AN OPTIMAL ADAPTIVE WAVELET METHOD WITHOUT COARSENING

TSOGTGEREL GANTUMUR, HELMUT HARBRECHT, ROB STEVENSON

Abstract. In this paper, an adaptive wavelet method for solving linear operator equations is constructed that is a modification of the method from [*Math. Comp*, 70 (2001), pp.27–75] by Cohen, Dahmen and DeVore, in the sense that there is no recurrent coarsening of the approximate solutions. Despite of this, it will be shown that the method has optimal computational complexity. Numerical results in a simple model problem indicate that the avoidance of coarsening results in a more efficient algorithm.

1. Preliminaries

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

$$Au = f.$$

As typical examples, we think of linear differential or integral equations of some order $2t$ in variational form. Furthermore, although systems of such equations also fit into the framework, usually we think of scalar equations, so that typically H is a Sobolev space H^t on the underlying domain or manifold, possibly incorporating essential boundary conditions.

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

$$\mathbf{A}\mathbf{u} = \mathbf{f},$$

where $\mathbf{A} := \langle \Psi, A\Psi\rangle : \ell_2 \to \ell_2$ is boundedly invertible and $\mathbf{f} := \langle \Psi, f\rangle \in \ell_2$. Here $\langle\cdot,\cdot\rangle$ denotes the duality product on (H^t, H^{-t}). In the following, we will also use $\langle\cdot,\cdot\rangle$ to denote $\langle\cdot,\cdot\rangle_{\ell_2}$, and use $\|\cdot\|$ to denote $\|\cdot\|_{\ell_2}$ as well as $\|\cdot\|_{\ell_2\to\ell_2}$. Throughout this paper, $\mathbf{u}$ and $\mathbf{f}$ will always denote the solution and right-hand side of this equation, respectively.

Let us denote by $\mathbf{u}_N$ a *best N-term approximation* for $\mathbf{u}$, i.e., a vector with at most N nonzero coefficients that has distance to $\mathbf{u}$ not larger than that of any vector with a support of that size. Note that $\|u - \mathbf{u}_N^T\Psi\|_{H^t} \eqsim \|\mathbf{u} - \mathbf{u}_N\|$. Considering bases Ψ of sufficiently smooth *wavelet* type, the theory of nonlinear approximation ([DeV98, Coh00]) tells us that if both

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

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

$$\sup_{N\in\mathbb{N}} N^s\|\mathbf{u} - \mathbf{u}_N\| < \infty. \tag{1.1}$$

The condition here involving Besov regularity is much milder that the corresponding condition $u \in H^{sn+t}$ involving Sobolev regularity that would be needed to guarantee the same rate of convergence with linear approximation in the span of N wavelets corresponding to the "coarsest levels." Indeed, assuming a sufficiently smooth right-hand side, for several boundary value problems it was proven that the solution has a much higher Besov than Sobolev regularity [DD97, Dah99]. Note that,

Date: March 22, 2005.

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

Key words and phrases. Adaptive methods, operator equations, wavelets, optimal computational complexity, best N-term approximation.

This work was supported by the Netherlands Organization for Scientific Research and by the EC-IHP project "Breaking Complexity".

regardless of the smoothness of the solution u, a rate higher than $\frac{d-t}{n}$ can never be expected with wavelets of order d, except when u happens to be exceptionally close to a finite linear combination of wavelets. On general domains or manifolds, suitable wavelet bases for H^t have been constructed in [DS99a, CTU99, CM00, DS99b, Ste04a, HS04].

Vectors $\mathbf{u} \in \ell_2$ that satisfy (1.1) can be characterized as follows (see [DeV98]): Let $\gamma_n(\mathbf{u})$ denote the nth largest coefficient in modulus of $\mathbf{u}$. For $0 < \tau < 2$, the space $\ell_\tau^w = \ell_\tau^w(\nabla)$ is defined by

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

It is easily verified that $\ell_\tau \hookrightarrow \ell_\tau^w \hookrightarrow \ell_{\tau+\delta}$ for any $\delta \in (0, 2-\tau]$, justifying why ℓ_τ^w is called *weak* ℓ_τ. The expression $|\mathbf{u}|_{\ell_\tau^w}$ defines only a quasi-norm since it does not necessarily satisfy the triangle inequality. With these ℓ_τ^w-spaces at hand, it can be shown that the property (1.1) is equivalent to $\mathbf{u} \in \ell_\tau^w$, with τ related to s according to $\tau = (\frac{1}{2} + s)^{-1}$. In particular, for each $\tau \in (0, 2)$,

$$\sup_N N^s \|\mathbf{u} - \mathbf{u}_N\| \eqsim |\mathbf{u}|_{\ell_\tau^w}, \tag{1.2}$$

see, e.g., [CDD01, Proposition 3.2]. Here and in the following, in order to avoid the repeated use of generic but unspecified constants, by $C \lesssim D$ we mean that C can be bounded by a multiple of D, independently of parameters which C and D may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

The aforementioned convergence rates under the mild Besov regularity assumption concern best N-term approximations, whose computation, however, requires full knowledge of the solution $\mathbf{u}$, which is only implicitly given. In [CDD01, CDD02], iterative methods for solving $\mathbf{A}\mathbf{u} = \mathbf{f}$ were developed that produce a sequence of approximations that converge with the same rate as that of the best N-term approximations, taking a number of operations that is equivalent to their support sizes. Both properties show that these methods are of *optimal computational complexity*. As a preparation for the results that will be derived in this paper, below we discuss both methods in some detail.

In each iteration of the methods, the matrix $\mathbf{A}$ has to be applied to some (finitely supported) vector. Since, generally, each column of $\mathbf{A}$ contains infinitely many non-zero entries, clearly this matrix-vector product cannot be computed exactly, and has to be approximated. For sufficiently smooth wavelets, that have sufficiently many vanishing moments, and for both differential operators with piecewise sufficiently smooth coefficients, or singular integral operators on sufficiently smooth manifolds, the results from [Ste04b, GS04, GS05] show that for some $s^* > \frac{d-t}{n}$, $\mathbf{A}$ is s^*-computable, meaning that for any $s < s^*$, for all $N \in \mathbb{N}$, there is an infinite matrix $\mathbf{A}_N$, having in each column $\mathcal{O}(N)$ non-zero entries, whose computations require $\mathcal{O}(N)$ operations, such that

$$\|\mathbf{A} - \mathbf{A}_N\| \lesssim N^{-s}. \tag{1.3}$$

Using this result, the adaptive approximate matrix-vector product **APPLY** from [CDD01] can be shown to have the following properties:

APPLY$[\mathbf{w}, \varepsilon] \to \mathbf{z}$. *The input satisfies $\varepsilon > 0$, and $\mathbf{w}$ is finitely supported. The output satisfies $\|\mathbf{A}\mathbf{w} - \mathbf{z}\| \le \varepsilon$, with for $s < s^*$, $\#\operatorname{supp}\mathbf{z} \lesssim \varepsilon^{-1/s} |\mathbf{w}|_{\ell_\tau^w}^{1/s}$, where, as always, $\tau = (\frac{1}{2} + s)^{-1}$, and the number of arithmetic operations and storage locations required by this call is bounded by some absolute multiple of $\varepsilon^{-1/s} |\mathbf{w}|_{\ell_\tau^w}^{1/s} + \#\operatorname{supp}\mathbf{w} + 1$.*

The construction of a sequence of approximations for $\mathbf{u}$ that converge with a certain rate requires the availability of a sequence of approximations for $\mathbf{f}$ that converge with at least that rate. It can be shown that for $s < s^*$, with $\tau = (\frac{1}{2} + s)^{-1}$, if $\mathbf{u} \in \ell_\tau^w$, then $\mathbf{f} \in \ell_\tau^w$, with $|\mathbf{f}|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$, and so $\sup_N N^s \|\mathbf{f} - \mathbf{f}_N\| \lesssim |\mathbf{u}|_{\ell_\tau^w}$, which, however does not tell how to *construct* an approximation $\mathbf{g}$ which is qualitatively as good as $\mathbf{f}_N$ with a comparable support size. We will assume the availability of the following routine, whose realization depends on the right-hand side at hand.

RHS$[\varepsilon] \to \mathbf{g}$ *with $\|\mathbf{f} - \mathbf{g}\| \le \varepsilon$, such that if $\mathbf{u} \in \ell_\tau^w$, and $s < s^*$, then $\#\operatorname{supp}\mathbf{g} \lesssim \varepsilon^{-1/s} |\mathbf{u}|_{\ell_\tau^w}^{1/s}$, and the number of arithmetic operations and storage locations required by this call is bounded by some absolute multiple of $\varepsilon^{-1/s} |\mathbf{u}|_{\ell_\tau^w}^{1/s} + 1$.*

The result concerning optimal computational complexity of the iterative methods from [CDD01, CDD02] requires the properties of **APPLY** and **RHS** mentioned above. Moreover, the methods apply under the condition that $\mathbf{A}$ is *symmetric, positive definite* (SPD), which, since $\mathbf{A} = \langle \Psi, A\Psi\rangle$, is equivalent to $\langle v, Aw\rangle = \langle Av, w\rangle$, $v, w \in H$, and $\langle v, Av\rangle \gtrsim \|v\|_H^2$, $v \in H$. For the case that A does not have both properties, in [CDD02] alternatives were sketched to reformulate $Au = f$ as an equivalent well-posed infinite matrix-vector problem with a symmetric, positive definite system matrix, as via the normal equations, or, in case the equation represents a saddle point problem, by using the reformulation as a positive definite system introduced in [BP88]. In these cases, $\mathbf{A}$, $\mathbf{f}$ are not given by $\langle \Psi, A\Psi\rangle$, $\langle \Psi, f\rangle$, respectively. We include these generalizations by, in those cases, simply *assuming* that we have routines **APPLY** and **RHS** as above, where for a discussion under which conditions this can be realized we refer to [CDD02]. Furthermore, to be able to conclude optimality of the iterative methods, in this generalized setting we now *assume* that the value s^* is larger than any s for which (1.1) can be expected.

The idea of the iterative method from [CDD02] is to apply Richardson iteration to $\mathbf{Au} = \mathbf{f}$. Of coarse, this iteration cannot be performed exactly, but by ensuring that the errors due to the inexact matrix-vector product and the approximation of $\mathbf{f}$ exhibit a proper decay when the iteration proceeds, a linear convergent method is obtained.

The principle behind the method from [CDD01] is to improve a given approximation $\mathbf{w}$ for $\mathbf{u}$ by realizing the saturation property: We set $\langle\!\langle \cdot, \cdot\rangle\!\rangle = \langle \mathbf{A}\cdot, \cdot\rangle$ and $|\!|\!| \cdot |\!|\!| = \langle\!\langle \cdot, \cdot\rangle\!\rangle^{\frac{1}{2}}$. For any $\Lambda \subset \nabla$, let $\mathbf{P}_\Lambda$ denote the ℓ_2-orthogonal projector onto $\ell_2(\Lambda)$, i.e., $\mathbf{P}_\Lambda$ replaces all coefficients outside Λ by zeros. With the notations $\mathbf{v}_\Lambda$, $\mathbf{z}_\Lambda$, etc., we will mean vectors in $\ell_2(\Lambda)$, i.e., vectors that are zero outside Λ. Using that $\mathbf{A}$ is SPD, one easily verifies that for any $\mathbf{v} \in \ell_2$, $\Lambda \subset \nabla$, $\mathbf{v}_\Lambda \in \ell_2(\Lambda)$,

$$\|\mathbf{A}^{-1}\|^{-\frac{1}{2}}\|\mathbf{v}\| \le |\!|\!|\mathbf{v}|\!|\!| \le \|\mathbf{A}\|^{\frac{1}{2}}\|\mathbf{v}\|, \quad \|\mathbf{Av}\| \le \|\mathbf{A}\|^{\frac{1}{2}}|\!|\!|\mathbf{v}|\!|\!|, \quad \|\mathbf{A}^{-1}\|^{-\frac{1}{2}}|\!|\!|\mathbf{v}_\Lambda|\!|\!| \le \|\mathbf{P}_\Lambda \mathbf{A}\mathbf{v}_\Lambda\|,$$

which properties will be often used in the following. The following lemma is well-known:

Lemma 1.1. *Let $\mu \in (0,1)$, $\mathbf{w} \in \ell_2$, $\nabla \supset \Lambda \supset \operatorname{supp} \mathbf{w}$ such that*

$$\|\mathbf{P}_\Lambda(\mathbf{f} - \mathbf{Aw})\| \ge \mu\|\mathbf{f} - \mathbf{Aw}\|. \tag{1.4}$$

Then, for $\mathbf{u}_\Lambda \in \ell_2(\Lambda)$ being the solution of the Galerkin system $\mathbf{P}_\Lambda \mathbf{A}\mathbf{u}_\Lambda = \mathbf{P}_\Lambda \mathbf{f}$, we have

$$|\!|\!|\mathbf{u} - \mathbf{u}_\Lambda|\!|\!| \le \left[1 - \kappa(\mathbf{A})^{-1}\mu^2\right]^{\frac{1}{2}} |\!|\!|\mathbf{u} - \mathbf{w}|\!|\!|.$$

Proof. We have

$$\begin{aligned} |\!|\!|\mathbf{u}_\Lambda - \mathbf{w}|\!|\!| &\ge \|\mathbf{A}\|^{-\frac{1}{2}}\|\mathbf{A}(\mathbf{u}_\Lambda - \mathbf{w})\| \ge \|\mathbf{A}\|^{-\frac{1}{2}}\|\mathbf{P}_\Lambda(\mathbf{f} - \mathbf{Aw})\| \\ &\ge \|\mathbf{A}\|^{-\frac{1}{2}}\mu\|\mathbf{f} - \mathbf{Aw}\| \ge \kappa(\mathbf{A})^{-\frac{1}{2}}\mu|\!|\!|\mathbf{u} - \mathbf{w}|\!|\!|, \end{aligned}$$

which, with $\kappa(\mathbf{A})^{-\frac{1}{2}}\mu$ reading as some arbitrary positive constant, is known as the saturation property of $\ell_2(\Lambda) \ni \mathbf{w}$. Using the Galerkin orthogonality $|\!|\!|\mathbf{u} - \mathbf{w}|\!|\!|^2 = |\!|\!|\mathbf{u} - \mathbf{u}_\Lambda|\!|\!|^2 + |\!|\!|\mathbf{u}_\Lambda - \mathbf{w}|\!|\!|^2$, the proof is completed. □

In this lemma it was assumed to have full knowledge about the exact residual, and that the arising Galerkin system is solved exactly, but as with the Richardson iteration, linear convergence is retained with an inexact evaluation of the residuals, and an inexact solution of the Galerkin systems, when the tolerances exhibit a proper decay as the iteration proceeds.

We remark that if, instead of being a Riesz basis, Ψ is only a frame for H, then the inexact Richardson method is still applicable (see [Ste03, DFR04]), whereas the other method is not since in that case the Galerkin systems can be arbitrarily badly conditioned.

Returning to the Riesz basis case, both the above iterative methods are linearly convergent, however, generally the rates are not as good as that of the best N-term approximations. Therefore in [CDD01, CDD02], these methods were extended with a so-called *coarsening* routine. After each K iterations, where K is a sufficiently large, fixed constant, the smallest coefficients from the current approximation vector are removed, increasing the upper bound for its error with some factor larger than 2, but with that restoring the optimal balance between accuracy and vector

length. Only after the extension with the coarsening routine, the resulting methods could be shown to be of optimal computational complexity.

In this paper, we reconsider the method from [CDD01]. Since for any subset $\Lambda \subset \nabla$, in energy norm, the best approximation from $\ell_2(\Lambda)$ is the Galerkin solution, which can be accurately approximated at relatively low cost, we expect that this method gives quantitatively the best results. The main point of the paper will be that we show that if μ is less than $\kappa(\mathbf{A})^{-\frac{1}{2}}$, and Λ is the *smallest* set containing $\operatorname{supp}\mathbf{w}$ that satisfies (1.4), then, *without coarsening* of the approximate solutions, these approximations converge with a rate as that of the best N-term approximations. As we will see, this result holds also true when the residuals and the Galerkin solutions are determined only inexactly, assuming a proper decay of the tolerances as the iteration proceeds, and when the cardinality of Λ is only minimal up to some constant factor, with which again a method of optimal computational complexity is obtained. Apart from the theoretical interest in that it is possible to construct an adaptive algorithm of optimal computational complexity without coarsening, it can be expected that the avoidance of coarsening also has a *quantitative advantage*. Indeed, in each coarsening step at least half, but often a much larger part of the current approximation vector is removed, which part has been involved in a number of computations.

Another difference with the method from [CDD01] is that for each call of **APPLY** or **RHS**, we will use a tolerance that is some fixed multiple of an a posteriori estimate of the current residual, instead of an a priori prescribed tolerance. Since it seems hard to avoid that a priori tolerances are increasingly either unnecessarily small, making the calls costly, or large so that the perturbed iteration due to the inexact evaluations converges significantly slower than the unperturbed one, also here we expect to obtain a quantitative improvement.

We consider approximations for $\mathbf{u}$ from $\ell_2(\Lambda)$, where Λ is any finite subset of ∇. In [CDD03], in the context of non-linear operators, a slightly restricted type of wavelet approximation is introduced, in the sense that only sets Λ are considered that are trees, meaning that if $\lambda \in \Lambda$, then for any $\lambda' \in \nabla$ with $\operatorname{supp}\psi_\lambda \subset \operatorname{supp}\psi_{\lambda'}$, also $\lambda' \in \Lambda$. Although, at least for linear operators, there is no real need to restrict to tree approximations, on the other hand, working with trees has advantages in view of obtaining an efficient implementation, whereas best tree N-term approximations converge towards $\mathbf{u}$ with a rate N^{-s} under regularity conditions that are only slightly stronger than that for unrestricted best N-term approximations. We note that by making obvious changes only, the results from this paper also apply to tree approximations.

We tested our adaptive wavelet solver for the Poisson equation on the interval. The results reported in the last section show that, in this simple example, the new method is indeed much more efficient than the inexact Richardson method with coarsening. In [DHS05], co-authored by the second author, numerical results based on tree approximations are given for singular integral equations on the boundary of three dimensional domains.

2. The adaptive method without coarsening

In the following lemma it is showed that for sufficiently small μ and $\mathbf{u} \in \ell_\tau^w$, for a set Λ as in Lemma 1.1 that has minimal cardinality, $\#(\Lambda\backslash\operatorname{supp}\mathbf{w})$ can be bounded in terms of $\|\mathbf{f}-\mathbf{A}\mathbf{w}\|$ and $|\mathbf{u}|_{\ell_\tau^w}$ only, i.e., *independently* of $|\mathbf{w}|_{\ell_\tau^w}$ and the value of s^* (cf. [CDD01, §4.2-4.3]).

Lemma 2.1. *Let $\mu \in (0, \kappa(\mathbf{A})^{-\frac{1}{2}})$ be a constant, $\mathbf{w} \in \ell_2$, and for some $s > 0$ and $\tau = (\frac{1}{2}+s)^{-1}$, $\mathbf{u} \in \ell_\tau^w$. Then the* smallest *set $\Lambda \supset \operatorname{supp}\mathbf{w}$ with*

$$\|\mathbf{P}_\Lambda(\mathbf{f}-\mathbf{A}\mathbf{w})\| \geq \mu\|\mathbf{f}-\mathbf{A}\mathbf{w}\|$$

satisfies

$$\#(\Lambda\backslash\operatorname{supp}\mathbf{w}) \lesssim \|\mathbf{f}-\mathbf{A}\mathbf{w}\|^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}.$$

Proof. Let $\lambda > 0$ be a constant with $\mu \leq \kappa(\mathbf{A})^{-\frac{1}{2}}(1-\|\mathbf{A}\|\lambda^2)^{\frac{1}{2}}$. Let N be such that a best N-term approximation $\mathbf{u}_N$ for $\mathbf{u}$ satisfies $\|\mathbf{u}-\mathbf{u}_N\| \leq \lambda|||\mathbf{u}-\mathbf{w}|||$. Since $|||\mathbf{u}-\mathbf{w}||| \leq \|\mathbf{A}^{-1}\|^{\frac{1}{2}}\|\mathbf{f}-\mathbf{A}\mathbf{w}\|$, we have

$$N \lesssim \|\mathbf{f}-\mathbf{A}\mathbf{w}\|^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}.$$

With $\breve{\Lambda} := \operatorname{supp}\mathbf{w} \cup \operatorname{supp}\mathbf{u}_N$, the solution of $\mathbf{P}_{\breve{\Lambda}}\mathbf{A}\mathbf{u}_{\breve{\Lambda}} = \mathbf{P}_{\breve{\Lambda}}\mathbf{f}$ satisfies

$$|||\mathbf{u} - \mathbf{u}_{\breve{\Lambda}}||| \le |||\mathbf{u} - \mathbf{u}_N||| \le \|\mathbf{A}\|^{\frac{1}{2}}\|\mathbf{u} - \mathbf{u}_N\| \le \|\mathbf{A}\|^{\frac{1}{2}}\lambda|||\mathbf{u} - \mathbf{w}|||,$$

and so by Galerkin orthogonality, $|||\mathbf{u}_{\breve{\Lambda}} - \mathbf{w}||| \ge (1 - \|\mathbf{A}\|\lambda^2)^{\frac{1}{2}}|||\mathbf{u} - \mathbf{w}|||$, giving

$$\begin{aligned}\|\mathbf{P}_{\breve{\Lambda}}(\mathbf{f} - \mathbf{A}\mathbf{w})\| &= \|\mathbf{P}_{\breve{\Lambda}}(\mathbf{A}\mathbf{u}_{\breve{\Lambda}} - \mathbf{A}\mathbf{w})\| \ge \|\mathbf{A}^{-1}\|^{-\frac{1}{2}}|||\mathbf{u}_{\breve{\Lambda}} - \mathbf{w}||| \\ &\ge \|\mathbf{A}^{-1}\|^{-\frac{1}{2}}(1 - \|\mathbf{A}\|\lambda^2)^{\frac{1}{2}}|||\mathbf{u} - \mathbf{w}||| \ge \kappa(\mathbf{A})^{-\frac{1}{2}}(1 - \|\mathbf{A}\|\lambda^2)^{\frac{1}{2}}\|\mathbf{f} - \mathbf{A}\mathbf{w}\| \\ &\ge \mu\|\mathbf{f} - \mathbf{A}\mathbf{w}\|.\end{aligned}$$

Since $\breve{\Lambda} \supset \operatorname{supp}\mathbf{w}$, by definition of Λ we conclude that

$$\#(\Lambda\backslash\operatorname{supp}\mathbf{w}) \le \#(\breve{\Lambda}\backslash\operatorname{supp}\mathbf{w}) \le N \lesssim \|\mathbf{f} - \mathbf{A}\mathbf{w}\|^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}.$$

□

Based on Lemmas 1.1 and 2.1, the following routine **GROW** provides a practical algorithm for extending the support of an approximation $\mathbf{w}$ for $\mathbf{u}$ to a set Λ, which cardinality can be bounded as in Lemma 2.1, but, on the other hand, which is sufficiently large such that $\ell_2(\Lambda)$ has the saturation property. First, inside a loop, the tolerances for the approximate matrix-vector product and the approximation of the right-hand side are decreased until either the computed approximate residual $\mathbf{r}$ has a sufficiently small relative error, or the norm of the residual is below the target tolerance meaning that $\mathbf{w}$ will be excepted as a valid approximation for $\mathbf{u}$. Second, in the first case, for a suitable constant α, the, up some constant factor, smallest $\Lambda \supset \operatorname{supp}\mathbf{w}$ is determined with $\|\mathbf{P}_\Lambda\mathbf{r}\| \ge \alpha\|\mathbf{r}\|$.

GROW$[\mathbf{w}, \bar{\nu}, \varepsilon] \to [\Lambda, \nu]$*:*
% Let α, ω *be constants with* $0 < \omega < \alpha \le 1$, $\frac{\alpha+\omega}{1-\omega} < \kappa(\mathbf{A})^{-\frac{1}{2}}$.
$\zeta := 2\frac{\omega\bar{\nu}}{1-\omega}$
`do` $\zeta := \zeta/2$, $\mathbf{r} := \mathbf{RHS}[\zeta/2] - \mathbf{APPLY}[\mathbf{w}, \zeta/2]$
`until` $\nu := \|\mathbf{r}\| + \zeta \le \varepsilon$ `or` $\zeta \le \omega\|\mathbf{r}\|$
`if` $\nu > \varepsilon$
`then` *determine a set* $\nabla \supset \Lambda \supset \operatorname{supp}\mathbf{w}$*, with, up to some absolute constant factor, minimal cardinality, such that* $\|\mathbf{P}_\Lambda\mathbf{r}\| \ge \alpha\|\mathbf{r}\|$
`else` $\Lambda := \emptyset$
`endif`

Remark 2.2. If $\bar{\nu} \notin [\frac{1-\omega}{1+\omega}\|\mathbf{f} - \mathbf{A}\mathbf{w}\|, \|\mathbf{f} - \mathbf{A}\mathbf{w}\|]$, then ζ at the first evaluation of $\mathbf{r}$ is outside $[\frac{\omega}{1+\omega}\|\mathbf{f} - \mathbf{A}\mathbf{w}\|, \frac{\omega}{1-\omega}\|\mathbf{f} - \mathbf{A}\mathbf{w}\|]$, and from $\omega\|\mathbf{f} - \mathbf{A}\mathbf{w}\| - \zeta \le \omega\|\mathbf{r}\| \le \omega\|\mathbf{f} + \mathbf{A}\mathbf{w}\| + \zeta$, one infers that in this case either the second test in the until-clause will fail anyway, meaning that the first iteration of the do-loop is not of any use, or that second test in the until-clause is always passed, but possibly with a tolerance that is unnecessarily small. We conclude that there is not much sense in calling **GROW** with a value of $\bar{\nu}$ that is far outside $[\frac{1-\omega}{1+\omega}\|\mathbf{f} - \mathbf{A}\mathbf{w}\|, \|\mathbf{f} - \mathbf{A}\mathbf{w}\|]$.

Remark 2.3. Selecting Λ in **GROW** with true minimal cardinality would require the sorting of all coefficients of $\mathbf{r}|_{\nabla\backslash\operatorname{supp}\mathbf{w}}$ by their modulus, which, with $N := \#\operatorname{supp}\mathbf{r}|_{\nabla\backslash\operatorname{supp}\mathbf{w}}$, requires $\mathcal{O}(N\log N)$ operations. Another $\mathcal{O}(\#\operatorname{supp}\mathbf{r})$ operations for computing $\|\mathbf{r}\|$ are unavoidable. In the following, we recall a procedure with which the above log-factor is avoided.

In view of our task to select $\Lambda \supset \operatorname{supp}\mathbf{w}$ with $\|\mathbf{P}_\Lambda\mathbf{r}\| \ge \alpha\|\mathbf{r}\|$, we may discard all coefficients of $\mathbf{r}|_{\nabla\backslash\operatorname{supp}\mathbf{w}}$ with modulus not larger than $\sqrt{(1-\alpha^2)}\|\mathbf{r}\|/\sqrt{N}$. With $M := \|\mathbf{r}|_{\nabla\backslash\operatorname{supp}\mathbf{w}}\|_\infty$, and q the smallest integer with $2^{-(q+1)/2}M \le \sqrt{(1-\alpha^2)}\|\mathbf{r}\|/\sqrt{N}$, we store the other coefficients of $\mathbf{r}|_{\nabla\backslash\operatorname{supp}\mathbf{w}}$ in $q+1$ bins corresponding whether it lies in $[M, \frac{1}{\sqrt{2}}M)$, $[\frac{1}{\sqrt{2}}M, \frac{1}{2}M), \ldots$, or $[2^{-q/2}M, 2^{-(q+1)/2}M)$. We then build Λ by extracting coefficients from the bins, starting with the first bin, and when it got empty moving to the second bin and so on until $\|\mathbf{P}_\Lambda\mathbf{r}\| \ge \alpha\|\mathbf{r}\|$ is satisfied. Let the resulting Λ now contains coefficients from the pth bin, but not from further bins. Then a minimal set $\tilde{\Lambda}$ that satisfies $\|\mathbf{P}_{\tilde{\Lambda}}\mathbf{r}\| \ge \alpha\|\mathbf{r}\|$ contains all coefficients from the bins up to the $(p-1)$th one. Since any two coefficients in the pth bin differ at most a factor $\sqrt{2}$, we infer that the cardinality

of the contribution from the pth bin to Λ is at most twice at large as that to $\tilde\Lambda$, so that $\#\Lambda \le 2\#\tilde\Lambda$. The number of operations and storage locations required by this procedure is $\mathcal{O}(\#\operatorname{supp}\mathbf{r}+q)$, where $q < 2\log_2(M\sqrt{N}/[\sqrt{1-\alpha^2}\|\mathbf{r}\|]) \le 2\log_2(\sqrt{N}/\sqrt{1-\alpha^2}) \lesssim \log_2(\sqrt{N}) < \#\operatorname{supp}\mathbf{r}$.

Theorem 2.4. *$[\Lambda,\nu] = \mathbf{GROW}[\mathbf{w},\bar\nu,\varepsilon]$ terminates, with $\nu \ge \|\mathbf{f}-\mathbf{A}\mathbf{w}\|$ and $\nu \gtrsim \min\{\bar\nu,\varepsilon\}$. If, for $s < s^*$ and with $\tau = (\frac{1}{2}+s)^{-1}$, $\mathbf{u} \in \ell^w_\tau$, then the number of arithmetic operations and storage locations required by the call is bounded by some absolute multiple of* $\min\{\bar\nu,\nu\}^{-1/s}\big[|\mathbf{w}|^{1/s}_{\ell^w_\tau} + |\mathbf{u}|^{1/s}_{\ell^w_\tau} + \bar\nu^{1/s}(\#\operatorname{supp}\mathbf{w}+1)\big]$.

If **GROW** *terminates with $\nu > \varepsilon$, then*

$$\tfrac{\alpha-\omega}{1+\omega}\nu \le \|\mathbf{P}_\Lambda(\mathbf{f}-\mathbf{A}\mathbf{w})\|, \tag{2.1}$$

and

$$\#(\Lambda\backslash\operatorname{supp}\mathbf{w}) \lesssim \nu^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau}. \tag{2.2}$$

Proof. If at evaluation of the until-clause, $\zeta > \omega\|\mathbf{r}\|$, then $\|\mathbf{r}\|+\zeta < (\omega^{-1}+1)\zeta$. Since ζ is halved in each iteration, we infer that, if not by $\zeta \le \omega\|\mathbf{r}\|$, **GROW** will terminate by $\|\mathbf{r}\|+\zeta \le \varepsilon$.

Since after any evaluation of $\mathbf{r}$ inside the algorithm, $\|\mathbf{r}-(\mathbf{f}-\mathbf{A}\mathbf{w})\| \le \zeta$, any value of ν determined inside the algorithm is an upper bound on $\|\mathbf{f}-\mathbf{A}\mathbf{w}\|$. If the do-loop terminates in the first iteration, or the algorithm terminates with $\nu > \varepsilon$, then $\nu \gtrsim \min\{\bar\nu,\varepsilon\}$. In the other case, let $\mathbf{r}^{\mathrm{old}} := \mathbf{RHS}[\zeta] - \mathbf{APPLY}[\mathbf{w},\zeta]$. We have $\|\mathbf{r}^{\mathrm{old}}\| + 2\zeta > \varepsilon$ and $2\zeta > \omega\|\mathbf{r}^{\mathrm{old}}\|$, so that $\nu \ge \zeta > (2\omega^{-1}+2)^{-1}(\|\mathbf{r}^{\mathrm{old}}\|+2\zeta) > \frac{\omega\varepsilon}{2+2\omega}$.

By the geometric decrease of ζ inside the algorithm, and in view of Remark 2.3, with ζ, $\mathbf{r}$ and ν having their values at termination, the properties of **RHS** and **APPLY** imply that the total cost of the call of **GROW** can be bounded by some multiple of $\zeta^{-1/s}(|\mathbf{w}|^{1/s}_{\ell^w_\tau} + |\mathbf{u}|^{1/s}_{\ell^w_\tau}) + K(\#\operatorname{supp}\mathbf{w}+1)$, with K being the number of calls of **APPLY** that were made. Taking into account its initial value, and the geometric decrease of ζ inside the algorithm, we have $K(\#\operatorname{supp}\mathbf{w}+1) = K\bar\nu^{-1/s}\bar\nu^{1/s}(\#\operatorname{supp}\mathbf{w}+1) \lesssim \zeta^{-1/s}\bar\nu^{1/s}(\#\operatorname{supp}\mathbf{w}+1)$. The proof of the first part of the theorem is completed once we have shown that $\zeta \gtrsim \min\{\bar\nu,\nu\}$. When the do-loop terminates in the first iteration, we have $\zeta \gtrsim \bar\nu$, and when the algorithm terminates with $\zeta \ge \omega\|\mathbf{r}\|$, we have $\zeta \gtrsim \nu$. In the other case, with $\mathbf{r}^{\mathrm{old}}$ as above, we have $\omega\|\mathbf{r}^{\mathrm{old}}\| < 2\zeta$, and so from $\|\mathbf{r}-\mathbf{r}^{\mathrm{old}}\| \le \zeta + 2\zeta$, we infer $\|\mathbf{r}\| \le \|\mathbf{r}^{\mathrm{old}}\| + 3\zeta < (2\omega^{-1}+3)\zeta$, so that $\nu < (2\omega^{-1}+4)\zeta$.

Now assume that **GROW** terminates with $\nu > \varepsilon$ and thus with $\zeta \le \omega\|\mathbf{r}\|$. With $\mathbf{g} = \mathbf{RHS}[\zeta/2]$ and $\mathbf{z} = \mathbf{APPLY}[\mathbf{w},\zeta/2]$, we have

$$\begin{aligned}\|\mathbf{P}_\Lambda(\mathbf{f}-\mathbf{A}\mathbf{w})\| &\ge \|\mathbf{P}_\Lambda\mathbf{r}\| - \|\mathbf{P}_\Lambda(\mathbf{A}\mathbf{w}-\mathbf{z})\| - \|\mathbf{P}_\Lambda(\mathbf{f}-\mathbf{g})\| \\ &\ge \alpha\|\mathbf{r}\| - \zeta \ge \tfrac{\alpha-\omega}{1+\omega}\nu,\end{aligned}$$

where the last inequality is a consequence of $\zeta \le \omega\|\mathbf{r}\|$, $\omega < \alpha$, and $\nu = \|\mathbf{r}\| + \zeta$.

To prove (2.2), with $\mu = \frac{\alpha+\omega}{1-\omega}$ let $\nabla \supset \hat\Lambda \supset \operatorname{supp}\mathbf{w}$ be the *smallest* set with

$$\|\mathbf{P}_{\hat\Lambda}(\mathbf{f}-\mathbf{A}\mathbf{w})\| \ge \mu\|\mathbf{f}-\mathbf{A}\mathbf{w}\|.$$

Then $\mu\|\mathbf{r}\| \le \mu\|\mathbf{f}-\mathbf{A}\mathbf{w}\| + \mu\zeta \le \|\mathbf{P}_{\hat\Lambda}(\mathbf{f}-\mathbf{A}\mathbf{w})\| + \mu\zeta \le \|\mathbf{P}_{\hat\Lambda}\mathbf{r}\| + (1+\mu)\omega\|\mathbf{r}\|$ or $\|\mathbf{P}_{\hat\Lambda}\mathbf{r}\| \ge \alpha\|\mathbf{r}\|$. By construction of Λ in **GROW**, we conclude that $\#(\Lambda\backslash\operatorname{supp}\mathbf{w}) \lesssim \#(\hat\Lambda\backslash\operatorname{supp}\mathbf{w})$. Since $\mu < \kappa(\mathbf{A})^{\frac{1}{2}}$ by the condition on ω and α, and $\|\mathbf{f}-\mathbf{A}\mathbf{w}\| \le \nu$, an application of Lemma 2.1 shows that $\#(\hat\Lambda\backslash\operatorname{supp}\mathbf{w}) \lesssim \nu^{-1/s}|\mathbf{u}|^{1/s}_{\ell^w_\tau}$ which completes the proof. □

When having extended $\operatorname{supp}\mathbf{w}$ to a set Λ such that $\ell_2(\Lambda)$ has the saturation property, the second ingredient of the iterative method is the approximate solution of the Galerkin system on $\ell_2(\Lambda)$. Given an approximation $\mathbf{g}_\Lambda$ for $\mathbf{P}_\Lambda\mathbf{f}$, there are various possibilities to approximately solving the system $\mathbf{P}_\Lambda\mathbf{A}\mathbf{u}_\Lambda = \mathbf{g}_\Lambda$, starting with some initial approximation $\mathbf{w}_\Lambda$ for $\mathbf{u}_\Lambda$, where obviously we will take $\mathbf{w}_\Lambda = \mathbf{w}$. After approximately computing the initial residual using the **APPLY** routine, instead of relying on the adaptive routine **APPLY** throughout the iteration, the following routine **GALSOLVE** iterates using some fixed, non-adaptive approximation for

$$\mathbf{A}_\Lambda := \mathbf{P}_\Lambda\mathbf{A}|_{\ell_2(\Lambda)}.$$

The accuracy of this approximation depends only on the *factor* with which one wants to reduce the norm of the residual. This approach can be expected to be particularly efficient when the approximate computation of the entries of $\mathbf{A}$ is relatively expensive, as with singular integral operators. As can be deduced from [ES04], it is even possible in the course of the iteration to gradually diminish the accuracy of the approximation for $\mathbf{A}_\Lambda$.

GALSOLVE$[\Lambda, \mathbf{g}_\Lambda, \mathbf{w}_\Lambda, \delta, \varepsilon] \to \tilde{\mathbf{w}}_\Lambda$*:*
% The input should satisfy $\delta \geq \|\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \mathbf{w}_\Lambda\|$.
% Let N *be such that, with* $\mathbf{A}_N$ *from* (1.3), $\tau := \|\mathbf{A} - \mathbf{A}_N\|\|\mathbf{A}^{-1}\| \leq \frac{\varepsilon}{3\varepsilon+3\delta}$.
% Set $\mathbf{B} := \mathbf{P}_\Lambda \frac{1}{2}(\mathbf{A}_N + \mathbf{A}_N^*)|_{\ell_2(\Lambda)}$.
$\mathbf{r}_0 := \mathbf{g}_\Lambda - \mathbf{P}_\Lambda(\mathbf{APPLY}[\mathbf{w}_\Lambda, \frac{\varepsilon}{3}])$
Apply a suitable iterative method for solving $\mathbf{Bx} = \mathbf{r}_0$*, e.g., Conjugate Gradients or Conjugate Residuals, to find an* $\mathbf{x}$ *with* $\|\mathbf{r}_0 - \mathbf{Bx}\| \leq \frac{\varepsilon}{3}$
$\tilde{\mathbf{w}}_\Lambda := \mathbf{w}_\Lambda + \mathbf{x}$

Theorem 2.5. $\tilde{\mathbf{w}}_\Lambda := \mathbf{GALSOLVE}[\Lambda, \mathbf{g}_\Lambda, \mathbf{w}_\Lambda, \delta, \varepsilon]$ *satisfies* $\|\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \tilde{\mathbf{w}}_\Lambda\| \leq \varepsilon$*. For any* $s < s^*$*, the number of arithmetic operations and storage locations required by the call is bounded by some absolute multiple of* $\varepsilon^{-1/s}(|\mathbf{w}_\Lambda|_{\ell_\tau^w}^{1/s} + |\mathbf{u}_\Lambda|_{\ell_\tau^w}^{1/s}) + c(\delta/\varepsilon)\#\Lambda$*, with* $c : \mathbb{R}_+ \to \mathbb{R}_+$ *being some non-decreasing function.*

Proof. Writing $\mathbf{B} = \mathbf{A}_\Lambda(\mathbf{I} + \mathbf{A}_\Lambda^{-1}(\mathbf{B} - \mathbf{A}_\Lambda))$, and using that $\|\mathbf{A}_\Lambda^{-1}\| \leq \|\mathbf{A}^{-1}\|$, $\|\mathbf{B} - \mathbf{A}_\Lambda\| \leq \|\mathbf{A} - \mathbf{A}_N\|$, and $\tau \leq \frac{1}{3} < 1$, we find that $\mathbf{B}$ is SPD with respect to the canonical scalar product on $\ell_2(\Lambda)$, with $\kappa(\mathbf{B}) \lesssim 1$ uniformly in ε and δ, and $\|\mathbf{B} - \mathbf{A}_\Lambda\|\|\mathbf{B}^{-1}\| \leq \frac{\tau}{1-\tau}$. We have $\|\mathbf{r}_0\| \leq \delta + \frac{\varepsilon}{3}$. Writing

$$\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \tilde{\mathbf{w}}_\Lambda = (\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \mathbf{w}_\Lambda - \mathbf{r}_0) + (\mathbf{r}_0 - \mathbf{Bx}) + (\mathbf{B} - \mathbf{A}_\Lambda)\mathbf{B}^{-1}(\mathbf{r}_0 + \mathbf{Bx} - \mathbf{r}_0),$$

we find

$$\|\mathbf{A}_\Lambda \tilde{\mathbf{w}}_\Lambda - \mathbf{g}_\Lambda\| \leq \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\tau}{1-\tau}(\delta + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3}) \leq \varepsilon.$$

The properties of **APPLY** and **RHS** show that the cost of the computation of $\mathbf{r}_0$ is bounded by some multiple of $\varepsilon^{-1/s}(|\mathbf{w}_\Lambda|_{\ell_\tau^w}^{1/s} + |\mathbf{u}_\Lambda|_{\ell_\tau^w}^{1/s})$. Since by (1.3), $\mathbf{B}$ is sparse and can be constructed in $\mathcal{O}(\#\Lambda)$ operations, and the required number of iterations of the iterative method is bounded, everything only dependent on an upper bound for δ/ε, the proof is completed. □

We have now the ingredients available to define our adaptive wavelet solver.

SOLVE$[\nu_{-1}, \varepsilon] \to \mathbf{w}_k$*:*
% Let γ *be a constant in* $\big(0, \frac{1}{6}\kappa(\mathbf{A})^{-\frac{1}{2}}\frac{\alpha-\omega}{1+\omega}\big)$*, with* α*,* ω *being the parameters inside* **GROW**.
% Let $\theta > 0$ *be a constant.*

$k := 0$; $\mathbf{w}_k := 0$
`while` *with* $[\Lambda_{k+1}, \nu_k] := \mathbf{GROW}[\mathbf{w}_k, \theta\nu_{k-1}, \varepsilon]$, $\nu_k > \varepsilon$ `do`
  $\mathbf{g}_{k+1} := \mathbf{P}_{\Lambda_{k+1}}(\mathbf{RHS}[\gamma\nu_k])$
  $\mathbf{w}_{k+1} := \mathbf{GALSOLVE}[\Lambda_{k+1}, \mathbf{g}_{k+1}, \mathbf{w}_k, (1+\gamma)\nu_k, \gamma\nu_k]$
  $k := k + 1$
`enddo`

Remark 2.6. As we will see, at the call of $\mathbf{GROW}[\mathbf{w}_k, \theta\nu_{k-1}, \varepsilon]$, we have that $\|\mathbf{f} - \mathbf{A}\mathbf{w}_k\| \lesssim \nu_{k-1}$. Although, for any fixed $\theta > 0$, **SOLVE** will be shown to be of optimal computational complexity, in view of Remark 2.2, a suitable tuning of θ will result in quantitatively better results. Ideally, θ has the largest value for which the do-loop inside **GROW** always terminates in one iteration.

Theorem 2.7. $\mathbf{w} := \mathbf{SOLVE}[\nu_{-1}, \varepsilon]$ *terminates with* $\|\mathbf{Aw} - \mathbf{f}\| \leq \varepsilon$*. If* $\nu_{-1} \eqsim \|\mathbf{f}\| \gtrsim \varepsilon$*, and for some* $s < s^*$*, and* $\tau = (\frac{1}{2}+s)^{-1}$*,* $\mathbf{u} \in \ell_\tau^w$*, then* $\#\operatorname{supp}\mathbf{w} \lesssim \varepsilon^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ *and the number of arithmetic operations and storage locations required by the call is bounded by some absolute multiple of the same expression.*

Proof. Before we come to the actual proof, first we indicate the need for the conditions involving ν_{-1}, $\|\mathbf{f}\|$ and ε. If $\nu_{-1} \not\gtrsim \varepsilon$, then the cost of the first call of **RHS** in the first call of **GROW** can be arbitrarily large. If $\nu_{-1} \not\lesssim \|\mathbf{f}\|$, then we cannot bound the number of iterations in the loop of the first call of **GROW**, each of them requiring in any case some arithmetic operations. Finally, if $\|\mathbf{f}\| \not\gtrsim \varepsilon$, then $\varepsilon^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ might be arbitrarily small, whereas **SOLVE** takes in any case some arithmetic operations.

Theorem 2.4 shows that $\nu_k \geq \|\mathbf{A}\mathbf{w}_k - \mathbf{f}\|$, and, as long as $\nu_k > \varepsilon$, that $\nu_k \lesssim \|\mathbf{f} - \mathbf{A}\mathbf{w}_k\|$. We have $\|\mathbf{g}_{k+1} - \mathbf{P}_{\Lambda_{k+1}}\mathbf{A}\mathbf{w}_k\| \leq (1+\gamma)\nu_k$, so that $(1+\gamma)\nu_k$ is a valid parameter for the $(k+1)$th call of **GALSOLVE**. Below we will prove that a constant $\xi < 1$ exists such that, as long as $\nu_k > \varepsilon$,

$$|||\mathbf{u} - \mathbf{w}_{k+1}||| \leq \xi |||\mathbf{u} - \mathbf{w}_k|||. \tag{2.3}$$

Because of $\|\mathbf{A}\mathbf{w}_k - \mathbf{f}\| \eqsim |||\mathbf{u} - \mathbf{w}_k|||$, this result shows that **SOLVE** terminates after finitely many iterations, say directly after the $(K+1)$th call of **GROW** that produces $[\Lambda_{K+1}, \nu_K]$, and furthermore that $\nu_k \lesssim \xi^{k-i}\nu_i$ for all $0 \leq i \leq k \leq K-1$. From, when $K > 0$, $\nu_K \leq \varepsilon < \nu_{K-1}$, and $\nu_0 \leq \max\{\varepsilon, \frac{1+\omega}{\alpha-\omega}\|\mathbf{f}\|\} \lesssim \nu_{-1}$, the latter inequality by assumption, we even have

$$\nu_k \lesssim \xi^{k-i}\nu_i, \qquad -1 \leq i \leq k \leq K. \tag{2.4}$$

Since, with $\Lambda_0 := \emptyset$, $\operatorname{supp}\mathbf{w}_i \subset \Lambda_i$ and $\Lambda_i \subset \Lambda_{i+1}$, by (2.2) for $1 \leq k \leq K$ we have

$$\#\operatorname{supp}\mathbf{w}_k \leq \#\Lambda_k = \sum_{i=0}^{k-1} \#(\Lambda_{i+1}\backslash\Lambda_i) \lesssim (\sum_{i=0}^{k-1} \nu_i^{-1/s})|\mathbf{u}|_{\ell_\tau^w}^{1/s} \lesssim \nu_{k-1}^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}. \tag{2.5}$$

From $|\mathbf{w}_k|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w} + (\#\operatorname{supp}\mathbf{w}_k)^s\|\mathbf{w}_k - \mathbf{u}\|$ ([CDD01, Lemma 4.11]), we infer that $|\mathbf{w}_k|_{\ell_\tau^w} \lesssim |\mathbf{u}|_{\ell_\tau^w}$.

By Theorem 2.4, for $k \leq K$ the cost of the $(k+1)$th call of **GROW** is bounded by an absolute multiple of

$$\min\{\nu_{k-1}, \nu_k\}^{-1/s}\big[|\mathbf{u}|_{\ell_\tau^w}^{1/s} + \nu_{k-1}^{1/s}(\nu_{k-1}^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s} + 1)\big] \lesssim \nu_k^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s},$$

where we used (2.5), $\min\{\nu_{k-1}, \nu_k\} \gtrsim \nu_k$ by (2.4), and $1 \lesssim \nu_{k-1}^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ by $\nu_{k-1} \lesssim \nu_{-1} \lesssim \|\mathbf{f}\| \lesssim |\mathbf{u}|_{\ell_\tau^w}$. For $k < K$, also the cost of the $(k+1)$th call of **RHS** or **GALSOLVE** is bounded by an absolute multiple of $\nu_k^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$ or $\nu_k^{-1/s}(|\mathbf{w}_k|_{\ell_\tau^w}^{1/s} + |\mathbf{u}|_{\ell_\tau^w}^{1/s}) + \#\Lambda_{k+1} \lesssim \nu_k^{-1/s}|\mathbf{u}|_{\ell_\tau^w}^{1/s}$, respectively. From (2.4) and $\nu_K \gtrsim \min\{\nu_{K-1}, \varepsilon\} \gtrsim \varepsilon$ by Theorem 2.4, where the second inequality follows from $\nu_{K-1} > \varepsilon$ when $K > 0$, and by assumption when $K = 0$, the proof is completed upon showing (2.3).

Abbreviating $\mathbf{P}_{\Lambda_{k+1}}$ as $\mathbf{P}_{k+1}$, for $0 \leq k < K$ let $\mathbf{u}_{k+1}$ be the solution of $\mathbf{P}_{k+1}\mathbf{A}\mathbf{u}_{k+1} = \mathbf{P}_{k+1}\mathbf{f}$. Because of $\|\mathbf{f} - \mathbf{A}\mathbf{w}_k\| \leq \nu_k$ and (2.1), that can be applied since $\nu_k > \varepsilon$, we have $\|\mathbf{P}_{k+1}(\mathbf{f} - \mathbf{A}\mathbf{w}_k)\| \geq \frac{\alpha-\omega}{1+\omega}\|\mathbf{f} - \mathbf{A}\mathbf{w}_k\|$, so that Lemma 1.1 shows that $|||\mathbf{u} - \mathbf{u}_{k+1}||| \leq [1 - \kappa(\mathbf{A})^{-1}(\frac{\alpha-\omega}{1+\omega})^2]^{\frac{1}{2}}|||\mathbf{u} - \mathbf{w}_k|||$.

Our $(k+1)$th iterand is, however, not $\mathbf{u}_{k+1}$ but $\mathbf{w}_{k+1}$, which contains errors because of the non-exact right-hand side and the inexact solution of the Galerkin system. One can simply estimate $|||\mathbf{u} - \mathbf{w}_{k+1}||| \leq |||\mathbf{u} - \mathbf{u}_{k+1}||| + |||\mathbf{u}_{k+1} - \mathbf{w}_{k+1}|||$, but a sharper result can be derived by using that $\mathbf{u} - \mathbf{w}_{k+1}$ is nearly $\langle\!\langle\cdot,\cdot\rangle\!\rangle$-orthogonal to $\ell_2(\Lambda_{k+1})$. With $\beta := \gamma\frac{2+2\omega}{\alpha-\omega}\kappa(\mathbf{A})^{\frac{1}{2}} < \frac{1}{3}$, we have

$$\begin{aligned}
|||\mathbf{u}_{k+1} - \mathbf{w}_{k+1}||| &\leq \|\mathbf{A}^{-1}\|^{\frac{1}{2}}\|\mathbf{P}_{k+1}\mathbf{A}(\mathbf{u}_{k+1} - \mathbf{w}_{k+1})\| \\
&\leq \|\mathbf{A}^{-1}\|^{\frac{1}{2}}\big(\|\mathbf{g}_{k+1} - \mathbf{P}_{k+1}\mathbf{A}\mathbf{w}_{k+1}\| + \|\mathbf{P}_{k+1}\mathbf{f} - \mathbf{g}_{k+1}\|\big) \\
&\leq \|\mathbf{A}^{-1}\|^{\frac{1}{2}}2\gamma\nu_k \leq \|\mathbf{A}^{-1}\|^{\frac{1}{2}}2\gamma\tfrac{1+\omega}{\alpha-\omega}\|\mathbf{P}_{k+1}(\mathbf{f} - \mathbf{A}\mathbf{w}_k)\| \leq \beta|||\mathbf{u}_{k+1} - \mathbf{w}_k|||.
\end{aligned}$$

Using $\mathbf{u} - \mathbf{u}_{k+1} \perp_{\langle\!\langle\,,\rangle\!\rangle} \ell_2(\Lambda_{k+1})$, we have

$$\begin{aligned}
|\langle\!\langle\mathbf{u} - \mathbf{w}_{k+1}, \mathbf{w}_{k+1} - \mathbf{w}_k\rangle\!\rangle| &= |\langle\!\langle\mathbf{u}_{k+1} - \mathbf{w}_{k+1}, \mathbf{w}_{k+1} - \mathbf{w}_k\rangle\!\rangle| \\
&\leq |||\mathbf{u}_{k+1} - \mathbf{w}_{k+1}|||\,|||\mathbf{w}_{k+1} - \mathbf{w}_k||| \leq \beta|||\mathbf{u}_{k+1} - \mathbf{w}_k|||\,|||\mathbf{w}_{k+1} - \mathbf{w}_k|||.
\end{aligned}$$

Now by writing

$$|||\mathbf{u} - \mathbf{w}_k|||^2 = |||\mathbf{u} - \mathbf{w}_{k+1}|||^2 + |||\mathbf{w}_{k+1} - \mathbf{w}_k|||^2 + 2\langle\!\langle\mathbf{u} - \mathbf{w}_{k+1}, \mathbf{w}_{k+1} - \mathbf{w}_k\rangle\!\rangle,$$

and, for obtaining the second line, two applications of

$$|||\mathbf{w}_{k+1} - \mathbf{w}_k||| \geq |||\mathbf{u}_{k+1} - \mathbf{w}_k||| - |||\mathbf{w}_{k+1} - \mathbf{u}_{k+1}||| \geq (1-\beta)|||\mathbf{u}_{k+1} - \mathbf{w}_k|||,$$

we find that

$$\begin{aligned}|||\mathbf{u}-\mathbf{w}_k|||^2 &\geq |||\mathbf{u}-\mathbf{w}_{k+1}|||^2 + |||\mathbf{w}_{k+1}-\mathbf{w}_k|||\big(|||\mathbf{w}_{k+1}-\mathbf{w}_k||| - 2\beta|||\mathbf{u}_{k+1}-\mathbf{w}_k|||\big)\\ &\geq |||\mathbf{u}-\mathbf{w}_{k+1}|||^2 + (1-\beta)(1-3\beta)|||\mathbf{u}_{k+1}-\mathbf{w}_k|||^2\\ &\geq |||\mathbf{u}-\mathbf{w}_{k+1}|||^2 + (1-\beta)(1-3\beta)\kappa(\mathbf{A})^{-1}(\tfrac{\alpha-\omega}{1+\omega})^2|||\mathbf{u}-\mathbf{w}_k|||^2,\end{aligned}$$

or

$$|||\mathbf{u}-\mathbf{w}_{k+1}||| \leq \big[1-(1-\beta)(1-3\beta)\kappa(\mathbf{A})^{-1}(\tfrac{\alpha-\omega}{1+\omega})^2\big]^{\frac{1}{2}}|||\mathbf{u}-\mathbf{w}_k|||,$$

which completes the proof. □

Remark 2.8. Inside the call of $\mathbf{GROW}[\mathbf{w}_k, \theta\nu_{k-1}, \varepsilon]$ made in **SOLVE**, we search an approximation $\mathbf{r}_{k,\zeta} := \mathbf{RHS}[\zeta/2] - \mathbf{APPLY}[\mathbf{w}_k, \zeta/2]$ for $\bar{\mathbf{r}}_k := \mathbf{f} - \mathbf{A}\mathbf{w}_k$ with a $\zeta \leq \omega\|\mathbf{r}_{k,\zeta}\|$ that is as large as possible in order to minimize $\operatorname{supp}\mathbf{r}_{k,\zeta}$. When $k > 0$, because of the preceding calls of **RHS** and **GALSOLVE**, we have available a set $\Lambda_k \supset \operatorname{supp}\mathbf{w}_k$ and a ν_{k-1} with $\|\mathbf{P}_{\Lambda_k}\bar{\mathbf{r}}_k\| \leq \delta_k := 2\gamma\nu_{k-1}$, whereas on the other hand, for relatively large ζ, $\|\mathbf{P}_{\Lambda_k}\mathbf{r}_{k,\zeta}\|$ might be larger than δ_k. In this remark, we investigate whether it is possible to benefit from the additional information concerning $\mathbf{P}_{\Lambda_k}\bar{\mathbf{r}}_k$.

Let $\mathbf{r}^I_{k,\zeta} := \mathbf{P}_{\Lambda_k}\mathbf{r}_{k,\zeta}$ and $\mathbf{r}^E_{k,\zeta} := \mathbf{P}_{\nabla\setminus\Lambda_k}\mathbf{r}_{k,\zeta}$, and similarly $\bar{\mathbf{r}}^I_k$ and $\bar{\mathbf{r}}^E_k$. From

$$\zeta^2 \geq \|\bar{\mathbf{r}}-\mathbf{r}_{k,\zeta}\|^2 = \|\bar{\mathbf{r}}^I-\mathbf{r}^I_{k,\zeta}\|^2 + \|\bar{\mathbf{r}}^E-\mathbf{r}^E_{k,\zeta}\|^2 \geq (\|\mathbf{r}^I_{k,\zeta}\|-\delta_k)^2 + \|\bar{\mathbf{r}}^E-\mathbf{r}^E_{k,\zeta}\|^2,$$

we have

$$\|\bar{\mathbf{r}}-\mathbf{r}^E_{k,\zeta}\| = (\|\bar{\mathbf{r}}^E-\mathbf{r}^E_{k,\zeta}\|^2 + \|\bar{\mathbf{r}}^I\|^2)^{\frac{1}{2}} \leq (\zeta^2 - (\|\mathbf{r}^I_{k,\zeta}\|-\delta_k)^2 + \delta_k^2)^{\frac{1}{2}} =: \breve{\zeta}.$$

So, alternatively, instead of $\mathbf{r}_{k,\zeta}$, we may use $\mathbf{r}^E_{k,\zeta}$ as an approximation for $\bar{\mathbf{r}}_k$, and thus stop the routine **GROW** as soon as $\nu_k := \|\mathbf{r}^E_{k,\zeta}\| + \breve{\zeta} \leq \varepsilon$ or $\breve{\zeta} \leq \omega\|\mathbf{r}^E_{k,\zeta}\|$, and use $\mathbf{r}^E_{k,\zeta}$ also for the determination of Λ_{k+1}. Since for any ζ and $\mathbf{r}_{k,\zeta}$ with $\mathbf{r}^I_{k,\zeta} \neq 0$ and $\zeta < \|\mathbf{r}_{k,\zeta}\|$, it holds that $\breve{\zeta}\|\mathbf{r}_{k,\zeta}\| < \zeta\|\mathbf{r}^E_{k,\zeta}\|$ if δ_k is small enough, under this condition the alternative test is passed more easily. This may even be a reason to decrease the parameter γ.

The approach discussed in this remark has been applied in the experiments reported in [DHS05].

3. Numerical experiment

We consider the variational formulation of the following problem of order $2t = 2$ on the interval $[0,1]$, i.e., $n = 1$, with periodic boundary conditions

$$-\Delta u + u = f \qquad \text{on } \mathbb{R}/\mathbb{Z}, \tag{3.1}$$

where the right-hand side f is defined by $f(v) = 4v(\frac{1}{2}) + \int_0^1 g(x)v(x)dx$, with

$$g(x) = (16\pi^2+1)\cos(4\pi x) - 4 + \begin{cases} 2x^2, & \text{if } x \in [0,1/2),\\ 2(1-x)^2, & \text{if } x \in [1/2,1],\end{cases} \tag{3.2}$$

so that the solution u is given by

$$u(x) = \cos(4\pi x) + \begin{cases} 2x^2, & \text{if } x \in [0,1/2),\\ 2(1-x)^2, & \text{if } x \in [1/2,1],\end{cases} \tag{3.3}$$

see Figure 1.

We use the periodized B-spline wavelets of order $d = 3$ with 3 vanishing moments from [CDF92]. The solution $u \in H^{s+1}(\mathbb{R}/\mathbb{Z})$ only for $s < \frac{1}{2}$. Since, on the other hand, u can be shown to be in $B^{s+1}_\tau(L_\tau(\mathbb{R}/\mathbb{Z}))$ for any $s > 0$, we deduce that the corresponding discrete solution $\mathbf{u}$ is in ℓ^w_τ for any $s < \frac{d-t}{n} = 2$, where $\tau = (\frac{1}{2}+s)^{-1}$.

We will compare the results of our adaptive wavelet algorithm **SOLVE** with those obtained with the Richardson iteration based method from [CDD02], which we refer as being the CDD2 method, and that reads as follows:

CDD2SOLVE$[\nu,\varepsilon] \to \mathbf{w}$*:*
% $\nu \geq \|\mathbf{u}\|$
% Define the parameters $\omega := \frac{2}{\|\mathbf{A}\|+\|\mathbf{A}^{-1}\|^{-1}}$, and $\rho := \frac{1-\kappa(\mathbf{A})}{1+\kappa(\mathbf{A})}$
% Let θ and K be constants with $2\rho^K < \theta < 1/2$.

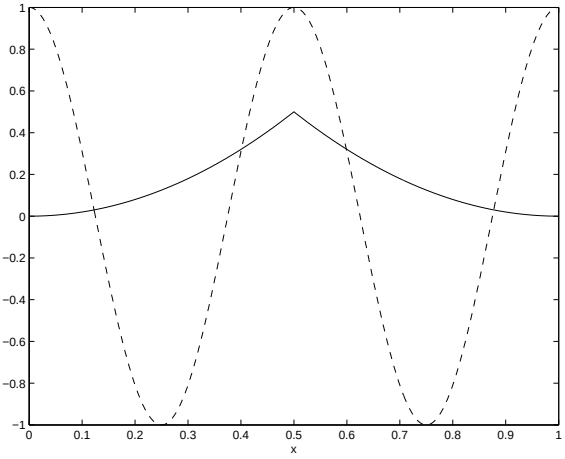

FIGURE 1. The solution u is the sum of both functions illustrated.

$\mathbf{w} := 0$
`while` $\nu > \varepsilon$ `do`
  `for` $j = 1$ `to` K
    $\mathbf{w} := \mathbf{w} + \omega \left(\mathbf{RHS}[\frac{\rho^j \nu}{2\omega K}] - \mathbf{APPLY}[\mathbf{w}, \frac{\rho^j \nu}{2\omega K}] \right)$
  `endfor`
  $\nu := 2\rho^K \nu / \theta$
  $\mathbf{w} := \mathbf{COARSE}[\mathbf{w}, (1-\theta)\nu]$
`enddo`

where the coarsening routine **COARSE** is defined by

$\mathbf{COARSE}[\mathbf{w}, \delta] \to \mathbf{w}_\delta$ *with* $\|\mathbf{w}_\delta - \mathbf{w}\| \leq \delta$*, where* $\#\operatorname{supp} \mathbf{w}_\delta$*, up to some absolute constant factor, is minimal.*

We tested our adaptive wavelet algorithm **SOLVE** or **CDD2SOLVE** with parameters $\mu = 0.4$, $\omega = 0.012618$, and $\gamma = 0.009581$, or $K = 5$ and $\theta = 2/7$, respectively. Inside the ranges where the methods are proven to be of optimal computational complexity, these parameters are close to the values that give quantitatively the best results. We implemented the routine **RHS** by computing all coefficients of $\mathbf{f}$ up to a sufficiently high level, and sorted them beforehand. Now in a call of **RHS**, the largest coefficients are gathered in a vector $\mathbf{g}$ until $\sqrt{\|\mathbf{f}\|^2 - \|\mathbf{g}\|^2}$ is less than the prescribed tolerance. The numerical results, given in Figure 2, illustrate the optimal computational complexity of both **SOLVE** and **CDD2SOLVE**, and show that, in this example, the new method needs less than a factor 10 computing time to achieve the same accuracy.

REFERENCES

[BP88] J.H. Bramble and J.E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Math. Comp.*, 50(181):1–17, 1988.

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

[CDD02] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods II - Beyond the elliptic case. *Found. Comput. Math.*, 2(3):203–245, 2002.

[CDD03] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet schemes for nonlinear variational problems. *SIAM J. Numer. Anal.*, 41:1785–1823, 2003.

[CDF92] A. Cohen, I. Daubechies, and J.C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pur. Appl. Math.*, 45:485–560, 1992.

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

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

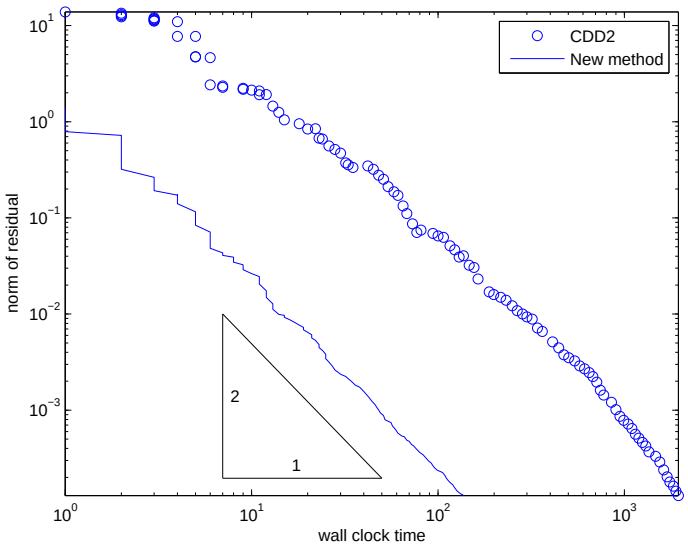

FIGURE 2. Convergence histories

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

[Dah99] S. Dahlke. Besov regularity for elliptic boundary value problems in polygonal domains. *Appl. Math. Lett.*, 12(6):31–36, 1999.

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

[DFR04] S. Dahlke, M. Fornasier, and T. Raasch. Adaptive frame methods for elliptic operator equations. Bericht Nr. 2004-3, Philipps-Universität Marburg, 2004.

[DHS05] W. Dahmen, H. Harbrecht, and R. Schneider. Adaptive application of global operators. IGPM report, RWTH Aachen, 2005.

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

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

[DeV98] R. DeVore. Nonlinear approximation. *Acta Numer.*, 7:51–150, 1998.

[ES04] J. van den Eshof and G.L.G. Sleijpen. Inexact Krylov subspace methods for linear systems. *SIAM J. Matrix Anal. Appl.*, 26(1):125–153, 2004.

[GS04] T. Gantumur and R.P. Stevenson. Computation of differential operators in wavelet coordinates. Technical Report 1306, Utrecht University, August 2004. To appear in *Math. Comp.*

[GS05] T. Gantumur and R.P. Stevenson. Computation of singular integral operators in wavelet coordinates. Technical Report 1321, Utrecht University, January 2005. Submitted.

[HS04] H. Harbrecht and R.P. Stevenson. Wavelets with patchwise cancellation properties. Technical Report 1311, Utrecht University, October 2004. Submitted.

[Ste03] R.P. Stevenson. Adaptive solution of operator equations using wavelet frames. *SIAM J. Numer. Anal.*, 41(3):1074–1100, 2003.

[Ste04a] R.P. Stevenson. Composite wavelet bases with extended stability and cancellation properties. Technical Report 1304, Utrecht University, July 2004. Submitted.

[Ste04b] R.P. Stevenson. On the compressibility of operators in wavelet coordinates. *SIAM J. Math. Anal.*, 35(5):1110–1132, 2004.

DEPARTMENT OF MATHEMATICS, UTRECHT UNIVERSITY, P.O. BOX 80.010, NL-3508 TA UTRECHT, THE NETHERLANDS

E-mail address: {gantumur,harbrecht,stevenson}@math.uu.nl