Adaptive Frame Methods for Nonlinear Elliptic Problems

Jens Kappei

30.03.2009

Abstract

This paper is concerned with adaptive numerical wavelet frame methods for nonlinear elliptic operator equations. The starting point of our considerations are the works [7, 8] of Cohen, Dahmen and DeVore, in which an adaptive, asymptotically optimal evaluation scheme for certain nonlinear expressions was established using wavelet Riesz bases. Due to further on existing problems in the construction of such bases on domains, we adjust their ideas to the case of wavelet frames. In particular, we show that the discretization of nonlinear problems by wavelet frames, which are constructed by an overlapping partition of the domain, allows us to establish a linear convergent, adaptive algorithm for the iterative solution of the discretized problem. We then focus on main aspects of the numerical implementation of this iterative scheme. More precisely, we point out that specific non-canonical frame expansions enable us to use tree approximation ideas known from the usage of wavelet Riesz bases for the estimation of the index set of significant frame expansion coefficients of nonlinear expressions. We then prove that the ideas from [1] can be used to approximate the afore selected expansion coefficients by quadrature in linear time. Finally, concerning the complexity of the approximation, we show that all building blocks needed in the algorithm can be realized with asymptotical optimality compared with a best N-term approximation respecting tree structures, which yields an asymptotically optimal algorithm for nonlinear problems discretized by wavelet frames.

Mathematics Subject Classification: 35A35, 35J60, 42C40, 65J15, 65N12, 65N15
Key Words: Adaptive algorithms, nonlinear approximation, nonlinear operator equations, optimal computational complexity, frames, Besov spaces, biorthogonal wavelets, tree structures, overlapping domain decomposition

1 Introduction

In the last years, wavelet methods have developed to a very powerful mathematical tool with many important applications, particularly in the field of numerical analysis. Here, the most exciting results have been obtained in the context of the numerical treatment of linear elliptic operator equations. Indeed, the strong analytical properties of wavelets yield uniformly bounded condition numbers of the associated stiffness matrices, they allow for efficient compression strategies and, moreover, open the possibility of adaptive numerical schemes of optimal order [2, 5, 10]. However, the classical approach to treat such an operator equation over a closed manifold or bounded domain by means of a wavelet bases still includes some problems worth mentioning: One is usually faced with relatively high condition numbers or lacking smoothness and the existing constructions are not easy to implement [4, 16, 17, 20]. One approach to ameliorate these problems is to use a weaker concept, i.e., to work with *frames* instead of bases [11, 12, 24]. Their construction is quite simple. One only has to construct an overlapping partition $\Omega = \bigcup_i \Omega_i$ of the underlying domain Ω by means of parametric images of the unite cube. Then, by lifting tensor product boundary adapted wavelet bases on the unit cube to each subdomain Ω_i and collecting everything together, a wavelet frame is obtained. Fortunately, it has been shown that all advantages of wavelet methods outlined above can also be established in the case of frames [12, 13, 24, 26].

Due to the encouraging results for linear elliptic operator equations, the numerical treatment of nonlinear operator equations has become a field of continuous mathematical research [1, 7, 8, 18]. The basic paradigm stated in [1] uses a two-step approach to the evaluation of nonlinear expressions $G(u)$, where u is a function in an adequate smoothness space and the $G(\cdot)$ the nonlinearity.

First, in the *prediction step*, the set of significant expansion coefficients of $G(u)$ is estimated, where "significance" means that the *prediction error*, i.e., the ℓ_2-norm of the dropped coefficients, is bounded by a given accuracy $\varepsilon > 0$. In [7, 8] it was shown for wavelet Riesz bases that this can be done by collecting the significant basis expansion coefficients of u and using then appropriate growth estimates for G on the coefficient level. The main structural tool in this context consists in imposing a *tree structure* on the index set of used wavelets. For this reason, we will adjust the concept of tree structures to an overlapping partition of the underlying domain by defining trees *locally* on each subdomain Ω_i first and unifying then those local trees to get a global tree structure on Ω, which we call an *aggregated tree*. Doing this, we show that using a special class of wavelet frames, the so-called *Gelfand frames*, which are, roughly speaking, Hilbert frames that in addition induce norm equivalences for associated smoothness spaces, together with a non-canonical dual frame allows essentially the same prediction strategy as for wavelet bases.

An important question in view of the numerical implementation of a prediction procedure concerns the *accuracy/work balance* that can be realized in the procedure. In particular, we are interested in the question, whether the prediction process can be designed to be *asymptotically optimal* in some sense. For elliptic linear problems adaptive wavelet methods have been established, see [5, 24], which are asymptotically optimal in comparison with the *best N-Term approximation*. To explain this, recall that in ℓ_2 the best N-term approximation $\mathbf{v}_N$ of a vector $\mathbf{v} \in \ell_2$ is just the vector consisting of the N largest coefficients of $\mathbf{v}$ in modulus (which is, in general, an unstructured coefficient set), since it provides the minimal error of all approximations to $\mathbf{v}$ with at most N elements, $\|\mathbf{v}-\mathbf{v}_N\|_{\ell_2} = \sigma_N(\mathbf{v}) := \inf\{\|\mathbf{v}-\mathbf{w}_N\|_{\ell_2} : \#\mathrm{supp}(\mathbf{w}_N) \leq N\}$. If this error of the best N-term approximation decays in N with a certain rate, $\sigma_N(\mathbf{v}) \leq C_1 N^{-s}$, with $s > 0$ and a constant C_1, the number of non-zero coefficients in $\mathbf{v}_N$ needed to approximate $\mathbf{v}$ up to an accuracy $\varepsilon > 0$ is bounded by $C_2\varepsilon^{-1/s}$. A numerical algorithm that calculates an approximation $\mathbf{w}_\varepsilon$ to $\mathbf{v}$ with $\|\mathbf{v} - \mathbf{w}_\varepsilon\|_{\ell_2} \leq \varepsilon$ is then called asymptotically optimal if the number of non-trivial coefficients in $\mathbf{w}_\varepsilon$ and the computational effort to determine them is, up to constants, also bounded by $\varepsilon^{-1/s}$.

Due to the structural demand of tree structures on the index set of used wavelets, we can not expect to find prediction procedures for nonlinear problems, which are asymptotically optimal in this way. But, using the *best N-term tree approximation* as benchmark, it has turned out [7, 8] that asymptotically optimal prediction procedures can be implemented using wavelet Riesz bases. As a main result of this paper we show that our prediction strategy for frames is asymptotically optimal compared to a best N-term approximation based on aggregated trees if the frame coefficients of G are in a suitable approximation space, which in turn follows from an adequate smoothness of the underlying function in a certain Besov scale. There, we will make intense use of the fact that the lifted basis systems on each subdomain form local wavelet bases for a corresponding scale of smoothness spaces. Working with the local bases offers the possibility of parallelization of at least parts of the prediction step.

Secondly, in the *approximation step*, the afore designated coefficients are approximated using quadrature. There are several approaches to this topic [1, 18] for wavelet bases. The RECOVER scheme developed in [1] does this approximation level by level starting with the finest appearing scale and working *top-to-bottom* to the coarsest scale exploiting certain properties of the underlying tree structure. A main advantage of RECOVER is that the resulting *approximation error* can be assured to stay the same size as the prediction error for sufficiently precise quadrature.

Of course, in the frame case, it suggests itself that we will apply the RECOVER scheme to the local trees estimated in the prediction step. The main question is then, if similar error estimates can be assured. We proof that this is actually possible, at least if some additions to the prediction step are made, which follow from structural assumptions of the local trees in order to give an adequate error estimation in dual Sobolev norms. Furthermore, if we assume that quadrature for scaling functions can be done with constant complexity, the approximation step shows the same complexity as the prediction step.

Combining these results concerning $G(u)$ with some thoughts about the realization of other building blocks needed for the implementation allows us finally to give an asymptotically optimal algorithm for the solution of nonlinear problems.

This paper is organized as follows. In Section 2 we recall the setting of adaptive wavelet frame discretizations as far as it is needed for our purposes. Due to the central role of wavelet bases indicated above, we also give a short outline of their useful properties. Then, in Section 3 the discretization of nonlinear problems via wavelet frames is addressed. The main result of this section is a linear convergent adaptive iterative scheme. Section 4 contains the necessary background in view of tree structures and the *best N-term tree approximation* as benchmark for the numerical algorithm. After this it deals primarily with the prediction step in the wavelet frame setting, while Section 5 introduces the RECOVER scheme and its application to Gelfand frames including error estimations for the approximation step. The final Section 6 combines the results from the previous sections in order to estimate the complexity of the algorithm given in Section 3 proving its asymptotical optimality.

2 Frame Discretization of Nonlinear Problems

2.1 The General Setting

In this subsection, we describe the basic concepts of frame discretization schemes for semi-nonlinear operator equations

$$Au + G(u) = f, \tag{2.1}$$

which include a linear part Au and a nonlinear part $G(u)$. For the first part we assume A to be a boundedly invertible *linear* operator from some separable Hilbert space H into its normed dual space H', i.e., there exist two constants $0 < c_A \le C_A < \infty$ with

$$c_A \|v\|_H \le \|Av\|_{H'} \le C_A \|v\|_H \quad \text{for all } v \in H. \tag{2.2}$$

We shall focus on the important special case that

$$a(v, w) := \langle Av, w \rangle_{H' \times H} \tag{2.3}$$

defines a *symmetric* bilinear form on H, where $\langle \cdot, \cdot \rangle_{H' \times H}$ denotes the dual pairing of H' and H. Note that in cases where $a(\cdot, \cdot)$ is not symmetric, it is possible to use the matrix $A^T A$ instead, see [7] for further details. Furthermore, $a(\cdot, \cdot)$ is assumed to be *elliptic*, i.e., there exist two constants $0 < c_e \le C_e < \infty$, so that

$$c_e \|v\|_H^2 \le a(v, v) \le C_e \|v\|_H^2 \quad \text{for all } v \in H, \tag{2.4}$$

which can be seen to imply (2.2). Concerning the nonlinear part $G(u)$, we assume $G : \mathbb{R} \to \mathbb{R}$ to be a real-valued nonlinear function mapping H to H'. It will be shown in the next subsection that this can be assured for a wide class of polynomial functions. We finally make the *monotonicity* assumption

$$(x-y)(G(x)-G(y)) \geq 0 \quad \text{for all } x,y \in \mathbb{R}. \tag{2.5}$$

Typical examples for this setting are variational formulations of second order elliptic boundary value problems over a bounded domain $\Omega \subset \mathbb{R}^d$ with a nonlinear reaction or perturbation term such as a disturbed Poisson equation

$$\begin{aligned} -\Delta u + u^3 &= f \quad \text{on } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.6}$$

In this case, $H = H_0^1(\Omega)$, $H' = H^{-1}(\Omega)$, the corresponding bilinear form is given by

$$a(v,w) = \int_\Omega \nabla v \cdot \nabla w \, \mathrm{d}x,$$

and $G(x) = x^3$ indeed maps $H_0^1(\Omega)$ to $H^{-1}(\Omega)$ if $d \leq 3$, see [9] for a proof. Thus, H is typically a Sobolev space of positive order. Therefore, from now on we always assume that H and H' form a *Gelfand triple* together with $L_2(\Omega)$, i.e.,

$$H \subset L_2(\Omega) \subset H' \tag{2.7}$$

with continuous and dense embeddings. Furthermore, problem (2.1) will be understood in a weak sense (variational form), i.e., our aim is to find some $u \in H$, so that for given data f it holds

$$\langle Au, v\rangle_{H'\times H} + \langle G(u), v\rangle_{H'\times H} = \langle f, v\rangle_{H'\times H} \quad \text{for all } v \in H. \tag{2.8}$$

2.2 Mapping Properties of Nonlinear Functions

As noted in the last subsection, for (2.8) to be a meaningful formulation, $G(\cdot)$ has to map H to H' and for the model problem (2.6) it was shown in [9] that $G(x) = x^3$ fulfills this assumption if $d \leq 3$ holds. In this subsection we want to get a more systematic view on the mapping properties of $G(\cdot)$. For the classical Sobolev spaces $H_0^t(\Omega), t > 0$, and their normed dual spaces $H^{-t}(\Omega)$ the following lemma shown in [8] can be used.

Lemma 2.1 *Assume that the real-valued nonlinear function $G(\cdot)$ satisfies the growth condition*

$$\left|G^{(n)}(x)\right| \leq C(1+|x|)^{\max\{0,p-n\}} \quad \textit{for all } x \in \mathbb{R},\ n = 0,1,\ldots,n^*, \tag{2.9}$$

for some $p \geq 0$ (not necessarily an integer) and an integer $n^* \geq 0$. Then, for $t > 0$, $G(\cdot)$ maps $H_0^t(\Omega)$ to $H^{-t}(\Omega)$ under the restriction $0 \leq p \leq p^* := \frac{d+2t}{d-2t}$ if $t < d/2$, and with no restriction otherwise. If in addition $n^* \geq 1$, the stability condition

$$\|G(v) - G(w)\|_{H^{-t}(\Omega)} \leq C_G(\max\{\|v\|_{H^t(\Omega)}, \|w\|_{H^t(\Omega)}\}) \cdot \|v-w\|_{H^t(\Omega)} \quad \textit{for all } v,w \in H_0^t(\Omega) \tag{2.10}$$

holds with $x \mapsto C_G(x)$ a non-decreasing non-negative function.

We assume that (2.9) holds in the following. This is particularly true for the model problem (2.6), since all polynomials $G(x) = x^p$ satisfy this assumption with $n^* := p$, if p is an integer, and $n^* := \lfloor p \rfloor$ otherwise, see also [9].

Looking for more general functions satisfying (2.9) leads to the theory of nonlinear Nemytskij-Operators. For a domain $\Omega \subset \mathbb{R}^d$ and $G(x,\xi) : \Omega \times \mathbb{R}^m \to \mathbb{C}$ a Carathéodory function, i.e., $G(\cdot,\xi)$ is measurable on Ω and $G(x,\cdot)$ is continuous for almost all $x \in \Omega$, a Nemytskij-Operator is of the form

$$T_G(f_1,\ldots,f_m)(x) := G(x, f_1(x), \ldots, f_m(x)), \quad x \in \Omega.$$

These operators and their mapping properties are addressed in detail in [23]. But only for real-valued L_p-functions and the Sobolev spaces $W^1(L_p(\Omega))$ there is a satisfying answer concerning their mapping properties in the framework of Besov-Triebel-Lizorkin spaces.

2.3 Well-Posed Nonlinear Variational Problem

Under the given assumptions on A and $G(\cdot)$ the problem (2.8) is, in principle, a meaningful variational formulation. For the numerical treatment we have to assure next that the problem is *well-posed.* The meaning of well-posed problems in the context of nonlinear functions can be found in [1] or [7] for instance.

According to this, we say that problem (2.8) is well-posed if

- There exists a unique solution $u \in H$ for (2.8) and
- For all v in a neighborhood $\mathcal{U}$ of u the mapping $v \mapsto DG(v)$ is continuous and there exist constants $0 < c_{G,v} \leq C_{G,v} < \infty$ depending on $G(\cdot)$ and v, so that

$$c_{G,v}\|w\|_H \leq \|DG(v)w\|_{H'} \leq C_{G,v}\|w\|_H \quad \text{for all } w \in H. \tag{2.11}$$

Here $DG(v)$ denotes the Fréchet derivative of $G(\cdot)$ at $v \in H$, which is a mapping from H to H' defined by

$$\langle z, DG(v)w\rangle_{H\times H'} = \lim_{t\to 0}\frac{1}{t}\langle z, G(v+tw) - G(v)\rangle_{H\times H'} \quad \text{for all } z \in H.$$

These assumptions might be difficult to verify for special nonlinearities. Under the given assumptions on $G(\cdot)$, however, they are an immediate consequence of the Browder-Minty theorem as shown in the proof of the following lemma (see [7]).

Lemma 2.2 *Assume that $G(\cdot)$ conforms assumption (2.5). Then, there exists a unique solution $u \in H$ of (2.8) for all $f \in H'$. Furthermore, assuming (2.10), problem (2.8) is well-posed with constants $c_{G,v} := c_e$ and $C_{G,v} := C_e + C_G(\|v\|_H)$, where c_e and C_e are the constants from (2.4) and $C_G(\cdot)$ is the function from (2.10).*

2.4 Wavelet Frames and their Construction

We are interested in the discretization of (2.8) using wavelet frames. To this end, recall first that for a countable index set $\mathcal{N}$ and a separable Hilbert space $\mathcal{H}$ with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}} := \langle\cdot,\cdot\rangle_{\mathcal{H}}^{1/2}$ a subset $\mathcal{F} = \{f_n\}_{n\in\mathcal{N}}$ of $\mathcal{H}$ is called a (Hilbert) *frame* for $\mathcal{H}$ if there exist constants $0 < A_{\mathcal{F}} \leq B_{\mathcal{F}} < \infty$ with

$$A_{\mathcal{F}}\|f\|_{\mathcal{H}}^2 \leq \sum_{n\in\mathcal{N}} |\langle f, f_n\rangle_{\mathcal{H}}|^2 \leq B_{\mathcal{F}}\|f\|_{\mathcal{H}}^2 \quad \text{for all } f \in \mathcal{H}. \tag{2.12}$$

The (ambiguous) constants $A_{\mathcal{F}}, B_{\mathcal{F}}$ are called *frame constants.* As a consequence of (2.12), any frame $\mathcal{F}$ is automatically dense in $\mathcal{H}$, which makes frames suitable for the numerical discretization of, e.g., the operator equation (2.8). Moreover, the corresponding *analysis* and *synthesis* operators given by

$$F : \mathcal{H} \to \ell_2(\mathcal{N}),\ f \mapsto (\langle f, f_n\rangle_{\mathcal{H}})_{n\in\mathcal{N}} \quad \text{and} \quad F^* : \ell_2(\mathcal{N}) \to \mathcal{H},\ \mathbf{c} \mapsto \sum_{n\in\mathcal{N}} c_n f_n$$

are bounded. In addition, F is injective and has closed range, while F^* is onto.

The composition $S := F^*F : \mathcal{H} \to \mathcal{H},\ Sf = \sum_{n\in\mathcal{N}}\langle f, f_n\rangle_{\mathcal{H}} f_n$, is a boundedly invertible, positive and self-adjoint operator, called the frame operator, and $\tilde{\mathcal{F}} := S^{-1}\mathcal{F}$ is again a frame for $\mathcal{H}$ with frame bounds $B_{\mathcal{F}}^{-1}$ and $A_{\mathcal{F}}^{-1}$, the so called *canonical dual frame* with analysis operator $\tilde{F} = FS^{-1}$ and synthesis operator $\tilde{F}^* = S^{-1}F^*$.

The importance of the canonical dual frame is its use in the reproduction of $\mathcal{H}$, since every $f \in \mathcal{H}$ possesses a (non-orthogonal) expansion

$$f = SS^{-1}f = \sum_{n\in\mathcal{N}} \langle f, S^{-1}f_n\rangle_{\mathcal{H}} f_n = S^{-1}Sf = \sum_{n\in\mathcal{N}} \langle f, f_n\rangle_{\mathcal{H}} S^{-1}f_n. \tag{2.13}$$

Since a frame is typically overcomplete in the sense that the coefficient functionals $\{c_n\}_{n\in\mathcal{N}}$ in the representation $f = \sum_{n\in\mathcal{N}} c_n(f)f_n$ are not unique, there may exist many possible non-canonical duals $\{\xi_n\}_{n\in\mathcal{N}}$ in $\mathcal{H}$, for which $f = \sum_{n\in\mathcal{N}}\langle f, \xi_n\rangle_{\mathcal{H}} f_n = \sum_{n\in\mathcal{N}}\langle f, f_n\rangle_{\mathcal{H}}\xi_n$ and $\|f\|_H^2 \eqsim \sum_{n\in\mathcal{N}} |\langle f, \xi_n\rangle_{\mathcal{H}}|^2$ hold. Here and in the following, '$a \eqsim b$' means that both quantities can be uniformly bounded by some constant multiple of each other, compare (2.12). A frame is a Riesz basis for $\mathcal{H}$ if and only if $\ker(F^*) = \{0\}$ and in such a case the dual frame is uniquely determined.

Concerning the discretization of (2.8), it remains to construct a suitable frame $\mathcal{F}$ for the solution space H of (2.1). Here we are particularly interested in the class of wavelet frames. Because wavelet systems are typically constructed in L_2 rather than in the solution space H, we will use frames that simultaneously allow expansions of the form (2.13) in a Hilbert space $\mathcal{H}$ and in a densely embedded Hilbert space $\mathcal{X} \subset \mathcal{H}$. Here we consider

$$\mathcal{X} \subset \mathcal{H} \simeq \mathcal{H}' \subset \mathcal{X}',$$

so that $(\mathcal{X}, \mathcal{H}, \mathcal{X}')$ is a Gelfand triple as in (2.7). Frames tailored to this Gelfand triple situation are given by the class of *Gelfand frames*, as introduced in [12].
A frame $\mathcal{F}$ for $\mathcal{H}$ with dual frame $\tilde{\mathcal{F}}$ is called a Gelfand frame for the Gelfand triple $(\mathcal{X}, \mathcal{H}, \mathcal{X}')$ if $\mathcal{F} \subset \mathcal{X}$, $\tilde{\mathcal{F}} \subset \mathcal{X}'$ and there exists a Gelfand triple $(b, \ell_2(\mathcal{N}), b')$ of sequence spaces, so that

$$F^* : b \to \mathcal{X},\ F^*\mathbf{c} = \sum_{n\in\mathcal{N}} c_n f_n \ \text{ and } \ \tilde{F} : \mathcal{X} \to b,\ \tilde{F}f = (\langle f, f_n\rangle_{\mathcal{X}\times\mathcal{X}'})_{n\in\mathcal{N}} \tag{2.14}$$

are bounded operators. By duality arguments in the Gelfand triple $(\mathcal{X}, \mathcal{H}, \mathcal{X}')$ we can infer from the identity $F^*\tilde{F} = id_{\mathcal{X}}$ that for a Gelfand frame $\mathcal{F}$ also the operators

$$\tilde{F}^* : b' \to \mathcal{X}',\ \tilde{F}^*\mathbf{c} = \sum_{n\in\mathcal{N}} c_n \tilde{f}_n \ \text{ and } \ F : \mathcal{X}' \to b',\ Ff = (\langle f, f_n\rangle_{\mathcal{X}'\times\mathcal{X}})_{n\in\mathcal{N}} \tag{2.15}$$

are bounded. In particular, the following reproducing formulas hold

$$\begin{aligned} f &= \sum_{n\in\mathcal{N}} \langle f, \tilde{f}_n\rangle_{\mathcal{X}\times\mathcal{X}'} f_n \ \text{ for all } f \in \mathcal{X}, \\ g &= \sum_{n\in\mathcal{N}} \langle g, f_n\rangle_{\mathcal{X}'\times\mathcal{X}} \tilde{f}_n \ \text{ for all } g \in \mathcal{X}'. \end{aligned}$$

In the following we assume that the abstract sequence space b and the space $\ell_2(\mathcal{N})$ can be identified via an isomorphism $D : b \to \ell_2(\mathcal{N})$. Consequently, the system $D^{-1}\mathcal{F}$ is indeed a frame for $\mathcal{X}$, see [22] for a proof. In all cases of practical interest b will be a suitable weighted $\ell_2(\mathcal{N})$-space.

For our applications the case $(\mathcal{X}, \mathcal{H}, \mathcal{X}') = (H, L_2(\Omega), H')$, where H denotes a Sobolev space of positive order, is the most important one. One prominent example [4, 14, 15, 16, 17, 20, 25] of a Gelfand frame is any *wavelet Riesz basis* $\Psi = \{\psi_\lambda\}_{\lambda\in K}$ in $L_2(\Omega)$, such that a range $s \in (0, \gamma)$ of Sobolev spaces $H_0^s(\Omega)$ can be characterized by weighted sequence norms of the primal wavelet coefficient arrays,

$$\|f\|_{H^s(\Omega)} \eqsim \left(\sum_{\lambda\in K} 2^{2|\lambda|s} |\langle f, \tilde{\psi}_\lambda\rangle_{L_2(\Omega)}|^2 \right)^{1/2},\ f \in H_0^s(\Omega),$$

where we have used that for wavelet systems the scaling mapping $\lambda \mapsto |\lambda| \in \mathbb{Z}_{\geq j_0}$ is well-defined. Any such wavelet Riesz basis Ψ is a Gelfand frame for $(H_0^s(\Omega), L_2(\Omega), H^{-s}(\Omega))$, $s \in (0,\gamma)$, with $b := \ell_{2,2^s}(K) := \ell_{2,2^{2|\cdot|s}}$. Here we use the weighted ℓ_2 space

$$\ell_{2,w}(K) := \left\{ \mathbf{c} = (c_\lambda)_{\lambda\in K} : \|\mathbf{c}\|_{\ell_{2,w}}^2 := \sum_{\lambda\in K} |c_\lambda|^2 w(\lambda) < \infty \right\}$$

for some weight function $w : K \to \mathbb{R}_+$.

Because wavelet Riesz bases in a separable Hilbert space $\mathcal{H}$ will play a central role in the following constructions, we will recall some of their basic properties referring to [19] for further details. A multiscale basis $\Psi = \{\psi_\lambda\}_{\lambda\in\Lambda} \subset \mathcal{H}$ is called a Riesz basis for $\mathcal{H}$ if every $h \in \mathcal{H}$ has a unique expansion $h = \sum_{\lambda\in\Lambda} h_\lambda \psi_\lambda =: \mathbf{h}^T\Psi$ with norm equivalence

$$\|\mathbf{h}\|_{\ell_2(\Lambda)} \eqsim \|h\|_{\mathcal{H}}, \tag{2.16}$$

where $\mathbf{h} = (\langle h, \tilde{\psi}_\lambda\rangle)_{\lambda\in\Lambda}$ are the expansion coefficients of h with respect to the dual Riesz basis $\tilde{\Psi} = \{\tilde{\psi}_\lambda\}_{\lambda\in\Lambda}$. Such bases are usually constructed by means of *multiresolution*, see [21]. To describe this, we take two sequences of linear, closed spaces $S_j, \tilde{S}_j \subset \mathcal{H}, j \geq j_0$, with a coarsest level j_0, so that

$$S_j \subset S_{j+1} \;\text{ for all } j \geq j_0, \qquad \overline{\bigcup_{j\geq j_0} S_j} = \mathcal{H}$$

and likewise for the dual spaces $\tilde{S}_j$. The spaces $S_j, \tilde{S}_j$ are usually defined as the span of corresponding stable *generator bases* $\Phi_j = \{\phi_{j,k} : k \in I_j\}$ and $\tilde{\Phi}_j = \{\tilde{\phi}_{j,k} : k \in I_j\}$, $j \geq j_0$. There, the functions $\phi_{j,k}$ (and analogously $\tilde{\phi}_{j,k}$) are usually derived from a single refinable function ϕ by dilation and translation, i.e.,

$$\phi_{j,k}(\cdot) = 2^{jd/2}\phi(2^{j/2}\cdot - k), \;\; j \geq j_0, \; k \in I_j.$$

Due to the nestedness of the spaces, we can establish a refinement relation expressing any coarse (lower) scale basis function by a linear combination of basis functions of finer (higher) scale. This means, there exist matrices $M_{j,\phi}, M_{j,\tilde{\phi}}$ with

$$\Phi_j^T = \Phi_{j+1}^T M_{j,\phi} \quad \text{and} \quad \tilde{\Phi}_j^T = \tilde{\Phi}_{j+1}^T M_{j,\tilde{\phi}}, \tag{2.17}$$

where the k-th columns of $M_{j,\phi}$ and $M_{j,\tilde{\phi}}$ contain the *mask* of $\phi_{j,k}$ and $\tilde{\phi}_{j,k}$ respectively.

We assume biorthogonality in the aforementioned construction, i.e., $\Phi_j, \tilde{\Phi}_j$ fulfill the duality relation $\langle \phi_{j,k}, \tilde{\phi}_{j,k'}\rangle = \delta_{k,k'}$, $k, k' \in I_j$, $j \geq j_0$, which implies $M_{j,\phi}^T M_{j,\tilde{\Phi}} = I$. Then, one identifies complement spaces $W_j, \tilde{W}_j$, such that

$$S_{j+1} = S_j \oplus W_j \quad \text{and} \quad \tilde{S}_{j+1} = \tilde{S}_j \oplus \tilde{W}_j \tag{2.18}$$

with biorthogonality $\tilde{W}_j \perp S_j$ and $W_j \perp \tilde{S}_j$. These $\mathcal{H}$-complement spaces have complement wavelet bases $\Psi_j = \{\psi_{j,k}\}_{k\in J_j}$ and $\tilde{\Psi}_j = \{\tilde{\psi}_{j,k}\}_{k\in J_j}$ respectively, where again every $\psi_{j,k}$ is typically obtained from a mother wavelet ψ by $\psi_{j,k}(\cdot) = 2^{jd/2}\psi(2^j\cdot - k)$ and biorthogonality again yields $\langle \psi_{j,k}, \tilde{\psi}_{j,k'}\rangle = \delta_{k,k'}$. In order to have a consistent notation, we always write ψ and distinguish wavelets and scaling functions by a type-parameter $e \in \{0,1\}^d$, where $e = \mathbf{0}$ is always reserved for the scaling functions. With this we can combine scale, translation and type in one index

$$\lambda = (j, k, e) \tag{2.19}$$

with $|\lambda| := j$, $k(\lambda) := k$ and $e(\lambda) := e$. For convenience, we will identify $(j, k, \mathbf{0})$ with $\lambda^\circ := (j, k) \in I_j$.

Furthermore, by (2.18) there exist matrices $M_{j,\psi}$ and $M_{j,\tilde{\psi}}$ with

$$\Psi_j^T = \Phi_{j+1}^T M_{j,\Psi} \quad \text{and} \quad \tilde{\Psi}_j^T = \tilde{\Phi}_{j+1}^T M_{j,\tilde{\Psi}}, \tag{2.20}$$

which implies that a change between the two bases Φ_{j+1} and $\Phi_j \cup \Psi_j$ of S_{j+1} can be described by the regular matrix $M_j := (M_{j,\phi}, M_{j,\psi})$ mapping $\ell_2(I_j) \oplus \ell_2(J_j)$ to $\ell_2(I_{j+1})$ and, likewise, by a regular matrix $\tilde{M}_j := (M_{j,\tilde{\phi}}, M_{j,\tilde{\psi}})$ for $\tilde{S}_{j+1}$. The inverse matrices are

$$M_j^{-1} =: G_j = \begin{pmatrix} G_{j,\phi} \\ G_{j,\psi} \end{pmatrix} = \begin{pmatrix} M_{j,\tilde{\phi}}^T \\ M_{j,\tilde{\psi}}^T \end{pmatrix} \quad \text{and} \quad \tilde{M}_j^{-1} =: \tilde{G}_j = \begin{pmatrix} G_{j,\tilde{\phi}} \\ G_{j,\tilde{\psi}} \end{pmatrix} = \begin{pmatrix} M_{j,\phi}^T \\ M_{j,\psi}^T \end{pmatrix}. \tag{2.21}$$

With these matrices the representation of S_{j+1} by $\Phi_j \cup \Psi_j$ can be expressed by the *two-scale relation*

$$\Phi_{j+1}^T = \Phi_j^T G_{j,\phi} + \Psi_j^T G_{j,\psi}. \tag{2.22}$$

We will always assume compactly supported wavelets and scaling functions, such that the number of wavelets on every level is bounded, $\# J_j \eqsim 2^{jd}$, where the constants depend on the underlying domain. Furthermore, compactness means that all matrices used above are uniformly sparse, i.e., in every row and column there is only a uniformly bounded number of non-zero elements.

The overall wavelet Riesz basis Ψ for $\mathcal{H}$ is obtained by aggregating the single complement bases Ψ_j, $j \geq j_0$, and the *generators* Φ_{j_0} from the coarsest level, i.e., the overall index set Λ is of the form

$$\Lambda = I_{j_0} \cup \bigcup_{j \geq j_0} J_j =: I_{j_0} \cup J. \tag{2.23}$$

In order to construct a wavelet Gelfand frame over a bounded domain $\Omega \subset \mathbb{R}^d$, we will follow the straightforward approach stated in [12, 22] using domain decomposition arguments. Hence, we assume Ω to be an overlapping union of $1 < M < \infty$ patches $\mathcal{C} := \{\Omega_i\}_{i=1}^M$, i.e.,

$$\Omega = \bigcup_{i=1}^M \Omega_i, \tag{2.24}$$

where each patch Ω_i is the image of the unit cube $\square := (0,1)^d$ under a suitable parametrization $\Omega_i = \kappa_i(\square)$. We assume that the parametrizations κ_i are C^k-diffeomorphisms and that

$$|\det D\kappa_i(x)| \eqsim 1 \quad \text{for } x \in \square.$$

The set of admissible domains Ω is restricted by raising these regularity conditions, the boundary of Ω has to be piecewise smooth enough. However, the particularly attractive case of polyhedral domains is still covered [11].

The construction of a Gelfand frame starts with a 'reference wavelet Riesz basis' $\Psi^\square = \{\psi_\lambda^\square\}_{\lambda \in K^\square} \subset L_2(\square)$ with dual Riesz basis $\tilde{\Psi} := \{\tilde{\psi}_\lambda^\square\}_{\lambda \in K^\square}$, such that appropriate rescaling yields a wavelet Riesz basis for the Sobolev Space $H_0^t(\square), t > 0$. The index set $K^\square$ is assumed to have the form of (2.23), i.e., $K^\square := I_{j_0}^\square \cup \bigcup_{j \geq j_0} J_j^\square$ with a coarsest level j_0. We then lift the system $\Psi^\square$ to patch Ω_i by setting

$$\psi_{i,\lambda} := \frac{\psi_\lambda^\square(\kappa_i^{-1}(\cdot))}{|\det D\kappa_i\left(\kappa_i^{-1}(\cdot)\right)|^{1/2}}. \tag{2.25}$$

The denominator is chosen such that $\|\psi_{i,\lambda}\|_{L_2(\Omega)} = \|\psi_\lambda^\square\|_{L_2(\square)}$. Analogously, we also lift the dual wavelets to Ω_i by

$$\tilde{\psi}_{i,\lambda} := \frac{\tilde{\psi}_\lambda^\square(\kappa_i^{-1}(\cdot))}{|\det D\kappa_i\left(\kappa_i^{-1}(\cdot)\right)|^{1/2}}. \tag{2.26}$$

It is immediate to see that the system $\Psi^{(i)} := \{\psi_{i,\lambda} : \lambda \in K^\square\}$ is a wavelet Riesz basis for $L_2(\Omega_i)$ with dual Riesz basis $\tilde{\Psi}^{(i)} := \{\tilde{\psi}_{i,\lambda} : \lambda \in K^\square\}$ characterizing the corresponding scale of Sobolev spaces over Ω_i. The most simple method to derive a wavelet Gelfand frame over the domain Ω is to aggregate the local bases $\Psi^{(i)}, i = 1, \ldots, M$, by means of the zero extension operator $E_i : \Omega_i \to \Omega$. Using the global index set

$$L := \bigcup_{i=1}^{M} \{i\} \times K^\square, \tag{2.27}$$

we will consider the family

$$\Psi := \{\psi_\lambda : \lambda \in L\}, \quad \lambda := (i,\mu), \quad \psi_{(i,\mu)} := E_i \psi_{i,\mu}, \quad \text{for } (i,\mu) \in L \tag{2.28}$$

with the convention $|\lambda| = |\mu|$ for $\lambda = (i,\mu) \in L$. For the system Ψ the following lemma from [22] holds.

Lemma 2.3 *The aggregated system Ψ from (2.28) is a Hilbert frame for $L_2(\Omega)$.*

Due to the construction of this frame by aggregating local basis systems, it is called an *aggregated Gelfand frame*. Unfortunately, the global canonical dual frame $S^{-1}\Psi$ of the Hilbert frame Ψ is only implicitly given and its properties are not obvious. In particular, it is not immediately clear how to prove the Gelfand frame properties in $(H_0^t(\Omega), L_2(\Omega), H^{-t}(\Omega))$. As an alternative, we therefore propose to work with a special non-canonical dual frame instead. This is possible, when the particular decomposition $\Omega = \bigcup_{i=1}^{M} \Omega_i$ admits the construction of a partition of unity $\{\sigma_i\}_{i=1}^{M}$ subordinate to the patches Ω_i, i.e.,

- supp $\sigma_i \subset \bar{\Omega}_i$,
- $\|\sigma_i u\|_{H^t(\Omega_i)} \lesssim \|u\|_{H^t(\Omega)}$ holds uniformly in $u \in H^t(\Omega)$,
- $\sigma_i u \in H_0^t(\Omega_i)$, for all $u \in H_0^t(\Omega),\ i = 1, \ldots, M.$

Note that the construction of a partition of unity satisfying these properties may not be a triviality, cf. [11]. In the following we assume the existence of a suffiently smooth partition of unity, which immediately allows the specification of a non-canonical global dual frame for Ψ as shown in the following lemma from [11].

Lemma 2.4 *Let Ψ be defined as in (2.28) and assume that $\{\sigma_i\}_{i=1}^{M}$ is a partition of unity subordinate to the patches Ω_i. Then, with $\tilde{\psi}_{i,\lambda}$ being the lifted local duals from (2.26), the system*

$$\tilde{\Psi} := \{\tilde{\psi}_\lambda : \lambda \in L\}, \quad \lambda = (i,\mu), \quad \tilde{\psi}_{(i,\mu)} := E_i(\sigma_i \tilde{\psi}_{i,\mu}), \quad \textit{for } (i,\mu) \in L \tag{2.29}$$

is a non-canonical dual frame for Ψ in $L_2(\Omega)$. More specific, Ψ and $\tilde{\Psi}$ form a wavelet Gelfand frame for $(H_0^t(\Omega), L_2(\Omega), H^{-t}(\Omega))$ with respect to the Gelfand triple of sequence spaces $(\ell_{2,2^t}(L), \ell_2(L), \ell_{2,2^{-t}}(L))$.

2.5 Frame discretization

With an overlapping partition of Ω according to (2.24) and the aggregated Gelfand frame Ψ from (2.28) for the Gelfand triple $(H_0^t(\Omega), L_2(\Omega), H^{-t}(\Omega))$ with non-canonical dual frame $\tilde{\Psi}$ from (2.29), we have all tools at hand to discretize problem (2.8). Due to the assumptions (2.3) and (2.4) on A and the boundedness of the operators in (2.14) and (2.15), the operator

$$\mathbf{A} := (D^*)^{-1} F A F^* D^{-1}$$

is bounded from $\ell_2(L)$ to $\ell_2(L)$ as composition of bounded operators. Moreover, $\mathbf{A}$ is self-adjoint, i.e., $\mathbf{A} = \mathbf{A}^*$, positive semi-definite and boundedly invertible on its range $\mathrm{ran}(\mathbf{A}) = \mathrm{ran}\left((D^*)^{-1}F\right)$, i.e., there exist constants $0 < c_{\mathrm{ran}(\mathbf{A})} \leq C_{\mathrm{ran}(\mathbf{A})} < \infty$, so that

$$c_{\mathrm{ran}(\mathbf{A})}\|\mathbf{v}\|_{\ell_2(L)} \leq \|\mathbf{A}\mathbf{v}\|_{\ell_2(L)} \leq C_{\mathrm{ran}(\mathbf{A})}\|\mathbf{v}\|_{\ell_2(L)} \quad \text{for all } \mathbf{v} \in \mathrm{ran}(\mathbf{A}), \tag{2.30}$$

see [12] for a proof. Furthermore, we have the orthogonal decomposition

$$\ell_2 = \ker(\mathbf{A}) \oplus \mathrm{ran}(\mathbf{A}), \tag{2.31}$$

where $\ker(\mathbf{A})$ denotes the kernel of $\mathbf{A}$. Analogously, we introduce the mapping

$$\mathbf{G}(\cdot) : \ell_2(L) \to \ell_2(L), \quad \mathbf{G}(\mathbf{c}) := (D^*)^{-1}FG\left(F^*D^{-1}\mathbf{c}\right).$$

Using $\mathbf{A}$, $\mathbf{G}(\cdot)$ and the definitions $\mathbf{u} := D\tilde{F}u$ and $\mathbf{f} := (D^*)^{-1}Ff$, equation (2.8) can be rewritten as

$$\mathbf{A}\mathbf{u} + \mathbf{G}(\mathbf{u}) = \mathbf{f}, \tag{2.32}$$

the solution u of (2.1) is then obtained by $u = F^*D^{-1}\mathbf{u}$ (see [12, 24] for the linear case). We want to solve (2.32) now by some iterative scheme using adaptively updated error tolerances. As the main result of the next chapter, we show that this is possible by a well-known damped Richardson iteration.

3 Iterative Solution of Nonlinear Problems

3.1 Uniqueness of the solution

In order to solve (2.32) by an iterative scheme, the first question to answer is, whether (2.32) has a unique solution. Unlike the basis case this is not clear a priori due to the fact that frames are in general redundant systems. Using the given assumptions on $G(\cdot)$ and the ideas from [12], where uniqueness of the solution in the range of $\mathbf{A}$ could be shown for linear systems $\mathbf{A}\mathbf{u} = \mathbf{f}$, the following result can be shown.

Lemma 3.1 *Assume that (2.1) has a unique solution u. Let $\mathbf{u} := D\tilde{F}u$ and $\mathbf{P} : \ell_2(L) \to \mathrm{ran}(\mathbf{A})$ be the orthogonal projection onto $\mathrm{ran}(\mathbf{A})$. Then $\mathbf{P}\mathbf{u}$ is the unique solution of (2.32) in $\mathrm{ran}(\mathbf{A})$ and the solution u of (2.1) can be computed by $u = F^*D^{-1}\mathbf{P}\mathbf{u}$.*

Proof: By definition, $\mathbf{u} = D\tilde{F}u \in \ell_2(L)$ is a solution of (2.32). By definition of $\mathbf{G}(\cdot)$ and using the fact shown in [12] that $\ker(\mathbf{A}) = \ker(F^*D^{-1})$, it follows that $\mathbf{P}\mathbf{u}$ is also a solution of (2.32). Next let $\mathbf{v}, \mathbf{w}$ be in $\ell_2(L)$ and note that due to

$$\begin{aligned}
(\mathbf{v}-\mathbf{w})^T(\mathbf{G}(\mathbf{v}) - \mathbf{G}(\mathbf{w})) &= (\mathbf{v}-\mathbf{w})^T\left((D^*)^{-1}FG(F^*D^{-1}\mathbf{v}) - (D^*)^{-1}FG(F^*D^{-1}\mathbf{w})\right) \\
&= \langle \mathbf{v}-\mathbf{w}, \left((D^*)^{-1}F\left[G(F^*D^{-1}\mathbf{v}) - G(F^*D^{-1}\mathbf{w})\right]\right)\rangle_{\ell_2(L)} \\
&= \langle F^*D^{-1}(\mathbf{v}-\mathbf{w}), G(F^*D^{-1}\mathbf{v}) - G(F^*D^{-1}\mathbf{w})\rangle_{H_0^t(\Omega)\times H^{-t}(\Omega)} \\
&= \langle F^*D^{-1}\mathbf{v} - F^*D^{-1}\mathbf{w}, G(F^*D^{-1}\mathbf{v}) - G(F^*D^{-1}\mathbf{w})\rangle_{H_0^t(\Omega)\times H^{-t}(\Omega)} \\
&=: \langle v-w, G(v) - G(w)\rangle_{H_0^t(\Omega)\times H^{-t}(\Omega)} \\
&\geq 0, \qquad (3.1)
\end{aligned}$$

where (2.5) was used in the last step, the monotonicity of $G(\cdot)$ carries over to $\ell_2(L)$. Now let $\tilde{\mathbf{u}} \in \mathrm{ran}(\mathbf{A})$ be a solution of (2.32) in the range of $\mathbf{A}$. Then it holds $\mathbf{A}\mathbf{P}\mathbf{u} - \mathbf{A}\tilde{\mathbf{u}} + \mathbf{G}(\mathbf{P}\mathbf{u}) - \mathbf{G}(\tilde{\mathbf{u}}) = \mathbf{0}$ and thus

$(\mathbf{Pu}-\tilde{\mathbf{u}})^T\mathbf{A}(\mathbf{Pu}-\tilde{\mathbf{u}})+(\mathbf{Pu}-\tilde{\mathbf{u}})^T(\mathbf{G}(\mathbf{Pu})-\mathbf{G}(\tilde{\mathbf{u}}))=0$. Because of the semi-definiteness of $\mathbf{A}$ and the monotonicity of $\mathbf{G}(\cdot)$ both summands must be equal to zero. From this and (2.30) it follows that

$$0=(\mathbf{Pu}-\tilde{\mathbf{u}})^T\mathbf{A}(\mathbf{Pu}-\tilde{\mathbf{u}})=\langle\mathbf{A}(\mathbf{Pu}-\tilde{\mathbf{u}}),\mathbf{Pu}-\tilde{\mathbf{u}}\rangle_{\ell_2(L)}\gtrsim\|\mathbf{Pu}-\tilde{\mathbf{u}}\|^2_{\ell_2(L)}.$$

Thus $(\mathbf{Pu}-\tilde{\mathbf{u}})=\mathbf{0}$ and $\mathbf{Pu}$ is the single solution of (2.32) from $\operatorname{ran}(\mathbf{A})$. □

3.2 A damped Richardson Iteration

Following the outline given at the end of Section 2, our next aim is to calculate $\mathbf{Pu}$ by an iterative scheme. Thus, we consider the basic iteration

$$\mathbf{u}^{n+1}:=\mathbf{u}^n-\alpha\cdot\mathbf{R}(\mathbf{u}^n),\quad n\in\mathbb{N},\tag{3.2}$$

where $\mathbf{R}(\mathbf{v}):=\mathbf{Av}+\mathbf{G}(\mathbf{v})-\mathbf{f}$ is the discrete *residual*. Note that $\mathbf{R}(\mathbf{u})=0$ and $\mathbf{R}(\mathbf{Pu})=0$ hold and that for a starting vector $\mathbf{u}^0\in\operatorname{ran}(\mathbf{A})$ all exact iterates $\mathbf{u}^n, n\in\mathbb{N}$, belong to $\operatorname{ran}(\mathbf{A})$ due to the definition of $\mathbf{G}(\cdot)$ and $\mathbf{f}$. The main task now is to determine a suitable damping parameter $\alpha>0$, such that (3.2) converges to $\mathbf{Pu}$. Hence, let us first note that

$$\begin{aligned}\mathbf{u}^{n+1}-\mathbf{Pu}&=\mathbf{u}^n-\mathbf{Pu}-\alpha\left(\mathbf{R}(\mathbf{u}^n)-\mathbf{R}(\mathbf{Pu})\right)\\&=\left(\mathbf{I}-\alpha\int_0^1\mathbf{A}+D\mathbf{G}(\mathbf{Pu}+s(\mathbf{u}^n-\mathbf{Pu}))\mathrm{d}s\right)\cdot(\mathbf{u}^n-\mathbf{Pu})\\&=:(\mathbf{I}-\alpha\mathbf{M}(\mathbf{u}^n,\mathbf{Pu}))\cdot(\mathbf{u}^n-\mathbf{Pu}).\end{aligned}\tag{3.3}$$

Due to (2.30) and (3.1), a lower estimate for the spectral radius of $\mathbf{M}(\mathbf{u}^n,\mathbf{Pu})$ follows by

$$\begin{aligned}\frac{(\mathbf{u}^n-\mathbf{Pu})^T\mathbf{M}(\mathbf{u}^n,\mathbf{Pu})(\mathbf{u}^n-\mathbf{Pu})}{\|\mathbf{u}^n-\mathbf{Pu}\|^2_{\ell_2(L)}}&=\frac{(\mathbf{u}^n-\mathbf{Pu})^T\mathbf{A}(\mathbf{u}^n-\mathbf{Pu})}{\|\mathbf{u}^n-\mathbf{Pu}\|^2_{\ell_2(L)}}+\frac{(\mathbf{u}^n-\mathbf{Pu})^T(\mathbf{G}(\mathbf{u}^n)-\mathbf{G}(\mathbf{Pu}))}{\|\mathbf{u}^n-\mathbf{Pu}\|^2_{\ell_2(L)}}\\&\geq\frac{(\mathbf{u}^n-\mathbf{Pu})^T\mathbf{A}(\mathbf{u}^n-\mathbf{Pu})}{\|\mathbf{u}^n-\mathbf{Pu}\|^2_{\ell_2(L)}}\geq c_{\operatorname{ran}(\mathbf{A})}.\end{aligned}\tag{3.4}$$

To show an upper estimate the following remark concerning a discrete version of the stability assumption (2.10) is needed.

Remark 3.1 *The stability condition (2.10) carries over to $\ell_2(L)$, i.e.,*

$$\|\mathbf{G}(\mathbf{v})-\mathbf{G}(\mathbf{w})\|_{\ell_2(L)}\leq\hat{C}_G\left(\max\{\|\mathbf{Pv}\|_{\ell_2(L)},\|\mathbf{Pw}\|_{\ell_2(L)}\}\right)\cdot\|\mathbf{v}-\mathbf{w}\|_{\ell_2(L)}\quad\textit{for all }\mathbf{v},\mathbf{w}\in\ell_2(L),\tag{3.5}$$

where $x\mapsto\hat{C}_G(x):=\|F\|\cdot\|(D^*)^{-1}\|\cdot\|F^*\|\cdot\|D^{-1}\|\cdot C_G(\|F^*\|\cdot\|D^{-1}\|\cdot x)$ is a non-decreasing function, which is essentially the function C_G from (2.10).

Proof: It is immediate to see that

$$\begin{aligned}\|\mathbf{G}(\mathbf{v})-\mathbf{G}(\mathbf{w})\|_{\ell_2(L)}&=\|(D^*)^{-1}FG(F^*D^{-1}\mathbf{v})-(D^*)^{-1}FG(F^*D^{-1}\mathbf{w})\|_{\ell_2(L)}\\&=\|(D^*)^{-1}FG(v)-(D^*)^{-1}FG(w)\|_{\ell_2(L)}\\&\leq\|(D^*)^{-1}\|\cdot\|F\|\cdot\|G(v)-G(w)\|_{H^{-t}(\Omega)}\\&\leq\|(D^*)^{-1}\|\cdot\|F\|\cdot C_G(\max\{\|v\|_{H^t(\Omega)},\|w\|_{H^t(\Omega)}\})\cdot\|v-w\|_{H^t(\Omega)}\\&\leq\hat{C}_G(\max\{\|\mathbf{Pv}\|_{\ell_2(L)},\|\mathbf{Pw}\|_{\ell_2(L)}\})\cdot\|\mathbf{v}-\mathbf{w}\|_{\ell_2(L)},\end{aligned}$$

where we have used $\|v\|_{H^t(\Omega)}=\|F^*D^{-1}\mathbf{v}\|_{H^t(\Omega)}=\|F^*D^{-1}\mathbf{Pv}\|_{H^t(\Omega)}\leq\|F^*\|\cdot\|D^{-1}\|\cdot\|\mathbf{Pv}\|_{\ell_2(L)}$ in the last step. □

Furthermore, note that

$$\begin{aligned}\|\mathbf{u}^{n+1}-\mathbf{P}\mathbf{u}\|_{\ell_2(L)} &= \|\mathbf{u}^n-\alpha\mathbf{R}(\mathbf{u}^n)-\mathbf{P}\mathbf{u}\|_{\ell_2(L)} = \|\mathbf{u}^n-\alpha\mathbf{A}\mathbf{u}^n-\alpha\mathbf{G}(\mathbf{u}^n)+\alpha\mathbf{f}-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\\ &= \|\mathbf{u}^n-\alpha\mathbf{A}\mathbf{u}^n-\alpha\mathbf{G}(\mathbf{u}^n)+\alpha\mathbf{A}\mathbf{P}\mathbf{u}+\alpha\mathbf{G}(\mathbf{P}\mathbf{u})-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\\ &= \|\mathbf{u}^n-\mathbf{P}\mathbf{u}-\alpha\left(\mathbf{A}(\mathbf{u}^n-\mathbf{P}\mathbf{u})+(\mathbf{G}(\mathbf{u}^n)-\mathbf{G}(\mathbf{P}\mathbf{u}))\right)\|_{\ell_2(L)}.\end{aligned}$$

Using now $\|\mathbf{A}(\mathbf{u}^n-\mathbf{P}\mathbf{u})\|_{\ell_2(L)} \leq C_{\mathrm{ran}(\mathbf{A})}\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}$, which follows from (2.30), $\|(\mathbf{G}(\mathbf{u}^n)-\mathbf{G}(\mathbf{P}\mathbf{u}))\|_{\ell_2(L)} \leq \hat{C}_G\left(\max\{\|\mathbf{u}^n\|_{\ell_2(L)},\|\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\}\right)\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}$, which follows from Remark 3.1, and the estimate $\max\{\|\mathbf{u}^n\|_{\ell_2(L)},\|\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\} \leq \|\mathbf{P}\mathbf{u}\|_{\ell_2(L)}+\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}$, we deduce by the monotonicity of $\hat{C}_G(\cdot)$ that

$$\begin{aligned}\|\mathbf{A}(\mathbf{u}^n-\mathbf{P}\mathbf{u})+(\mathbf{G}(\mathbf{u}^n)-\mathbf{G}(\mathbf{P}\mathbf{u}))\|_{\ell_2(L)} &\leq \hat{C}_G\left(\|\mathbf{P}\mathbf{u}\|_{\ell_2(L)}+\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\right)\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\\ &\quad + C_{\mathrm{ran}(\mathbf{A})}\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}. \qquad (3.6)\end{aligned}$$

If we know (or can estimate) the behavior of the stability function $\hat{C}_G(\cdot)$, we can estimate the value $\hat{C}_G\left(\|\mathbf{P}\mathbf{u}\|_{\ell_2(L)}+\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\right)$ by a priori estimates for $\|\mathbf{P}\mathbf{u}\|_{\ell_2(L)}$ and $\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}$. Therefore the following remark is useful.

Remark 3.2 *For any $\mathbf{v}\in\ell_2(L)$ it holds*

$$\|\mathbf{P}\mathbf{u}-\mathbf{P}\mathbf{v}\|_{\ell_2(L)} \leq \frac{1}{c_{ran(\mathbf{A})}}\|\mathbf{R}(\mathbf{v})\|_{\ell_2(L)}. \qquad (3.7)$$

Proof: As a consequence of the Cauchy-Schwarz inequality, (2.30) and (3.1) it holds for any $\mathbf{v}\in\ell_2(L)$

$$\begin{aligned}\|\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}\cdot\|\mathbf{R}(\mathbf{v})\|_{\ell_2(L)} &\geq (\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u})^T(\mathbf{R}(\mathbf{v})) = (\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u})^T(\mathbf{R}(\mathbf{P}\mathbf{v}))\\ &= (\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u})^T(\mathbf{R}(\mathbf{P}\mathbf{v})-\mathbf{R}(\mathbf{P}\mathbf{u}))\\ &= (\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u})^T(\mathbf{A}(\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u}))+(\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u})^T(\mathbf{G}(\mathbf{P}\mathbf{v})-\mathbf{G}(\mathbf{P}\mathbf{u}))\\ &\geq c_{\mathrm{ran}(\mathbf{A})}\|\mathbf{P}\mathbf{v}-\mathbf{P}\mathbf{u}\|_{\ell_2(L)}^2.\end{aligned}$$

Inequation (3.7) is then an immediate consequence. □

Choosing any starting vector $\mathbf{u}^0\in\mathrm{ran}(\mathbf{A})$ for the iteration, we can deduce

$$\begin{aligned}\|\mathbf{P}\mathbf{u}-\mathbf{u}^0\|_{\ell_2(L)} &\leq \frac{1}{c_{\mathrm{ran}(\mathbf{A})}}\|\mathbf{R}(\mathbf{u}^0)\|_{\ell_2(L)} \leq \frac{1}{c_{\mathrm{ran}(\mathbf{A})}}\left(\|\mathbf{G}(\mathbf{u}^0)\|_{\ell_2(L)}+\|\mathbf{A}\mathbf{u}^0\|_{\ell_2(L)}+\|\mathbf{f}\|_{\ell_2(L)}\right)\\ &\leq \frac{1}{c_{\mathrm{ran}(\mathbf{A})}}\left(\|\mathbf{G}(\mathbf{u}^0)\|_{\ell_2(L)}+\|\mathbf{G}(\mathbf{0})\|_{\ell_2(L)}+\|\mathbf{A}\mathbf{u}^0\|_{\ell_2(L)}+\|\mathbf{f}\|_{\ell_2(L)}\right) =: \varepsilon_0, \qquad (3.8)\end{aligned}$$

which allows to formulate the following lemma concerning the choice of a suitable damping parameter α.

Lemma 3.2 *Assume that α from (3.2) satisfies*

$$0<\alpha<\frac{2}{C_{ran(\mathbf{A})}+\hat{C}_G(3\varepsilon_0)}. \qquad (3.9)$$

Then, it follows

$$\|\mathbf{u}^{n+1}-\mathbf{P}\mathbf{u}\|_{\ell_2(L)} \leq \varrho_n\cdot\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|_{\ell_2(L)} \quad \textit{for all } n\geq 0 \qquad (3.10)$$

for the iterates from (3.2), where

$$\begin{aligned}\varrho_{n+1} &\leq \varrho_n := \max\left\{|1-\alpha c_{ran(\mathbf{A})}|,|1-\alpha(C_{ran(\mathbf{A})}+\hat{C}_G(\|\mathbf{P}\mathbf{u}\|+\|\mathbf{u}^n-\mathbf{P}\mathbf{u}\|))|\right\}\\ &\leq \varrho := \max\left\{|1-\alpha c_{ran(\mathbf{A})}|,|1-\alpha(C_{ran(\mathbf{A})}+\hat{C}_G(3\varepsilon_0))|\right\} \qquad (3.11)\\ &< 1\end{aligned}$$

for all $n\geq 0$, which means that we have a linearly convergent iterative scheme.

Proof: As a consequence of (3.7) and (3.8) we have

$$\|\mathbf{Pu}\|_{\ell_2(L)} \leq \frac{1}{c_{\mathrm{ran}(\mathbf{A})}} \|\mathbf{R}(\mathbf{0})\|_{\ell_2(L)} \leq \frac{1}{c_{\mathrm{ran}(\mathbf{A})}} \left(\|\mathbf{G}(\mathbf{0})\|_{\ell_2(L)} + \|\mathbf{f}\|_{\ell_2(L)} \right) \leq \varepsilon_0.$$

Hence, it follows with (3.8) and the monotonicity of $\hat{C}_G(\cdot)$ that

$$C_{\mathrm{ran}(\mathbf{A})} + \hat{C}_G \left(\|\mathbf{Pu}\| + \|\mathbf{u}^0 - \mathbf{Pu}\| \right) \leq C_{\mathrm{ran}(\mathbf{A})} + \hat{C}_G(2\varepsilon_0) \leq C_{\mathrm{ran}(\mathbf{A})} + \hat{C}_G(3\varepsilon_0).$$

Thus, $\varrho_0 \leq \varrho < 1$ and combining (3.3), (3.4) and (3.6) yields $\|\mathbf{u}^1 - \mathbf{Pu}\|_{\ell_2(L)} \leq \varrho_0 \|\mathbf{u}^0 - \mathbf{Pu}\|_{\ell_2(L)} < \|\mathbf{u}^0 - \mathbf{Pu}\|_{\ell_2(L)}$. Using now (3.6) and the monotonity of $\hat{C}_G(\cdot)$ again, which says that

$$C_{\mathrm{ran}(\mathbf{A})} + \hat{C}_G \left(\|\mathbf{Pu}\|_{\ell_2(L)} + \|\mathbf{u}^1 - \mathbf{Pu}\|_{\ell_2(L)} \right) \leq C_{\mathrm{ran}(\mathbf{A})} + \hat{C}_G \left(\|\mathbf{Pu}\|_{\ell_2(L)} + \|\mathbf{u}^0 - \mathbf{Pu}\|_{\ell_2(L)} \right),$$

we immediately get, using the same arguments, $\|\mathbf{u}^2 - \mathbf{Pu}\|_{\ell_2(L)} \leq \varrho_1 \|\mathbf{u}^1 - \mathbf{Pu}\|_{\ell_2(L)}$ with $\varrho_1 \leq \varrho_0$. By induction follows then

$$\|\mathbf{u}^{n+1} - \mathbf{Pu}\|_{\ell_2(L)} \leq \varrho_n \|\mathbf{u}^n - \mathbf{Pu}\|_{\ell_2(L)} \quad \text{for all } n \geq 0.$$

□ We are now able to formulate an adaptive iteration scheme $\mathbf{SOLVE}[\varepsilon, \alpha, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}]$, which allows the approximation of $\mathbf{Pu}$ up to any accuracy $\varepsilon > 0$. We assume (for the moment) the existence of two methods $\mathbf{RES}[\eta, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}, \mathbf{v}]$ and $\mathbf{COARSE}[\eta, \mathbf{v}] \to \mathbf{w}_\eta$ with the following properties:

RES $[\eta, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}, \mathbf{v}] \to \mathbf{w}_\eta$:
Determines for any tolerance $\eta > 0$ and every $\mathbf{v} \in \ell_2(L)$ with finite support a vector $\mathbf{w}_\eta \in \ell_2(L)$ with finite support so that $\|\mathbf{R}(\mathbf{v}) - \mathbf{w}_\eta\|_{\ell_2(L)} \leq \eta$.

COARSE $[\eta, \mathbf{v}] \to \mathbf{w}_\eta$:
Determines for any tolerance $\eta > 0$ and every $\mathbf{v} \in \ell_2(L)$ with finite support a vector $\mathbf{w}_\eta \in \ell_2(L)$ with minimal support so that $\|\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(L)} \leq \eta$.

Then, based on [5, 12, 24], the following Algorithm 1 can be formulated:

Algorithm 1
$\mathbf{SOLVE}[\varepsilon, \alpha, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}] \to \mathbf{u}_\varepsilon$:
`Let` $\varepsilon_0 > 0$ `be defined as in` (3.8), `set` $\mathbf{u}^0 := \mathbf{0}$ `and` $j := 0$.
`Furthermore, let` $C^* \geq 1$ `be an absolute constant, whose relevance will be explained later, and` $k^* \in \mathbb{N}$ `the smallest integer satisfying`

$$3\varrho^{k^*} < \frac{1}{2C^* + 1},$$

`with` ϱ `from` (3.11).

`While` $\varepsilon_j > \varepsilon$ `do`
  `Set` $j := j + 1$ `and` $\varepsilon_j := 3\varrho^{k^*}(2C^* + 1)\varepsilon_{j-1}$.
  `Set` $\eta_j := \frac{\varepsilon_j}{3\alpha k^* \cdot (2C^*+1)}$, $k := 0$ `and` $\mathbf{v}^{j,0} := \mathbf{u}^{j-1}$.
  `For` $k = 0, \ldots, k^* - 1$ `do`
    `Compute` $\mathbf{r}^k := \mathbf{RES}[\eta_j, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}, \mathbf{v}^{j,k}]$ `and` $\mathbf{v}^{j,k+1} := \mathbf{v}^{j,k} - \alpha \cdot \mathbf{r}^k$.
  `od`
  `Set` $\mathbf{u}^j := \mathbf{COARSE}\left[\frac{2C^*}{2C^*+1}\varepsilon_j, \mathbf{v}^{j,k^*}\right]$.
`od`
`return` $\mathbf{u}_\varepsilon := \mathbf{u}^j$.

This algorithm refines the one given in [12, 24]. Recall that in these algorithms a fixed parameter $\theta < \frac{1}{3}$ with $3\varrho^{k^*} < \theta$ is crucial for the complexity estimates of the algorithms. By $C^* \geq 1$ and $\frac{1}{2C^*+1} \leq \frac{1}{3}$ this assumption is included implicitly in the above algorithm. The main result concerning Algorithm 1 reads as follows.

Lemma 3.3 *For the iterates* $\mathbf{u}^j, j = 0, 1, \ldots,$ *from Algorithm 1 it holds*

$$\|\mathbf{P}(\mathbf{u} - \mathbf{u}^j)\|_{\ell_2(L)} \leq \varepsilon_j. \tag{3.12}$$

Particularly, it holds $\|\mathbf{P}(\mathbf{u} - \mathbf{u}_\varepsilon)\|_{\ell_2(L)} \leq \varepsilon$ *for the last iterate. Furthermore, for* $j \geq 1$*, it holds*

$$\|\mathbf{P}\mathbf{u} + (\mathrm{id} - \mathbf{P})\mathbf{u}^{j-1} - \mathbf{v}^{j,k^*}\|_{\ell_2(L)} \leq \frac{2}{3(2C^*+1)}\varepsilon_j. \tag{3.13}$$

Proof: For $j = 0$ assertion (3.12) is true by definition of ε_0, see (3.8). Now, let (3.12) be true for $j \geq 0$. For $j+1$ we use

$$\begin{aligned} \|\mathbf{P}(\mathbf{u} - \mathbf{u}^{j+1})\|^2_{\ell_2(L)} &\leq \|\mathbf{P}(\mathbf{u} - \mathbf{u}^{j+1})\|^2_{\ell_2(L)} + \|(\mathrm{id} - \mathbf{P})(\mathbf{u}^j - \mathbf{u}^{j+1})\|^2_{\ell_2(L)} \\ &= \|\mathbf{P}(\mathbf{u} - \mathbf{u}^{j+1}) + (\mathrm{id} - \mathbf{P})(\mathbf{u}^j - \mathbf{u}^{j+1})\|^2_{\ell_2(L)} \\ &= \|\mathbf{P}\mathbf{u} + (\mathrm{id} - \mathbf{P})\mathbf{u}^j - \mathbf{u}^{j+1}\|^2_{\ell_2(L)}. \end{aligned}$$

The last term is bounded by

$$\begin{aligned} \|\mathbf{P}\mathbf{u} + (\mathrm{id} - \mathbf{P})\mathbf{u}^j - \mathbf{u}^{j+1}\|_{\ell_2(L)} &\leq \|\mathbf{P}\mathbf{u} + (\mathrm{id} - \mathbf{P})\mathbf{u}^j - \mathbf{v}^{j+1,k^*}\|_{\ell_2(L)} + \|\mathbf{v}^{j+1,k^*} - \mathbf{u}^{j+1}\|_{\ell_2(L)} \\ &\leq \|\mathbf{P}\mathbf{u} + (\mathrm{id} - \mathbf{P})\mathbf{u}^j - \mathbf{v}^{j+1,k^*}\|_{\ell_2(L)} + \frac{2C^*}{2C^*+1}\varepsilon_{j+1}. \end{aligned}$$

Thus, since $\frac{2}{3(2C^*+1)}\varepsilon_{j+1} + \frac{2C^*}{2C^*+1}\varepsilon_{j+1} \leq \varepsilon_{j+1}$, (3.12) is proven, once (3.13) is shown. For that purpose, let $\mathbf{w}^{j+1,k}$, $0 \leq k \leq k^*$, be the exact Richardson iterates from (3.2) starting with $\mathbf{w}^{j+1,0} := \mathbf{u}^j$ and let $\mathbf{z}^{j+1,k}$, $0 \leq k \leq k^*$, be the exact Richardson iterates starting with $\mathbf{z}^{j+1,0} := \mathbf{P}\mathbf{u}^j$. Because $\mathbf{R}(\mathbf{u}^j) = \mathbf{R}(\mathbf{P}\mathbf{u}^j)$, it holds $(\mathbf{P}\mathbf{u} - \mathbf{z}^{j+1,k^*}) - (\mathbf{P}\mathbf{u} - \mathbf{w}^{j+1,k^*}) = (\mathrm{id} - \mathbf{P})\mathbf{u}^j$. This yields

$$\begin{aligned} \|\mathbf{P}\mathbf{u} + (\mathrm{id} - \mathbf{P})\mathbf{u}^j - \mathbf{v}^{j+1,k^*}\|_{\ell_2(L)} &= \|\mathbf{P}\mathbf{u} + (\mathbf{P}\mathbf{u} - \mathbf{z}^{j+1,k^*}) - (\mathbf{P}\mathbf{u} - \mathbf{w}^{j+1,k^*}) - \mathbf{v}^{j+1,k^*}\|_{\ell_2(L)} \\ &\leq \|\mathbf{P}\mathbf{u} - \mathbf{v}^{j+1,k^*} - (\mathbf{P}\mathbf{u} - \mathbf{w}^{j+1,k^*})\|_{\ell_2(L)} + \|\mathbf{P}\mathbf{u} - \mathbf{z}^{j+1,k^*}\|_{\ell_2(L)}. \end{aligned}$$

Since $\mathbf{v}^{j+1,0} = \mathbf{w}^{j+1,0} = \mathbf{u}^j$ and by definition of $\mathbf{w}^{j+1,k}$, $0 \leq k \leq k^*$, we have $\|\mathbf{P}\mathbf{u} - \mathbf{v}^{j+1,k^*} - (\mathbf{P}\mathbf{u} - \mathbf{w}^{j+1,k^*})\|_{\ell_2(L)} \leq \alpha k^* \frac{\varepsilon_{j+1}}{3\alpha k^* \cdot (2C^*+1)} = \frac{\varepsilon_{j+1}}{3(2C^*+1)}$. Secondly, by induction hypothesis, $\|\mathbf{P}(\mathbf{u} - \mathbf{u}^j)\|_{\ell_2(L)} \leq \varepsilon_j$, by definition of $\mathbf{z}^{j+1,k}$, $0 \leq k \leq k^*$, and (3.10), it follows $\|\mathbf{P}\mathbf{u} - \mathbf{z}^{j+1,k^*}\|_{\ell_2(L)} \leq \varrho^{k^*}\|\mathbf{P}(\mathbf{u} - \mathbf{u}^j)\|_{\ell_2(L)} \leq \frac{\varepsilon_{j+1}}{3(2C^*+1)}$, which completes the proof. □

Regardless of the convergence of Algorithm 1, there is a problem in its realization using frames. Each application of **RES** and **COARSE** causes an error, which might have a component in the kernel of $\mathbf{A}$, see (2.31). These errors are not reduced in the following iterations in general and might cause, although they do not affect the overall solution $u = F^* D^{-1}\mathbf{u}$, major computational effort, which inhibits an effective numerical implementation of Algorithm 1.

As a first resort, in [24] a condition on $\mathbf{P}$ (and thus on the frame) was assumed that allows the proof of optimal complexity. But, since the properties of $\mathbf{P}$ remain widely unclear, we follow the second approach given in [24] and introduce a second algorithm, which contains the explicit application of a projector $\mathbf{Q}$ to reduce error components in $\ker(\mathbf{A})$.

Therefore, let $\mathbf{Q} : \ell_2(L) \to \ell_2(L)$ be some bounded projector with $\ker(\mathbf{Q}) = \ker(\mathbf{A})$ and $\ell_2(L) = \mathrm{ran}(\mathbf{Q}) \oplus \ker(\mathbf{Q})$. Note that, since $\ker(\mathbf{P}) = \ker(\mathbf{A})$, it holds $\mathbf{Q} = \mathbf{QP}$, and, since $\mathbf{Q}$ is a projector, $\mathbf{QPQ} = \mathbf{Q}$.

The application of Remark 3.2 and Lemma 3.2 in connection with $\mathbf{Q}$ requires some modifications, which we summarize in the following remark.

Remark 3.3 *For any $\mathbf{v} \in \ell_2(L)$ and any projector $\mathbf{Q}$ with the above properties it holds*

$$\|\mathbf{PQu} - \mathbf{Pv}\|_{\ell_2(L)} \leq \frac{1}{c_{ran(\mathbf{A})}} \|\mathbf{R}(\mathbf{v})\|_{\ell_2(L)}. \tag{3.14}$$

Furthermore, if the damping parameter α satisfies

$$0 < \alpha < \frac{2}{C_{ran(\mathbf{A})} + \hat{C}_G(3\|\mathbf{Q}\|\varepsilon_0)}, \tag{3.15}$$

it follows for $\mathbf{u}^0 \in ran(\mathbf{A})$ that

$$\|\mathbf{u}^{n+1} - \mathbf{PQu}\|_{\ell_2(L)} \leq \tilde{\varrho}_n \cdot \|\mathbf{u}^n - \mathbf{PQu}\|_{\ell_2(L)},\ n \geq 0, \tag{3.16}$$

for the iterates from (3.2), where

$$\begin{aligned} \tilde{\varrho}_{n+1} &\leq \tilde{\varrho}_n := \max\left\{|1 - \alpha c_{ran(\mathbf{A})}|, |1 - \alpha(C_{ran(\mathbf{A})} + \hat{C}_G(\|\mathbf{PQu}\| + \|\mathbf{u}^n - \mathbf{PQu}\|))|\right\} \\ &\leq \tilde{\varrho} := \max\left\{|1 - \alpha c_{ran(\mathbf{A})}|, |1 - \alpha(C_{ran(\mathbf{A})} + \hat{C}_G(3\|\mathbf{Q}\|\varepsilon_0))|\right\} \\ &< 1 \end{aligned} \tag{3.17}$$

for all $n \geq 0$.

Proof: Assertion (3.14) is proven analogously to (3.7), using $\mathbf{R}(\mathbf{PQu}) = \mathbf{0}$. This is in turn true since $\mathbf{R}(\mathbf{PQu}) = \mathbf{R}(\mathbf{PQu} - \mathbf{Pu} + \mathbf{Pu})$ and $\mathbf{P}(\mathbf{Q} - \mathrm{id})\mathbf{u} = \mathbf{0}$, because $\ker(\mathbf{Q}) = \ker(\mathbf{A}) = \ker(\mathbf{P})$.

Assertion (3.16) is proven analogously to (3.10) using $\|\mathbf{PQu}\|_{\ell_2(L)} + \|\mathbf{u}^0 - \mathbf{PQu}\|_{\ell_2(L)} \leq \|\mathbf{Q}\|\varepsilon_0 + \varepsilon_0 \leq 2\|\mathbf{Q}\|\varepsilon_0$, where $\mathbf{Q} = \mathbf{QP}$ and (3.14) were used in the first step. □

For the application of $\mathbf{Q}$ in the iteration we assume that a method **APPLY** is available with

APPLY $[\eta, \mathbf{B}, \mathbf{v}] \to \mathbf{w}_\eta$:

Determines for any tolerance $\eta > 0$, a linear $\mathbf{B} : \ell_2(L) \to \ell_2(L)$ and every $\mathbf{v} \in \ell_2(L)$ with finite support a vector $\mathbf{w}_\eta \in \ell_2(L)$ with finite support so that $\|\mathbf{Bv} - \mathbf{w}_\eta\|_{\ell_2(L)} \leq \eta$.

Then, for a damping parameter α satisfying (3.15) consider the following

Algorithm 2

modSOLVE$[\varepsilon, \alpha, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}] \to \mathbf{u}_\varepsilon$:

`Let` $\varepsilon_0 > 0$ `be defined as in` (3.8), `set` $\tilde{\varepsilon}_0 := \|\mathbf{Q}\|\varepsilon_0$, $\mathbf{u}^0 := \mathbf{0}$ and $j := 0$.

`Furthermore, let` $C^* \geq 1$ `be the same absolute constant as in Algorithm 1 and let now` $k^* \in \mathbb{N}$ `be the smallest integer satisfying`

$$3\|\mathbf{Q}\|\tilde{\varrho}^{k^*} < \frac{1}{2C^*+1},$$

`with` $\tilde{\varrho}$ `from` (3.17).

```
While ε̃_j > ε do
    Set j := j+1 and ε̃_j := 3‖Q‖ϱ^{k*}(2C*+1)ε̃_{j-1}.
    Set η_j := ϱ^{k*} ε̃_{j-1} / (α k*), k := 0 and v^{j,0} := u^{j-1}.
    For k = 0,...,k*-1 do
        Compute r^k := RES[η_j, A, G(·), f, v^{j,k}] and v^{j,k+1} := v^{j,k} - α·r^k.
    od
    Set z^j := APPLY[ ε̃_j / (3(2C*+1)), Q, v^{j,k*} ].
    Set u^j := COARSE[ (2C*/(2C*+1)) ε̃_j, z^j ].
od
return u_ε := u^j.
```

The analogon of Lemma 3.3 reads as follows.

Lemma 3.4 *For* $\mathbf{z}^j$ *and* $\mathbf{u}^j$ *from Algorithm 2 it holds*

$$\|\mathbf{Q}\mathbf{u} - \mathbf{u}^j\|_{\ell_2(L)} \leq \tilde{\varepsilon}_j, \; j \geq 0, \tag{3.18}$$

and, thus, particularly $\|\mathbf{Q}\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell_2(L)} \leq \varepsilon$. *Furthermore, it holds*

$$\|\mathbf{Q}\mathbf{u} - \mathbf{z}^j\|_{\ell_2(L)} \leq \frac{1}{2C^*+1}\tilde{\varepsilon}_j, \; j \geq 1. \tag{3.19}$$

Proof: For $j = 0$ we have $\|\mathbf{Q}\mathbf{u} - \mathbf{u}^0\|_{\ell_2(L)} \leq \|\mathbf{Q}\| \cdot \|\mathbf{P}\mathbf{u}\|_{\ell_2(L)} \leq \|\mathbf{Q}\|\varepsilon_0 = \tilde{\varepsilon}_0$ by definition of $\tilde{\varepsilon}_0$. Let now $\mathbf{w}^{1,k}$, $1 \leq k \leq k^*$, be the exact iterates $\mathbf{w}^{1,k} = \mathbf{w}^{1,k-1} - \alpha\mathbf{R}(\mathbf{w}^{1,k-1})$ with $\mathbf{w}^{1,0} := \mathbf{u}^0$. It holds

$$\mathbf{Q}\mathbf{u} - \mathbf{Q}\mathbf{w}^{1,k} = \mathbf{Q}\mathbf{P}\mathbf{Q}\mathbf{u} - \mathbf{Q}\mathbf{P}\mathbf{w}^{1,k} = \mathbf{Q}(\mathbf{P}\mathbf{Q}\mathbf{u} - \mathbf{P}\mathbf{w}^{1,k}) = \mathbf{Q}\left(\mathbf{P}\mathbf{Q}\mathbf{u} - \mathbf{P}\mathbf{u}^0 + \alpha\sum_{i=0}^{k-1}\mathbf{R}(\mathbf{w}^{1,i})\right).$$

Because $\|\mathbf{P}\mathbf{Q}\mathbf{u}\|_{\ell_2(L)} + \|\mathbf{P}\mathbf{Q}\mathbf{u} - \mathbf{P}\mathbf{u}^0\|_{\ell_2(L)} \leq \|\mathbf{P}\mathbf{Q}\mathbf{u}\|_{\ell_2(L)} + \|\mathbf{Q}\mathbf{u} - \mathbf{u}^0\|_{\ell_2(L)} \leq 2\tilde{\varepsilon}_0$, (3.15) and (3.16) yield

$$\|\mathbf{Q}\mathbf{u} - \mathbf{Q}\mathbf{w}^{1,k}\|_{\ell_2(L)} \leq \|\mathbf{Q}\|\tilde{\varrho}^k\|\mathbf{P}\mathbf{Q}\mathbf{u} - \mathbf{P}\mathbf{u}^0\|_{\ell_2(L)} \leq \|\mathbf{Q}\|\tilde{\varrho}^k\|\mathbf{Q}\mathbf{u} - \mathbf{u}^0\|_{\ell_2(L)}. \tag{3.20}$$

Then, by definition of $\mathbf{v}^{1,k}$, $1 \leq k \leq k^*$, we have $\mathbf{v}^{1,k} = \mathbf{v}^{1,k-1} - \alpha\mathbf{R}(\mathbf{v}^{1,k-1}) + \delta_{1,k}$ with $\|\delta_{1,k}\|_{\ell_2(L)} \leq \frac{\tilde{\varrho}^{k^*}\tilde{\varepsilon}_0}{k^*}$. Therefore, we have $\mathbf{Q}\mathbf{u} - \mathbf{v}^{1,k^*} = \mathbf{Q}\mathbf{u} - \mathbf{v}^{1,0} + \alpha\sum_{i=0}^{k^*-1}\mathbf{R}(\mathbf{v}^{1,i}) + \sum_{i=1}^{k^*}\delta_{1,i}$ and, using (3.20), we get

$$\|\mathbf{Q}\mathbf{u} - \mathbf{Q}\mathbf{v}^{1,k^*}\|_{\ell_2(L)} \leq \|\mathbf{Q}\|\left(\tilde{\varrho}^{k^*}\|\mathbf{Q}\mathbf{u} - \mathbf{u}^0\|_{\ell_2(L)} + \sum_{i=1}^{k^*}\|\delta_{1,i}\|_{\ell_2(L)}\right) \leq 2\|\mathbf{Q}\|\tilde{\varrho}^{k^*}\tilde{\varepsilon}_0.$$

By definition of $\mathbf{z}^1$ we can now conclude $\|\mathbf{Q}\mathbf{u} - \mathbf{z}^1\|_{\ell_2(L)} \leq \|\mathbf{Q}\mathbf{u} - \mathbf{Q}\mathbf{v}^{1,k^*}\|_{\ell_2(L)} + \|\mathbf{Q}\mathbf{v}^{1,k^*} - \mathbf{z}^1\|_{\ell_2(L)} \leq 2\|\mathbf{Q}\|\tilde{\varrho}^{k^*}\tilde{\varepsilon}_0 + \frac{\tilde{\varepsilon}_1}{3(2C^*+1)} = \frac{2\tilde{\varepsilon}_1}{3(2C^*+1)} + \frac{\tilde{\varepsilon}_1}{3(2C^*+1)} = \frac{\tilde{\varepsilon}_1}{2C^*+1}$ and $\|\mathbf{Q}\mathbf{u} - \mathbf{u}^1\|_{\ell_2(L)} \leq \|\mathbf{Q}\mathbf{u} - \mathbf{z}^1\|_{\ell_2(L)} + \|\mathbf{z}^1 - \mathbf{u}^1\|_{\ell_2(L)} \leq \frac{\tilde{\varepsilon}_1}{2C^*+1} + \frac{2C^*\tilde{\varepsilon}_1}{2C^*+1} = \tilde{\varepsilon}_1$.

Using now $\|\mathbf{Q}\mathbf{u} - \mathbf{u}^1\|_{\ell_2(L)} \leq \tilde{\varepsilon}_1 < \tilde{\varepsilon}_0$, the fact that $\mathbf{P}\mathbf{w}^{j,k} \in \mathrm{ran}(\mathbf{A})$ for all $j \geq 1$ and $0 \leq k \leq k^*$, where $\mathbf{w}^{j,k}$ is defined analogously to $\mathbf{w}^{1,k}$, and the monotonicity of $\hat{C}_G(\cdot)$ allows the iteration of the above arguments in j and thus yields the assertion. □

Up to now, we did not care about the numerical realization of the abstract **APPLY**, **COARSE** and **RES** routines. We will turn to this in the next chapters before we come back to Algorithm 2 in the last chapter in order to analyze its complexity.

4 Adaptive Numerical Frame Schemes for Nonlinear Operator Equations

For the practical realization of the iteration schemes discussed in the last section numerically implementable versions of **APPLY**, **RES** and **COARSE** have to be established. In detail we need four building blocks:

- **RHS**$[\varepsilon, \mathbf{f}] \rightarrow \mathbf{f}_\varepsilon$
  determines for $\mathbf{f} \in \ell_2(L)$ a finitely supported $\mathbf{f}_\varepsilon \in \ell_2(L)$ so that $\|\mathbf{f} - \mathbf{f}_\varepsilon\|_{\ell_2(L)} \leq \varepsilon$.

- **APPLY**$[\varepsilon, \mathbf{B}, \mathbf{v}] \rightarrow \mathbf{w}_\varepsilon$
  determines for linear $\mathbf{B} : \ell_2(L) \rightarrow \ell_2(L)$ and any finitely supported $\mathbf{v} \in \ell_2(L)$ a finitely supported $\mathbf{w}_\varepsilon \in \ell_2(L)$ so that $\|\mathbf{B}\mathbf{v} - \mathbf{w}_\varepsilon\|_{\ell_2(L)} \leq \varepsilon$.

- **COARSE**$[\varepsilon, \mathbf{v}] \rightarrow \mathbf{w}_\varepsilon$
  determines for a finitely supported $\mathbf{v} \in \ell_2(L)$ a finitely supported $\mathbf{w}_\varepsilon \in \ell_2(L)$ with at most N nonzero coefficients so that $\|\mathbf{v} - \mathbf{w}_\varepsilon\|_{\ell_2(L)} \leq \varepsilon$. Here, $N \lesssim N_{\min}$ is required, where $N_{\min}$ is the minimal number of entries for which this is valid.

- **EVAL**$[\varepsilon, \mathbf{G}(\cdot), \mathbf{v}] \rightarrow \mathbf{w}_\varepsilon$
  determines for any finitely supported $\mathbf{v} \in \ell_2(L)$ a finitely supported $\mathbf{w}_\varepsilon$ so that $\|\mathbf{G}(\mathbf{v}) - \mathbf{w}_\varepsilon\|_{\ell_2(L)} \leq \varepsilon$.

For the first three methods it is known how to implement them numerically in order to achieve asymptotically optimal procedures for linear problems in comparison with the best N-term approximation of the solution u as well for bases [5] as for frames [12, 24].

But, as noted before, the evaluation of nonlinear terms by wavelet bases requires the usage of a structural additive, a so called *tree structure*, on the index set of used wavelets [1, 7, 8]. Thus, the natural benchmark is no longer the best (unrestricted) N-term approximation, but a *best N-Term tree approximation* respecting the constraints imposed by tree structures.

As a preparation we show first in this chapter how the concept of tree approximation can be carried over to the case of aggregated Gelfand frames resulting in the concept of *aggregated trees*. We then use the ideas from [7] to prove the existence of an asymptotically optimal **COARSE** routine in this context and give a short hint that the **APPLY** and **RHS** routines developed and enhanced in [1, 5, 6] show - under slightly stronger assumptions - the same complexity. Finally, we turn to the main point of this chapter, the first part, i.e., the prediction step, of the development of an adaptive evaluation scheme **EVAL**, which shows asymptotically optimal complexity concerning the aggregated tree approximation.

4.1 Tree Structured Index Sets

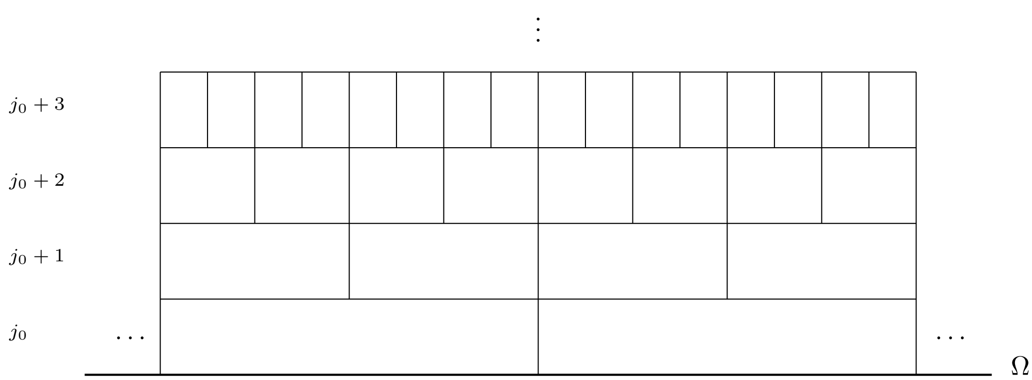
Let us first explain what we mean by a tree structure associated to a set of wavelet indices. A straightforward illustration can be given by means of a one-dimensional wavelet basis for the real line. In this case, every *parent* index $\mu = (j, k)$ has two *children* $\mu_1 = (j+1, 2k)$ and $\mu_2 = (j+1, 2k+1)$ and an index set Σ is called a *tree* if for every $\sigma = (j, k) \in \Sigma$ also its parent index $\eta = (j-1, \lfloor k \rfloor)$ belongs to Σ.

A similar tree structure can be associated to all available constructions of wavelet bases $\Psi = \{\psi_\lambda\}_{\lambda\in\Lambda}$ on a multidimensional domain $\Omega \subset \mathbb{R}^d$, see [7]. Recalling that we identify $\lambda = (j, k, \mathbf{0})$ with $\lambda^\circ = (j, k) \in I_j$, we can associate to every λ° a *reference cube* $\square_{(j,k)} = \square_{\lambda^\circ} \subset \mathbb{R}^d$ with the following properties

- $\square_{\lambda^\circ} \subset \operatorname{supp} \psi_\lambda$,
- $\bar{\Omega} = \bigcup_{\lambda^\circ \in I_j} \square_{\lambda^\circ}$ for all $j \geq j_0$,
- Each $\square_{\lambda^\circ}$ is an essentially disjoint union of reference cubes $\square_{\eta_1^\circ}, \ldots, \square_{\eta_N^\circ}$ of the next level, $|\eta_s^\circ| = |\lambda^\circ| + 1$ for all $s = 1, \ldots, N$. We allow the parameter $N \geq 2$ to vary from one reference cube to another, but we assume N to be bounded uniformly in j, see [7]. Nevertheless, the intuitive case $N = 2^d$ is the most common.

Note that by this definition all ψ_λ with the same level and spatial location are associated to the same reference cube $\square_{\lambda^\circ}$. Thus, in the following we will often identify $\lambda \in J_j$ with the reference cube $\square_{\lambda^\circ}$ and with Σ° we will denote the index set of reference cubes associated with an index set $\Sigma \subset \Lambda$.

Due to the compact support of the used wavelets and the boundedness of Ω, the number of reference cubes on each level is finite and for $d = 1$ they can be intuitively displayed in the following way.



The reference cubes $\square_{\lambda^\circ}$ induce a natural antecessor-successor-relation by means of set inclusion over Ω.

Definition 4.1 *If* $\square_{\lambda^\circ} \subsetneq \square_{\mu^\circ}, |\mu^\circ| \geq j_0$, *then* λ° *is called a* descendant *or* successor *of* μ°,

$$\lambda^\circ \succ \mu^\circ. \tag{4.1}$$

μ° *is called an* antecessor *of* λ° *obversely. Moreover,* $\lambda^\circ \succeq \mu^\circ$ *means that* λ° *is a descendant or equal to* μ°. *We define the set of all descendants of* μ° *by*

$$\Gamma_{\mu^\circ} := \{\lambda^\circ : \lambda^\circ \succeq \mu^\circ\}. \tag{4.2}$$

If $\lambda^\circ \succ \mu^\circ$ *and* $|\lambda^\circ| = |\mu^\circ| + 1$, *then* λ° *is called a* child *of* μ° *and* μ° *is named the* parent *of* λ°, *which will be expressed by the notations* $\lambda^\circ = \mathcal{C}(\mu^\circ)$ *and* $\mu^\circ = \mathcal{P}(\lambda^\circ)$.

An index set $\mathcal{T}^\circ \subset I := \cup_{j \geq j_0} I_j$ is called a tree*, if $\lambda^\circ \in \mathcal{T}^\circ$ implies $\mu^\circ \in \mathcal{T}^\circ$, whenever $\lambda^\circ \succ \mu^\circ$. An item $\lambda^\circ \in \mathcal{T}^\circ$ is also called a* knot *in the following. The knots on level j_0 have some special importance and are called* roots *or* root knots*.*

Note that the tree structure is defined on the index set I of scaling function indices. But since all wavelet indices $\lambda = (j, k, e) \in J, e \neq \mathbf{0}$, are also associated with $\Box_{\lambda^\circ}$, the tree structure on I induces a *tree-like structure* on J by defining that $\mathcal{T} \subset J$ has tree structure if $\mathcal{T}^\circ \subset I$ is a tree in the above sense. For convenience, we will always speak about 'trees'. It will simplify data structures in case the used tree is *complete* in the following sense.

Definition 4.2 *A tree $\mathcal{T}^\circ \subset I$ (or rather the corresponding tree $\mathcal{T} \subset J$) is called* complete *if*

- *$\lambda^\circ \in \mathcal{T}^\circ$ implies that all* siblings *of λ° belong to the tree, i.e., all μ° with $\mathcal{P}(\mu^\circ) = \mathcal{P}(\lambda^\circ)$,*
- *all roots belong to $\mathcal{T}^\circ$ and*
- *the corresponding tree $\mathcal{T} \subset J$ has the property that $\lambda \in \mathcal{T}$ implies that all μ with $|\mu| = |\lambda|$ and $k(\mu) = k(\lambda)$ belong to $\mathcal{T}$.*

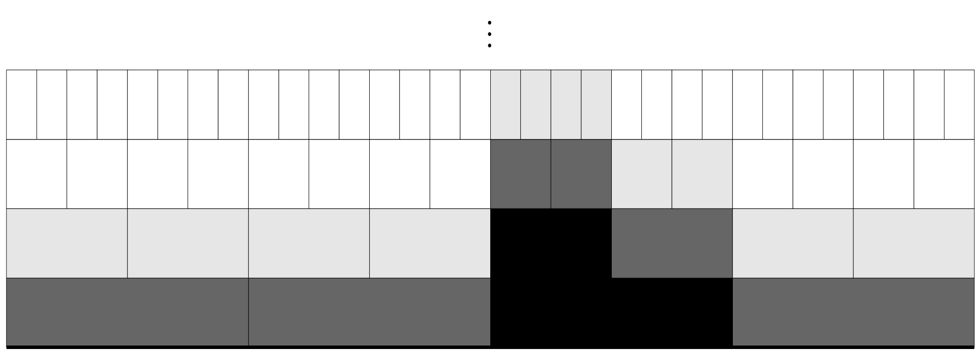
We will work exclusively with complete trees in the following. Comparing the number of knots $\#\mathcal{T}$ of a tree $\mathcal{T}$ and its 'completed tree' $\mathcal{T}_c$, we have $\#\mathcal{T}_c \lesssim \#\mathcal{T}$ due to the uniformly bounded number of siblings. Thus the assumption of a complete tree does not affect asymptotic results. Besides the roots of a tree we have to take into account other important index sets. For a finite tree $\mathcal{T}^\circ \subset I$ the set $\mathcal{D} = \mathcal{D}(\mathcal{T}^\circ)$ of *outer leaves* is defined as the set of those indices outside the tree whose parents belong to the tree,

$$\mathcal{D}(\mathcal{T}^\circ) := \{\lambda^\circ \in I : \lambda^\circ \notin \mathcal{T}^\circ, \mathcal{P}(\lambda^\circ) \in \mathcal{T}^\circ\}. \tag{4.3}$$

The set $\mathcal{B} = \mathcal{B}(\mathcal{T}^\circ)$ of *leaves* of the finite tree $\mathcal{T}^\circ$ is defined as the set of all indices in the tree so that none of their children (and thus none of their successors at all) is a member of the tree,

$$\mathcal{B}(\mathcal{T}^\circ) = \{\lambda^\circ \in I : \lambda \in \mathcal{T}^\circ, \mu^\circ \succ \lambda^\circ \Longrightarrow \mu^\circ \notin \mathcal{T}^\circ\}. \tag{4.4}$$

All tree knots, which are no leaves, are also called *interior knots.* The following picture shows a complete tree T of eight knots with black colored interior knots, grey colored leaves and light-grey colored outer leaves.



So far tree structures are defined on the index set of a single wavelet basis $\Psi = \{\psi_\lambda\}_{\lambda \in \Lambda}$. But for the construction of aggregated Gelfand frames (2.28) an overlapping decomposition $\mathcal{C} = \{\Omega_i\}_{i=1}^M$ of Ω is used, where the overall index set L is of the form $L = \bigcup_{i=1}^M \{i\} \times K^\Box$. The natural extension of tree structures to this setting are unions of locally defined trees.

Definition 4.3 *Assume that an index set $\mathcal{L} := \bigcup_{i=1}^{m} \{i\} \times \Lambda_i$, $m \in \mathbb{N}$, is given, where each Λ_i is of the form (2.23), $i = 1, \ldots, m$. A subset $\mathcal{T}^\circ := \bigcup_{i=1}^{m} \mathcal{T}_i^\circ \subset \mathcal{L}^\circ$ is called an* aggregated tree *if for all $i = 1, \ldots, m$ it holds $\mathcal{T}_i^\circ \subset \{i\} \times \Lambda_i^\circ$ and $\mathcal{T}_i^\circ$ is a tree in the sense of Definition 4.1. For convenience, we will call $\mathcal{T}_i^\circ$ a* local tree *in the following.*

It is important to understand that all concepts described above - leaves, successors and others - are understood locally when we talk of an aggregated tree $\mathcal{T}^\circ$. An index $\lambda^\circ \in \mathcal{T}^\circ$ for example is called a leaf of $\mathcal{T}^\circ$ if there exists an $i \in \{1, \ldots, m\}$ so that $\lambda^\circ = (i, \mu^\circ) \in \mathcal{T}_i^\circ$ is a leaf of the local tree $\mathcal{T}_i^\circ$.

Note that aggregated trees are defined on the index set $\mathcal{L}$, which is less restrictive than L. However, for convenience, we keep on using L in the following.

4.2 Best N-term Tree Approximation

After the introduction of trees for overlapping partitions of Ω we are ready to deal with the adequate standard of comparison, the best N-term tree approximation. Therefore, we introduce two different concepts corresponding to (local) trees and aggregated trees. For $N = 1, 2, \ldots$, let $\Sigma_{N,\mathcal{T}}$ denote the nonlinear subspace of $\ell_2(\Lambda)$ consisting of all vectors with at most N nonzero entries whose support has (local) tree structure. Given $\mathbf{v} \in \ell_2(\Lambda)$, we introduce the error of approximation

$$\sigma_{N,\mathcal{T}}(\mathbf{v}) := \inf_{\mathbf{w} \in \Sigma_{N,\mathcal{T}}} \|\mathbf{v} - \mathbf{w}\|_{\ell_2(\Lambda)}$$

and define for $s > 0$ the corresponding approximation space $\mathcal{A}_{\mathcal{T}}^s$ as

$$\mathcal{A}_{\mathcal{T}}^s := \{\mathbf{v} \in \ell_2(\Lambda) : \sigma_{N,\mathcal{T}} \lesssim N^{-s}\},$$

i.e., the space of all vectors from $\ell_2(\Lambda)$ whose best N-term tree approximation is of the order $\mathcal{O}(N^{-s})$ with the quasi-seminorm

$$\|\mathbf{v}\|_{\mathcal{A}_{\mathcal{T}}^s} := \sup_{N>0} N^s \cdot \sigma_{N,\mathcal{T}}(\mathbf{v}).$$

The concept of best N-term approximation is closely related to the *weak ℓ_τ spaces* ℓ_τ^w, see [5]. Their counterpart in the context of tree structures are the *weak ℓ_τ tree spaces* $\ell_{\tau,\mathcal{T}}^w$. Defining for $\mathbf{v} \in \ell_2(\Lambda)$ the vector $\tilde{\mathbf{v}}$ by

$$\tilde{v}_\lambda := \left(\sum_{\mu \in \Gamma_\lambda} |v_\mu|^2 \right)^{1/2}, \tag{4.5}$$

the space $\ell_{\tau,\mathcal{T}}^w(\Lambda)$ is defined by

$$\ell_{\tau,\mathcal{T}}^w(\Lambda) := \{\mathbf{v} \in \ell_2(\Lambda) : \tilde{\mathbf{v}} \in \ell_\tau^w(\Lambda)\}$$

with the quasi-seminorm

$$|\mathbf{v}|_{\ell_{\tau,\mathcal{T}}^w} := |\tilde{\mathbf{v}}|_{\ell_\tau^w}. \tag{4.6}$$

Further details on the spaces $\ell_{\tau,\mathcal{T}}^w(\Lambda)$ can be found in [7, 8]. Note that the spaces $\mathcal{A}_{\mathcal{T}}^s$ and $\ell_{\tau,\mathcal{T}}^w(\Lambda)$ do not coincide, $\ell_{\tau,\mathcal{T}}^w(\Lambda)$ is a real subspace of $\mathcal{A}_{\mathcal{T}}^s$.

For the Gelfand frame setting we finally define for $N = 1, 2, \ldots$ the nonlinear subspace $\Sigma_{N,\mathcal{AT}}$ of $\ell_2(L)$ consisting of all vectors with at most N nonzero entries whose support has the structure of an aggregated tree $\mathcal{T}^\circ = \bigcup_{i=1}^{M} \mathcal{T}_i^\circ$. Given $\mathbf{v} \in \ell_2(L)$, we introduce the error of approximation

$$\sigma_{N,\mathcal{AT}}(\mathbf{v}) := \inf_{\mathbf{w} \in \Sigma_{N,\mathcal{AT}}} \|\mathbf{v} - \mathbf{w}\|_{\ell_2(L)}$$

and define for $s > 0$ the corresponding approximation space $\mathcal{A}^s_{\mathcal{AT}}$ by

$$\mathcal{A}^s_{\mathcal{AT}} := \{\mathbf{v} \in \ell_2(L) : \sigma_{N,\mathcal{AT}} \lesssim N^{-s}\},$$

i.e., the space of all vectors whose best N-term aggregated tree approximation is of the order $\mathcal{O}(N^{-s})$ with quasi-seminorm

$$\|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} := \sup_{N>0} N^s \cdot \sigma_{N,\mathcal{AT}}(\mathbf{v}). \tag{4.7}$$

Again, there is also a concept of corresponding *weak ℓ_τ aggregated tree spaces* $\ell^w_{\tau,\mathcal{AT}}$ related to the concept of best N-term aggregated tree approximation. Therefore, note that any sequence in $\ell_2(L)$ can be uniquely rearranged as an M-tuple $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_M)$, where $\mathbf{v}_i$ contains the elements from $\{i\} \times K^\square$. The space $\ell^w_{\tau,\mathcal{AT}}(L)$ is defined by

$$\ell^w_{\tau,\mathcal{AT}}(L) := \{\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_M) : \mathbf{v}_i \in \ell^w_{\tau,\mathcal{T}}(\{i\} \times K^\square) \text{ for all } i = 1, \ldots, M\}$$

with corresponding quasi-seminorm

$$|\mathbf{v}|_{\ell^w_{\tau,\mathcal{AT}}(L)} := \max_{i=1}^{M} |\mathbf{v}_i|_{\ell^w_{\tau,\mathcal{T}}(\{i\}\times K^\square)}.$$

The relationship between the different concepts of weak ℓ_τ spaces is obvious: $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_M) \in \ell_2(L)$ is in $\ell^w_{\tau,\mathcal{AT}}(L)$ if and only if $\mathbf{v}_i$ is in $\ell^w_{\tau,\mathcal{T}}(\{i\} \times K^\square)$ for all $i = 1, \ldots, M$. For the approximation spaces $\mathcal{A}^s$ a similar result holds.

Lemma 4.1 *For $s > 0$ and $\mathbf{v} \in \ell_2(L)$ holds: $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$ if and only if $\mathbf{v}_i \in \mathcal{A}^s_{\mathcal{T}}$ for all $i = 1, \ldots, M$.*

Proof: Assume that $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$. Therefore, $\sigma^2_{N,\mathcal{AT}}(\mathbf{v}) = \inf_{\mathbf{w}\in\Sigma_{N,\mathcal{AT}}} \|\mathbf{v} - \mathbf{w}\|^2_{\ell_2(L)} \lesssim N^{-2s}$. Then, it is clear that $\sum_{i=1}^{M} \inf_{\mathbf{w}_i\in\Sigma_{N,\mathcal{T}}} \|\mathbf{v}_i - \mathbf{w}_i\|^2_{\ell_2(\{i\}\times K^\square)} \lesssim \inf_{\mathbf{w}\in\Sigma_{N,\mathcal{AT}}} \|\mathbf{v} - \mathbf{w}\|^2_{\ell_2(L)} \lesssim N^{-2s}$, from which it follows that $\sigma^2_{N,\mathcal{T}}(\mathbf{v}_i) = \inf_{\mathbf{w}_i\in\Sigma_{N,\mathcal{T}}} \|\mathbf{v}_i - \mathbf{w}_i\|^2_{\ell_2(\{i\}\times K^\square)} \lesssim N^{-2s}$ for all $i = 1, \ldots, M$. This means $\mathbf{v}_i \in \mathcal{A}^s_{\mathcal{T}}$ for all $i = 1, \ldots, M$.
If otherwise $\mathbf{v}_i \in \mathcal{A}^s_{\mathcal{T}}$ for all $i = 1, \ldots, M$, it follows that $\sigma^2_{N/M,\mathcal{T}}(\mathbf{v}_i) = \inf_{\mathbf{w}_i\in\Sigma_{N/M,\mathcal{T}}} \|\mathbf{v}_i - \mathbf{w}_i\|^2_{\ell_2(\{i\}\times K^\square)} \lesssim \left(\frac{N}{M}\right)^{-2s}$ for all $i = 1, \ldots, M$. Therefore, $\sigma^2_{N,\mathcal{AT}}(\mathbf{v}) = \inf_{\mathbf{w}\in\Sigma_{N,\mathcal{AT}}} \|\mathbf{v} - \mathbf{w}\|^2_{\ell_2(L)} \leq \sum_{i=1}^{M} \inf_{\mathbf{w}_i\in\Sigma_{N/M,\mathcal{T}}} \|\mathbf{v}_i - \mathbf{w}_i\|^2_{\ell_2(\{i\}\times K^\square)} \lesssim \sum_{i=1}^{M} \left(\frac{N}{M}\right)^{-2s} \lesssim N^{-2s}$. □

A classical result from the theory of nonlinear tree approximation with wavelet Riesz bases reads as follows, see [7, 9]. Let $(\Psi, \tilde{\Psi})$ be a pair of biorthogonal wavelet Riesz bases for $L_2(\Omega)$ such that, with the diagonal matrix $D = \mathrm{diag}(2^{|\lambda|t})_{\lambda\in\Lambda}$, $D^{-1}\Psi$ is a wavelet Riesz basis for $H^t_0(\Omega), t > 0$. Furthermore, let Ψ be of order n, meaning that locally polynomials up to degree $n-1$ can be represented exactly. Then, for sufficiently smooth wavelets and $0 < s < (n-t)/d$, for the solution u of (2.1) holds

$$\text{If } u \in B^{sd+t}_{\tau'}(L_{\tau'}(\Omega)), \text{ then } D\tilde{F}u = \left(2^{|\lambda|t}\langle u, \tilde{\psi}_\lambda\rangle_{L_2(\Omega)}\right)_{\lambda\in\Lambda} \in \ell^w_{\tau,\mathcal{T}}(\Lambda), \tag{4.8}$$

where $1/\tau' < 1/\tau = s + 1/2$ and $D\tilde{F}u$ are the unique expansion coefficients of u with respect to the wavelet Riesz basis $D^{-1}\Psi$ in $H^t_0(\Omega)$. As usual, $B^\alpha_{\tau'}(L_{\tau'}(\Omega))$ denotes the classical Besov space measuring smoothness up to order α in $L_{\tau'}(\Omega)$.

Due to their redundancy, such a complete characterization can not be expected generally for frames. But, in [11] it was shown for the Gelfand frame expansion concerning the non-canonical dual frame from (2.29) that these frame coefficients show a certain decay for elliptic linear problems $Au = f$ over the two-dimensional L-domain $\Omega := (-1,1)^2 \setminus [0,1)^2 \subset \mathbb{R}^2$ decomposed into two patches $\Omega_1 := (-1,0) \times (-1,1)$ and $\Omega_2 := (-1,1) \times (-1,0)$. Thus, for this setting, a result analogous to (4.8) can be established, essentially based on a sufficiently smooth partition of unity. We refer to [11] for further details.

4.3 Tree Approximation and Coarsening

In view of the implementation of a **COARSE** routine respecting (aggregated) tree structures, we face the problem to determine a tree $\mathcal{T}$, which approximates a finitely supported vector $\mathbf{v} \in \ell_2(\Lambda)$ up to a target accuracy $\varepsilon > 0$ and whose support is - due to computational effort - as small as possible, i.e.,

$$\|\mathbf{v} - \mathbf{v}_\mathcal{T}\|_{\ell_2(\Lambda)} \le \varepsilon \;\text{ and }\; \#\mathcal{T} = \min\{\#\mathcal{T}' : \|\mathbf{v} - \mathbf{v}_{\mathcal{T}'}\|_{\ell_2(\Lambda)},\, \mathcal{T}' \text{ is a tree}\}. \tag{4.9}$$

Here $\mathbf{v}_\mathcal{T}$ denotes the vector, where all entries outside $\mathcal{T}$ are set to zero. Such a tree $\mathcal{T} = \mathcal{T}^*(\varepsilon, \mathbf{v})$ is called *ε-best*, cf. [3, 7]. Unfortunately, the determination of an ε-best tree is very expensive, because the number of possible trees grows exponentially in $\#\mathcal{T}$. Hence, the only thing we can hope to achieve practically is the determination of an *ε-near-best tree* $\mathcal{T} = \mathcal{T}(\varepsilon, \mathbf{v})$, i.e., a tree such that for a constant $C > 0$ it is

$$\|\mathbf{v} - \mathbf{v}_\mathcal{T}\|_{\ell_2(\Lambda)} \le \varepsilon \;\text{ and }\; \#\mathcal{T} \le C \cdot \#\mathcal{T}^*(\varepsilon/C, \mathbf{v}). \tag{4.10}$$

Note that (4.10) is equivalent to the existence of two constants $C_1, C_2 > 0$ with $\|\mathbf{v} - \mathbf{v}_\mathcal{T}\|_{\ell_2(\Lambda)} \le C_1\varepsilon$ and $\#\mathcal{T} \le C_2 \#\mathcal{T}^*(\varepsilon, \mathbf{v})$. In [3] it was shown that the determination of an ε-near-best tree is possible in linear time by an algorithm based on the evaluation of (4.5). The main idea is the following:

- Define a non-negative functional e on the set I by $e(\lambda^\circ) := \tilde{v}^2_{\lambda^\circ}$, where $\tilde{v}_{\lambda^\circ}$ is defined by (4.5).
- Given $\mathbf{v} \in \ell_2(\Lambda)$ with finite support determine $e(\lambda^\circ)$ for all $\lambda^\circ \in \bar{\mathcal{T}}^\circ(\mathbf{v})$, where $\bar{\mathcal{T}}(\mathbf{v})$ is the smallest tree containing $\operatorname{supp}(\mathbf{v})$. This tree $\bar{\mathcal{T}}^\circ(\mathbf{v})$ can be constructed in linear time by identifying $j_{\max}$, the highest coefficient level appearing in $\mathbf{v}$ and determination of

$$\begin{aligned}
U_{j_{\max}}(\mathbf{v}) &:= \{\lambda^\circ : |\lambda^\circ| = j_{max}, v_\lambda \neq 0\},\\
U_j(\mathbf{v}) &:= \{\mu^\circ : \mu^\circ = \mathcal{P}(\lambda^\circ), \lambda^\circ \in U_{j+1}\} \cup \{\lambda^\circ : |\lambda^\circ| = j, v_\lambda \neq 0\},\; j_{max} > j > j_0,\\
U_{j_0}(\mathbf{v}) &:= I_{j_0},\\
\bar{\mathcal{T}}^\circ(\mathbf{v}) &:= \bigcup_{j=j_0}^{j_{\max}} U_j(\mathbf{v}).
\end{aligned} \tag{4.11}$$

- Determine a sequence of trees $\mathcal{T}_j, j \ge 0$, starting with the tree $\mathcal{T}_0$, which consists of all root knots. After $\mathcal{T}_{j-1}$ has been defined, examine the set of leaves $\mathcal{B}(\mathcal{T}_{j-1})$ of $\mathcal{T}_{j-1}$ and subdivide all leaves λ° with the largest value of $e(\lambda^\circ)$. Stop if $\sum_{\lambda^\circ \in \mathcal{B}(\mathcal{T}_j)} e(\lambda^\circ) \le \varepsilon$.

The number of operations and computations needed to determine $\bar{\mathcal{T}}^\circ(\mathbf{v})$ remains proportional to $\#\bar{\mathcal{T}}^\circ(\mathbf{v})$. It was shown in [3] that this algorithm provides (ε, C^*)-near best trees, where $C^* \ge 1$ is an absolute constant, independent from $\mathbf{v}$ and ε. Later on, we will assume near-best trees to be constructed by (4.11). Then, C^* is exactly the constant mentioned in Algorithm 1 and 2. We refer to [3, 7] for further details.

For the Gelfand frame setting we introduce *ε-best aggregated trees* $\mathcal{T}^* \subset L$ and *ε-near-best aggregated trees* $\mathcal{T} \subset L$ analogously to (4.9) and (4.10) for $\ell_2(L)$ and refer to the following remark.

Remark 4.1 *Let $\varepsilon > 0$ and assume that $\mathbf{v} \in \ell_2(L)$ has finite support, where $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_M)$ with $\mathbf{v}_i \in \ell_2(\{i\} \times K^\square)$. Then, an ε-near-best aggregated tree $\mathcal{T}(\varepsilon, \mathbf{v})$ for $\mathbf{v}$ is obtained by any union of ε/M-near-best trees $\mathcal{T}(\frac{\varepsilon}{M}, \mathbf{v}_i), i = 1, \ldots, M$, for $\mathbf{v}_i$.*

Proof: Let $\mathcal{T}(\frac{\varepsilon}{M}, \mathbf{v}_i)$ be ε/M-near-best trees for $\mathbf{v}_i, i = 1, \ldots, M$, with $\#\mathcal{T}(\frac{\varepsilon}{M}, \mathbf{v}_i) \le C_i \#\mathcal{T}^*(\frac{\varepsilon}{C_i M}, \mathbf{v}_i)$ for constants $C_i > 0$, $i = 1, \ldots, M$, and set $\mathcal{T} := \bigcup_{i=1}^M \mathcal{T}(\frac{\varepsilon}{M}, \mathbf{v}_i)$. It immediately follows that $\|\mathbf{v} - \mathbf{v}_\mathcal{T}\|_{\ell_2(L)} \le$

$\sum_{i=1}^M \|\mathbf{v}_i - \mathbf{v}_{i\,\mathcal{T}(\varepsilon/M,\mathbf{v}_i)}\|_{\ell_2(\{i\}\times K^\square)} \le \varepsilon$. Furthermore, for $C := \max_{i=1}^M C_i$ we get $\#\mathcal{T} = \sum_{i=1}^M \#\mathcal{T}(\frac{\varepsilon}{M}, \mathbf{v}_i) \le \sum_{i=1}^M C_i \cdot \#\mathcal{T}^*(\frac{\varepsilon}{C_i M}, \mathbf{v}_i) \le \sum_{i=1}^M C \cdot \#\mathcal{T}^*(\frac{\varepsilon}{CM}, \mathbf{v}_i) \le \sum_{i=1}^M C \cdot \#\mathcal{T}^*(\frac{\varepsilon}{CM}, \mathbf{v}) = CM \cdot \#\mathcal{T}^*(\frac{\varepsilon}{CM}, \mathbf{v})$. Thus, $\mathcal{T}$ is an ε-near best aggregated tree. □

Remark 4.1 has important consequences for the practical determination of near-best aggregated trees. It tells us that it suffices to determine local near-best trees $\mathcal{T}(\xi, \mathbf{v}_i)$ over patch Ω_i for $i = 1, \dots, M$, (which are independent of each other, so that this can be done in parallel) and merge these local trees. If all local near-best trees are constructed by (4.11), the constants C_i, $i = 1, \dots, M$, coincide with C^*. This allows us to specify the abstract **COARSE** routine in the following way.

ATCOARSE$[\varepsilon, \mathbf{v}] \to \mathbf{w}_\varepsilon$

determines for any tolerance $\varepsilon > 0$ and any finitely supported input vector $\mathbf{v} \in \ell_2(L)$ an ε-near-best aggregated tree $\mathcal{T}(\varepsilon, \mathbf{v})$ using the algorithm from [3] and Remark 4.1 and sets $\mathbf{w}_\varepsilon := \mathbf{v}_{\mathcal{T}(\varepsilon,\mathbf{v})}$.

Analogous to the remark given in [7] we show the following lemma.

Lemma 4.2 *Let $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$ and $\|\mathbf{v}-\mathbf{w}\|_{\ell_2(L)} \le \frac{\varepsilon}{2C^*+1}$ with C^* the constant from [3] and $\#\mathrm{supp}(\mathbf{w}) < \infty$. Then, for $\mathbf{w}_\varepsilon :=$**ATCOARSE**$[\frac{2C^*}{2C^*+1}\varepsilon, \mathbf{w}]$ it holds*

$$\begin{aligned} \#\mathrm{supp}(\mathbf{w}_\varepsilon) &\lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} \varepsilon^{-1/s}, \quad (4.12)\\ \|\mathbf{v} - \mathbf{w}_\varepsilon\|_{\ell_2(L)} &\le \varepsilon, \\ \|\mathbf{w}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} &\lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}. \quad (4.13)\end{aligned}$$

Furthermore, as noted above, the number of operations and computations needed to determine $\mathbf{w}_\varepsilon$ is bounded by $C\#\mathrm{supp}(\mathbf{w}_\varepsilon) \le C\|\mathbf{v}\|^{1/s}\varepsilon^{-1/s}$. The involved constants depend only on s, when $s \to 0$, and on C^*.

Proof: The proof is analogous to [7] using $\mathcal{A}^s_{\mathcal{AT}}$. Thus, let $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$ and $\mathcal{T}^*_N(\mathbf{v})$ be the best aggregated tree approximation to $\mathbf{v}$ with at most N elements, where $N \in \mathbb{N}$ will be chosen below. It is

$$\begin{aligned} \|\mathbf{w} - \mathbf{w}_{\mathcal{T}^*_N(\mathbf{v})}\|_{\ell_2(L)} &= \|\mathbf{w}_{L\setminus\mathcal{T}^*_N(\mathbf{v})} + \mathbf{v} - \mathbf{v}_{\mathcal{T}^*_N(\mathbf{v})} - \mathbf{v}_{L\setminus\mathcal{T}^*_N(\mathbf{v})}\|_{\ell_2(L)} \\ &\le \|(\mathbf{w}-\mathbf{v})_{L\setminus\mathcal{T}^*_N(\mathbf{v})}\|_{\ell_2(L)} + \|\mathbf{v} - \mathbf{v}_{\mathcal{T}^*_N(\mathbf{v})}\|_{\ell_2(L)} \\ &\le \frac{\varepsilon}{2C^*+1} + \|\mathbf{v} - \mathbf{v}_{\mathcal{T}^*_N(\mathbf{v})}\|_{\ell_2(L)}. \end{aligned}$$

By definition of $\|\cdot\|_{\mathcal{A}^s_{\mathcal{AT}}}$ we have $\|\mathbf{v} - \mathbf{v}_{\mathcal{T}^*_N(\mathbf{v})}\|_{\ell_2(L)} \le \frac{\varepsilon}{2C^*+1}$ for an $N \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}(\frac{\varepsilon}{2C^*+1})^{-1/s} \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\varepsilon^{-1/s}$. For this N we have $\|\mathbf{w} - \mathbf{w}_{\mathcal{T}^*_N(\mathbf{v})}\|_{\ell_2(L)} \le \frac{2\varepsilon}{2C^*+1}$. By definition of the $(\frac{2C^*\varepsilon}{2C^*+1})$-near-best aggregated tree $\mathcal{T}(\frac{2C^*\varepsilon}{2C^*+1}, \mathbf{w})$ it follows

$$\#\mathcal{T}\left(\frac{2C^*\varepsilon}{2C^*+1}, \mathbf{w}\right) \le C^*\#\mathcal{T}^*\left(\frac{2\varepsilon}{2C^*+1}, \mathbf{w}\right) \le C^*\#\mathcal{T}^*_N(\mathbf{v}) \le C^*N \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\varepsilon^{-1/s}, \quad (4.14)$$

which shows (4.12). The second assertion is an immediate consequence of the triangle inequality, while for the proof of (4.13) it suffices to show that for every $\delta > 0$ there exists an aggregated tree $\mathcal{T}_\delta$ with $\|\mathbf{w}_\varepsilon - (\mathbf{w}_\varepsilon)_{\mathcal{T}_\delta}\|_{\ell_2(L)} \le \delta$ and $\#\mathcal{T}_\delta \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\delta^{-1/s}$. Because $\|\mathbf{w}_\varepsilon - (\mathbf{w}_\varepsilon)_{\mathcal{T}_\delta}\|_{\ell_2(L)} \le \delta$ means $\sigma_{\bar{N},\mathcal{A}^s_{\mathcal{AT}}}(\mathbf{w}_\varepsilon) \le \delta$ with $\bar{N} := \#\mathcal{T}_\delta$, (4.13) follows then by $\|\mathbf{w}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} = \sup_{N>0} N^s \cdot \sigma_{N,\mathcal{AT}}(\mathbf{w}_\varepsilon) \le (\#\mathrm{supp}(\mathbf{w}_\varepsilon))^s \cdot \delta \lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}\delta^{-1}\delta = \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}$, using (4.12).

At first, let $\delta \le 2\varepsilon$. In this case, choose $\mathcal{T}_\delta := \mathrm{supp}(\mathbf{w}_\varepsilon) = \mathcal{T}(\frac{2C^*}{2C^*+1}\varepsilon, \mathbf{w})$. It is $\|\mathbf{w}_\varepsilon - (\mathbf{w}_\varepsilon)_{\mathcal{T}_\delta}\|_{\ell_2(L)} = 0$ by construction and $\#\mathcal{T}_\delta = \#\mathcal{T}(\frac{2C^*}{2C^*+1}\varepsilon, \mathbf{w}) \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\varepsilon^{-1/s} \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\delta^{-1/s}$ on delta's choice.

Assume now $\delta > 2\varepsilon$. Let $\mathcal{T}_\delta := \mathcal{T}_N^*(\mathbf{v})$ be a best N-term aggregated tree, where again N will be chosen below. Then, we have

$$\|\mathbf{w}_\varepsilon - (\mathbf{w}_\varepsilon)_{\mathcal{T}_\delta}\|_{\ell_2(L)} \le \|(\mathbf{w}_\varepsilon - \mathbf{v})_{\mathcal{T}_\delta}\|_{\ell_2(L)} + \|\mathbf{v} - \mathbf{v}_{\mathcal{T}_\delta}\|_{\ell_2(L)} \le \varepsilon + \|\mathbf{v} - \mathbf{v}_{\mathcal{T}_\delta}\|_{\ell_2(L)}.$$

Thus, $\|\mathbf{w}_\varepsilon - (\mathbf{w}_\varepsilon)_{\mathcal{T}_\delta}\|_{\ell_2(L)} \le \delta \iff \delta - \varepsilon \ge \|\mathbf{v} - \mathbf{v}_{\mathcal{T}_\delta}\|_{\ell_2(L)}$. This is true for an $N \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}(\delta - \epsilon)^{-1/s} \le 2^{1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\delta^{-1/s} \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\delta^{-1/s}$, since $\delta - \varepsilon > \frac{1}{2}\delta$, which completes the proof. □

After the description of **ATCOARSE** the question after homologous **ATAPPLY** and **ATRHS** preserving aggregated tree structures naturally arises. We won't go into this in detail, but give some general remarks.

Concerning the right-hand side $\mathbf{f}$ of (2.32) we assume, as customary, to be able to compute approximate wavelet expansions of f in the dual frame up to any given accuracy, i.e., for any $\eta > 0$ there exists a computable, finitely supported coefficient array $\mathbf{f}_\eta \in \ell_2(L)$ such that $\|\mathbf{f} - \mathbf{f}_\eta\|_{\ell_2(L)} \le \frac{\eta}{2C^*+1}$. Then, if we assume that the exact right-hand side $\mathbf{f}$ is in $\mathcal{A}^s_{\mathcal{AT}}$ for some $s > 0$, **ATRHS** can be implemented using **ATCOARSE** by

$$\textbf{ATRHS}[\eta, \mathbf{f}] := \textbf{ATCOARSE}[\tfrac{2C^*}{2C^*+1}\eta, \mathbf{f}_\eta].$$

We further assume the computational effort for the determination of $\mathbf{f}_\eta$ to be bounded by at most a multiple of the effort of **ATCOARSE**, so that the overall computational effort of **ATRHS** is - up to constants - determined by **ATCOARSE**.

For the numerical implementation of **ATAPPLY** we want to use the **APPLY** routine developed in [1, 5, 6] for s^*-compressible matrices $\mathbf{B}$ (we need this **APPLY** routine for $\mathbf{B} = \mathbf{A}$ and $\mathbf{B} = \mathbf{Q}$). For convenience, recall

Definition 4.4 *Let $s^* > 0$. A bounded $\mathbf{B} : \ell_2(L) \to \ell_2(L)$ is called* s^*-compressible*, when for each $j \in \mathbb{N}$ there exist constants α_j and C_j, and an infinite matrix $\mathbf{B}_j$ having at most $\alpha_j 2^j$ non-zero entries in each column, such that $\|\mathbf{B} - \mathbf{B}_j\| \le C_j$, $(\alpha_j)_{j\in\mathbb{N}}$ is summable and for any $s < s^*$, $(C_j 2^{sj})_{j\in\mathbb{N}}$ is summable.*

We leave the details of **APPLY** to the reader, see [1, 7, 24], and confine ourselves to point out two main facts. First, for s^*-compressible matrices the following remark holds.

Remark 4.2 *Let $\mathbf{B}$ be s^*-compressible. Then, for any $0 < s < s^*$,* **APPLY** *can be implemented such that for $\mathbf{w}_\varepsilon$:=***APPLY***$[\varepsilon, \mathbf{B}, \mathbf{v}]$ the following estimates are valid.*

(i) $\|\mathbf{B}\mathbf{v} - \mathbf{w}_\varepsilon\|_{\ell_2(L)} \le \varepsilon$,

(ii) $\#\mathrm{supp}(\mathbf{w}_\varepsilon) \lesssim \varepsilon^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}$,

(iii) the number of arithmetic operations used to compute $\mathbf{w}_\varepsilon$ is bounded by at most a fixed multiple of $\varepsilon^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#\mathrm{supp}(\mathbf{v})$.

Proof: See the proof in [24] and use $|\mathbf{v}|_{\ell^w_\tau} \le \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}$ for all $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$, $\frac{1}{\tau} = s + \frac{1}{2}$. □

Secondly, note that the resulting approximation $\mathbf{w}_\varepsilon$ to $\mathbf{B}\mathbf{v}$ cannot be expected to have aggregated tree structure in general. But, this can be achieved by combining **APPLY** with **ATCOARSE** described above by

$$\textbf{ATAPPLY}[\varepsilon, \mathbf{B}, \mathbf{v}] := \textbf{ATCOARSE}\Big[\tfrac{2C^*}{2C^*+1}\varepsilon, \textbf{APPLY}\,[\tfrac{1}{2C^*+1}\varepsilon, \mathbf{B}, \mathbf{v}]\Big].$$

The asymptotic optimality of this **ATAPPLY** routine follows from Lemma 4.2 if we can assure $\mathbf{B}$ to be bounded on $\mathcal{A}^s_{\mathcal{AT}}$. For $\mathbf{B} = \mathbf{A}$ this is indeed the case for $\mathbf{A}$ with certain off-diagonal decay.

Lemma 4.3 *Assume in addition to Chapter 2 that A is a local operator, i.e., $\langle Av, w\rangle = 0$ if $\mathrm{supp}(v) \cap \mathrm{supp}(w) = \emptyset$, and let for any $v = \mathbf{v}^T\Psi$, $\mathbf{v} \in \ell_2(L)$, with $\#\mathrm{supp}(\mathbf{v}) < \infty$*

$$|(\mathbf{A}\mathbf{v})_\lambda| \leq C(\|\mathbf{v}\|_{\ell_2(L)}) \sup_{\mu : S_\mu \cap S_\lambda \neq \emptyset} v_\mu \cdot 2^{-\gamma(|\lambda|-|\mu|)}, \tag{4.15}$$

for all $|\lambda| > j_0$, where $S_\mu := \mathrm{supp}(\psi_\mu)$, $\gamma > d/2$ is an absolute constant and $x \to C(x)$ is a positive non-decreasing function. Then, $\mathbf{A}$ is bounded from $\mathcal{A}^s_{\mathcal{AT}}$ to $\mathcal{A}^s_{\mathcal{AT}}$.

Proof: Use Assumption 4.1 for A and Theorem 4.1, both given in the next chapter. □

We conclude our notes on **ATAPPLY** with the remark that (4.15) holds, for instance, if the stiffness matrix of a local operator A belongs to the important *Lemarié* class, which is in turn the case for elliptic differential operators discretized by the aggregated Gelfand frame (2.28). We refer to [12] for further details.

We assume in the following the existence of a near-best aggregated tree $\mathcal{T} = \bigcup_{i=1}^M \mathcal{T}_i$ for $\mathbf{v}$, where each $\mathcal{T}_i$ is constructed by (4.11). Our next goal is the usage of the aggregated tree concept for the efficient determination of the support of an approximation $\mathbf{w}_\varepsilon$ of $\mathbf{G}(\mathbf{v})$, i.e., the realization of **EVAL**. According to this we show in the next subsection that the ideas from [8] for wavelet Riesz bases can be transferred to the case of aggregated Gelfand frames.

4.4 Evaluation of Nonlinear Expressions I: Support Prediction

Given an aggregated tree $\mathcal{T} \subset \mathcal{L}$ suppose the following *expansion process*. For any $\lambda \in \mathcal{T}$ define the sets

$$\begin{aligned} \Phi_0(\lambda) &:= \{\lambda\}, \\ \Phi_k(\lambda) &:= \{\mu \in \mathcal{L} : |\mu| = |\lambda| - k,\ \exists \xi \in \Phi_{k-1}(\lambda) : S_\xi \cap S_\mu \neq \emptyset\},\ 1 \leq k \leq |\lambda|, \\ \Phi(\lambda) &:= \bigcup_{k=0}^{|\lambda|} \Phi_k(\lambda), \end{aligned} \tag{4.16}$$

where $S_\lambda := \mathrm{supp}(\psi_\lambda)$, and

$$\tilde{\mathcal{T}} := \bigcup_{\lambda \in \mathcal{T}} \Phi(\lambda). \tag{4.17}$$

For $\lambda \in \mathcal{T}$ the expansion process collects all wavelets from level $j-1$, whose supports have a non-empty intersection with the support of wavelets collected on level j beginning with Φ_0. Doing this, we can be sure that $\Phi(\lambda)$ contains all wavelets below level $|\lambda|$ whose support have a non-trivial intersection with S_λ. Thus, the final set $\tilde{\mathcal{T}}$ contains at least all wavelets of lower levels that intersect the support of $\mathcal{T}$ in a non-trivial way. Thus, more precisely, we have established a *top-down expansion process*. In addition $\tilde{\mathcal{T}}$ has the following properties.

Lemma 4.4 *(i) For any $\mu \in \tilde{\mathcal{T}}$ and $\eta \in J$ it holds: If $|\eta| < |\mu|$ and $S_\eta \cap S_\mu \neq \emptyset$, then $\eta \in \tilde{\mathcal{T}}$.*

(ii) The index set $\tilde{\mathcal{T}}$ from (4.17) is an aggregated tree and there exists a constant C independent of $\tilde{\mathcal{T}}$ so that $\#\tilde{\mathcal{T}} \leq C\#\mathcal{T}$.

As in [7], we call the first property expansion property*.*

Proof: The expansion property is an immediate consequence of the expansion process described above. For the second part write $\tilde{\mathcal{T}}$ as $\tilde{\mathcal{T}} = \bigcup_{i=1}^m \tilde{\mathcal{T}}_i$ with $\tilde{\mathcal{T}}_i \subset \{i\} \times \Lambda_i$. It suffices then to show that $\tilde{\mathcal{T}}_i$ is a tree for all $i = 1, \ldots, m$. Suppose that $\lambda \in \tilde{\mathcal{T}}_i$, i.e., there exists an $\eta \in \mathcal{T}$ and $0 \le k \le |\eta|$ with $\lambda \in \Phi_k(\eta)$, and let μ° be any antecessor of λ°, $\lambda^\circ \succ \mu^\circ$. It is an immediate consequence of the construction of the reference cubes that

$$\lambda^\circ \succ \mu^\circ \quad \Longrightarrow \quad \text{supp } \psi_\lambda \subsetneq \text{supp } \psi_\mu. \tag{4.18}$$

By this $S_\lambda \cap S_\mu \neq \emptyset$. Thus, by construction of $\Phi_{k+1}(\eta), \ldots, \Phi_{k+|\lambda|-|\mu|}(\eta)$ it follows that $\mu \in \Phi_{k+|\lambda|-|\mu|}(\eta)$ and therefore, $\mu \in \tilde{\mathcal{T}}_i$, i.e., $\tilde{\mathcal{T}}_i$ is a tree.
To proof the estimate concerning the size of $\tilde{\mathcal{T}}$ it suffices to show that for all $\mu \in \tilde{\mathcal{T}}$ there is a $\lambda \in \mathcal{T}$ so that $|\mu| = |\lambda|$ and $\text{dist}(S_\mu, S_\lambda) \le C_0 2^{-|\mu|}$. The assertion follows then by the compactness of the support of the used wavelets and the finite number of patches that can overlap at a given point. For any $\mu \in \tilde{\mathcal{T}}$ there exists a $\lambda \in \mathcal{T}$ and $0 \le k \le |\lambda|$ so that $\mu \in \Phi_k(\lambda)$ by construction. For $k = 0$ there is nothing to proof since $\Phi_0(\lambda) = \{\lambda\}$, which means $\mu = \lambda$. Using induction let μ be contained in $\Phi_{k+1}(\lambda) \setminus \Phi_k(\lambda)$. Again by construction, there exist $\mu' \in \Phi_k(\lambda)$ with $S_\mu \cap S_{\mu'} \neq \emptyset$ and, by induction hypothesis, $\bar{\lambda} \in \mathcal{T}$ with $\text{dist}(S_{\bar{\lambda}}, S_{\mu'}) \le C_0 2^{-|\mu'|}$. Therefore $\text{dist}(S_\mu, S_{\bar{\lambda}}) \le C_0 2^{-|\mu'|} + \text{diam}(S_{\mu'}) \le C_0 2^{-|\mu'|} + C_0 2^{-|\mu'|} = C_0 2^{-|\mu|}$. Due to $\bar{\lambda} \in \mathcal{T}$ we know $\mathcal{P}(\bar{\lambda}) \in \mathcal{T}$ with $|\mathcal{P}(\bar{\lambda})| = |\mu|$ and by (4.18) we get $\text{dist}(S_\mu, S_{\mathcal{P}(\bar{\lambda})}) \le C_0 2^{-|\mu|}$. □
Concerning wavelets with higher levels we assume the following decay property.

Assumption 4.1 *Assume that $\Psi = \{\psi_\lambda\}_{\lambda \in \mathcal{L}}$ is an aggregated Gelfand frame for $(H_0^t(\Omega), L_2(\Omega), H^{-t}(\Omega))$ with dual frame $\tilde{\Psi} = \{\tilde{\psi}_\lambda\}_{\lambda \in \mathcal{L}}$ and let for any $v = \sum_{\mu \in V} \langle v, \tilde{\psi}_\mu \rangle \psi_\mu = \mathbf{v}^T \Psi$ with $\#V < \infty$ the estimate*

$$2^{-t|\lambda|} |\langle G(v), \psi_\lambda \rangle| \le C(\|\mathbf{v}\|_{\ell_2(\mathcal{L})}) \sup_{\mu : S_\mu \cap S_\lambda \neq \emptyset} 2^{t|\mu|} |\langle v, \tilde{\psi}_\mu \rangle| 2^{-\gamma(|\lambda| - |\mu|)} \tag{4.19}$$

hold for all $|\lambda| > j_0$, where $\gamma > d/2$ is an absolute constant and $x \to C(x)$ is a positive non-decreasing function.

Note the importance of (4.19) for numerical handling of the nonlinearity $G(v)$. It tells us that the wavelet coefficients of $G(v)$ are decreasing exponentially with growing scale so that we can hope that it suffices to examine a finite number of higher levels in order to achieve a suitable approximation of $G(v)$. It can therefore be interpreted as the counterpart of (4.17) concerning higher order wavelets.

Combining (4.17) and (4.19) we construct for a given vector $\mathbf{v} \in \ell_2(\mathcal{L})$ with finite support and $\varepsilon > 0$ an index set $\breve{\mathcal{T}}$ by the following steps:

```
Let J := min { j ∈ ℕ : 2^j ε / (M(j+1)) ≥ max_{i=1}^M ‖v_i‖_{ℓ2({i}×Λ_i)} }, α := 2/(2γ−d) and T_0 := ⋃_{i=1}^M T(ε/M, v_i) with
    expanded aggregated tree T̃_0.
For j = J,...,1 do
    Determine T_j := ⋃_{i=1}^M T(2^j ε/(M(j+1)), v_i) and T̃_j.
    Set Δ_{j−1} := T̃_{j−1} \ T̃_j.
    For all μ ∈ Δ_{j−1} determine
```

$$\Theta_{\varepsilon,\mu} := \{\lambda : S_\lambda \cap S_\mu \neq \emptyset, |\mu| \le |\lambda| \le |\mu| + (j-1)\alpha\}. \tag{4.20}$$

```
od
Set
```

$$\breve{\mathcal{T}} := I_{j_0} \cup \tilde{\mathcal{T}}_0 \cup \left(\bigcup_{\mu \in \tilde{\mathcal{T}}_0} \Theta_{\varepsilon,\mu} \right). \tag{4.21}$$

As the main result of this subsection we proof the following properties of $\breve{\mathcal{T}}$.

Theorem 4.1 *The index set $\breve{\mathcal{T}}$ from (4.21) is an aggregated tree satisfying*

$$\|\mathbf{G}(\mathbf{v}) - \mathbf{G}(\mathbf{v})_{\breve{\mathcal{T}}}\|_{\ell_2(L)} \lesssim \varepsilon. \tag{4.22}$$

Moreover, if $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}, 0 < s < \frac{2\gamma-d}{2d}$, *we have the estimate*

$$\#\breve{\mathcal{T}} \lesssim \varepsilon^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#(I_{j_0}). \tag{4.23}$$

Hence, we have $\mathbf{G}(\mathbf{v}) \in \mathcal{A}^s_{\mathcal{AT}}$ *and*

$$\|\mathbf{G}(\mathbf{v})\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim 1 + \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}. \tag{4.24}$$

The number of computations needed to find $\breve{\mathcal{T}}$ is bounded by $C\left(\varepsilon^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + (\#I_{j_0}) + \#\mathcal{T}(\mathrm{supp}(\mathbf{v}))\right)$, *where $\mathcal{T}(\mathrm{supp}(\mathbf{v}))$ is the smallest aggregated tree containing* $\mathrm{supp}(\mathbf{v})$.

The constants in the above inequalities depend on $\|\mathbf{v}\|_{\ell_2(L)}$, the space dimension d, the parameter s and the number of used patches M.

Proof: To proof the aggregated tree property it suffices again to show that for $\eta \in \breve{\mathcal{T}}$ all antecessors of η are in $\breve{\mathcal{T}}$. For the elements in I_{j_0} there is nothing to proof and for $\eta \in \tilde{\mathcal{T}}_0$ the statement follows from Lemma 4.4. So, assume that there are μ, j with $\mu \in \Delta_j$ and $\eta \in \Theta_{\varepsilon,\mu}$. Thus, $S_\eta \cap S_\mu \neq \emptyset$ and $|\mu| \leq |\eta| \leq |\mu| + \alpha j$. Therefore, we have $S_{\mathcal{P}(\eta)} \cap S_\mu \neq \emptyset$. Thus, if $|\mu| \leq |\mathcal{P}(\eta)|$, $\mathcal{P}(\eta) \in \breve{\mathcal{T}}$ follows by definition of $\Theta_{\varepsilon,\mu}$, while for $|\mathcal{P}(\eta)| < |\mu|$, we have $\mathcal{P}(\eta) \in \breve{\mathcal{T}}$ by construction of $\tilde{\mathcal{T}}$.
The remaining parts of the proof follow the lines of the proof given in [7] for the basis case. In order to proof (4.22) note first that

$$\|\mathbf{G}(\mathbf{v}) - \mathbf{G}(\mathbf{v})_{\breve{\mathcal{T}}}\|_{\ell_2(\mathcal{L})} \leq \|\mathbf{G}(\mathbf{v}) - \mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0})\|_{\ell_2(\mathcal{L})} + \|\mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0}) - \mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0})_{\breve{\mathcal{T}}}\|_{\ell_2(\mathcal{L})} + \|\mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0})_{\breve{\mathcal{T}}} - \mathbf{G}(\mathbf{v})_{\breve{\mathcal{T}}}\|_{\ell_2(\mathcal{L})}.$$

Using (3.5) and the fact that $\mathcal{T}_0 \subset \tilde{\mathcal{T}}_0$, the first summand can be estimated by $\|\mathbf{G}(\mathbf{v}) - \mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0})\|_{\ell_2(\mathcal{L})} \leq C(\|\mathbf{v}\|_{\ell_2(\mathcal{L})}) \cdot \varepsilon$. Because $\mathbf{v}$ restricted to $\breve{\mathcal{T}}$ has less non-trivial entries than $\mathbf{v}$ itself, we can immediately deduce $\|\mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0})_{\breve{\mathcal{T}}} - \mathbf{G}(\mathbf{v})_{\breve{\mathcal{T}}}\|_{\ell_2(\mathcal{L})} \leq C(\|\mathbf{v}\|_{\ell_2(\mathcal{L})}) \cdot \varepsilon$. It remains to estimate $\|\mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0}) - \mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0})_{\breve{\mathcal{T}}}\|_{\ell_2(\mathcal{L})}$. To this end, note first that for $v_{\tilde{\mathcal{T}}_0} := \mathbf{v}^T_{\tilde{\mathcal{T}}_0}\Psi$ Assumption 4.1 yields the estimate

$$2^{-2t|\lambda|}|\langle G(v_{\tilde{\mathcal{T}}_0}), \psi_\lambda\rangle|^2 \lesssim \sum_{\substack{\mu \in \tilde{\mathcal{T}}_0, \\ S_\mu \cap S_\lambda \neq \emptyset}} 2^{2t|\mu|}|\langle v, \tilde{\psi}_\mu\rangle|^2 2^{-2\gamma(|\lambda|-|\mu|)}. \tag{4.25}$$

Furthermore, by definition of Δ_j it follows

$$\begin{aligned}
\sum_{\mu \in \Delta_j} 2^{2t|\mu|}|\langle v, \tilde{\psi}_\mu\rangle|^2 &= \|\mathbf{v}_{\tilde{\mathcal{T}}_j} - \mathbf{v}_{\tilde{\mathcal{T}}_{j+1}}\|^2_{\ell_2(L)} \leq \|\mathbf{v}_{\mathcal{T}_j} - \mathbf{v}\|^2_{\ell_2(L)} + \|\mathbf{v} - \mathbf{v}_{\mathcal{T}_{j+1}}\|^2_{\ell_2(L)} \\
&= \sum_{i=1}^M \|\mathbf{v}_i - \mathbf{v}_{i,\mathcal{T}(\frac{2^j\varepsilon}{M(j+1)})}\|^2_{\ell_2(\{i\}\times K^\square)} + \sum_{i=1}^M \|\mathbf{v}_i - \mathbf{v}_{i,\mathcal{T}(\frac{2^{j+1}\varepsilon}{M(j+2)})}\|^2_{\ell_2(\{i\}\times K^\square)} \\
&\leq \frac{2^{2j}\varepsilon^2}{M(j+1)^2} + \frac{2^{2(j+1)}\varepsilon^2}{M(j+2)^2} \lesssim \frac{2^{2j}\varepsilon^2}{(j+1)^2}.
\end{aligned} \tag{4.26}$$

Using (4.25) and (4.26) we finally get

$$
\begin{aligned}
\|\mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0}) - \mathbf{G}(\mathbf{v}_{\tilde{\mathcal{T}}_0})_{\breve{\mathcal{T}}}\|^2_{\ell_2(L)} &= \sum_{\lambda\notin\breve{\mathcal{T}}} 2^{-2t|\lambda|}|\langle G(v_{\tilde{\mathcal{T}}}),\psi_\lambda\rangle|^2 \lesssim \sum_{\lambda\notin\breve{\mathcal{T}}} \sum_{\substack{\mu\in\tilde{\mathcal{T}}_0,\\ S_\mu\cap S_\lambda\neq\emptyset}} 2^{2t|\mu|}|\langle v,\tilde{\psi}_\mu\rangle|^2 \cdot 2^{-2\gamma(|\lambda|-|\mu|)} \\
\lesssim & \sum_{\mu\in\tilde{\mathcal{T}}_0} 2^{2t|\mu|}|\langle v,\tilde{\psi}_\mu\rangle|^2 \sum_{\substack{\lambda\notin\breve{\mathcal{T}},\\ S_\mu\cap S_\lambda\neq\emptyset}} 2^{-2\gamma(|\lambda|-|\mu|)} = \sum_{j\geq 0}\sum_{\mu\in\Delta_j} 2^{2t|\mu|}|\langle v,\tilde{\psi}_\mu\rangle|^2 \sum_{\substack{\lambda\notin\breve{\mathcal{T}},\\ S_\mu\cap S_\lambda\neq\emptyset}} 2^{-2\gamma(|\lambda|-|\mu|)} \\
\lesssim & \sum_{j\geq 0}\sum_{\mu\in\Delta_j} 2^{2t|\mu|}|\langle v,\tilde{\psi}_\mu\rangle|^2 \sum_{k\geq 0} 2^{-2\gamma(\alpha j+k)}\cdot 2^{d(\alpha j+k))} = \sum_{j\geq 0}\sum_{\mu\in\Delta_j} 2^{2t|\mu|}|\langle v,\tilde{\psi}_\mu\rangle|^2 \cdot 2^{(d-2\gamma)\alpha j}\sum_{k\geq 0} 2^{(d-2\gamma)k} \\
\lesssim & \sum_{j\geq 0} 2^{(d-2\gamma)\alpha j}\sum_{\mu\in\Delta_j} 2^{2t|\mu|}|\langle v,\tilde{\psi}_\mu\rangle|^2 \lesssim \sum_{j\geq 0} 2^{(d-2\gamma)\alpha j}\frac{2^{2j}\varepsilon^2}{(j+1)^2} = \sum_{j\geq 0}\frac{\varepsilon^2}{(j+1)^2} \lesssim \varepsilon^2.
\end{aligned}
$$

In order to proof (4.23) note that by definition the number of elements in $\Theta_{\varepsilon,\mu}$ can be bounded by

$$
\sum_{k=|\mu|}^{|\mu|+\alpha j} K2^{kd} \lesssim \sum_{k=|\mu|}^{|\mu|+\alpha j} (2^d)^k \leq (2^d)^{|\mu|+\alpha j} \lesssim 2^{d\alpha j},
$$

independent of $|\mu|$ due to the finite support of $\mathbf{v}$. Furthermore, by Lemma 4.4 and (4.7) we can infer $\#(\tilde{\mathcal{T}}_0) \lesssim \#(\mathcal{T}_0) \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\cdot\varepsilon^{-1/s}$ and $\#(\Delta_j) \leq \#(\tilde{\mathcal{T}}_j) \lesssim \#(\mathcal{T}_j) \lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}\cdot\varepsilon^{-1/s}\cdot 2^{-j/s}(j+1)^{1/s}$. Thus, we can deduce

$$
\begin{aligned}
\#(\breve{\mathcal{T}}) &\leq \#(I_{j_0}) + \#(\tilde{\mathcal{T}}_0) + \sum_{j\geq 0} 2^{d\alpha j}\#(\Delta_j) \\
&\lesssim \#(I_{j_0}) + \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\cdot\varepsilon^{-1/s} + \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\cdot\varepsilon^{-1/s}\left(\sum_{j\geq 0} 2^{d\alpha j - j/s}(j+1)^{1/s}\right) \\
&= \#(I_{j_0}) + \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\cdot\varepsilon^{-1/s}\left(1+\sum_{j\geq 0} 2^{(d\alpha-1/s)j}(j+1)^{1/s}\right) \lesssim \#(I_{j_0}) + \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}\cdot\varepsilon^{-1/s},
\end{aligned}
$$

where the last inequality holds if $d\alpha - \frac{1}{s} < 0$, which is equivalent to $0 < s < \frac{2\gamma-d}{2d}$. Using (4.23), inequality (4.24) is an immediate consequence of (4.22).
To estimate the number of used computations note that the construction of $\breve{\mathcal{T}}$ requires the calculation of the array $\tilde{\mathbf{v}}$, which can be determined from the leaves of $\bar{\mathcal{T}}(\mathbf{v})$ down to the roots. The number of operations for this task stays proportional to $\#\mathcal{T}(\mathbf{v})$. Because $\mathcal{T}_{j+1}\subset\mathcal{T}_j$, the previously computed aggregated tree $\mathcal{T}_{j+1}$ can be used to determine $\mathcal{T}_j$. The additional cost is bounded by $C\|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}\cdot\varepsilon^{-1/s}\cdot 2^{-j/s}(j+1)^{1/s}$. Summing up this from $j = 1,\ldots,J$ yields the last assertion. □

Note that $\tilde{\mathcal{T}}$ is usually not a complete tree by construction. Thus it may be necessary to 'complete' it adding all necessary reference cubes to $\tilde{\mathcal{T}}$. Because the number of siblings of a reference cube is uniformly bounded, this completion does not affect asymptotic results.

It remains to validate Assumption 4.1 for aggregated Gelfand frames, which is not a trivial task in general, because the properties of dual frames are often given only implicitly. Nonetheless, a positive result can be formulated for the frame pair $(\Psi,\tilde{\Psi})$, where $\Psi = \{\psi_\lambda\}_{\lambda\in L}$ is the aggregated Gelfand frame from (2.28) and $\tilde{\Psi} = \{\tilde{\psi}_\lambda\}_{\lambda\in L}$ is the non-canonical dual from (2.29).

Lemma 4.5 *Let $(\Psi, \tilde{\Psi})$ be the aggregated Gelfand frame from (2.28), the non-canonical dual frame from (2.29) respectively. Assume that the wavelets $\psi_\lambda \in \Psi$ belong to C^m and have m vanishing moments, i.e., are orthogonal to $\mathbb{P}_{m-1}$, the space of polynomials of total degree at most $m-1$. Then (4.19) holds for $\gamma := r + t + \frac{d}{2}$ with*

$$r := \begin{cases} \lceil \min\{m, p, n^*\} \rceil, & \text{if } t < \frac{d}{2} \text{ and } G(\cdot) \text{ satisfies (2.9) for } 0 \le p < p^* \\ \min\{m, n^*\}, & \text{if } t \ge \frac{d}{2} \text{ and } G(\cdot) \text{ satisfies (2.9) for some } p \ge 0 \end{cases} .$$

Proof: Again, the proof can be done following the lines of [7]. The crucial point in this proof is the inequality

$$\|v\|_{B^s_\infty(L_\infty(S_\lambda))} \lesssim \sup_{\mu: S_\mu \cap S_\lambda \neq \emptyset} \left(2^{(s+\frac{d}{2}-t)|\mu|} \cdot 2^{t|\mu|} |\langle v, \tilde{\xi}_\mu \rangle| \right), \tag{4.27}$$

which is true if $\Xi = \{\xi_\mu\}$ and $\tilde{\Xi} = \{\tilde{\xi}_\mu\}$ are a pair of biorthogonal wavelet Riesz bases and $\langle v, \tilde{\xi}_\mu \rangle$ are the unique expansion coefficients of v with respect to the dual basis, but it is not true for all dual frames in general. But the special structure of the dual frame (2.29) and a sufficient smooth partition of unity allow us to establish it again by

$$\begin{aligned} \|v\|_{B^s_\infty(L_\infty(S_\lambda))} &= \| \sum_{i=1}^{M} \sum_{\mu \in \{i\} \times K^\square} 2^{t|\mu|} \langle v, \sigma_i \tilde{\psi}_{i,\mu} \rangle \cdot 2^{-t|\mu|} \psi_{i,\mu} \|_{B^s_\infty(L_\infty(S_\lambda))} \\ \le \quad & \sum_{i=1}^{M} \| \sum_{\mu \in \{i\} \times K^\square} 2^{t|\mu|} \langle \sigma_i v, \tilde{\psi}_{i,\mu} \rangle \cdot 2^{-t|\mu|} \psi_{i,\mu} \|_{B^s_\infty(L_\infty(S_\lambda))} \\ \lesssim \quad & \sum_{i=1}^{M} \sup_{\substack{\mu \in \{i\} \times K^\square \\ S_\mu \cap S_\lambda \neq \emptyset}} \left(2^{(s+\frac{d}{2}-t)|\mu|} \cdot 2^{t|\mu|} |\langle \sigma_i v, \tilde{\psi}_{i,\mu} \rangle| \right) \lesssim \sup_{\substack{\mu \in L \\ S_\mu \cap S_\lambda \neq \emptyset}} \left(2^{(s+\frac{d}{2}-t)|\mu|} \cdot 2^{t|\mu|} |\langle v, \sigma_i \tilde{\psi}_{i,\mu} \rangle| \right) \end{aligned} \tag{4.28}$$

To establish (4.28), we used (4.27) for the local dual Riesz bases $\tilde{\Psi}_i$ from (2.26), so that the suprema are taken over finite numbers of expansion coefficients. $\square$

5 Evaluation of Nonlinear Expressions II: Tree-wise calculation of wavelet coefficients

As the main result of the previous section, we haven given a satisfactory reply to the question how to implement the prediction step of our problem in the case of aggregated Gelfand frames. In particular, for any given target accuracy $\varepsilon > 0$ we are able to determine an aggregated tree-structured index set $\mathcal{T} \subset L$ so that $\|\mathbf{G}(\mathbf{v}) - \mathbf{G}(\mathbf{v})_{\mathcal{T}}\| \lesssim \varepsilon$. Thus, we turn now to the second part of the problem, that is the numerical approximation of the coefficients in $\mathbf{G}(\mathbf{v})_{\mathcal{T}}$. Hence, the task is the calculation of a vector $\mathbf{w} \in \ell_2(L)$ with $\mathrm{supp}(\mathbf{w}) \subset \mathcal{T}$ and $\|\mathbf{G}(\mathbf{v})_{\mathcal{T}} - \mathbf{w}\|_{\ell_2(L)} \lesssim \varepsilon$. This means, the approximation error $\|\mathbf{G}(\mathbf{v})_{\mathcal{T}} - \mathbf{w}\|_{\ell_2(L)}$ stays (up to constants) the same size as the prediction error, such that we finally have an overall *evaluation error* of $\|\mathbf{G}(\mathbf{v}) - \mathbf{w}\|_{\ell_2(L)} \lesssim \varepsilon$.

This problem has been studied (and solved) by several authors [1, 18] for a (single) pair of biorthogonal wavelet Riesz bases $(\Psi, \tilde{\Psi})$. As the main result of this section we show that it is possible to exploit the structure of the aggregated Gelfand frame (2.28) and its non-canonical dual frame (2.29) to convey the so called RECOVER algorithm from [1] to the case of aggregated Gelfand frames. This algorithm uses quadrature to approximate certain scaling function coefficients $\langle \phi_\lambda, G(v) \rangle$, which in turn allow the approximation of $\mathbf{G}(\mathbf{v})_{\mathcal{T}}$. Of course, this quadrature will be realized inexactly in general and we have to look carefully to the consequences of quadrature errors for the approximation. In [1] two error estimates were given depending on whether the underlying function space is $L_2(\Omega)$ or $H^{-t}(\Omega)$. We show that both error can be transferred to the Gelfand frame situation with little modifications.

5.1 The RECOVER algorithm

In this subsection we describe the idea of the RECOVER algorithm in the Riesz basis setting and focus on the case $\mathcal{H} = L_2(\Omega)$ first. To this end, suppose we have a tree $\mathcal{T} \subset \Lambda$ containing the significant expansion coefficients of $G(v)$ concerning the Riesz basis $\Psi = \{\psi_\lambda\}_{\lambda\in\Lambda}$, i.e., we take $\sum_{\lambda\in\mathcal{T}} \langle\psi_\lambda, G(v)\rangle\tilde{\psi}_\lambda$ as approximation to $G(v)$ and assume $\langle\psi_\lambda, G(v)\rangle = 0$ outside $\mathcal{T}$. We shall abbreviate in the following

$$c_{j,k} := \langle\phi_{j,k}, G(v)\rangle, \quad (j,k) \in I_j, \qquad d_\lambda := \langle\psi_\lambda, G(v)\rangle, \quad \lambda \in J.$$

Note that the aforementioned assumption then means $d_\lambda = 0$ outside $\mathcal{T}$. For any subset S of I_j or J we will write

$$\mathbf{c}_j(S) := (c_{j,k} : (j,k) \in S), \qquad \mathbf{d}(S) := (d_\lambda : \lambda \in S),$$

where we simply set $\mathbf{c}_j := \mathbf{c}_j(I_j)$ and $\mathbf{d}_j = \mathbf{d}(J_j)$. Furthermore, let $t_j := \mathcal{T} \cap J_j$ denote the index set on the j-th level of the tree.

The starting point for the following considerations is the equation $\tilde{\phi}_{j+1}^T \mathbf{c}_{j+1} = \tilde{\phi}_j^T \mathbf{c}_j + \tilde{\Psi}_j^T \mathbf{d}_j$, which is a direct consequence of (2.22). Using $\tilde{\phi}_j^T = \tilde{\phi}_{j+1}^T M_{j,\tilde{\phi}}$ and $\tilde{\Psi}_j^T = \tilde{\phi}_{j+1}^T M_{j,\tilde{\Psi}}$ and the biorthogonality of $\Psi, \tilde{\Psi}$ we can infer that

$$\mathbf{c}_{j+1} - M_{j,\tilde{\phi}}\mathbf{c}_j = M_{j,\tilde{\Psi}}\mathbf{d}_j = M_{j,\tilde{\Psi}}\mathbf{d}_j(t_j). \tag{5.1}$$

This means, the vector $\mathbf{c}_{j+1} - M_{j,\tilde{\phi}}\mathbf{c}_j$ vanishes outside the finite union of columns of $M_{j,\tilde{\Psi}}$ selected by t_j. By the uniform sparsity of $M_{j,\tilde{\Psi}}$ we conclude that a finite number of coefficients of $\mathbf{c}_j$ and $\mathbf{c}_{j+1}$ suffices for an exact representation of $\mathbf{c}_{j+1} - M_{j,\tilde{\phi}}\mathbf{c}_j$. Developing this idea the following results from [1] can be formulated.

Lemma 5.1 *Let $M_{j,\tilde{\psi}}^{|\lambda}$ and $M_{j,\tilde{\phi}}^{|\lambda}$ denote the λ-th column of $M_{j,\tilde{\psi}}$ and $M_{j,\tilde{\phi}}$ respectively. Defining*

$$H_{j+1} := \bigcup_{\lambda:\lambda\in t_j} \operatorname{supp} M_{j,\tilde{\psi}}^{|\lambda} \cup \bigcup_{\lambda:\lambda\in t_j^\circ} \operatorname{supp} M_{j,\tilde{\phi}}^{|\lambda}, \tag{5.2}$$

the two-scale relation $\tilde{\phi}_{j+1}^T \mathbf{c}_{j+1} = \tilde{\phi}_j^T \mathbf{c}_j + \tilde{\Psi}_j^T \mathbf{d}_j$ from (2.22) remains valid in the form

$$\tilde{\phi}_{j+1}^T \left(\mathbf{c}_{j+1}(H_{j+1}) - (M_{j,\tilde{\phi}}\mathbf{c}_j(I_j \setminus t_j^\circ))(H_{j+1})\right) = \tilde{\phi}_j^T \mathbf{c}_j(t_j^\circ) + \tilde{\Psi}_j^T \mathbf{d}_j(t_j). \tag{5.3}$$

Furthermore, applying $G_{j,\tilde{\phi}}$ to the vector $\mathbf{c}_{j+1}(H_{j+1}) - (M_{j,\tilde{\phi}}\mathbf{c}_j(I_j \setminus t_j^\circ))(H_{j+1})$ yields the coefficient array $\mathbf{c}_j(t_j^\circ)$ and applying $G_{j,\tilde{\psi}}$ to the vector $\mathbf{c}_{j+1}(H_{j+1}) - (M_{j,\tilde{\phi}}\mathbf{c}_j(I_j \setminus t_j^\circ))(H_{j+1})$ yields the coefficient array $\mathbf{d}_j(t_j)$. Moreover, in (5.3) and the above calculations we do not need the complete vector $\mathbf{c}_j(I_j \setminus t_j^\circ))$, but only $\mathbf{c}_j(H_j \setminus t_j^\circ))$ because the latter contains all information needed for the next lower level.

Lemma 5.1 immediately brings out the core ingredient of the RECOVER scheme: It suffices to approximate the vector $\mathbf{c}_{j+1}(H_{j+1}) - (M_{j,\tilde{\phi}}\mathbf{c}_j(H_j \setminus t_j^\circ))(H_{j+1})$ to get $\mathbf{d}_j(t_j)$, where $\mathbf{c}_{j+1}(t_{j+1}^\circ)$ itself can be determined from the next higher level by $G_{j+1,\tilde{\phi}}(\mathbf{c}_{j+2}(H_{j+2}) - (M_{j+1,\tilde{\phi}}\mathbf{c}_{j+1}(H_{j+1} \setminus t_{j+1}^\circ))(H_{j+2}))$. By this, only the scaling function coefficients coefficients $\mathbf{c}_j(H_j \setminus t_j^\circ)$ outside the tree must be approximated by quadrature, while all wavelet coefficients inside the tree can be calculated by the above decomposition. Starting now with the highest level j^* so that $t_j = \emptyset$ for all $j > j^*$ (which exists because $\mathcal{T}$ is a finite index set), we can calculate $\mathbf{d}_j$ for all $j^*, j^* - 1, \ldots, j_0$ from top to bottom. Additionally, at the bottom we can also calculate $\mathbf{c}_{j_0}$, the scaling function coefficients for the coarsest level. Note that this procedure uses only Lemma 5.1 and its consequences from the two-scale relation (2.22). This allows the application of RECOVER to the local Riesz bases $(\Psi^{(i)}, \tilde{\Psi}^{(i)})$ from (2.25) and (2.26) respectively and by

$$\langle G(v), E_i\psi_{i,\lambda}\rangle = \langle E_i^* G(v), \psi_{i,\lambda}\rangle,$$

where $E_i^* : H^{-t}(\Omega) \to H^{-t}(\Omega_i)$ is the adjoint operator of E_i, we can use RECOVER for the calculation of the frame expansion coefficients $\langle G(v), \psi_\lambda\rangle$ concerning the aggregated Gelfand frame $(\Psi, \tilde{\Psi})$ from (2.28) and (2.29).

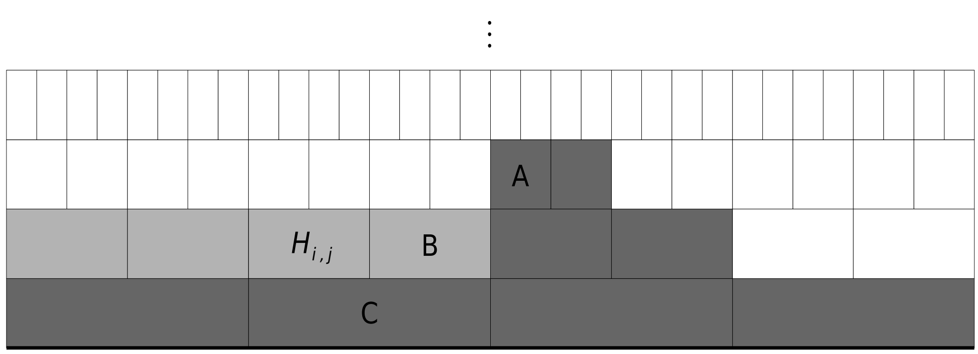
Before we formulate the basic algorithm of RECOVER for aggregated Gelfand frames of the form (2.28) we make a structural assumption on the underlying aggregated tree $\mathcal{T}$, which does not only relieve the algorithm but is also needed for an effective error analysis.

Assumption 5.1 *We assume the aggregated tree $\mathcal{T} = \cup_{i=1}^M \mathcal{T}_i$ to be* well-graded*, i.e., we assume all local trees $\mathcal{T}_i, i = 1, \ldots, M$, to be well-graded in the sense of [1], i.e.,*

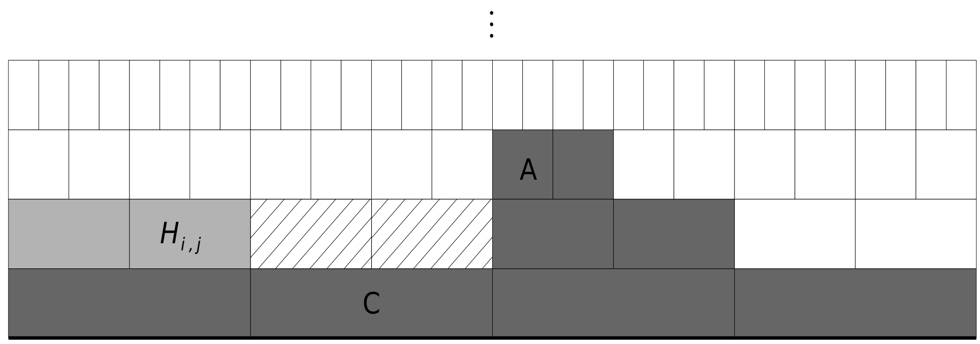
$$\left(\bigcup_{\lambda:\lambda^\circ \in H_{i,j} \setminus t_{i,j}^\circ} M_{i,j,\tilde{\phi}}^{|\lambda} \right) \cap t_{i,j+1}^\circ = \emptyset \quad \textit{for all } i = 1, \ldots, M. \tag{5.4}$$

Here, $M_{i,j,\tilde{\phi}}$ and $H_{i,j}$ are defined analogously to (2.20) and (5.2) for the local Riesz bases. This means, no scaling functions from level j of $\mathcal{T}_i$ are needed to represent the scaling functions in $H_{i,j} \setminus t_{i,j}^\circ$ (outside the tree) by the two-scale relation.

Any (local) tree $\mathcal{T}$ satisfying the expansion property can be transformed into a well-graded tree by at most a finite uniformly bounded number of subdivisions of the leaves, see [1, 7] for further details. By this, assuming a well-graded aggregated tree in the following will not affect asymptotic results. It is intuitively clear that this process 'broadens' the tree, which will be an important aspect in the error estimation later. For an illustration recall the tree example from Section 4.



If we assume $\square_{\lambda^\circ} \subsetneq \psi_\lambda$, the scaling functions belonging to the reference cubes A and B will overlap, which contradicts (5.4). This can be avoided by adding B (and an adequate number of neighbors of B, depending on the used wavelet basis) to the tree, i.e., by subdividing leaf C.



With this assumption consider the following basic RECOVER algorithm for the approximation of $(\langle \psi_\lambda, G(v)\rangle)_{\lambda\in\mathcal{T}}$.

RECOVER: $(G(v),\mathcal{T}) \to \mathbf{d}^R(\mathcal{T})$

For $i = 1,\dots,M$ do
    Determine the minimal $j_i^* \in \mathbb{N}$ so that $t_{i,j} = \emptyset$ for all $j > j_i^*$.
    Determine $H_{j_i^*+1}$ and approximate (by quadrature) the array $\mathbf{c}_{j_i^*+1}(H_{j_i^*+1})$.
    Call this approximation $\mathbf{q}_{j_i^*+1}$. Let $\hat{\mathbf{c}}_{j_i^*+1} := \mathbf{0}$.
    For $j = j_i^*, j_i^*-1,\dots,j_0$ do
        Determine $H_{i,j} \setminus t_{i,j}^\circ$ and $\mathbf{q}_{i,j}(H_{i,j} \setminus t_{i,j}^\circ)$ by quadrature serving as approximation for $\mathbf{c}_{i,j}(H_{i,j} \setminus t_{i,j}^\circ)$.
        Set

$$\begin{aligned}\bar{\mathbf{c}}_{i,j+1} &:= \hat{\mathbf{c}}_{i,j+1} + \mathbf{q}_{i,j+1}(H_{i,j+1}\setminus t_{i,j+1}^\circ) - (M_{j,\tilde{\phi}}\mathbf{q}_{i,j}(H_{i,j}\setminus t_{i,j}^\circ))(H_{i,j+1}\setminus t_{i,j+1}^\circ)\\ \hat{\mathbf{c}}_{i,j} &:= G_{i,j,\tilde{\phi}}\bar{\mathbf{c}}_{i,j+1}\\ \mathbf{d}_{i,j}^R &:= G_{i,j,\tilde{\Psi}}\bar{\mathbf{c}}_{i,j+1}\end{aligned}$$

    od
    Set $\mathbf{d}_{i,j_0-1}^R := \hat{\mathbf{c}}_{i,j_0}$.
    Set $\mathbf{d}_i^R := \bigcup_{j=j_0}^{j_i^*} \mathbf{d}_{i,j}^R$.
od
Set $\mathbf{d}^R := \bigcup_{i=1}^M \mathbf{d}_i^R$.

The following remark is a direct consequence of the results in [1].

Remark 5.1 *The RECOVER algorithm requires computing the approximations $q_{i,j,k}$ to $c_{i,j,k}$ only on the set*

$$H^- := \bigcup_{i=1}^{M} \bigcup_{j=j_0}^{j_i^*+1} H_{i,j}^-. \tag{5.5}$$

Furthermore, it holds $\#H^- \lesssim \#\mathcal{T}$. Thus, whenever the quadrature requires at most a constant cost per entry, the computational effort for RECOVER stays proportional to $\#\mathcal{T}$.

Note that in the RECOVER algorithm there is no intersection on processes of different patches. Hence, the above algorithm can be completely parallelized. A second note concerns the possibility of tresholding the resulting coefficient arrays. Because the wavelet coefficients in $\mathbf{d}_{i,j}^R$ are not needed for further calculations they can be tresholded on the fly during the algorithm.

5.2 Error analysis

Of course, the resulting coefficient vectors $\mathbf{d}^R_{j_0-1} := \bigcup_{i=1}^M \mathbf{d}^R_{i,j_0-1}$ and $\mathbf{d}^R$ from RECOVER will generally not coincide with the exact vectors $\mathbf{c}_{j_0}$ and $\mathbf{d}$. In this chapter we will give two error estimates for $\|\mathbf{c}_{j_0} - \mathbf{d}^R_{j_0-1}\|^2_{\ell_2(L)} + \|\mathbf{d}(\mathcal{T}) - \mathbf{d}^R(\mathcal{T})\|^2_{\ell_2(L)}$ and its rescaled version respectively depending on the underlying function space. Therefore, we have to take carefully into account that the quadrature error is caused outside the tree and then transferred into it during the algorithm.

For $(i, \lambda^\circ) = (i, j, k)$ we will abbreviate the quadrature error $q_{i,j,k} - c_{i,j,k}$ by

$$e_{i,\lambda^\circ} := e_{i,j,k} := q_{i,j,k} - c_{i,j,k} \tag{5.6}$$

in the following. The corresponding vectors $\mathbf{e}_{i,\lambda^\circ} := (e_{i,\lambda^\circ})_{\lambda \in H_{i,j} \setminus t^\circ_{i,j}}$ are always supported in $H^-_{i,j} := H_{i,j} \setminus t^\circ_{i,j}$ with $H^-_{i,j^*_i+1} := H_{i,j^*_i+1}$ for the highest level needed to generate $\mathbf{d}(\mathcal{T}_i)$. Summing up over all local trees $\mathcal{T}_i$, $i = 1, \ldots, M$, and all levels $j_0, \ldots, j^*_i$ the RECOVER scheme can be interpreted to produce the approximation

$$P^R_{\mathcal{T}}(G(v)) := \sum_{i=1}^M \sum_{j=j_0+1}^{j^*_i+1} \tilde{\phi}^T_{i,j} \left(\mathbf{q}_{i,j} - M_{i,j-1,\tilde{\phi}} \mathbf{q}_{i,j-1}(H^-_{i,j-1}) \right) (H^-_{i,j}),$$

from which we get the error representation

$$E_{\mathcal{T}} := \sum_{i=1}^M E_{\mathcal{T}_i} := \sum_{i=1}^M \sum_{j=j_0+1}^{j^*_i} \tilde{\phi}^T_{i,j} \left(\mathbf{e}_{i,j} - M_{i,j-1,\tilde{\phi}} \mathbf{e}_{i,j-1}(H^-_{i,j-1}) \right) (H^-_{i,j}). \tag{5.7}$$

Measuring the error $E_{\mathcal{T}}$ in $L_2(\Omega)$ first allows the following error estimation.

Lemma 5.2 *There exists a constant C so that the computed arrays $\mathbf{d}^R$ and $\mathbf{d}^R_{j_0-1}$ satisfy*

$$\|\mathbf{c}_{j_0} - \mathbf{d}^R_{j_0-1}\|^2_{\ell_2(L)} + \|\mathbf{d}(\mathcal{T}) - \mathbf{d}^R(\mathcal{T})\|^2_{\ell_2(L)} \le C \cdot \sum_{i=1}^M \sum_{j=j_0}^{j^*_i+1} \|\mathbf{e}_{i,j}(H^-_{i,j})\|^2_{\ell_2(\{i\} \times K^\square)}. \tag{5.8}$$

Proof: The proof is analogous to the proof in [1] for Riesz bases. Nevertheless we give a short view on the main steps of the proof in order to show that the overlapping decomposition $\{\Omega_i\}_{i=1}^M$ of Ω causes no problems. It is

$$\begin{aligned} & \|\mathbf{c}_{j_0} - \mathbf{d}^R_{j_0-1}\|^2_{\ell_2(L)} + \|\mathbf{d}(\mathcal{T}) - \mathbf{d}^R(\mathcal{T})\|^2_{\ell_2(L)} \\ = \; & \sum_{i=1}^M \|\mathbf{c}_{i,j_0} - \mathbf{d}^R_{i,j_0-1}\|^2_{\ell_2(\{i\}\times K^\square)} + \sum_{i=1}^M \|\mathbf{d}(\mathcal{T}_i) - \mathbf{d}^R(\mathcal{T}_i)\|^2_{\ell_2(\{i\}\times K^\square)} \lesssim \sum_{i=1}^M \|E_{\mathcal{T}_i}\|^2_{L_2(\Omega_i)}, \end{aligned}$$

where we have used (2.16). Next, we use the fact that the reference cubes of the outer leaves $\mathcal{D}(\mathcal{T}^\circ_i)$ of $\mathcal{T}_i$ form a non-overlapping (aside from sets with zero measure) decomposition of Ω_i. With this we can go on

$$\begin{aligned} \sum_{i=1}^M \|E_{\mathcal{T}_i}\|^2_{L_2(\Omega)_i} &= \sum_{i=1}^M \sum_{\lambda^\circ \in \mathcal{D}(\mathcal{T}_i)} \|E_{\mathcal{T}_i}\|^2_{L_2(\square_{\lambda^\circ})} \\ &\le \sum_{i=1}^M \sum_{\lambda^\circ \in \mathcal{D}(\mathcal{T}_i)} \sum_{j=j_0+1}^{j^*_i+1} \|\tilde{\phi}^T_{i,j} \left(\mathbf{e}_{i,j} - M_{i,j-1,\tilde{\phi}} \mathbf{e}_{i,j-1}(H^-_{i,j-1}) \right) (H^-_{i,j})\|^2_{L_2(\square_{\lambda^\circ})} \end{aligned}$$

The well-gradedness of the local trees $\mathcal{T}_i$ and the finite number of patches imply that $\|\tilde{\phi}^T_{i,j} \left(\mathbf{e}_{i,j} - M_{i,j-1,\tilde{\phi}} \mathbf{e}_{i,j-1}(H^-_{i,j-1}) \right) (H^-_{i,j})\|^2_{L_2(\square_{\lambda^\circ})} \neq 0$ holds only for a uniformly bounded number of levels. Using this and the fact that $M_{i,j,\tilde{\phi}}$ defines a bounded mapping it follows

$$\sum_{i=1}^{M} \|E_{\mathcal{T}_i}\|^2_{L_2(\Omega_i)} \lesssim \sum_{i=1}^{M} \sum_{j=j_0+1}^{j_i^*+1} \|\mathbf{e}_{i,j}(H^-_{i,j})\|^2_{\ell_2(\{i\}\times K^\square)}.$$

□

However, as noted earlier, the relevant function space for adaptive schemes is most often not $L_2(\Omega)$, but the Sobolev space $H^t_0(\Omega), t > 0$. Thus the error concerning the rescaled dual frame elements will be measured in the negative Sobolev norm $\|\cdot\|_{H^{-t}}$. In this situation we can still use the RECOVER scheme to approximate the rescaled coefficients $\left(2^{-t|\lambda|}\langle G(v), \psi_\lambda\rangle\right)_{\lambda\in L}$, since the scaling can be directly included in the algorithm using the diagonal matrix D with $(D)_{\lambda,\lambda} := 2^{|\lambda|}$:

$$\mathbf{g}^R_{i,j} := D^{-t}\mathbf{d}^R_{i,j} = D^{-t} G_{i,j,\tilde{\Psi}} \bar{\mathbf{c}}_{i,j+1}. \tag{5.9}$$

Our goal is to estimate $\|\mathbf{g}_\mathcal{T} - \mathbf{g}^R_\mathcal{T}\|_{\ell_2(L)}$ and with this finally $\|\mathbf{g} - \mathbf{g}^R_\mathcal{T}\|_{\ell_2(L)}$. The native idea to do this is to use the arguments from Lemma 5.2 and relate the coefficient error not to the error in L_2 but H^{-t}. But, estimating the error in H^{-t} one is faced with a principal problem, which can interpreted as inherent missing locality, since the negative Sobolev norms are not localized in a simple way. Thus this intuitive idea fails and it is less obvious how to estimate the resulting coefficient array in a proper way, see [1] for a detailed derivation. For the following basic error estimation we will use the localization of the primal norm $\|\cdot\|_{H^t}$, i.e., restrictions of $f \in H^t_0(\Omega)$ to subdomains $\Omega' \subset \Omega$ (of suitable regularity) belong to a localized version $H^t_0(\Omega')$ and for any partition P of Ω into subdomains Δ we have

$$\sum_{\Delta\in P} \|f\|^2_{H^t(\Delta)} \lesssim \|f\|^2_{H^t(\Omega)}, \quad f \in H^t_0(\Omega). \tag{5.10}$$

Using the following definitions we can give some general estimate for the error in the dual norm.

Definition 5.1 *Let $\Omega_{i,j}$ be the joint support of the dual wavelets from $H^-_{i,j}$, i.e., $\Omega_{i,j} := \bigcup_{(j,k)\in H^-_{i,j}} \tilde{S}_{i,j,k} := \bigcup_{(j,k)\in H^-_{i,j}} \operatorname{supp} \tilde{\psi}_{i,j,k}$. Furthermore, consider the following operators*

$$\begin{aligned} P_{i,j}(f) &:= P_{i,j}(f, H^-_{i,j}) := \sum_{(j,k)\in H^-_{i,j}} \langle \phi_{i,j,k}, f\rangle \tilde{\phi}_{i,j,k}, \\ L_{i,j}(f) &:= L_{i,j}(f, H^-_{i,j}) := \sum_{(j,k)\in H^-_{i,j}} q_{i,j,k} \tilde{\phi}_{i,j,k}. \end{aligned}$$

With $P^*_{i,j}(\cdot)$ we will denote the adjoint operator of $P_{i,j}(\cdot)$, and by

$$\|f\|_{\tilde{H}^{-t}(\Omega')} := \sup_{w\in \tilde{H}^t_0(\Omega')} \frac{\langle w, f\rangle_{L_2(\Omega')}}{\|w\|_{H^t(\Omega')}}, \tag{5.11}$$

where $\tilde{H}^t_0(\Omega')$ consists for any $\Omega' \subset \Omega$ of those $w \in H^t_0(\Omega')$, whose zero extension to Ω is still in $H^t_0(\Omega)$.

Lemma 5.3 *The error $E_\mathcal{T}$ can be estimated by*

$$\|E_\mathcal{T}\|_{H^{-t}(\Omega)} \lesssim \left(\sum_{i=1}^{M} \sum_{j=j_0}^{j_i^*+1} \|\tilde{\phi}^T_{i,j}\mathbf{e}_{i,j}\|^2_{\tilde{H}^{-t}(\Omega_{i,j})}\right)^{1/2} \tag{5.12}$$

$$\|E_\mathcal{T}\|_{H^{-t}(\Omega)} \lesssim \left(\sum_{i=1}^{M} \sum_{j=j_0}^{j_i^*+1} \|L_{i,j}G(v) - P_{i,j}G(v)\|^2_{\tilde{H}^{-t}(\Omega_{i,j})}\right)^{1/2} \tag{5.13}$$

Proof: It is

$$\begin{aligned}
\|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} &= \sup_{f\in H_0^t(\Omega)} \frac{\langle E_{\mathcal{T}}, f\rangle_{H^{-t}(\Omega)\times H_0^t(\Omega)}}{\|f\|_{H^t(\Omega)}} \\
&= \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1} \langle \tilde{\phi}_{i,j}^T(\mathbf{e}_{i,j} - M_{i,j-1,\tilde{\phi}}\mathbf{e}_{i,j-1}(H_{i,j-1}^-))(H_{i,j}^-), f\rangle_{L_2(\Omega_{i,j})}}{\|f\|_{H^t(\Omega)}} \\
&= \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1} \langle \tilde{\phi}_{i,j}^T(\mathbf{e}_{i,j} - M_{i,j-1,\tilde{\phi}}\mathbf{e}_{i,j-1}(H_{i,j-1}^-))(H_{i,j}^-), P_{i,j}^*(f, H_{i,j}^-)\rangle_{L_2(\Omega_{i,j})}}{\|f\|_{H^t(\Omega)}},
\end{aligned}$$

where we have used that by the biorthogonality of $\phi_i, \tilde{\phi}_i, i = 1, \ldots, M$, it holds $P_{i,j}(E_{\mathcal{T}}) = E_{\mathcal{T}}$. Because $P_{i,j}^*(f, H_{i,j}^-)$ and $\mathbf{e}_{i,j-1}$ are by definition supported in $H_{i,j}^-$ and $H_{i,j-1}^-$ respectively, it follows

$$\begin{aligned}
\|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} &= \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1} \langle \tilde{\phi}_{i,j}^T(\mathbf{e}_{i,j} - M_{i,j-1,\tilde{\phi}}\cdot\mathbf{e}_{i,j-1}(H_{i,j-1}^-)), P_{i,j}^*(f, H_{i,j}^-)\rangle_{L_2(\Omega_{i,j})}}{\|f\|_{H^t(\Omega)}} \\
&= \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1} \langle \tilde{\phi}_{i,j}^T\mathbf{e}_{i,j} - \tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}, P_{i,j}^*(f, H_{i,j}^-)\rangle_{\Omega_{i,j}}}{\|f\|_{H^t(\Omega)}} \\
&\leq \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1} \|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j} - \tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}\|_{\tilde{H}^{-t}(\Omega_{i,j})} \cdot \|P_{i,j}^*(f, H_{i,j}^-)\|_{H^t(\Omega_{i,j})}}{\|f\|_{H^t(\Omega)}},
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality in the last step. Using the uniform boundedness of $P_{i,j}^*(\cdot)$, which is an implication from the boundedness of F from (2.15) and from (2.16), and Cauchy-Schwarz it follows

$$\begin{aligned}
\|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} &\lesssim \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1} \|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j} - \tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}\|_{\tilde{H}^{-t}(\Omega_{i,j})} \cdot \|f\|_{H^t(\Omega_{i,j})}}{\|f\|_{H^t(\Omega)}} \\
&\leq \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\left(\sum_{j=j_0}^{j_i^*+1} \|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j} - \tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}\|_{\tilde{H}^{-t}(\Omega_{i,j})}^2\right)^{1/2} \cdot \left(\sum_{j=j_0}^{j_i^*+1} \|f\|_{H^t(\Omega_{i,j})}^2\right)^{1/2}}{\|f\|_{H^t(\Omega)}} \\
&\lesssim \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\left(\sum_{j=j_0}^{j_i^*+1} \|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j} - \tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}\|_{\tilde{H}^{-t}(\Omega_{i,j})}^2\right)^{1/2} \cdot \|f\|_{H^t(\Omega_i)}}{\|f\|_{H^t(\Omega)}},
\end{aligned}$$

where we have also made use of (5.10). Using now

$$\|\tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}\|_{H^{-t}(\Omega_{i,j})} = \|\tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}\|_{H^{-t}(\Omega_{i,j}\cap\Omega_{i,j-1})} \leq \|\tilde{\phi}_{i,j-1}^T\mathbf{e}_{i,j-1}\|_{H^{-t}(\Omega_{i,j-1})},$$

which is true because the dual norm is monotone in the support size and a tree gets 'leaner' with growing scale, we get

$$
\begin{aligned}
\|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} &\lesssim \sup_{f\in H_0^t(\Omega)} \frac{\sum_{i=1}^{M}\left(\sum_{j=j_0}^{j_i^*+1}\|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j}\|^2_{\tilde{H}^{-t}(\Omega_{i,j})}\right)^{1/2}\cdot\|f\|_{H^t(\Omega_i)}}{\|f\|_{H^t(\Omega)}}\\
&\leq \sup_{f\in H_0^t(\Omega)} \frac{\left(\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1}\|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j}\|^2_{\tilde{H}^{-t}(\Omega_{i,j})}\right)^{1/2}\cdot\left(\sum_{i=1}^{M}\|f\|^2_{H^t(\Omega_i)}\right)^{1/2}}{\|f\|_{H^t(\Omega)}}\\
&\lesssim \left(\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1}\|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j}\|^2_{\tilde{H}^{-t}(\Omega_{i,j})}\right)^{1/2}.
\end{aligned}
$$

To show the second estimate, note that $e_{i,j,k} = q_{i,j,k} - c_{i,j,k} = \langle\phi_{i,j,k}, L_{i,j}(G(v)) - P_{i,j}(G(v))\rangle$, so that by the biorthogonality of the local scaling functions it holds $\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j} = L_{i,j}(G(v)) - P_{i,j}(G(v))$. Thus, we can infer

$$
\begin{aligned}
\|\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j}\|_{\tilde{H}^{-t}(\Omega_{i,j})} &= \sup_{w\in H^t(\Omega_{i,j})}\frac{\langle\tilde{\phi}_{i,j}^T\mathbf{e}_{i,j}, P_{i,j}^*(w)\rangle_{L_2(\Omega_{i,j})}}{\|w\|_{H^t(\Omega_{i,j})}} = \sup_{w\in H^t(\Omega_{i,j})}\frac{\langle L_{i,j}(G(v)) - P_{i,j}(G(v)), P_{i,j}^*(w)\rangle_{L_2(\Omega_{i,j})}}{\|w\|_{H^t(\Omega_{i,j})}}\\
&\lesssim \sup_{w\in H^t(\Omega_{i,j})}\frac{\|L_{i,j}(G(v)) - P_{i,j}(G(v))\|_{\tilde{H}^{-t}(\Omega_{i,j})}\cdot\|w\|_{H^t(\Omega_{i,j})}}{\|w\|_{H^t(\Omega)_{i,j}}}\\
&= \|L_{i,j}(G(v)) - P_{i,j}(G(v))\|_{\tilde{H}^{-t}(\Omega_{i,j})}.
\end{aligned}
$$

Thus, we finally have $\|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} \lesssim \left(\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1}\|L_{i,j}(G(v)) - P_{i,j}(G(v))\|^2_{\tilde{H}^{-t}(\Omega_{i,j})}\right)^{1/2}$. □

In order to convey an estimate for $\|\mathbf{g} - \mathbf{g}_{\mathcal{T}}\|_{\ell_2(L)}$ from Lemma 5.3 we need the following assumption.

Assumption 5.2 *For all $i = 1,\ldots,M$ and all $j = j_0,\ldots,j_i^*+1$ it holds*

$$
V := V(\mathcal{T}) := \{\lambda\in\mathcal{T} : |\lambda| > j, \tilde{S}_\lambda\cap\Omega_{i,j}\neq\emptyset\} = \emptyset. \tag{5.14}
$$

Thus, the support of the used scaling functions outside the aggregated tree does not intersect with wavelets from higher levels in the tree. Using this assumption the general error estimate from Lemma 5.3 yields the aimed error estimation for $\|\mathbf{g} - \mathbf{g}_{\mathcal{T}}\|_{\ell_2(L)}$.

Lemma 5.4 *Assume that Assumption 5.2 holds. Then, Lemma 5.3 yields*

$$
\|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} \lesssim \|\mathbf{g} - \mathbf{g}_{\mathcal{T}}\|_{\ell_2(L)}. \tag{5.15}
$$

By (5.7), we can immediately deduce $\|\mathbf{g}_{\mathcal{T}} - \mathbf{g}_{\mathcal{T}}^R\|_{\ell_2(L)} \lesssim \|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} \lesssim \|\mathbf{g} - \mathbf{g}_{\mathcal{T}}\|_{\ell_2(L)}$ and thus $\|\mathbf{g} - \mathbf{g}_{\mathcal{T}}^R\|_{\ell_2(L)} \lesssim \|\mathbf{g} - \mathbf{g}_{\mathcal{T}}\|_{\ell_2(L)}$.

Proof: By the boundedness of $\tilde{F}^*$ from (2.15) we have $\|g\|^2_{\tilde{H}^{-t}(\Omega')} \lesssim \sum_{\lambda:\tilde{S}_\lambda\cap\Omega'\neq\emptyset} 2^{2t|\lambda|}|\langle\psi_\lambda, g\rangle|^2$ for all $\Omega'\subset\Omega, g\in\tilde{H}^{-t}(\Omega')$. Using this for $g - P_{i,j}(g)$ gives $\|L_{i,j}(G(v)) - P_{i,j}(G(v))\|_{\tilde{H}^{-t}(\Omega_{i,j})} \lesssim \sum_{\substack{|\lambda|>j\\ \tilde{S}_\lambda\cap\Omega_{i,j}\neq\emptyset}} |g_\lambda|^2$, from which it follows by Assumption 5.2, the finite number of patches and the finiteness of $\mathcal{T}$ that

$$\begin{aligned}\|E_{\mathcal{T}}\|_{H^{-t}(\Omega)} &\lesssim \left(\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1}\|L_{i,j}(g)-P_{i,j}(g)\|_{\tilde{H}^{-t}(\Omega_{i,j})}\right)^{1/2}\\ &\lesssim \left(\sum_{i=1}^{M}\sum_{j=j_0}^{j_i^*+1}\sum_{\substack{|\lambda|>j\\ \tilde{S}_\lambda\cap\Omega_{i,j}\neq\emptyset}}|g_\lambda|^2\right)^{1/2} \lesssim \left(\sum_{\lambda\notin\mathcal{T}}|g_\lambda|^2\right)^{1/2} = \|\mathbf{g}-\mathbf{g}_{\mathcal{T}}\|_{\ell_2(L)}.\end{aligned}$$

□

It remains to study the meaning of Assumption 5.2. Note that in the case of a single wavelet Riesz basis Ψ for Ω it follows basically from (5.4), see [1]. But using an overlapping partition $\mathcal{C}=\{\Omega_i\}_{i=1}^M$ of Ω and the concept of aggregated trees we have to check carefully if during the construction of $\breve{\mathcal{T}}=\bigcup_{i=1}^M\breve{\mathcal{T}}_i$ in (4.21) from $\mathcal{T}=\bigcup_{i=1}^M\mathcal{T}_i$ the assumption may be violated. The crucial point here are the intersection areas of two or more patches. We restrict ourselves to a more detailed view on the case $M=2$ in the following, examining the intersection $R:=\Omega_1\cap\Omega_2\neq\emptyset, \Omega_1\neq\Omega_2$, and local trees $\mathcal{T}_1$ over Ω_1 and $\mathcal{T}_2$ over Ω_2 respectively. The main question is: Is there for any $j=j_0+1,\dots,j_1^*+1$ an index $\lambda\in\breve{\mathcal{T}}_2$, so that $\tilde{S}_\lambda\cap\Omega_{1,j}\neq\emptyset$ and vice versa?

Of course, this question is trivial to answer if there is no intersection between wavelets or rather reference cubes of $\mathcal{T}_1$ and $\mathcal{T}_2$ during the tree expansion process and the following completion and well-grading of the local trees. If an intersection appears, we have to distinguish the two following situations:

1. The intersection appears during the tree expansion itself, i.e., by construction of (4.16) and (4.20) respectively.

2. The intersection appears during the completion and well-grading of the local trees in (4.21).

The first case is uncomplicated, since by construction of $\Phi_k(\lambda)$ and $\Theta_{\varepsilon,\mu}$ intersecting trees 'cross' each other and thus (5.14) remains valid. Unfortunately, in the second case local completion and well-grading the local trees may, depending on diam(R) and the used wavelets, cause $\breve{\mathcal{T}}_1$ to extend into $\breve{\mathcal{T}}_2$ or vice versa, such that Assumption 5.2 may not hold. A possible resort to assure (5.14) is to establish a third part of tree expansion in the following way

```
For i = 1,...,M do
    For j = j_i^*,...j_0 do
        For k ∈ {l : Ω_l ∩ Ω_i ≠ ∅} do
            Determine W_{i,j,k} := {λ ∈ J_k : ∃μ ∈ t_{i,j}, |λ| = |μ|, S̃_λ ∩ S̃_μ ≠ ∅}.
        od
    od
od
Set W := ⋃_{i,j,k} W_{i,j,k} and T̊ := T̆ ∪ W.
```

By this 're-grading' we assure that both trees are enlarged in the intersection area avoiding $\Omega_{1,j}$ and $\tilde{S}_\lambda, \lambda\in\breve{\mathcal{T}}_2$, to overlap. Again, by compactness of the used wavelets, the finite number of patches, and the fact that a well-graded local tree can be achieved by a uniformly bounded number of leave refinements, asymptotic results remain untouched, i.e., $\#(\mathring{\mathcal{T}})\lesssim\#\breve{\mathcal{T}}$.

6 Optimality of modSOLVE

Connecting the results from Sections 3, 4 and 5, we are finally ready to analyze the complexity of Algorithm 2.

Theorem 6.1 *Let $s^* > 0$ and assume that for all $s \in (0, s^*)$ the variational problem (2.8) has a solution $\mathbf{u} \in \mathcal{A}^s_{\mathcal{AT}}$. Furthermore, let $\mathbf{A}$ be s^*-compressible and bounded on $\mathcal{A}^s_{\mathcal{AT}}$. Then, for $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$ the routine* **RES** *can be implemented by*

$$\mathbf{RES}[\eta, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}, \mathbf{v}] := \mathbf{ATAPPLY}[\tfrac{\eta}{3}, \mathbf{A}, \mathbf{v}] + \mathbf{EVAL}[\tfrac{\eta}{3}, \mathbf{G}(\cdot), \mathbf{v}] - \mathbf{ATRHS}[\tfrac{\eta}{3}, \mathbf{f}],$$

where **EVAL** *is done by Theorem 4.1 and RECOVER. For $\mathbf{w}_\eta := \mathbf{RES}[\eta, \mathbf{A}, \mathbf{G}(\cdot), \mathbf{f}, \mathbf{v}]$ it holds*

$$\begin{aligned} \#\mathrm{supp}(\mathbf{w}_\eta) &\leq C\eta^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1) && (6.1)\\ \|\mathbf{w}_\eta\|_{\mathcal{A}^s_{\mathcal{AT}}} &\leq C(\|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1) && (6.2) \end{aligned}$$

Furthermore, under the assumption given in Remark 5.1 concerning quadrature, the computational effort for **RES** *is bounded by $C\left(\eta^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1) + \#\mathcal{T}(\mathrm{supp}(\mathbf{v}))\right)$, where $\mathcal{T}(\mathrm{supp}(\mathbf{v}))$ is again the smallest aggregated tree containing* $\mathrm{supp}(\mathbf{v})$.

Proof: By $\mathbf{u} \in \mathcal{A}^s_{\mathcal{AT}}$, the boundedness of $\mathbf{A}$ on $\mathcal{A}^s_{\mathcal{AT}}$ and (4.24), we have $\mathbf{f} \in \mathcal{A}^s_{\mathcal{AT}}$ and $\|\mathbf{f}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1$. Using Remark 4.2, Lemma 4.2 and Theorem 4.1 we get

$$\begin{aligned} \#\mathrm{supp}(\mathbf{w}_\eta) &\leq C(\eta^{-1/s}\|\mathbf{A}\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \eta^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#(I_{j_0}) + \eta^{-1/s}\|\mathbf{f}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}})\\ &\leq C(\eta^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#(I_{j_0}) + \eta^{-1/s}(\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1))\\ &\leq C\eta^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1) \end{aligned}$$

and

$$\|\mathbf{w}_\eta\|_{\mathcal{A}^s_{\mathcal{AT}}} \leq C(\|\mathbf{A}\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1) \leq C(\|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1).$$

Finally, the computational effort for **RES** can be bounded using Remark 4.2, the computational effort for building the aggregated trees (4.11), Lemma 4.2, (4.14), Theorem 4.1 and Remark 5.1 by

$$\begin{aligned} & C\Big(\eta^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#\mathrm{supp}(\mathbf{v}) + \eta^{-1/s}\|\mathbf{A}\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \eta^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#(I_{j_0}) + \#\mathcal{T}(\mathrm{supp}(\mathbf{v}))\\ & + \eta^{-1/s}(\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1) + \eta^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#(I_{j_0})\Big)\\ \leq\ & C\Big(\eta^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \#\mathcal{T}(\mathrm{supp}(\mathbf{v})) + \#(I_{j_0}) + \eta^{-1/s}(\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1)\Big)\\ \leq\ & C\Big(\eta^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1) + \#\mathcal{T}(\mathrm{supp}(\mathbf{v}))\Big). \end{aligned}$$

□

Using Theorem 6.1 we are now able to prove the optimality of Algorithm 2.

Theorem 6.2 *Assume that additionally to the assumptions of Theorem 6.1, $\mathbf{Q}$ is s^*-compressible and bounded on $\mathcal{A}^s_{\mathcal{AT}}$, $s \in (0, s^*)$. Then, the following estimates for $\mathbf{u}_\varepsilon$, the output of* **modSOLVE** *in Algorithm 2, are valid.*

$$\begin{aligned} \#\mathrm{supp}(\mathbf{u}_\varepsilon) &\leq C\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}},\\ \|\mathbf{u}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} &\leq C\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}. \end{aligned}$$

Furthermore, the computational effort for the determination of $\mathbf{u}_\varepsilon$ is bounded by $C\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}$.

Proof: Using the notation from Algorithm 2 and Lemma 3.4, we have, by (3.19), $\|\mathbf{Qu}-\mathbf{z}^j\|_{\ell_2(L)} \le \frac{1}{2C^*+1}\tilde{\varepsilon}_j$ for $j \ge 1$. By (4.12) we get

$$\#\text{supp}(\mathbf{u}^j) \le C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{Qu}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s} \le C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}, \tag{6.3}$$

using the boundedness of $\mathbf{Q}$ on $\mathcal{A}^s_{\mathcal{AT}}$, and thus finally $\#\text{supp}(\mathbf{u}_\varepsilon) \le C\varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}$. By (4.13), we also have

$$\|\mathbf{u}^j\|_{\mathcal{A}^s_{\mathcal{AT}}} \le C\|\mathbf{Qu}\|_{\mathcal{A}^s_{\mathcal{AT}}} \le C\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}, \tag{6.4}$$

yielding $\|\mathbf{u}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} \le C\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}$.

Thus, for $j \ge 1$, the estimates (6.3) and (6.4) are valid for the first input $\mathbf{v}^{j+1,0} = \mathbf{u}^j$ of **RES**. By this and Theorem 6.1, the effort for the determination of $\mathbf{v}^{j+1,1}$ is bounded by $C(\tilde{\varepsilon}_j^{-1/s}(\|\mathbf{u}^j\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s} + \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s} + 1) + \#\text{supp}(\mathbf{u}^j)) \le C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}$. Due to (6.1) and (6.2), the estimates (6.3) and (6.4) are also valid for $\mathbf{v}^{j+1,1}$. Iterating this argument k^* times (note that k^* is uniformly bounded in j), yields that the overall effort for the k^* applications of **RES** remains bounded by $C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}$ and the result, $\mathbf{v}^{j+1,k^*}$, satisfies (6.3) and (6.4). Combining this with Remark 4.2 shows that the effort of the **APPLY** step in **ATAPPLY** is bounded by $C(\tilde{\varepsilon}_j^{-1/s}\|\mathbf{v}^{j+1,k^*}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s} + \#\text{supp}(\mathbf{v}^{j+1,k^*})) \le C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}$, the effort for the following **ATCOARSE** step is by Lemma 4.2 bounded by $C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{v}^{j+1,k^*}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s} \le C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}$. Finally, again using Remark 4.2 and Lemma 4.2, the effort for calculating $\mathbf{u}^{j+1}$ from $\mathbf{z}^{j+1}$ is also bounded by $C\tilde{\varepsilon}_j^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}$. It remains to look after $j = 0$. Since we start with $\mathbf{u}^0 = \mathbf{0}$, (6.3) and (6.4) are also valid for $j = 0$ yielding the same error estimates for $\mathbf{v}^{1,k^*}$. The assertion follows now by the fact that the algorithm terminates after a finite number of steps with decreasing accuracy $\tilde{\varepsilon}_j$. □

Of course, an open question are the s^*-compressibility of $\mathbf{Q}$ and its boundedness on $\mathcal{A}^s_{\mathcal{AT}}$ assumed in Theorem 6.2. We won't go into this here, but refer to [5, 12, 24] and the references given there for the topic s^*-compressibility. Concerning the boundedness of $\mathbf{Q}$ on $\mathcal{A}^s_{\mathcal{AT}}$, Lemma 4.3 seems a promising starting point for the verification of this assumption.

References

[1] A. Barinka, *Fast computation tools for adaptive wavelet schemes*, Ph.D. thesis, RWTH Aachen, 2005.

[2] A. Barinka, T. Barsch, P. Charton, A. Cohen, S. Dahlke, W. Dahmen, and K. Urban, *Adaptive wavelet schemes for elliptic problems: Implementation and numerical experiments*, SIAM J. Sci. Comput. **23** (2001), no. 3, 910–939.

[3] P. Binev and R. DeVore, *Fast computation in adaptive tree approximation*, Numer. Math. **97** (2004), no. 2, 193–217.

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

[5] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations – Convergence rates*, Math. Comput. **70** (2001), no. 233, 27–75.

[6] ———, *Adaptive wavelet methods II: Beyond the elliptic case*, Found. Comput. Math. **2** (2002), no. 3, 203–245.

[7] ______, *Adaptive wavelet schemes for nonlinear variational problems*, SIAM J. Numer. Anal. **41** (2003), no. 5, 1785–1823.

[8] ______, *Sparse evaluation of compositions of functions using multiscale expansions*, SIAM J. Math. Anal. **35** (2003), no. 2, 279–303.

[9] ______, *Adaptive wavelet techniques in numerical simulation*, Encyclopedia of Computational Mechanics, vol. 1, John Wiley & Sons, 2004, pp. 1–64.

[10] S. Dahlke, W. Dahmen, R. Hochmuth, and R. Schneider, *Stable multiscale bases and local error estimation for elliptic problems*, Appl. Numer. Math. **23** (1997), no. 1, 21–48.

[11] S. Dahlke, M. Fornasier, M. Primbs, T. Raasch, and M. Werner, *Nonlinear and adaptive frame approximation schemes for elliptic PDEs: Theory and numerical experiments*, Bericht 2007-7, Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, 2007, To appear in *Numer. Methods Partial Differ. Equations.*

[12] S. Dahlke, M. Fornasier, and T. Raasch, *Adaptive frame methods for elliptic operator equations*, Adv. Comput. Math. **27** (2007), no. 1, 27–63.

[13] S. Dahlke, M. Fornasier, T. Raasch, R. Stevenson, and M. Werner, *Adaptive frame methods for elliptic operator equations: The steepest descent approach*, IMA J. Numer. Anal. **27** (2007), no. 4, 717–740.

[14] W. Dahmen, A. Kunoth, and K. Urban, *Biorthogonal spline-wavelets on the interval — Stability and moment conditions*, Appl. Comput. Harmon. Anal. **6** (1999), 132–196.

[15] W. Dahmen and R. Schneider, *Wavelets with complementary boundary conditions — Function spaces on the cube*, Result. Math. **34** (1998), no. 3–4, 255–293.

[16] ______, *Composite wavelet bases for operator equations*, Math. Comput. **68** (1999), 1533–1567.

[17] ______, *Wavelets on manifolds I. Construction and domain decomposition*, SIAM J. Math. Anal. **31** (1999), 184–230.

[18] W. Dahmen, R. Schneider, and Y. Xu, *Nonlinear functionals of wavelet expansions — Adaptive reconstruction and fast evaluation*, Numer. Math. **86** (2000), no. 1, 49–101.

[19] I. Daubechies, *Ten lectures on wavelets*, CBMS–NSF Regional Conference Series in Applied Math., vol. 61, SIAM, Philadelphia, 1992.

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

[21] S. Mallat, *Multiresolution approximation and wavelet orthonormal bases of* $L_2(\mathbb{R}^d)$, Trans. Amer. Math. Soc. **315** (1989), 69–87.

[22] T. Raasch, *Adaptive wavelet and frame schemes for elliptic and parabolic equations*, Dissertation, Philipps-Universität Marburg, 2007.

[23] T. Runst and R. Sickel, *Sobolev spaces of fractional order, Nemytskij operators and nonlinear partial differential equations*, De Gruyter, Berlin, 1996.

[24] R. Stevenson, *Adaptive solution of operator equations using wavelet frames*, SIAM J. Numer. Anal. **41** (2003), no. 3, 1074–1100.

[25] ______, *Composite wavelet bases with extended stability and cancellation properties*, SIAM J. Numer. Anal. **45** (2007), no. 1, 133–162.

[26] R. Stevenson and M. Werner, *Computation of differential operators in aggregated wavelet frame coordinates*, IMA J. Numer. Anal. **28** (2008), no. 2, 354–381.