{\psi_\lambda^{(i)} : \lambda \in J^{(i)}\}$ and $\tilde{\Psi}^{(i)} = \{\tilde{\psi}_\lambda^{(i)} : \lambda \in J^{(i)}\}$, where we think as λ consisting of two coordinates referring to scale and location respectively. Denoting the scale associated to λ as $|\lambda| \in \mathbb{N}$, we assume that the primal wavelets are *local* in the sense that $\operatorname{supp} \psi_\lambda^{(i)}$ intersects a uniformly bounded number of boxes $2^{-|\lambda|}(\alpha + [0,1]^n)$ ($\alpha \in \mathbb{Z}^n$), and conversely, that any box $2^{-|\lambda|}(\alpha + [0,1]^n)$ intersects the support of at most a uniformly bounded number of $\psi_\lambda^{(i)}$ with this level $|\lambda|$.

We put

$$\gamma = \sup\{s \in \mathbb{R} : \|\psi_\lambda^{(i)}\|_{H^s(\Omega_i)} \lesssim 2^{|\lambda|s}\|\psi_\lambda^{(i)}\|_{L_2(\Omega_i)},\ \lambda \in J^{(i)},\ 1 \le i \le m\},$$

with an analogous definition of $\tilde{\gamma}$ involving dual wavelets. Necessarily, it holds that $t \in (-\tilde{\gamma}, \gamma)$. It is known that if the primal wavelets are piecewise smooth globally C^r-functions for some $r \in \mathbb{N} \cup \{-1\}$, then $\gamma = r + \frac{3}{2}$. It holds that $r \le d - 2$ with equality sign for spline wavelets.

Now let $L : \mathcal{H}^t \to (\mathcal{H}^t)'$ be boundedly invertible. Then $\mathbf{M} = FLF'$ is represented by an $m \times m$ blockmatrix with $(i,\check{\imath})$-th block equal to the infinite matrix $\langle \omega_i \Psi^{(i)}, L\omega_{\check{\imath}}\Psi^{(\check{\imath})}\rangle_{L_2(\Omega)}$. Assuming that the weights ω_i vanish at the internal boundaries so that they are globally smooth, the analysis of the compressibility of each of these blocks can follow exactly the same lines as that of the compressibility of L with respect to a biorthogonal wavelet *basis* characterized by the same tuple $(d, \gamma, \tilde{d}, \tilde{\gamma})$.

If for some $\sigma > 0$, $L, L' : \mathcal{H}^{t+\sigma} \to \mathcal{H}^{-t+\sigma}$ are bounded, then by substituting the estimates [Dah97, (9.4.5), (9.4.8)] into [CDD01, Proposition 6.6.2] we infer that $\mathbf{M}$ is s^*-compressible with

$$s^* = \frac{\min\{\sigma, \gamma - t, t + \tilde{d}\}}{n} - \frac{1}{2}, \tag{4.4}$$

at least when this value is positive. (We have used the fact that the condition $\sigma < t + \tilde{\gamma}$ imposed for [Dah97, (9.4.8)] can actually be relaxed to $\sigma \leq t + \tilde{d}$.) Above result holds true for local operators L, i.e., $\langle v, Lw\rangle_{L_2(\Omega)} = 0$ when $\operatorname{supp} v \cap \operatorname{supp} w = \emptyset$, as well as for non-local L of the form

$$(Lv)(x) = \int_\Omega K(x,y)v(y)dy,$$

with a Schwartz kernel that has the Calderon-Zygmund property

$$|\partial_x^\alpha \partial_y^\beta K(x,y)| \lesssim |x-y|^{-(n+2t+|\alpha|+|\beta|)}, \qquad (n+2t+|\alpha|+|\beta|>0).$$

The spaces $\mathcal{H}^r$ we used to formulate above continuity assumptions on L and L' are defined for $r \geq 0$ by $\mathcal{H}^r = \begin{cases} [L_2(\Omega), H^{|t|}_{0,\Gamma^D}(\Omega)]_{r/t} & r \leq |t|, \\ H^{|t|}_{0,\Gamma^D}(\Omega) \cap H^r(\Omega) & r \geq |t|, \end{cases}$ and $\mathcal{H}^{-r} = (\mathcal{H}^r)'$.

The result given in (4.4) is not completely satisfactory. Indeed, in any case when the primal wavelets are sufficiently smooth, in §4.2 we learned that if the solution u is in $B^{sn+t}_\tau(L_\tau(\Omega))$ for some $s \in (0, \frac{d-t}{n})$, then it has a representation $u = \mathbf{u}^T\Psi$ such that the best N-term approximation for $\mathbf{u}$ converges with a rate N^{-s}. On the other hand, the convergence rate of the solutions yielded by **(mod)SOLVE** is not only bounded by above value of s but also by s^*. Since s^* given in (4.4) is less or equal to $\frac{\gamma-t}{n} - \frac{1}{2}$, and moreover $\gamma < d$, on basis of this result we may only conclude that **(mod)SOLVE** has optimal computational complexity for solutions *with limited regularity*.

Yet, in a forthcoming paper ([Ste02]) it will be shown that (4.4) is actually too pessimistic, and that, for $\sigma \geq d - t$, with suitable wavelets for local as well as for non-local operators s^*-compressibility with $s^* \geq \frac{d-t}{n}$ can be shown.

Since the use of the projector $\mathbf{P}$ applied in **modSOLVE** seems restricted to the frame construction from this paper, we discuss its compressibility here. Recall that $\mathbf{P}$ is given by an $m \times m$ block matrix with $(i,\breve{\imath})$-th block being equal to the infinite matrix $\mathbf{P}^{(i,\breve{\imath})} = \langle \hat{\chi}_i \omega_i \tilde{\Psi}^{(i)}, \omega_{\breve{\imath}} \Psi^{(\breve{\imath})}\rangle_{L_2(\Omega)}$. So it is sufficient to investigate the compressibility of any of these blocks.

For biorthogonal wavelet bases it can be shown that for $r \in [-\tilde{d}, \gamma)$, $s < \gamma$,

$$\|\cdot\|_{\mathcal{H}^r_i} \lesssim 2^{\breve{\ell}(r-s)} \|\cdot\|_{\mathcal{H}^s_{\breve{\imath}}} \quad \text{on} \quad W^{(\breve{\imath})}_{\breve{\ell}} := \operatorname{span}\{\psi^{(\breve{\imath})}_{\breve{\lambda}} : |\breve{\lambda}| = \breve{\ell}\}, \tag{4.5}$$

and analogously for $r \in [-d, \tilde{\gamma})$, $s < \tilde{\gamma}$,

$$\|\cdot\|_{\mathcal{H}^r_i} \lesssim 2^{\ell(r-s)} \|\cdot\|_{\mathcal{H}^s_i} \quad \text{on} \quad \tilde{W}^{(i)}_\ell = \operatorname{span}\{\tilde{\psi}^{(i)}_\lambda : |\lambda| = \ell\}. \tag{4.6}$$

Here $\mathcal{H}^r_{\breve{\imath}}$ is defined as $\mathcal{H}^r$ with (Ω, Γ^D) replaced by $(\Omega_{\breve{\imath}}, \Gamma^D_{\breve{\imath}})$.

So assuming that the weights vanish at the internal boundaries, for $\tilde{w}^{(i)}_\ell \in \tilde{W}^{(i)}_\ell$, $w^{(\breve{\imath})}_{\breve{\ell}} \in W^{(\breve{\imath})}_{\breve{\ell}}$, and $-d \leq s - t < \tilde{\gamma}$, $-\tilde{d} \leq t - s < \gamma$, i.e., $s \in (t-\gamma, t+\tilde{\gamma})$, we have

$$\begin{aligned} |\langle \hat{\chi}_i \omega_i^{-1} \tilde{w}^{(i)}_\ell, \omega_{\breve{\imath}} w^{(\breve{\imath})}_{\breve{\ell}}\rangle_{L_2(\Omega)}| &\lesssim \|\hat{\chi}_i \omega_i^{-1} \tilde{w}^{(i)}_\ell\|_{\mathcal{H}^{s-t}} \|\omega_{\breve{\imath}} w^{(\breve{\imath})}_{\breve{\ell}}\|_{\mathcal{H}^{t-s}} \\ &\lesssim \|\tilde{w}^{(i)}_\ell\|_{\mathcal{H}^{s-t}_i} \|w^{(\breve{\imath})}_{\breve{\ell}}\|_{\mathcal{H}^{t-s}_{\breve{\imath}}} \lesssim 2^{s(\ell-\breve{\ell})} \|\tilde{w}^{(i)}_\ell\|_{\mathcal{H}^{-t}_i} \|w^{(\breve{\imath})}_{\breve{\ell}}\|_{\mathcal{H}^t_{\breve{\imath}}}. \end{aligned}$$

Let us now define $\hat{\mathbf{P}}_j^{(i,\breve{\imath})}$ by removing all blocks $\left[\langle\hat{\chi}_i\omega_i\tilde{\psi}_\lambda^{(i)},\omega_{\breve{\imath}}\psi_{\breve{\lambda}}^{(\breve{\imath})}\rangle_{L_2(\Omega)}\right]_{|\lambda|=\ell,\,|\breve{\lambda}|=\breve{\ell}}$ from $\mathbf{P}^{(i,\breve{\imath})}$ for which $\ell>\breve{\ell}+\frac{k(j)}{n-1}$ or $\breve{\ell}>\ell+\frac{\gamma-t}{\tilde{\gamma}+t}\frac{k(j)}{n-1}$ (here and in the next sentence $\frac{k(j)}{n-1}$ should read as $2^{k(j)}$ when $n=1$). Then using the fact that the $\tilde{\Psi}^{(i)}$ or $\Psi^{(\breve{\imath})}$ are Riesz bases for $\mathcal{H}_i^{-t}$ or $\mathcal{H}_{\breve{\imath}}^t$ respectively, we infer that for any $0<s<\gamma-t$,

$$\|\mathbf{P}^{(i,\breve{\imath})}-\hat{\mathbf{P}}_j^{(i,\breve{\imath})}\|\lesssim 2^{-s\frac{k(j)}{n-1}}.$$

For the next step, we will assume that also *the dual wavelets are local* and have connected supports. Furthermore, we assume that the primal wavelets are piecewise smooth. With this we mean that $\operatorname{sing\,supp}\psi_{\breve{\lambda}}^{(\breve{\imath})}$ is $(n-1)$-dimensional, i.e., for $\ell\geq|\breve{\lambda}|$ the number of boxes $2^{-\ell}(\alpha+[0,1]^n)$ that intersect $\operatorname{sing\,supp}\psi_{\breve{\lambda}}^{(\breve{\imath})}$ is of the order $2^{(\ell-|\breve{\lambda}|)(n-1)}$, and furthermore that $\operatorname{supp}\psi_{\breve{\lambda}}^{(\breve{\imath})}\backslash\operatorname{sing\,supp}\psi_{\breve{\lambda}}^{(\breve{\imath})}$ is the union of k sets $\Xi_{\breve{\lambda}}^{(\breve{\imath},1)},\dots,\Xi_{\breve{\lambda}}^{(\breve{\imath},k)}$, mutually separated by $\operatorname{sing\,supp}\psi_{\breve{\lambda}}^{(\breve{\imath})}$, and that for each $1\leq q\leq k$, there exists a smooth function $\xi_{\breve{\lambda}}^{(\breve{\imath},q)}$ on $\Omega_{\breve{\imath}}$ with $\psi_{\breve{\lambda}}^{(\breve{\imath})}=\xi_{\breve{\lambda}}^{(\breve{\imath},q)}$ on $\Xi_{\breve{\lambda}}^{(\breve{\imath},q)}$, $\operatorname{diam}\operatorname{supp}\xi_{\breve{\lambda}}^{(\breve{\imath},q)}\lesssim 2^{-|\breve{\lambda}|}$, $\sup_{x\in\Omega_{\breve{\imath}}}|\partial^\beta\xi_{\breve{\lambda}}^{(\breve{\imath},q)}|\lesssim 2^{(|\beta|+\frac{n}{2}-t)|\breve{\lambda}|}$ ($\beta\in I\!N^n$), and so in particular

$$\|\omega_{\breve{\imath}}\xi_{\breve{\lambda}}^{(\breve{\imath},q)}\|_{\mathcal{H}^d}\lesssim 2^{|\breve{\lambda}|(d-t)}. \tag{4.7}$$

Given $\breve{\lambda}$, $1\leq q\leq k$ and $\ell>|\breve{\lambda}|$, let $A_{\breve{\lambda},q,\ell}^{(i,\breve{\imath})}=\{|\lambda|=\ell:\operatorname{supp}\tilde{\psi}_\lambda^{(i)}\subset\Xi_{\breve{\lambda}}^{(\breve{\imath},q)}\}$. We define $\mathbf{P}_j^{(i,\breve{\imath})}$ by removing all entries $\langle\hat{\chi}_i\omega_i\tilde{\psi}_\lambda^{(i)},\omega_{\breve{\imath}}\psi_{\breve{\lambda}}^{\breve{\imath}}\rangle_{L_2(\Omega)}$ from $\hat{\mathbf{P}}_j^{(i,\breve{\imath})}$ when $|\lambda|-|\breve{\lambda}|>\frac{k(j)}{n}$, $\operatorname{supp}\tilde{\psi}_\lambda^{(i)}\cap\operatorname{supp}\psi_{\breve{\lambda}}^{(\breve{\imath})}\neq\emptyset$ and $\operatorname{supp}\tilde{\psi}_\lambda^{(i)}\cap\operatorname{sing\,supp}\psi_{\breve{\lambda}}^{(\breve{\imath})}=\emptyset$, and so $\lambda\in A_{\breve{\lambda},q,|\lambda|}^{(i,\breve{\imath})}$ for some $1\leq q\leq k$. Then by using (4.7), and (4.6) with $(r,s)=(-d,-t)$, for any $\mathbf{c},\mathbf{d}\in\ell_2$ we have

$$\begin{aligned}
\langle\mathbf{c},(\hat{\mathbf{P}}_j^{(i,\breve{\imath})}-\mathbf{P}_j^{(i,\breve{\imath})})\mathbf{d}\rangle_{\ell_2}&=|\sum_{\ell-\breve{\ell}>\frac{k(j)}{n}}\sum_{|\breve{\lambda}|=\breve{\ell}}\mathbf{d}_{\breve{\lambda}}\langle\sum_{q=1}^k\sum_{\lambda\in A_{\breve{\lambda},q,\ell}^{(i,\breve{\imath})}}\mathbf{c}_\lambda\hat{\chi}_i\omega_i^{-1}\tilde{\psi}_\lambda^{(i)},\omega_{\breve{\imath}}\xi_{\breve{\lambda}}^{(\breve{\imath},q)}\rangle|\\
&\lesssim\sum_{\ell-\breve{\ell}>\frac{k(j)}{n}}2^{-(d-t)(\ell-\breve{\ell})}\sum_{|\breve{\lambda}|=\breve{\ell}}|\mathbf{d}_{\breve{\lambda}}|\sum_{q=1}^k\|\sum_{\lambda\in A_{\breve{\lambda},q,\ell}^{(i,\breve{\imath})}}\mathbf{c}_\lambda\tilde{\psi}_\lambda^{(i)}\|_{\mathcal{H}_i^{-t}}\\
&\lesssim\sum_{\ell-\breve{\ell}>\frac{k(j)}{n}}2^{-(d-t)(\ell-\breve{\ell})}\sqrt{\sum_{|\breve{\lambda}|=\breve{\ell}}|\mathbf{d}_\lambda|^2}\sqrt{\sum_{|\breve{\lambda}|=\breve{\ell}}\left(\sum_{q=1}^k\sqrt{\sum_{\lambda\in A_{\breve{\lambda},q,\ell}^{(i,\breve{\imath})}}|\mathbf{c}_\lambda|^2}\right)^2}\\
&\lesssim\sum_{\ell-\breve{\ell}>\frac{k(j)}{n}}2^{-(d-t)(\ell-\breve{\ell})}\sqrt{\sum_{|\breve{\lambda}|=\breve{\ell}}|\mathbf{d}_\lambda|^2}\sqrt{\sum_{|\lambda|=\ell}|\mathbf{c}_\lambda|^2}\lesssim 2^{-(d-t)\frac{k(j)}{n}}\|\mathbf{d}\|\|\mathbf{c}\|,
\end{aligned}$$

where for the last line we have used that, by the locality of the primal wavelets, each λ is contained in at most a uniformly bounded number of sets $A^{(i,\check{\imath})}_{\check{\lambda},q,|\lambda|}$. We conclude that $\|\hat{\mathbf{P}}_j^{(i,\check{\imath})} - \mathbf{P}_j^{(i,\check{\imath})}\| \lesssim 2^{(d-t)\frac{k(j)}{n}}$.

By the locality of both primal and dual wavelets and the piecewise smoothness of the primal wavelets, the number of non-zeros in each *column* of $\mathbf{P}_j^{(i,\check{\imath})}$ is of the order $2^{n\frac{k(j)}{n}} + 2^{(n-1)\frac{k(j)}{n-1}} = 2^{k(j)}$ ($2^{k(j)} + 2^{k(j)} = 2^{k(j)}$ when $n = 1$). By substituting $k(j) = j + \log(\alpha_j)$, with for example $\alpha_j = j^{-(1+\varepsilon)}$ for some $\varepsilon > 0$, we infer that $\mathbf{P}$ is s^*-compressible with

$$s^* = \min\{\tfrac{\gamma-t}{n-1}, \tfrac{d-t}{n}\}.$$

($s^* = d - t$ for $n = 1$). So, if, when $n > 1$, $r \geq -\frac{3}{2} + d + (t-d)/n$, which was also assumed in (4.3), then $s^* = \frac{d-t}{n}$.

Acknowledgement

The author would like to thank Karsten Urban (RWTH Aachen) for many inspiring dicussions, and for his valuable comments on an earlier version of this manuscript.

References

[Bar02] A. Barinka. *Fast Evaluation Tools for Adaptive Wavelet Schemes*. PhD thesis, RTWH Aachen, 2002. To appear.

[BL76] J. Bergh and J. Löfström. *Interpolation Spaces*. Grundlehren der mathematischen Wissenschaften 223. Springer-Verlag, Berlin, 1976.

[CDD00] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods II - Beyond the elliptic case. IGPM preprint 199, RWTH Aachen, November 2000. To appear in Found. Comput. Math.

[CDD01] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations – Convergence rates. *Math. Comp*, 70:27–75, 2001.

[CES00] A. Cohen, L.M. Echeverry, and Q. Sun. Finite element wavelets. Technical report, Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, 2000.

[CM00] A. Cohen and R. Masson. Wavelet adaptive method for second order elliptic problems: Boundary conditions and domain decomposition. *Numer. Math.*, 86:193–238, 2000.

[Coh00] A. Cohen. Wavelet methods in numerical analysis. In P.G. Ciarlet and J. L. Lions, editors, *Handbook of numerical analysis. Vol. VII.*, pages 417–711. North-Holland, Amsterdam, 2000.

[CTU99] C. Canuto, A. Tabacco, and K. Urban. The wavelet element method part I: Construction and analysis. *Appl. Comput. Harmonic Anal*, 6:1–52, 1999.

[Dah97] W. Dahmen. Wavelet and multiscale methods for operator equations. *Acta Numerica*, 55:55–228, 1997.

[Dah99a] S. Dahlke. Besov regularity for elliptic boundary value problems on polygonal domains. *Appl. Math. Lett.*, 12(6):31–36, 1999.

[Dah99b] S. Dahlke. Besov regularity for the Stokes problem. In W. Haussmann, K. Jetter, and M. Reimer, editors, *Advances in Multivariate Approximation*, Mathematical Research 107, pages 129–138, Berlin, 1999. Wiley VCH.

[Dau92] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

[DD97] S. Dahlke and R. DeVore. Besov regularity for elliptic boundary value problems. *Comm. Partial Differential Equations*, 22(1 & 2):1–16, 1997.

[DDU01] S. Dahlke, W. Dahmen, and K. Urban. Adaptive wavelet methods for saddle point problems - Optimal convergence rates. IGPM report, RWTH Aachen, 2001. To appear in SIAM J. Numer. Anal.

[DeV98] R. DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.

[DKU99] W. Dahmen, A. Kunoth, and K. Urban. Biorthogonal spline-wavelets on the interval - Stability and moment conditions. *Appl. Comp. Harm. Anal.*, 6:132–196, 1999.

[DS99a] W. Dahmen and R. Schneider. Composite wavelet bases for operator equations. *Math. Comp.*, 68:1533–1567, 1999.

[DS99b] W. Dahmen and R. Schneider. Wavelets on manifolds I: Construction and domain decomposition. *SIAM J. Math. Anal.*, 31:184–230, 1999.

[DS99c] W. Dahmen and R.P. Stevenson. Element-by-element construction of wavelets satisfying stability and moment conditions. *SIAM J. Numer. Anal.*, 37(1):319–352, 1999.

[DUV02] W. Dahmen, K. Urban, and J. Vorloeper. Adaptive wavelet methods - Basic concepts and applications to the Stokes problem. IGPM report, RWTH Aachen, May 2002.

[Met02] A. Metselaar. *Handling Wavelet Expansions in Numerical Methods*. PhD thesis, University of Twente, 2002. To appear.

[Ste00] R.P. Stevenson. Locally supported, piecewise polynomial biorthogonal wavelets on non-uniform meshes. Technical Report 1157, University of Utrecht, September 2000. To appear in Constr. Approx.

[Ste02] R.P. Stevenson. On the compressibility of operators in wavelet coordinates. In preparation, 2002.

Department of Mathematics, Utrecht University, P.O. Box 80.010, NL-3508 TA Utrecht, The Netherlands

E-mail address: stevenso@math.uu.nl