An adaptive wavelet Newton-Schwarz method for nonlinear elliptic PDEs

Dominik Lellek

October 14, 2015

Abstract

This paper is concerned with an adaptive wavelet method for nonlinear elliptic partial differential equations. The suggested approach combines Newton's method with overlapping domain decomposition techniques. It is shown that the method is convergent and asymptotically optimal. These theoretical results are supported by conclusive numerical experiments.

1 Introduction

Wavelets are a powerful tool in numerical analysis. Their multiscale structure makes them suitable for applications wherever local phenomena are to be resolved efficiently. This includes not only signal processing but also partial differential equations (PDEs). After the groundbreaking paper [3], much progress was made in adaptive numerical methods for such problems. One common property of these methods is that they can often be guaranteed to converge with an asymptotically optimal rate. Most of the work has been done on linear problems, but recently nonlinear PDEs have increasingly been investigated, see for instance [5]. One natural approach considered there is Newton's method, which reduces the nonlinear problem to a sequence of linear problems.

A difficulty in the application of these methods lies in the need for an appropriate wavelet basis. Although there are suitable well-conditioned wavelet bases on the unit cube, see for instance [10, 18], the construction on non-convex domains is technically much more involved and may lead to conditioning problems. In many cases, however, it is possible to avoid these difficulties by decomposing the domain into overlapping subdomains, which in turn are convex. On the subdomains, wavelet bases with favorable properties can then be constructed more easily. Under mild conditions, simply collecting the wavelet bases on the subdomains leads to a possibly overcomplete generating system, called wavelet frame.

Perhaps surprisingly, even though this generating system may be overcomplete, possible redundancies caused by this approach can be controlled for and adaptive wavelet methods based on frames can again be proven to be convergent and asymptotically optimal. First results of this kind for linear problems were shown in [19]. These results were generalized to nonlinear problems in [13]. Because the wavelets frames applied in these methods are constructed by an overlapping domain decomposition, the use of Schwarz methods fits very well into this setting. For linear problems, these methods were analyzed and tested in [20], where a significant improvement of the quantitative behavior compared to standard methods could be observed. These methods could also be extended to a class of nonlinear problems, see [15].

To be able to treat a wider range of nonlinear problems, the combination of wavelet Schwarz methods with Newton's method is a natural further step. In this paper, we construct such a Newton-Schwarz method using wavelet frames. We show that the method is convergent and asymptotically optimal, i.e., that it converges with the same rate as the best N-term approximation in the wavelet frame. Moreover, we present numerical tests to support our theoretical results. This paper partly reports results previously published in [16].

2 Setting

Let Ω be a Lipschitz domain in $\mathbb{R}^d$. In this paper, we are concerned with problems of the form

$$Au + G(u) = f, \tag{1}$$

or, equivalently, $R(u) := Au + G(u) - f = 0$, where $A \in L(H_0^t(\Omega), H^{-t}(\Omega))$ is an operator corresponding to an elliptic symmetric bilinear form, $Av = a(v, \cdot)$, $f \in H^{-t}(\Omega)$ is a given right-hand side and G is a monotonous Nemitsky operator. By this we mean that G is simply given by a monotonously increasing function from $\mathbb{R}$ to $\mathbb{R}$. Let us furthermore assume that G fulfills the polynomial growth condition

$$|G^{(k)}(x)| \lesssim (1 + |x|)^{\max\{0, p-k\}},\ x \in \mathbb{R},\ k = 0, \dots, n^*,\ n^* \geq 2$$

for some $p \geq 0$ with $p < \frac{d+2t}{d-2t}$ if $d > 2t$. Then, it is known from [6, Prop. 4.1] that $G(v) \in H^{-t}(\Omega)$ for all $v \in H_0^t(\Omega)$ and there exists a monotonous, nondecreasing real function C_G such that

$$\|G(v) - G(w)\|_{H^{-t}(\Omega)} \leq C_G(\max\{\|v\|_{H^t(\Omega)}, \|w\|_{H^t(\Omega)}\})\|v - w\|_{H^t(\Omega)}, \quad v, w \in H_0^t(\Omega). \tag{2}$$

Moreover, see [5], it is shown there that for each right-hand side $f \in H^{-t}(\Omega)$ there exists a unique solution $u \in H_0^t(\Omega)$ and a neighborhood $U = B_\delta(u)$ such that for each $v \in U$ the Fréchet derivative $DR(v)$ is an isomorphism from $H_0^t(\Omega)$ to $H^{-t}(\Omega)$ and for any $\bar{u}^{(0)} \in U$ the iterates

$$\bar{u}^{(n+1)} = \bar{u}^{(n)} - DR(\bar{u}^{(n)})^{-1} R(\bar{u}^{(n)}) \tag{3}$$

from the Newton scheme remain in U and are quadratically convergent, $\bar{u}^{(n)} \to u$ with

$$\|\bar{u}^{(n+1)} - u\|_{H^t(\Omega)} \leq \omega \|\bar{u}^{(n)} - u\|_{H^t(\Omega)}^2, \tag{4}$$

and $\|\bar{u}^{(n+1)} - u\|_{H^t(\Omega)} \leq \|\bar{u}^{(n)} - u\|_{H^t(\Omega)}$. The following sections are concerned with the preliminaries needed to formulate an adaptive scheme combining this Newton approach with domain decomposition ideas and wavelet discretization.

2.1 Wavelet bases and frames

To discretize equation (1), we make use of a wavelet frame, which is a possibly overcomplete generating system. First, let us define the basic terms for a general Hilbert space $\mathcal{H}$, see [2] for more details. In the following, we denote by $\langle \cdot, \cdot \rangle_{\mathcal{H} \times \mathcal{H}'}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}' \times \mathcal{H}}$ the dual pairing on $\mathcal{H} \times \mathcal{H}'$ and $\mathcal{H}' \times \mathcal{H}$, respectively, and omit the indices when clear from the context.

Definition 2.1 *A countable set $\Psi := \{\psi_\lambda\}_{\lambda \in \Lambda} \subset \mathcal{H}$ is called a* Riesz basis *for $\mathcal{H}$ if $\overline{\operatorname{span} \Psi} = \mathcal{H}$ and $\|\mathbf{c}^\top \Psi\|_{\mathcal{H}} \eqsim \|\mathbf{c}\|_{\ell_2(\Lambda)}$ for $\mathbf{c} \in \ell_2(\Lambda)$, where $\mathbf{c}^\top \Psi := \sum_\lambda c_\lambda \psi_\lambda$. The set Ψ is called a* frame *for $\mathcal{H}$ if $\|f\|_{\mathcal{H}'} \eqsim \|(\langle f, \psi_\lambda \rangle_{\mathcal{H}' \times \mathcal{H}})_{\lambda \in \Lambda}\|_{\ell_2(\Lambda)}$ holds uniformly in $f \in \mathcal{H}'$.*

It can be shown that for each Riesz basis $\Psi = \{\psi_\lambda\}_{\lambda\in\Lambda}$, there exists a *dual Riesz basis* $\tilde{\Psi} = \{\tilde{\psi}_\lambda\}_{\lambda\in\Lambda}$ for $\mathcal{H}'$ such that the reconstruction formula

$$f = \sum_{\lambda\in\Lambda} \langle f, \tilde{\psi}_\lambda\rangle_{\mathcal{H}\times\mathcal{H}'}\psi_\lambda$$

holds. In the same way, one can define *dual frames*. Contrary to the dual Riesz basis, dual frames are not unique in general. An equivalent characterization of frames is that Ψ is a frame if and only if the operator $F^* : \ell_2(\Lambda) \to \mathcal{H}$, $\mathbf{c} \mapsto \mathbf{c}^\top\Psi$ is well-defined, bounded and surjective. Note that this operator is not required to be injective, hence overcompleteness is allowed for.

In this paper, we make use of wavelet bases and frames for Sobolev spaces. Detailed constructions of wavelet bases can, for instance, be found in [10] and [18], the latter being the construction used in our numerical experiments. By an overlapping domain decomposition approach following the lines from [8, 19], we obtain wavelet frames. Here, we only sketch the construction principles. The starting point usually is a wavelet basis in $L_2([0,1])$ of the form

$$\Psi_{[0,1]} = \{\psi_{j,k}, j \geq j_0 - 1, k \in \nabla_j \subset \mathbb{Z}\},$$

with j being called the *level* of $\psi_{j,k}$. Typically, j is a dilation parameter, k is a translation parameter and the wavelets $\psi_{j,k}$, $j \geq j_0$ are related to a *mother wavelet* ψ by $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ except for modifications at the boundary. By a tensor product approach, from the basis on the interval one can construct a wavelet basis on the unit cube, $\Psi_\square = \{\psi_\lambda, \lambda \in \Lambda\}$, where by $|\lambda|$, we denote the level of the wavelets. Moreover, the wavelet basis can be constructed in a way that norm equivalences with Sobolev spaces hold and that Dirichlet boundary conditions are incorporated, meaning that for a range of parameters s, a scaled version $\Psi_{H_0^s(\square)} = \{2^{-s|\lambda|}\psi_\lambda, \lambda \in \Lambda\}$ is a Riesz basis for $H_0^s(\square)$. In addition, for suitable pairs $(l, \tilde{l})$ of natural numbers with $\tilde{l} \geq l$, the basis can be designed so that it has order l, i.e., the ψ_λ, $|\lambda| = j_0 - 1$ can reproduce polynomials up to degree $l - 1$, and the wavelets ψ_λ, $|\lambda| \geq j_0$ have $\tilde{l}$ vanishing moments. Such a basis can easily be transformed to affine images of the unit cube, but the construction on more complicated, especially non-convex, domains becomes substantially more difficult.

Instead, let us assume we can decompose the domain into *overlapping* images of the unit cube $\Omega = \bigcup_{i=0}^{m-1}\Omega_i$ such that there exists a *partition of unity*, that is functions $\sigma_i : \Omega \to \mathbb{R}$, $i = 0, \ldots, m-1$ such that

- $\sum_{i=0}^{m-1}\sigma_i \equiv 1$ in Ω,
- $\operatorname{supp}\sigma_i \subset \bar{\Omega}_i$,
- $\sigma_i v \in H_0^t(\Omega_i)$ for all $v \in H_0^t(\Omega)$,
- $\|\sigma_i v\|_{H^t(\Omega)} \lesssim \|v\|_{H^t(\Omega)}$ uniformly for all $v \in H^t(\Omega)$.

Under this condition, a frame is obtained simply by collecting wavelet bases on the subdomains.

Proposition 2.2 *([21], Proposition 2.8) Assume there exists a partition of unity subject to the above conditions, let Ψ_i, $i = 0, \ldots, m-1$ be frames or Riesz bases for $H_0^t(\Omega_i)$. Then, $\Psi := \bigcup_{i=0}^{m-1}\Psi_i$ is a frame for $H_0^t(\Omega)$, called* aggregated frame. *Here and in the following, a function in $H_0^t(\Omega_i)$ is also understood as a function in $H_0^t(\Omega)$ by continuing it with zero outside Ω_i.*

In the following, we will write the index set belonging to the frame Ψ as $\Lambda = \bigcup_{i=0}^{m-1}\{i\} \times \Lambda_i$, where Λ_i is the index set belonging to Ψ_i. Moreover, with the help of the partition of unity, a dual frame can be constructed. To do so, let $\tilde{\Psi}_i = \{\tilde{\psi}_\lambda, \lambda \in \Lambda_i\}$ be the dual Riesz bases on the subdomains Ω_i. Setting

$$\tilde{\Psi} := \bigcup_{i=0}^{m-1} \{\sigma_i \tilde{\psi}_\lambda, \lambda \in \Lambda_i\}, \tag{5}$$

we indeed obtain a dual frame to Ψ, see [8].

Let us remark that the constraints on the partition of unity are not too restrictive, because the functions σ_i do not necessarily have to be smooth. For the prototype of a non-convex polygonal domain in $\mathbb{R}^2$, the L-shaped domain $\Omega = (-1,1)^2 \setminus [0,1)^2$ decomposed by the rectangles $\Omega_0 = (-1,0) \times (-1,1)$ and $\Omega_1 = (-1,1) \times (-1,0)$, a suitable partition of unity is constructed in [8].

2.2 Discretization with wavelet frames

Having at hand a wavelet frame Ψ for $H_0^t(\Omega)$ constructed as sketched in Section 2.1, we can discretize equation (1), compare, e.g., [3, 19]. By rewriting the original equation in the coordinates of the wavelet frame, we obtain the discrete equation

$$\mathbf{A}\mathbf{u} + \mathbf{G}(\mathbf{u}) = \mathbf{f}, \tag{6}$$

in $\ell_2(\Lambda)$, where $\mathbf{A} := \{a(\psi_\lambda, \psi_\mu)\}_{\lambda,\mu\in\Lambda}$ is the infinite-dimensional stiffness matrix, $\mathbf{G}(\mathbf{v}) := \{\langle G(\mathbf{v}^\top\Psi), \psi_\lambda\rangle\}_{\lambda\in\Lambda}$ is the discrete nonlinearity, and $\mathbf{f} := \{\langle f, \psi_\lambda\rangle\}_{\lambda\in\Lambda}$ is the right-hand side. Alternatively, Equation 6 will also be formulated as $\mathbf{R}(\mathbf{u}) := \mathbf{A}\mathbf{u} + \mathbf{G}(\mathbf{u}) - \mathbf{f} = \mathbf{0}$. The solutions of Equation (6) are only unique up to elements in $\ker F^*$. However, from each discrete solution $\mathbf{u} \in \ell_2(\Lambda)$, the unique solution $u \in H_0^t(\Omega)$ can be retrieved by $u = \mathbf{u}^\top\Psi$, see [14] for proofs. Nevertheless, it will sometimes be necessary to restrict ourselves to one particular representation of the solution $u \in H_0^t(\Omega)$. To do so, according to [19, 21], with the help of the special dual frame from Section 2.1, we define the linear operator $\mathbf{P} : \ell_2(\Lambda) \to \ell_2(\Lambda)$ by

$$(\mathbf{P}\mathbf{v})_{|\Lambda_i} = \{\langle \mathbf{v}^\top\Psi, \sigma_i\tilde{\psi}_\lambda\rangle_{H_0^t(\Omega)\times H^{-t}(\Omega)}\}_{\lambda\in\Lambda_i}. \tag{7}$$

Then, $\mathbf{P}$ is a projector onto its range and has the same kernel as $\mathbf{A}$. The application of $\mathbf{P}$ to $\mathbf{v}$ corresponds to calculating the expansion coefficients $\mathbf{w}_i$ of $\sigma_i(\mathbf{v}^\top\Psi) \in H_0^t(\Omega_i)$ in the basis Ψ_i for $H_0^t(\Omega_i)$ and setting $\mathbf{P}\mathbf{v} = (\mathbf{w}_0, \dots, \mathbf{w}_{m-1})$. In particular, it holds that $(\mathbf{P}\mathbf{v})^\top\Psi = \mathbf{v}^\top\Psi$, hence for any solution $\mathbf{u}$ of the discrete equation (6), $\mathbf{P}\mathbf{u}$ is a solution as well. For a more detailed discussion, we refer to [19, 21]. Moreover, as for instance in [15], a short calculation using the Riesz basis property of Ψ_i and the partition of unity shows the useful norm equivalence

$$\|\mathbf{P}\mathbf{v}\|_{\ell_2(\Lambda)} \eqsim \sum_{i=0}^{m-1} \|(\mathbf{P}\mathbf{v})_{|\Lambda_i}\|_{\ell_2(\Lambda)} \eqsim \sum_{i=0}^{m-1} \|\sigma_i v\|_{H^t(\Omega_i)} \eqsim \|v\|_{H^t(\Omega)} = \|\mathbf{v}^\top\Psi\|_{H^t(\Omega)}. \tag{8}$$

Since we aim at an application of Newtons method, in addition to the above-mentioned matrices and nonliner mappings we also need the derivative of the residual. Written as a matrix in wavelet frame coordinates, this operator is given by

$$D\mathbf{R}(\mathbf{v}) := \mathbf{A} + D\mathbf{G}(\mathbf{v}) = \mathbf{A} + (\langle G'(\mathbf{v}^\top\Psi)\psi_\lambda, \psi_\mu\rangle)_{\lambda,\mu\in\Lambda}.$$

2.3 Nonlinear tree approximation

The aim of this paper is to show that the adaptive Newton-Schwarz algorithm later defined in Section 3 is both convergent and asymptotically optimal. To state more clearly what is meant by asymptotic optimality, we first have to define the appropriate approximation spaces following the lines from [5, 13].

Let $\mathbf{v} \in \ell_2(\Lambda)$ be given. A *best N-term approximation* to $\mathbf{v}$ is a vector $\mathbf{w}_N \in \ell_2(\Lambda)$ with at most N nonzero entries such that $\|\mathbf{v} - \mathbf{w}_N\|_{\ell_2(\Lambda)}$ is minimal. We can say that $\mathbf{v}$ belongs to an approximation space $\mathcal{A}^s$ for a parameter $s > 0$, if $\|\mathbf{v} - \mathbf{w}_N\|_{\ell_2(\Lambda)}$ decays with rate s as N grows. When dealing with nonlinear equations, an additional technical restriction becomes necessary, which we will now explain.

For isotropic wavelet bases such as described in Section 2.1, it is possible to impose a partial ordering on the index sets Λ_i. Roughly speaking, a wavelet index $\mu \in \Lambda_i$ is called a *predecessor* of $\lambda \in \Lambda_i$ if the level of λ is higher than that of μ, $|\lambda| > |\mu|$, and if $\operatorname{supp}\psi_\lambda \subset \operatorname{supp}\psi_\mu$. Then, a set $T_i \subset \Lambda_i$ is called a *tree* if for all $\lambda \in \Lambda_i$ all of its predecessors are also contained in T_i. This concept can easily be generalized to aggregated wavelet frames by saying that $T = \bigcup_{i=0}^{m-1}\{i\} \times T_i \subset \Lambda$ is an *aggregated tree*, if all the $T_i \subset \Lambda_i$ are trees in the above sense. Considering now the error of best N-term approximation in aggregated tree structure

$$\sigma_{N,\mathcal{AT}}(\mathbf{v}) := \inf\{\|\mathbf{v} - \mathbf{w}\|_{\ell_2(\Lambda)},\, \mathbf{w} \in \ell_2(\Lambda),\, \operatorname{supp}(\mathbf{w}) \text{ aggregated tree },\, \#\operatorname{supp}\mathbf{w} \le N\},$$

we can define the approximation space

$$\mathcal{A}^s_{\mathcal{AT}} := \{\mathbf{v} \in \ell_2(\Lambda),\, \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} := \sup_{N\in\mathbb{N}} (N+1)^s \sigma_{N,\mathcal{AT}}(\mathbf{v}) < \infty\},$$

which contains all vectors in $\ell_2(\Lambda)$ that can be approximated with rate s in aggregated tree structure. This space is only slightly smaller than the approximation space $\mathcal{A}^s$ that corresponds to unrestricted N-term approximation, see [5, 13].

Our aim will be to show that if a wavelet frame representation $\mathbf{u} \in \ell_2(\Lambda)$ of the solution $u = \mathbf{u}^\top\Psi$ to (1) is contained in $\mathcal{A}^s_{\mathcal{AT}}$, our algorithm asymptotically realizes the same rate s for parameters in a certain range. Of course, we would also like to know how to determine if and for which range of parameters s a discrete representation of u lies in $\mathcal{A}^s_{\mathcal{AT}}$. One would expect that the higher the smoothness of u, the better it can be approximated in the wavelet frame. Since we are dealing with a possibly overcomplete generating system, one cannot expect to obtain results for an arbitrary representation $u = \mathbf{u}^\top\Psi$. But it turns out that, when the special representation $\mathbf{Pu}$ is considered and smoothness is measured in an appropriate Besov scale, such a characterization can be shown.

Proposition 2.3 *([14], Proposition 3.4.6) Assume there exists an arbitrarily smooth partition of unity corresponding to $\Omega = \bigcup_{i=0}^{m-1}\Omega_i$. Denote by $B^s_{p,q}(\partial\Omega, k)$ the closure of the C^∞-functions that vanish with order k at $\partial\Omega$ in $B^s_{p,q}(\Omega)$. Let l be the order of the wavelet bases. Then, from $\mathbf{v}^\top\Psi \in B^{sd+t}_{\bar\tau,\bar\tau}(\partial\Omega, t-1)$ it follows that $\mathbf{Pv} \in \mathcal{A}^s_{\mathcal{AT}}$ if the parameters fulfill $s \in (0, \frac{l-t}{d})$, $\frac{1}{\bar\tau} < \frac{1}{\tau} = s + \frac{1}{2}$ and $sd + t - \frac{1}{\tau} \notin \{0, \dots, l-2\}$.*

In Proposition 2.3, an arbitrarily smooth partition of unity is required, which unfortunately does not exist for the standard example of the L-shaped domain decomposed by two rectangles. However, even for this special case, a similar result can be shown, see ([14], Proposition 3.4.7). Nevertheless, even for arbitrarily smooth functions, an approximation rate higher than $\frac{l-t}{d}$ cannot be expected.

3 The Newton-Schwarz algorithm

One step in the Newton iteration (3) amounts to solving for $v^{(n)} := \bar{u}^{(n+1)} - \bar{u}^{(n)} \in H_0^t(\Omega)$ the linear problem

$$DR(\bar{u}^{(n)})v^{(n)} = -R(\bar{u}^{(n)}).$$

Defining $a_x(y,z) := \langle DR(x)y, z\rangle = \langle Ay, z\rangle + \langle G'(x)y, z\rangle$, this problem is the same as finding $v^{(n)} \in H_0^t(\Omega)$ such that

$$a_{\bar{u}^{(n)}}(v^{(n)}, w) = -\langle R(\bar{u}^{(n)}), w\rangle, \quad w \in H_0^t(\Omega). \tag{9}$$

The idea of the additive Schwarz method proposed in [11] is to decompose Problem (9) into a sequence of problems on overlapping subdomains. In our context, it is of course natural to make use of the same overlapping domain decomposition $\Omega = \bigcup_{i=0}^{m-1} \Omega_i$ that was applied for the construction of the wavelet frame. Hence, starting from some approximation $v^{(n,k)}$ for $v^{(n)}$, we calculate *local* correction terms $e_i^{(n,k)} \in H_0^t(\Omega_i)$ from the problems

$$a_{\bar{u}^{(n)}}(v^{(n,k)} + e_i^{(n,k)}, w) = -\langle R(\bar{u}^{(n)}), w\rangle, \quad w \in H_0^t(\Omega_i)$$

or, equivalently

$$a_{\bar{u}^{(n)}}(e_i^{(n,k)}, w) = -a_{\bar{u}^{(n)}}(v^{(n,k)}, w) - \langle R(\bar{u}^{(n)}), w\rangle, \quad w \in H_0^t(\Omega_i). \tag{10}$$

It is important to note that the subproblems are indeed local problems on $H_0^t(\Omega_i)$, since we solve for and test against functions from $H_0^t(\Omega_i)$. Moreover, the problems on the subdomains are independent from each other, so that they can easily be solved in parallel. Having calculated the local correction terms $e_i^{(n,k)}$ on each subdomain, the global iterate is updated,

$$v^{(n,k+1)} = v^{(n,k)} + \alpha \sum_{i=0}^{m-1} e_i^{(n,k)}, \tag{11}$$

where $\alpha > 0$ is a relaxation parameter. To discuss the convergence of the iteration (11), an appropriate norm is required. Hence, define $|\!|\!| w |\!|\!|_v^2 := a_v(w,w)$. From the following lemma we see that $|\!|\!| \cdot |\!|\!|_w$, $w \in U$ is indeed a family of equivalent norms to the standard Sobolev norm or the ℓ_2-norm of projected wavelet coefficients.

Lemma 3.1 *It holds that $\|w\|_{H^t(\Omega)} \eqsim |\!|\!| w |\!|\!|_v$ uniformly in $v \in U$. Moreover, there exist constants c_E, $C_E > 0$ independent from $v \in U$ such that*

$$c_E \|\mathbf{P}\mathbf{w}\|_{\ell_2(\Lambda)} \le |\!|\!| \mathbf{w}^\top \Psi |\!|\!|_v \le C_E \|\mathbf{P}\mathbf{w}\|_{\ell_2(\Lambda)}, \quad \mathbf{w} \in \ell_2(\Lambda).$$

Proof. With $w := \mathbf{w}^\top \Psi$, from Remark 3.1 in [5] we obtain

$$c\|w\|_{H^t(\Omega)} \le \|DR(v)w\|_{H^{-t}(\Omega)} \le C\|w\|_{H^t(\Omega)}, \tag{12}$$

where c is a constant and C is only dependent on $\|v\|_{H^t(\Omega)}$. Since we take v from the bounded set U, it holds that $\|w\|_{H^t(\Omega)} \eqsim \|DR(v)w\|_{H^{-t}(\Omega)}$ uniformly for $v \in U$. Hence, on the one hand, we have

$$|\!|\!| w |\!|\!|_v^2 = a_v(w,w) = \langle DR(v)w, w\rangle \le \|DR(v)w\|_{H^{-t}(\Omega)} \|w\|_{H^t(\Omega)} \eqsim \|w\|_{H^t(\Omega)}^2.$$

On the other hand, using that G is monotonous and A is elliptic, we conclude

$$|\!|\!| w |\!|\!|_v^2 = \langle Aw, w\rangle + \langle G'(v)w, w\rangle \ge \langle Aw, w\rangle \eqsim \|w\|_{H^t(\Omega)}^2.$$

The norm equivalence (8) completes the proof. □

The above lemma allows us to relate the convergence of the inner Schwarz iteration to the convergence of the outer Newton iteration. That is because from the standard theory of additive Schwarz methods, we know that for the iterates defined in (11), it holds that

$$|||v^{(n,k+1)} - v^{(n)}|||_{\bar{u}^{(n)}} \leq \rho_n |||v^{(n,k)} - v^{(n)}|||_{\bar{u}^{(n)}}, \tag{13}$$

where the factor ρ_n is given by

$$\rho_n = |||I - \alpha(P_{0,\bar{u}^{(n)}} + \ldots + P_{m-1,\bar{u}^{(n)}})|||_{\bar{u}^{(n)}}.$$

Here, $|||\cdot|||_{\bar{u}^{(n)}}$ is to be understood as the operator norm induced by the identically named norm on $H_0^t(\Omega)$ and $P_{i,v}$ is the $a_v(\cdot,\cdot)$-orthogonal projector on $H_0^t(\Omega_i)$. Moreover, it can be shown that $\rho_n < 1$ if α is sufficiently small, see for instance [21] for a proof in our context. Hence, iteration (11) is geometrically convergent in the $|||\cdot|||_{\bar{u}^{(n)}}$-norm with rate ρ_n.

3.1 Building blocks for the algorithm

To construct an adaptive wavelet-based variant of the Newton-Schwarz iteration sketched before, some elementary building blocks are necessary. Considering the form of the problems (9) to be solved in each inner Schwarz iteration, we should at least be able to approximately evaluate for $\mathbf{z} \in \ell_2(\Lambda)$ the discrete residual

$$\mathbf{R}(\mathbf{z}) := \mathbf{A}\mathbf{z} + \mathbf{G}(\mathbf{z}) - \mathbf{f}.$$

Hence, we must ensure we have at hand methods for the efficient approximate evaluation of the infinite-dimensional matrix-vector product $\mathbf{A}\mathbf{z}$, the discrete nonlinearity $\mathbf{G}(\mathbf{z})$ and the discrete right-hand side $\mathbf{f}$. Fortunately, these building blocks have been designed and optimized in many works before, compare [1, 3, 4, 6, 13] as an incomplete sample. Thus, we just collect them here and state their main properties.

The first building block for the matrix-vector product is called

$$\mathbf{ATAPPLY}[\mathbf{A}, \mathbf{z}, \varepsilon] \mapsto \mathbf{w}_\varepsilon$$

and for any prescribed tolerance $\varepsilon > 0$ computes a sufficiently good approximation $\mathbf{w}_\varepsilon$ in aggregated tree structure, $\|\mathbf{A}\mathbf{z} - \mathbf{w}_\varepsilon\|_{\ell_2(\Lambda)} \leq \varepsilon$. If the matrix $\mathbf{A}$ is s^*-*compressible*, $s^* > s$, meaning that it can be well approximated by sparse matrices, such a method can be constructed, see [13]. For the case of differential operators such as the Laplace operator, the required assumptions on $\mathbf{A}$ are fulfilled, see for instance [20]. This property is also valid for the projector $\mathbf{P}$, see [19]. Hence, the method can also be used for the application of the projector defined in (7). If, in addition, we assume that the matrix is bounded on $\mathcal{A}^s_{\mathcal{AT}}$, we have the sparsity and complexity estimates

$$\|\mathbf{w}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{z}\|_{\mathcal{A}^s_{\mathcal{AT}}}, \quad \#\operatorname{supp}\mathbf{w}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{z}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}, \quad \#\text{flops} \lesssim \varepsilon^{-1/s}\|\mathbf{z}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \mathcal{T}(\mathbf{z}), \tag{14}$$

where $\mathcal{T}(\mathbf{z})$ denotes the size of the smallest aggregated tree that contains the support of $\mathbf{z}$. This boundedness assumption has been verified for a broad range of matrices $\mathbf{A}$ stemming from the discretization of linear elliptic operators, see Lemma 4.2 in [13]. Although expected to hold, for case of the projector $\mathbf{P}$, the assumption has not yet been strictly verified, compare loc.cit.. Alternatively, there are other, slightly more technical constructions of such a projector, for which the assumptions can strictly be verified, see [9].

For the evaluation of the nonlinearity, we make use of a method

$$\mathbf{EVAL}[\mathbf{G}, \mathbf{z}, \varepsilon] \mapsto \mathbf{w}_\varepsilon$$

as constructed in [13], based on [1, 5]. The output of this routine is a vector $\mathbf{w}_\varepsilon$ in aggregated tree structure such that $\|\mathbf{G}(\mathbf{z}) - \mathbf{w}_\varepsilon\|_{\ell_2(\Lambda)} \le \varepsilon$. Moreover, we have the estimates

$$\|\mathbf{w}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1, \quad \#\operatorname{supp}\mathbf{w}_\varepsilon \lesssim \varepsilon^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1). \tag{15}$$

For the right-hand side, as usual, we assume we have at hand a method

$$\mathbf{ATRHS}[\mathbf{f}, \varepsilon] \mapsto \mathbf{f}_\varepsilon$$

such that $\|\mathbf{f} - \mathbf{f}_\varepsilon\|_{\ell_2(\Lambda)} \le \varepsilon$, and

$$\|\mathbf{f}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{f}\|_{\mathcal{A}^s_{\mathcal{AT}}}, \quad \#\operatorname{supp}\mathbf{f}_\varepsilon \le \varepsilon^{-1/s}\|\mathbf{f}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}. \tag{16}$$

Now these three building blocks can be combined. In [13], it is shown that under the above-mentioned conditions the discrete residual approximation

$$\mathbf{w}_\varepsilon := \mathbf{RES}[\mathbf{A}, \mathbf{G}, \mathbf{f}, \mathbf{z}, \varepsilon] := \mathbf{ATAPPLY}[\mathbf{A}, \mathbf{z}, \varepsilon/3] + \mathbf{EVAL}[\mathbf{G}, \mathbf{z}, \varepsilon/3] - \mathbf{ATRHS}[\mathbf{f}, \varepsilon/3]$$

is sufficiently accurate, $\|\mathbf{R}(\mathbf{z}) - \mathbf{w}_\varepsilon\|_{\ell_2(\Lambda)} \le \varepsilon$, and that for a solution $\mathbf{u} \in \mathcal{A}^s_{\mathcal{AT}}$ to (6) and $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$, the estimates

$$\|\mathbf{w}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1, \tag{17}$$

$$\#\operatorname{supp}\mathbf{w}_\varepsilon \lesssim \varepsilon^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1), \tag{18}$$

$$\#\operatorname{flops} \lesssim \varepsilon^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1) + \mathcal{T}(\mathbf{v}) \tag{19}$$

hold.

Moreover, in view of the subproblems (9), an evaluation routine for the application of the derivative of the discrete residual $D\mathbf{R}(\mathbf{v}) = \mathbf{A} + D\mathbf{G}(\mathbf{v})$, $\mathbf{v} \in \ell_2(\Lambda)$ to vectors in $\ell_2(\Lambda_i)$, $0 \le i \le m-1$ becomes necessary. Although in principle this is the sum of two matrix-vector products, the computation of the matrix coefficients in $D\mathbf{G}(\mathbf{v})$ is too expensive for a direct application of **ATAPPLY**, which therefore is only used for the linear part $\mathbf{A}$. Hence, following [5], we consider the term $(D\mathbf{R}(\mathbf{v}))\mathbf{z}$ as a nonlinearity in two arguments and make use of a method

$$\mathbf{EVAL}[D\mathbf{R}, \mathbf{v}, \mathbf{z}, \varepsilon] \mapsto \mathbf{w}_\varepsilon \in \ell_2(\Lambda_i).$$

Besides from $\|D\mathbf{R}(\mathbf{v})\mathbf{z} - \mathbf{w}_\varepsilon\|_{\ell_2(\Lambda_i)} \le \varepsilon$, this routine is supposed to fulfill $\|\mathbf{w}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{z}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1$ and $\#\operatorname{supp}\mathbf{w}_\varepsilon \lesssim \varepsilon^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{z}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1)$. The computational effort is bounded by a constant times $\varepsilon^{-1/s}(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{z}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1) + \mathcal{T}(\mathbf{v}) + \mathcal{T}(\mathbf{z})$ operations. Although the construction of such a method in [5] is so far only carried through in detail for wavelet bases, we expect that these properties can be generalized to wavelet frames following the lines from [13].

An additional routine will be required to obtain an asymptotically optimal balance between the degrees of freedom and the residual norm. The principal idea of the coarsening method

$$\mathbf{ATCOARSE}[\mathbf{z}, \varepsilon] \mapsto \mathbf{w}_\varepsilon$$

is to remove the smallest entries from $\mathbf{z} \in \ell_2(\Lambda)$ to obtain a sparser, tree-structured vector $\mathbf{w}_\varepsilon \in \ell_2(\Lambda)$ with $\|\mathbf{z} - \mathbf{w}_\varepsilon\|_{\ell_2(\Lambda)} \le \varepsilon$. The following property will be useful in the proof of optimality. It basically says that if a vector is close enough to a vector in $\mathcal{A}^s_{\mathcal{AT}}$, the result from coarsening the first vector is also in $\mathcal{A}^s_{\mathcal{AT}}$ and we have an upper estimate for its support.

Lemma 3.2 *([13, Lemma 4.1], [5, Prop. 6.2, 6.3]) Let $\varepsilon > 0$, $\mathbf{v} \in \mathcal{A}^s_{\mathcal{AT}}$, $\mathbf{w} \in \ell_2(\Lambda)$ with $\#\operatorname{supp}\mathbf{w} < \infty$. There exists a constant $C^* > 1$ such that if $\|\mathbf{w}-\mathbf{v}\|_{\ell_2(\Lambda)} \le \frac{\varepsilon}{2C^*+1}$ and $\mathbf{w}_\varepsilon := \mathbf{ATCOARSE}[\mathbf{w}, \frac{2C^*}{2C^*+1}]$, we have*

$$\|\mathbf{w}_\varepsilon\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\mathcal{AT}}}, \quad \#\operatorname{supp}\mathbf{w}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}},$$

and the number of operations needed to compute $\mathbf{w}_\varepsilon$ is bounded by a constant times $\mathcal{T}(\mathbf{w})$.

3.2 The adaptive algorithm

We have now collected all building blocks needed to formulate the adaptive Newton-Schwarz algorithm. However, we first have to fix some constants to simplify the notation.

By definition of a frame and by (8), there exist constants c_Ψ, $C_\Psi > 0$ such that

$$c_\Psi\|\mathbf{P}\mathbf{v}\|_{\ell_2(\Lambda)} \le \|\mathbf{v}^\top\Psi\|_{H^t(\Omega)} \le C_\Psi\|\mathbf{v}\|_{\ell_2(\Lambda)}.$$

We have seen in (13) that the inner Schwarz iteration converges with rate $\rho_n < 1$. To allow for an inexact solution of the subproblems, we will aim a rate $\tilde\rho_n := (1+\rho_n)/2 < 1$. Now consider the following algorithm that combines the solution of the local subproblems with the application of the projector $\mathbf{P}$ and coarsening. The procedure for solving the subproblems will be discussed in more detail afterwards.

Algorithm 1 NewtonSchwarz[ε]

Choose tolerances $\varepsilon_n \searrow 0$ such that $\delta > \varepsilon_0 \ge \|u-u^{(0)}\|_{H^t(\Omega)}$ and
$\varepsilon_{n+1} \ge \frac{1}{1-\eta}(2C^*+1)\frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2$ for some fixed $\eta \in (0,1)$.
while $\varepsilon_n \le \varepsilon$ **do**
  $v^{(n,0)} = 0$
  Choose $k^* \in \mathbb{N}$ such that $\tilde\rho_n^{k^*} \le \frac{1}{4\varepsilon_n}\frac{c_E}{C_E}\frac{c_\Psi}{C_\Psi}(\frac{1}{2C^*+1}\varepsilon_{n+1} - \frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2)$
  for $k = 0,\dots,k^*-1$ **do**
    for $i = 0,\dots,m-1$ **do**
      Compute $\tilde e_i^{(n,k)} = (\tilde{\mathbf{e}}_i^{(n,k)})^\top\Psi_i \in H^t_0(\Omega_i)$ as approximation to $e_i^{(n,k)} \in H^t_0(\Omega_i)$ from
        $a_{u^{(n)}}(w, v^{(n,k)} + e_i^{(n,k)}) = -R(u^{(n)})(w)$, $w \in H^t_0(\Omega_i)$
      up to tolerance $\|\!|e_i^{(n,k)} - \tilde e_i^{(n,k)}|\!\|_{u^{(n)}} \le \frac{1-\rho_n}{m\alpha}\frac{C_E}{c_\Psi}\varepsilon_n\tilde\rho_n^k$.
    end for
    $\mathbf{v}^{(n,k+1)} = \mathbf{v}^{(n,k)} + \alpha\sum_{i=0}^{m-1}\tilde{\mathbf{e}}_i^{(n,k)}$
  end for
  $\hat{\mathbf{u}}^{(n+1)} := \mathbf{ATAPPLY}[\mathbf{P}, \mathbf{u}^{(n)} + \mathbf{v}^{(n,k^*)}, \frac{1}{2C_\Psi}(\frac{1}{2C^*+1}\varepsilon_{n+1} - \frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2)]$
  $\mathbf{u}^{(n+1)} := \mathbf{ATCOARSE}[\hat{\mathbf{u}}^{(n+1)}, \frac{1}{C_\Psi}\frac{2C^*}{2C^*+1}\varepsilon_{n+1}]$
end while
$\mathbf{u}_\varepsilon = \mathbf{u}^{(n+1)}$

Let us first show that the algorithm is indeed locally convergent.

Proposition 3.3 *For the iterates $u^{(n)}$ from Algorithm 1, it holds that*

$$\|u-u^{(n)}\|_{H^t(\Omega)} \le \varepsilon_n.$$

Proof. We use induction over n. For $n = 0$, the claim is valid by definition. To proceed the induction, let $v^{(n)}$ be the exact Newton step, $\bar{u}^{(n+1)} = u^{(n)} + v^{(n)}$, $DR(u^{(n)})v^{(n)} = -R(u^{(n)})$. Then

$$\|\!|v^{(n)}|\!\|_{u^{(n)}} \leq \|\!|\bar{u}^{(n+1)} - u|\!\|_{u^{(n)}} + \|\!|u^{(n)} - u|\!\|_{u^{(n)}} \leq \frac{C_E}{c_\Psi}(\|\bar{u}^{(n+1)} - u\|_{H^t(\Omega)} + \|u^{(n)} - u\|_{H^t(\Omega)}) \leq 2\frac{C_E}{c_\Psi}\varepsilon_n.$$

Because the local subproblems are only solved inexactly, estimate (13) now reads

$$\begin{aligned}\|\!|v^{(n,k+1)} - v^{(n)}|\!\|_{u^{(n)}} \leq& \rho_n \|\!|v^{(n,k)} - v^{(n)}|\!\|_{u^{(n)}} + \alpha \sum_{i=0}^{m-1} \|\!|e_i^{(n,k)} - \tilde{e}_i^{(n,k)}|\!\|_{u^{(n)}} \\ \leq& \rho_n \|\!|v^{(n,k)} - v^{(n)}|\!\|_{u^{(n)}} + (1-\rho_n)\frac{C_E}{c_\Psi}\varepsilon_n \tilde{\rho}_n^k.\end{aligned}$$

By an inner induction over k, using $\|\!|v^{(n,0)} - v^{(n)}|\!\|_{u^{(n)}} = \|\!|v^{(n)}|\!\|_{u^{(n)}} \leq 2\frac{C_E}{c_\Psi}\varepsilon_n$, we obtain

$$\|\!|v^{(n,k^*)} - v^{(n)}|\!\|_{u^{(n)}} \leq 2\tilde{\rho}_n^{k^*}\frac{C_E}{c_\Psi}\varepsilon_n \leq \frac{1}{2}\frac{c_E}{C_\Psi}\left(\frac{1}{2C^*+1}\varepsilon_{n+1} - \frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2\right).$$

Then, we infer $\|\mathbf{P}\mathbf{v}^{(n,k)} - \mathbf{P}\mathbf{v}^{(n)}\|_{\ell_2(\Lambda)} \leq \frac{1}{2C_\Psi}\left(\frac{1}{2C^*+1}\varepsilon_{n+1} - \frac{\omega}{c_\Psi}\varepsilon_n^2\right)$. Let $\bar{u}^{(n+1)} = (\bar{\mathbf{u}}^{(n+1)})^\top \Psi$ be the result of the exact Newton step. From the above and the tolerance for the call of **ATAPPLY**, we have

$$\begin{aligned}\|\hat{\mathbf{u}}^{(n+1)} - \mathbf{P}\bar{\mathbf{u}}^{(n+1)}\|_{\ell_2(\Lambda)} &\leq \|\mathbf{P}(\mathbf{u}^{(n)} + \mathbf{v}^{(n,k^*)}) - \mathbf{P}(\mathbf{u}^{(n)} + \mathbf{v}^{(n)})\|_{\ell_2(\Lambda)} + \frac{1}{2C_\Psi}\left(\frac{1}{2C^*+1}\varepsilon_{n+1} - \frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2\right) \\ &\leq \frac{1}{C_\Psi}\left(\frac{1}{2C^*+1}\varepsilon_{n+1} - \frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2\right). \end{aligned} \tag{20}$$

Using the induction hypothesis, it follows that

$$\|\mathbf{P}\bar{\mathbf{u}}^{(n+1)} - \mathbf{P}\mathbf{u}\|_{\ell_2(\Lambda)}^2 \leq \frac{1}{c_\Psi}\|\bar{u}^{(n+1)} - u\|_{H^t(\Omega)} \leq \frac{1}{c_\Psi}\omega\|u^{(n)} - u\|_{H^t(\Omega)}^2 \leq \frac{1}{c_\Psi}\omega\varepsilon_n^2. \tag{21}$$

Combining the estimates (20) and (21), we have

$$\begin{aligned}\|\hat{\mathbf{u}}^{(n+1)} - \mathbf{P}\mathbf{u}\|_{\ell_2(\Lambda)} \leq& \|\hat{\mathbf{u}}^{(n+1)} - \mathbf{P}\bar{\mathbf{u}}^{(n+1)}\|_{\ell_2(\Lambda)} + \|\mathbf{P}\bar{\mathbf{u}}^{(n+1)} - \mathbf{P}\mathbf{u}\|_{\ell_2(\Lambda)} \\ \leq& \frac{1}{C_\Psi}\left(\frac{1}{2C^*+1}\varepsilon_{n+1} - \frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2\right) + \omega\frac{1}{c_\Psi}\varepsilon_n^2 \\ =& \frac{1}{C_\Psi}\frac{1}{2C^*+1}\varepsilon_{n+1}. \end{aligned} \tag{22}$$

With the tolerance for the coarsening step, it holds that

$$\begin{aligned}\|\mathbf{u}^{(n+1)} - \mathbf{P}\mathbf{u}\|_{\ell_2(\Lambda)} \leq& \|\mathbf{u}^{(n+1)} - \hat{\mathbf{u}}^{(n+1)}\|_{\ell_2(\Lambda)} + \|\hat{\mathbf{u}}^{(n+1)} - \mathbf{P}\mathbf{u}\|_{\ell_2(\Lambda)} \\ \leq& \frac{1}{C_\Psi}\frac{2C^*}{2C^*+1}\varepsilon_{n+1} + \frac{1}{C_\Psi}\frac{1}{2C^*+1}\varepsilon_{n+1} \\ =& \frac{1}{C_\Psi}\varepsilon_n.\end{aligned}$$

Finally, we conclude

$$\|u^{(n+1)} - u\|_{H^t(\Omega)} \leq C_\Psi\|\mathbf{u}^{(n+1)} - \mathbf{P}\mathbf{u}\|_{\ell_2(\Lambda)} \leq C_\Psi\frac{1}{C_\Psi}\varepsilon_{n+1} = \varepsilon_{n+1}.$$

□

Now, using the properties of the coarsening method it is easy to see that the outer iterates $\mathbf{u}^{(n)}$ converge with an asymptotically optimal rate with respect to the degrees of freedom.

Proposition 3.4 *Let $\mathbf{u} \in \mathcal{A}^s_{\mathcal{AT}}$ be a solution to problem (6). Then, all $\mathbf{u}^{(n)}$, $n \geq 1$, are contained in $\mathcal{A}^s_{\mathcal{AT}}$ with*

$$\|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}$$

and

$$\#\operatorname{supp}\mathbf{u}^{(n)} \lesssim \varepsilon_n^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}.$$

Proof. As seen in (22), it holds that $\|\hat{\mathbf{u}}^{(n+1)} - \mathbf{Pu}\|_{\ell_2(\Lambda)} \leq \frac{1}{C_\Psi}\frac{1}{2C^*+1}\varepsilon_{n+1}$. Then, from Lemma 3.2, we infer that $\|\mathbf{u}^{(n+1)}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{Pu}\|_{\mathcal{A}^s_{\mathcal{AT}}}$ and $\#\operatorname{supp}\mathbf{u}^{(n+1)} \lesssim (\frac{1}{C_\Psi}\varepsilon_{n+1})^{-1/s}\|\mathbf{Pu}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \varepsilon_{n+1}^{-1/s}\|\mathbf{Pu}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}}$. The assertion follows from the assumption that $\mathbf{P}$ is bounded on $\mathcal{A}^s_{\mathcal{AT}}$. □

3.3 Approximate solution of the subproblems

To construct an implementable version of Algorithm 1, it remains to specify how to solve the linear subproblems

$$a_{u^{(n)}}(w, e_i^{(n,k)}) = -R(u^{(n)})(w) - a_{u^{(n)}}(w, v^{(n,k)}), \quad w \in H^t_0(\Omega_i)$$

on the subdomains Ω_i up to the given tolerance. In discretized form, the subproblems read as

$$(D\mathbf{R}(\mathbf{u}^{(n)})^{(i)})\mathbf{e}_i^{(n,k)} = -\mathbf{R}(\mathbf{u}^{(n)})_{|\Lambda_i} - (D\mathbf{R}(\mathbf{u}^{(n)}))\mathbf{v}^{(n,k)}_{|\Lambda_i},$$

where $(D\mathbf{R}(\mathbf{u}^{(n)})^{(i)})$ denotes the diagonal block of $(D\mathbf{R}(\mathbf{u}^{(n)}))$ belonging to Λ_i. Because of (12) and because the Ψ_i are Riesz bases, the matrices $D\mathbf{R}(\mathbf{u}^{(n)})^{(i)}$ are symmetric positive definite and boundedly invertible with constants independent from $u^{(n)} = (\mathbf{u}^{(n)})^\top\Psi \in U$. Moreover $\|\mathbf{v}\|^2_{i,\mathbf{u}^{(n)}} := \langle (D\mathbf{R}(\mathbf{u}^{(n)}))^{(i)}\mathbf{v}, \mathbf{v}\rangle_{\ell_2(\Lambda_i)}$ is a norm on $\ell_2(\Lambda_i)$ equivalent to the standard ℓ_2-norm. In addition, there exists a relaxation parameter $\gamma > 0$ such that in the corresponding operator norm, for all iterates $\mathbf{u}^{(n)}$ it holds that

$$\|I - \gamma D\mathbf{R}(\mathbf{u}^{(n)})^{(i)}\|_{i,\mathbf{u}^{(n)}} \leq \xi < 1,$$

and we can apply a perturbed Richardson iteration following the lines from [4] and [21]. For convenience, we state the algorithm explicitly.

Algorithm 2 LocSolve[tol]

Let $S \geq \|\mathbf{e}_i^{(n,k)}\|_{i,\mathbf{u}^{(n)}}$
Let $p \in \mathbb{N}$ such that $2\xi^p(S/tol + \frac{1}{2}) \leq \frac{1}{2}$
Let $tol_{\mathbf{r}} := \frac{1}{4}\|(D\mathbf{R}(\mathbf{u}^{(n)})^{(i)})^{-1}\|^{-1/2}_{\ell_2(\Lambda_i)\to\ell_2(\Lambda_i)}tol$
$\mathbf{r} := -\mathbf{RES}[\mathbf{A}, \mathbf{G}, \mathbf{f}, \mathbf{u}^{(n)}, tol_{\mathbf{r}}]_{|\Lambda_i} - \mathbf{EVAL}[D\mathbf{R}, \mathbf{u}^{(n)}, \mathbf{v}^{(n,k)}, tol_{\mathbf{r}}]_{|\Lambda_i}$
$\mathbf{z}^{(0)} := \mathbf{0}$
for $r = 0, \ldots, p-1$ **do**
  $\mathbf{z}^{(r+1)} = \mathbf{z}^{(r)} + \gamma(\mathbf{r} - \mathbf{EVAL}[D\mathbf{R}^{(i)}, \mathbf{u}^{(n)}, \mathbf{z}^{(r)}, \frac{\xi^p(S/tol+\frac{1}{2})tol}{\gamma p\|(D\mathbf{R}(\mathbf{u}^{(n)})^{(i)})\|^{1/2}_{\ell_2(\Lambda_i)\to\ell_2(\Lambda_i)}}])$
end for
$\tilde{\mathbf{e}}_i^{(n,k)} = \mathbf{z}^{(p)}$

The following lemma states that the local solver indeed achieves the desired tolerance. The proof is analogous to Lemma 6.3 in [21], so we omit it.

Lemma 3.5 *For the output $\tilde{\mathbf{e}}_i^{(n,k)}$ from Algorithm 2, it holds that $\|\tilde{\mathbf{e}}_i^{(n,k)} - \mathbf{e}_i^{(n,k)}\|_{i,\mathbf{u}^{(n)}} \leq tol$.*

Note that because of the Riesz basis property of the Ψ_i, we have $\|\!| e_i^{(n,k)} - \tilde{e}_i^{(n,k)} |\!\|_{u^{(n)}} \eqsim \|\!| \mathbf{e}_i^{(n,k)} - \tilde{\mathbf{e}}_i^{(n,k)} |\!\|_{i,\mathbf{u}^{(n)}}$, so it is sufficient to achieve a given error tolerance in the discrete norm. Therefore this local solver can be applied within Algorithm 1.

Remark 3.6 *Since the local subproblems are formulated in $H_0^t(\Omega_i)$, for which we have the non-redundant local Riesz bases Ψ_i, it is also possible to solve them using an adaptive wavelet Galerkin method, see, e.g., [3]. This approach might lead to better quantitative results in practice. However, the theoretical analysis is more transparent when considering a Richardson iteration. We expect that the following complexity estimates, albeit with a more tedious notation, also carry over with a Galerkin approach to the local problems.*

3.4 Complexity estimates

To estimate the overall complexity of the algorithm, two steps will be necessary. First of all, we have to analyze the number of calls of the local solvers. Then, we have to estimate the complexity of the local solver itself. For the first part, the following lemma, that gives a uniform bound for the convergence rate of the outer Schwarz algorithm, will be useful.

Lemma 3.7 *There exists a constant $\rho < 1$ such that for all $v \in U$, it holds that*

$$\rho(v) := \|\!| I - \alpha(P_{0,v} + \ldots + P_{m-1,v}) |\!\|_v \leq \rho.$$

In particular, we have $\rho_n \leq \rho$ for all $n \in \mathbb{N}$.

Proof. Since the subdomains are assumed to be overlapping, each $w \in H_0^t(\Omega)$ can be decomposed as $w = \sum_{i=0}^{m-1} w_i$ with $w_i \in H_0^t(\Omega_i)$ and $\sum_{i=0}^{m-1} \|w_i\|_{H^t(\Omega)} \lesssim \|w\|_{H^t(\Omega)}$. This can be realized for instance by choosing $w_i = \sigma_i w$. Moreover, because $\|\!| \cdot |\!\|_v \eqsim \|\cdot\|_{H^t(\Omega)}$ holds uniformly in $v \in U$, there is a constant $a \in \mathbb{R}$ such that $\sum_{i=0}^{m-1} \|\!| w_i |\!\|_v \leq a \|\!| w |\!\|_v$.

Define the operator $T_v := \sum_{i=0}^{m-1} P_{i,v}$. This operator is symmetric and positive definite with respect to the inner product $a_v(\cdot,\cdot)$, see [21, Thm. 6.5] for details. As in the proof of Theorem 8.3 in [17], we see that a lower bound for its eigenvalues is given by $\lambda_{\min}(T_v) \geq \frac{1}{a^2}$. Since T_v is a sum of m orthogonal projectors, an upper $\lambda_{\max}(T_v)$ for its eigenvalues is given by m. By standard arguments, we see that

$$\rho(v) = \|\!| I - \alpha T |\!\|_v \leq \max\{1 - \alpha\lambda_{\min}(T_v), \alpha\lambda_{\max}(T_v) - 1\} \leq \max\{1 - \frac{\alpha}{a^2}, \alpha m - 1\} =: \rho,$$

which is smaller than one for all $\alpha \in (0, \frac{2}{m})$. □

Because of Lemma 3.7, we can replace all the ρ_n in Algorithm 1 with the constant ρ. Doing so, we see that the condition on the number k^* of inner iterations within Algorithm 1, namely

$$\rho^{k^*} \leq \frac{1}{4\varepsilon_n} \frac{c_E}{C_E} \frac{c_\Psi}{C_\Psi} \left(\frac{1}{2C^* + 1} \varepsilon_{n+1} - \frac{C_\Psi}{c_\Psi} \omega \varepsilon_n^2\right),$$

where $\frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2 \leq (1-\eta)\frac{1}{2C^*+1}\varepsilon_{n+1}$, is fulfilled whenever $k^* \geq \frac{\log(K\varepsilon_{n+1}/\varepsilon_n)}{\log\rho}$ with constant $K = \frac{c_E c_\Psi \eta}{4C_E C_\Psi (2C^*+1)}$. In particular, if the tolerances are chosen in a geometrically decreasing fashion, i.e., $\varepsilon_{n+1} = q\varepsilon_n$, $q \in (0,1)$, the number of k^* of inner iterations is independent of n. In this case, we can show that the inner iterates $v^{(n,k)}$ are in the same sparsity class as the exact solution u. This will allow us to adapt the optimality proof from [20] to our situation.

Lemma 3.8 *If the tolerances ε_n are chosen to be geometrically decreasing, then for all n and k it holds that*

$$\|\mathbf{v}^{(n,k)}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1$$

and

$$\#\operatorname{supp}\mathbf{v}^{(n,k)} \lesssim \varepsilon_n^{-1/s}(\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1).$$

Proof. For each n, we show the result by induction over k. The case $k = 0$ is clear by definition of $v^{(n,0)}$. Let us now proceed from k to $k+1$. By the properties (15), (17) and (18) of the methods **RES** and **EVAL**, we have

$$\|\mathbf{r}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{v}^{(n,k)}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1. \tag{23}$$

and

$$\#\operatorname{supp}\mathbf{r} \lesssim \varepsilon_n^{-1/s}(\|\mathbf{u}^{(n)}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{v}^{(n,k)}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1). \tag{24}$$

Hence, from the induction hypothesis and Proposition 3.4, it holds that $\|\mathbf{r}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1$. In the same way as in Equation (10), for all $w \in H^t_0(\Omega_i)$ it follows that

$$a_{u^{(n)}}(e_i^{(n,k)}, w) = -a_{u^{(n)}}(v^{(n,k)}, w) - \langle R(u^{(n)}), w\rangle = a_{u^{(n)}}(-v^{(n,k)} - (DR(u^{(n)})^{-1}R(u^{(n)}), w).$$

Hence, $e_i^{(n,k)} = P_{i,u^{(n)}}(-v^{(n,k)} - (DR(u^{(n)})^{-1}R(u^{(n)}))$. Because $P_{i,u^{(n)}}$ is an orthogonal projector with respect to $-a_{u^{(n)}}(\cdot,\cdot)$ and $DR(u^{(n)})$ is an isomorphism with constants independent from $u^{(n)}$, see (12), it holds that $\|\!|e_i^{(n,k)}|\!\|_{u^{(n)}} \lesssim \|\!|v^{(n,k)}|\!\|_{u^{(n)}} + \|\!|R(u^{(n)})|\!\|_{u^{(n)}}$. Moreover, by the stability estimate (2) and the boundedness of A, we have

$$\|\!|R(u^{(n)})|\!\|_{u^{(n)}} \eqsim \|R(u^{(n)})\|_{H^t(\Omega)} = \|R(u^{(n)}) - R(u)\|_{H^t(\Omega)} \lesssim \|u^{(n)} - u\|_{H^t(\Omega)} \le \varepsilon_n.$$

In particular, because of $u^{(n)} \in B_\delta(u)$, the implicit constants in the above estimates can be chosen independently from $u^{(n)}$. Moreover, because of $v^{(n,k)} = u^{(n+1)} - u^{(n)}$, we have

$$\|\!|v^{(n,k)}|\!\|_{u^{(n)}} \eqsim \|v^{(n,k)}\|_{H^t(\Omega)} \le \|u^{(n+1)} - u\|_{H^t(\Omega)} + \|u^{(n)} - u\|_{H^t(\Omega)} \le \varepsilon_{n+1} + \varepsilon_n \lesssim \varepsilon_n.$$

Hence, it follows that $\|\!|e_i^{(n,k)}|\!\|_{u^{(n)}} \lesssim \varepsilon_n$. Because the $\Psi^{(i)}$ are Riesz bases for the subdomains and the energy norms are equivalent to the Sobolev and the ℓ_2-norm, we have a similar estimate $\|\!|\mathbf{e}_i^{(n,k)}|\!\|_{u^{(n)}} \lesssim \varepsilon_n$ for the discrete solutions of the subproblems. Thus, the parameter S within LocSolve can be chosen as a constant times ε_n. Because the tolerance *tol* for the local solvers fulfills $tol \eqsim \varepsilon_{n+1}$, for the number p of iterations within the local solvers it holds that $p \lesssim \log(\frac{\varepsilon_n}{\varepsilon_{n+1}}) + 1$. In particular, for geometrically decreasing ε_n, p can be chosen as a constant independent of n, i and k. By definition of $\mathbf{z}^{(r+1)}$, the properties (15) of **EVAL** and the estimates (23) and (24) we have

$$\|\mathbf{z}^{(r+1)}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{z}^{(r)}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{v}^{(n,k)}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1, \quad r = 0, \dots, p-1. \tag{25}$$

and

$$\#\operatorname{supp}\mathbf{z}^{(r+1)} \lesssim \#\operatorname{supp}\mathbf{z}^{(r)} + \varepsilon_n^{-1/s}(\|\mathbf{u}^{(n)}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + \|\mathbf{v}^{(n,k)}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1), \quad r = 0, \dots, p-1. \tag{26}$$

Since p is constant $\mathbf{z}^{(0)} = \mathbf{0}$, by induction and Proposition 3.4, it follows that $\|\tilde{\mathbf{e}}_i^{(n,k)}\|_{\mathcal{A}^s_{\mathcal{AT}}} = \|\mathbf{z}^{(p)}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1$ and $\#\operatorname{supp}\tilde{\mathbf{e}}_i^{(n,k)} \lesssim \varepsilon_n^{-1/s}(\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1)$. By definition of $\mathbf{v}^{(n,k+1)}$ and the induction hypothesis, we conclude that $\|\mathbf{v}^{(n,k+1)}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}} + 1$ and $\#\operatorname{supp}\mathbf{v}^{(n,k+1)} \lesssim \varepsilon_{n+1}^{-1/s}(\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\mathcal{AT}}} + 1)$ Because $0 \le k \le k^* - 1$ and k^* is a constant, the implicit constants in the above estimates can be chosen independently of k and n. □

With these key results at hand, we can proceed following lines from [15, 20, 21] and estimate the computational effort needed to compute the local correction terms.

Lemma 3.9 *The computational effort for the computation of* $\tilde{\mathbf{e}}_i^{(n,k)}$ *within LocSolve is bounded by a constant times* $\varepsilon_n^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$ *operations.*

Proof. By the properties of **EVAL** and the estimate (19) for **RES**, the computation of the right-hand side $\mathbf{r}$ requires a constant times

$$tol_{\mathbf{r}}^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+\|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)+\mathcal{T}(\mathbf{u}^{(n)})+tol_{\mathbf{r}}^{-1/s}(\|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+\|\mathbf{v}^{(n,k)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)+\mathcal{T}(\mathbf{v}^{(n,k)})$$

operations. Because $tol_{\mathbf{r}} \eqsim \varepsilon_n$ and the vectors $\mathbf{u}^{(n)}$ and $\mathbf{v}^{(n,k)}$ are tree-structured, from Proposition 3.4 and Lemma 3.8, we conclude that this quantity can be bounded by a constant times $\varepsilon_n^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$. Let us now consider the computation of $\mathbf{z}^{(r+1)}$ from $\mathbf{z}^{(r)}$. Since the tolerance for the call of **EVAL** is a constant times ε_n, this step requires a computational effort of the order

$$\varepsilon_n^{-1/s}(\|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+\|\mathbf{z}^{(r)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)+\mathcal{T}(\mathbf{u}^{(n)})+\mathcal{T}(\mathbf{z}^{(r)}).$$

Since $\mathbf{z}^{(r)}$ and $\mathbf{u}^{(n)}$ are tree-structured and $\|\mathbf{z}^{(r)}\|_{\mathcal{A}^s_{\mathcal{AT}}} \lesssim \|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}}+1$ and by Proposition 3.4, similarly to the proof of Lemma 3.8, we see that the effort for the computation of $\mathbf{z}^{(p)}$ is bounded by a constant times $\varepsilon_n^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$ operations, which shows the assertion. □

With the estimates shown so far, we are now ready to prove that the Newton-Schwarz algorithm is asymptotically optimal with respect to the computational effort.

Proposition 3.10 *If the tolerances* ε_n *are chosen to be geometrically decreasing, the computational effort for the calculation of* $\mathbf{u}^{(n)}$ *is a constant multiple of* $\varepsilon_n^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$ *operations.*

Proof. Let us consider the calculation of $\mathbf{u}^{(n+1)}$ from $\mathbf{u}^{(n)}$ within Algorithm 1. By Lemma 3.9 and the fact that k^* is constant, the computational effort for the calculation of $\mathbf{v}^{(n,k^*)}$ is of the order $\varepsilon_n^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$. Because of $\varepsilon_{n+1} \geq \frac{1}{1-\eta}(2C^*+1)\frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2$, it holds that $\frac{1}{2C_\Psi}(\frac{1}{2C^*+1}\varepsilon_{n+1}-\frac{C_\Psi}{c_\Psi}\omega\varepsilon_n^2) \eqsim \varepsilon_{n+1}$. Hence, using that both $\mathbf{u}^{(n)}$ and $\mathbf{v}^{(n,k^*)}$ are tree-structured, the computational effort for the application of the projector via **ATAPPLY** is bounded by a constant times $\varepsilon_{n+1}^{-1/s}(\|\mathbf{u}^{(n)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+\|\mathbf{v}^{(n,k^*)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s})+\#\operatorname{supp}\mathbf{u}^{(n)}+\#\operatorname{supp}\mathbf{v}^{(n,k^*)}$. By Proposition 3.4 and Lemma 3.8, this quantity is of the order $\varepsilon_{n+1}^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$. By Lemma 3.2, the effort for the final coarsening step is a multiple of $\#\operatorname{supp}\tilde{\mathbf{u}}^{(n+1)}$, which, by (14), Proposition 3.4 and Lemma 3.8, is of the order $\varepsilon_{n+1}^{-1/s}\|\mathbf{u}^{(n)}+\mathbf{v}^{(n,k^*)}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+\mathcal{T}(\mathbf{u}^{(n)}+\mathbf{v}^{(n,k^*)}) \lesssim \varepsilon_{n+1}^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$. Hence, the overall costs for the computation of $\mathbf{u}^{(n+1)}$ from $\mathbf{u}^{(n)}$ is of the order $\varepsilon_{n+1}^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}+1)$. From this, the assertion follows by a standard geometric series argument. □

Summing up the preceding results, we have seen that the algorithm is convergent and asymptotically optimal. This will be summarized in the following main theorem of this paper.

Theorem 3.11 *Let the assumptions from Subsection 3.1 hold and the* ε_n *be geometrically decreasing. Let* $u=\mathbf{u}^\top\Psi$ *be the exact solution of problem* (1) *with* $\mathbf{u}\in\mathcal{A}^s_{\mathcal{AT}}$. *Then, for any* $\varepsilon>0$, *the adaptive Newton-Schwarz algorithm with LocSolve as a solver for the local subproblems computes an approximation* $u_\varepsilon := (\mathbf{u}_\varepsilon)^\top\Psi$ *to* u *with*

$$\|u-u_\varepsilon\|_{H^t(\Omega)} \leq \varepsilon, \quad \#\operatorname{supp}\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{AT}}}^{1/s}.$$

The overall computational cost is bounded by a constant times $\varepsilon^{-1/s}(\|\mathbf{u}\|_{\mathcal{A}^s_{\mathcal{A}\mathcal{T}}}^{1/s} + 1)$ *operations.*

Proof. The convergence follows from Proposition 3.3. The support estimate follows from Proposition 3.4, using that $\varepsilon_{n+1} \eqsim \varepsilon$ in the iteration when the algorithm terminates. The bound on the costs follows from Proposition 3.10. □

4 Numerical tests

To observe whether the predicted asymptotic behavior of the algorithm can be observed in practical implementations at moderate scales, it is necessary to run numerical tests. The following results are taken from [16]. We apply our algorithm to the standard test problem

$$-\Delta u + u^3 = f \tag{27}$$

with Dirichlet boundary conditions on the L-shaped domain with the domain decomposition defined in Section 2.1. In [5], it is shown that this problem indeed fits into our general setting described in Section 2. As the exact solution, we fix the function

$$u(r,\theta) := \eta(r) r^{2/3} \sin(\frac{2}{3}\theta) \tag{28}$$

written in polar coordinates, where η is a C^∞-function with $\xi(r) = 1$ for $r \leq 0$ and $\xi(r) = 0$ for $r \geq 1$. The exact solution is depicted in Figure 1. The right-hand side is calculated accordingly. This choice is motivated by the general theory for linear problems in [12], which shows that the solution of a linear Poisson equation can be written as a sum of a regular and a singular part of the form (28). Moreover, it is shown there that the singular part has a limited Sobolev regularity, $u \in H^s(\Omega)$ only for $s < \frac{5}{3}$, which is due to the reentrant corner at the origin. In [7], however, it is shown that u belongs to all Besov spaces $B^s_{\tau,\tau}(\Omega)$, $\frac{1}{\tau} = \frac{s}{2} + \frac{1}{2}$. Therefore, the application of an adaptive method is justified.

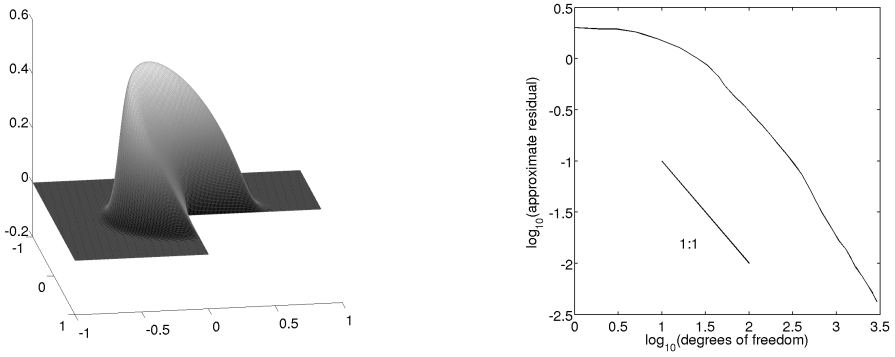

Figure 1: Exact solution u and convergence history.

The algorithm is implemented using the Wavelet software at the Numerics group at Philipps-Universität Marburg. We employ spline wavelets of order $l = 3$ as constructed in [18] to discretize our problem. For the practical realization, we make a few modifications to the methods stated above. The tolerances in the algorithm are worst-case estimates which we partly replace by more realistic numbers. Moreover, for technical reasons only, we set a maximum level up to which wavelets are considered

and we leave out the application of the projector $\mathbf{P}$. Although the implementation of this projection step can be realized in practice, see the discussion in [21], it turns out that this step does not pay off from a computational point of view. The condition on the tolerances ε_n allows them to converge to zero at an order between linear and quadratic convergence, but the latter does not necessarily lead to the best performance. This is because this choice leads to a very fast increase in the support sizes of the iterates. Therefore, we choose geometrically decreasing tolerances, $\varepsilon_{n+1} = \frac{4}{5}\varepsilon_n$. The results in Figure 1 show the exact solution and the convergence history, i.e., the degrees of freedom versus an approximation of the residual in ℓ_2. We see that, after an initial phase, the maximally expected rate $\frac{l-t}{d} = \frac{3-1}{2} = 1$ is indeed approximately realized.

Recall that a solution $u = \mathbf{u}^\top \Psi$ can be written as a sum of local parts $u = \mathbf{u}_{|\Lambda_0}\Psi_0 + \mathbf{u}_{|\Lambda_1}\Psi_1$. Since the subdomains overlap, it is not clear in advance how the local parts $\mathbf{u}_{|\Lambda_0}\Psi_0 \in H_0^1(\Omega_0)$ and $\mathbf{u}_{|\Lambda_1}\Psi_1 \in H_0^1(\Omega_1)$ look like. We are interested in whether or not the approximate solutions produced by our algorithm show oscillations in the overlapping region, because these oscillations would indicate unnecessary computational effort. However, in Figure 2, we see that even without the application of the projector $\mathbf{P}$, which explicitly removes these oscillations, the representation in the overlapping region is not very redundant. This supports our view that, from a practical point of view, the application of the projection step can be omitted.

To sum up, we may conclude that the numerical test supports the theoretical results stated in Section 3.

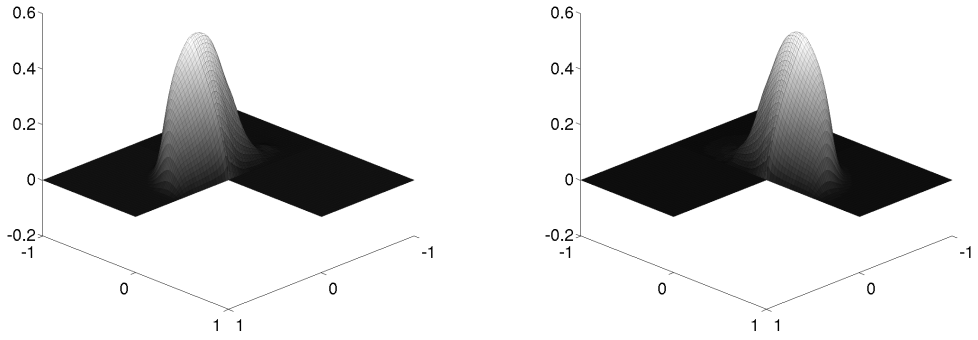

Figure 2: Local parts of the approximate solution.

References

[1] A. Barinka, W. Dahmen, and R. Schneider, *Fast computation of adaptive wavelet expansions*, Numer. Math. **105** (2007), no. 4, 549–589.

[2] O. Christensen, *An introduction to frames and Riesz bases*, Birkhäuser, Boston, 2003.

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

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

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

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

[7] S. Dahlke, *Besov regularity for elliptic boundary value problems in polygonal domains*, Appl. Math. Lett. **12** (1999), no. 6, 31–36.

[8] S. Dahlke, M. Fornasier, M. Primbs, T. Raasch, and M. Werner, *Nonlinear and adaptive frame approximation schemes for elliptic PDEs: Theory and numerical experiments*, Numer. Methods Partial Differ. Equations **25** (2009), no. 6, 1366–1401.

[9] S. Dahlke, D. Lellek, S. H. Lui, and R. Stevenson, *Adaptive wavelet Schwarz methods for the Navier-Stokes equation*, DFG-SPP 1324 Preprint series (2013), no. 140.

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

[11] M. Dryja and O. Widlund, *An additive variant of the Schwarz alternating method for the case of many subregions*, New York University, Courant Institute of Mathematical Sciences, Computer Science Technical Report (1987), no. 339.

[12] P. Grisvard, *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics, vol. 24, Pitman (Advanced Publishing Program), Boston, MA, 1985.

[13] J. Kappei, *Adaptive frame methods for nonlinear elliptic problems*, Appl. Anal. **90** (2011), no. 8, 1323–1353.

[14] ———, *Adaptive wavelet frame methods for nonlinear elliptic problems*, Logos Verlag Berlin, 2011.

[15] D. Lellek, *Adaptive wavelet frame domain decomposition methods for nonlinear elliptic equations*, Numer. Methods Partial Differ. Equations **29** (2013), no. 1, 297–319.

[16] ———, *Adaptive wavelet schwarz for nonlinear elliptic partial differential equations*, Logos Verlag Berlin, 2015.

[17] S. H. Lui, *Numerical analysis of partial differential equations*, Pure and Applied Mathematics (Hoboken), John Wiley & Sons Inc., Hoboken, NJ, 2011. MR 2895081

[18] M. Primbs, *New stable biorthogonal spline-wavelets on the interval*, Results Math. **57** (2010), 121–162.

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

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

[21] M. Werner, *Adaptive wavelet frame domain decomposition methods for elliptic operator equations*, Logos Verlag Berlin, 2009.