Preprint No. 1271, Dept. of Math., Univ. of Utrecht, Submitted to SIAM J. Numer. Anal., March 2003

AN OPTIMAL ADAPTIVE FINITE ELEMENT METHOD

ROB STEVENSON

Abstract. Although existing adaptive finite element methods for solving second order elliptic equations often perform better in practical computations than non-adaptive ones, usually they are even not proven to converge. Only recently in the works of Dörfler ([Dör96]) and that of Morin, Nochetto and Siebert ([MNS00]), adaptive methods were constructed for which converge could be demonstrated. However, convergence alone does not imply that the method is more efficient than its non-adaptive counterpart. In [BDD02], Binev, Dahmen and DeVore added a coarsening step to the routine from [MNS00]. They could prove that the resulting method is quasi-optimal in the following sense: If the solution is such that for some $s > 0$, the errors in energy norm of the best continuous piecewise linear approximations subordinate to any partition with n triangles are $\mathcal{O}(n^{-s})$, then given an $\varepsilon > 0$, the adaptive method produces an approximation with an error less than ε subordinate to a partition with $\mathcal{O}(\varepsilon^{-1/s})$ triangles, in addition taking only $\mathcal{O}(\varepsilon^{-1/s})$ operations.

In this paper, employing a different type of adaptive partitions, we develop an adaptive method with similar properties as in [BDD02]. Unlike [BDD02], our coarsening routine will be based on a transformation to a wavelet basis, and we expect it to have better quantitative properties. Furthermore, all our results are valid uniformly in the size of possible jumps of the diffusion coefficients. We allow right-hand sides in $H^{-1}(\Omega)$, whereas the methods from [MNS00, BDD02] are restricted to right-hand sides in $L_2(\Omega)$. In our final adaptive algorithm, all tolerances depend on a posteriori estimate of the current error instead an a priori one, which can be expected to give quantitative advantages.

1. Introduction

For solving elliptic boundary value problems for which the solution has singularities, the use of adaptive finite element methods potentially has the advantage of a strong reduction of the computational costs compared to non-adaptive methods. Although the adaptive methods that can be found in the literature indeed often exhibit such a reduction, they are usually even not proven to converge, let alone that they are shown to outperform the non-adaptive methods. Only quite recently in the work of Dörfler ([Dör96]), that was later extended by Morin, Nochetto and Siebert in ([MNS00]), adaptive methods were constructed that are proven to converge. These methods are based on an adaptive refinement strategy that guarantees the so-called saturation property, saying that the difference between the solutions on two consecutive partitions is greater than some multiple of the error in the

Date: March 21, 2003.

2000 *Mathematics Subject Classification.* 65N30, 65N50, 65N15, 65Y20, 65T60, 41A25.

Key words and phrases. Adaptive finite element method, optimal computational complexity, wavelet basis, a posteriori error estimator, discontinuous coefficients.

solution on the first partition. Exploiting Galerkin orthogonality, convergence now easily follows.

In [BDD02], Binev, Dahmen and DeVore added a coarsening step to the method from [MNS00]. Basically the idea of such a step, that has to applied after each fixed number of refinement steps, is to undo refinements that in the end turn out hardly to contribute to a better approximation. Thanks to this coarsening, under some conditions on the right-hand side the resulting adaptive method was proven to be quasi-optimal in the following sense: If the solution is such that for some $s > 0$, the errors in energy norm of the best continuous piecewise linear approximations subordinate to any partition with n triangles are $\mathcal{O}(n^{-s})$, then given an $\varepsilon > 0$, the adaptive method produces an approximation with an error less than ε subordinate to a partition with $\mathcal{O}(\varepsilon^{-1/s})$ triangles, in addition taking only $\mathcal{O}(\varepsilon^{-1/s})$ operations.

In this paper, we consider a method as developed in [MNS00] and extended with a coarsening in [BDD02] in slightly different context: Instead of considering conforming partitions produced with the so-called newest vertex bisection method, we consider generally non-conforming partitions produced by only 'red-refinement' steps, i.e., splittings of triangles into four congruent subtriangles. In this setting we generalize or improve the findings from [MNS00] and [BDD02] on a number of points:

• Following [MNS00], we consider the model problem of Poisson's equation on a two-dimensional domain, generalized in the sense that a piecewise constant diffusion tensor is allowed. Although in the discussion and in the numerical examples quite some attention has been paid to the situation that this tensor has jumps, the estimates from [MNS00] are not valid uniformly in the size of such jumps. Assuming a so-called quasi-monotone distribution of these jumps, all results from this paper will be proven to hold uniformly in the size of the jumps. Among other things, we therefore had to modify the a posteriori error estimator from [Ver96, BM87] to include terms that depend on the diffusion tensor.
• With an adaptive method, together the right-hand side and the discrete Galerkin solution subordinate to the current partition determine the next partition via an a posteriori error estimator. In [MNS00] it was assumed that the exact Galerkin solutions are available for this goal. Aiming at proving optimal computational complexity, in [BDD02] inexact solutions of the discrete systems were considered. In addition to that, in this paper we allow inexact right-hand sides to be used for both setting up the discrete systems and the evaluation of the a posteriori error estimator. This generalization can be used to model the effect of the application of quadrature, and it allows us to analyze the scheme exactly as it is performed in practical computations. Furthermore, it will allow us to prove quasi-optimality of the adaptive method for right-hand sides in $H^{-1}(\Omega)$, whereas the methods from [MNS00, BDD02] are restricted to right-hand sides in $L_2(\Omega)$.
• We introduce a new coarsening procedure, that unlike the procedure from [BDD02] is based on a transformation to a wavelet basis for the space of continuous piecewise linears subordinate to the adaptively refined partition. We expect our procedure to have better quantitative properties, although admittedly a final answer can only be given after

performing numerical tests. In any case, with our procedure the additional uniform refinement steps needed in [BDD02] are avoided. Moreover, our procedure does not only yield a quasi-uniform partition, but at the same time also a quasi-optimal continuous piecewise linear approximation subordinate to this partition, which simplifies the development of the adaptive method. Both the coarsening from [BDD02] as our coarsening rely on an adaptive tree approximation algorithm developed by Binev and DeVore in [BD02].

• The adaptive finite element method from [BDD02] and our first routine **SOLVE1** require as input an a priori upperbound μ of the convergence rate of the algorithm without coarsening. When one supplies a μ that is too small, quasi-optimality as a consequence of the coarsening is not guaranteed. On the other hand, taking a μ that is unnecessarily close to one will result in a quantitatively less attractive algorithm, since due to the coarsening, the convergence rate will be limited by this μ. In this paper, we develop an a second routine **SOLVE2** in which the tolerances allowed in the inexact Galerkin solutions and in the approximations of the right-hand side, and those required in the coarsening are all some fixed multiples of an a posteriori estimate of the current error. Apart that this releases the user from the task supplying this most critical parameter, the new algorithm benefits from any better convergence rate than appears from an a priori worst case analysis.

Finally, let us comment on the necessity of applying a coarsening routine. In the numerical experiments reported in [MNS00], the partitions, although produced without coarsening, already seem to have a quasi-optimal cardinality. Of course, this does not exclude the possibility that there are other examples for which coarsening is necessary. Indeed, in [MNS00] only right-hand sides in $L_2(\Omega)$ are considered, meaning that singularities can only be caused by the shape of the domain or jumping diffusion coefficients. Allowing $f \in H^{-1}(\Omega)$ as we do may give rise to singularities that require a coarsening routine. On the other hand, it is also possible that coarsening is not a necessary ingredient of a quasi-optimal adaptive algorithm for solving these elliptic problems, but that a proof of such a fact is missing up to now. However, even in the latter case, we do not consider our construction of a coarsening routine as only being relevant for theory, since we expect that it might be very useful inside adaptive routines for solving non-stationary problems, where an efficient coarsening is definitely of practical importance.

This paper is organized as follows: In §2 our model boundary value problem is described.

In §3, we introduce the class of admissible partitions, which is the subclass of all partitions that can be generated by red-refinements, for which the generations of neighbouring triangles differ at most one. We show that any partition can be refined to an admissible one by increasing the number of triangles by at most a constant factor.

In §4, we introduce a wavelet basis for the space of continuous piecewise linears subordinate to any admissible partition. We show that both the basis transformation from wavelet to nodal basis and its inverse can be performed in optimal computational complexity.

Our coarsening routine is defined in §5. It is based on a transformation to wavelet basis, an application of the adaptive tree approximation routine from [BD02], and finally a construction of a reduced partition subordinate to which the remaining terms in the wavelet expansion are continuous piecewise linear.

In §6, an a posteriori error estimator is derived. A refinement strategy is developed that also for inexact, but sufficiently accurate, right-hand sides and discrete solutions is shown to be convergent.

In §7, the coarsening routine and the convergent adaptive refinement strategy are combined to an optimal adaptive finite element method.

Finally, in §8 we derive such a method in which the tolerances for the errors in the right-side and in the discrete solution, and for the coarsening routine are determined by an a posteriori estimate of the current error.

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

2. Boundary value problem

Let Ω be a polygonal bounded domain in $\mathbb{R}^2$. We consider the following model boundary value problem in variational form: Given $f \in H^{-1}(\Omega)$, find $u \in H_0^1(\Omega)$ such that

$$a(u,w) := \int_\Omega \mathbf{A}\nabla u \cdot \nabla w = f(w), \quad (w \in H_0^1(\Omega)), \tag{2.1}$$

where $\mathbf{A} \in L_\infty(\Omega)$ is a 2×2 matrix with $\mathbf{A}^*(x) = \mathbf{A}(x) > 0$ a.e.. Further assumptions on $\mathbf{A}$ are collected in the forthcoming Assumption 3.8. Defining $L : H_0^1(\Omega) \to H^{-1}(\Omega) = (H_0^1(\Omega))'$ by $(Lu)(w) = a(u,w)$, (2.1) can be rewritten as

$$Lu = f.$$

On some places we will assume that the right-hand side $f \in L_2(\Omega)$, in which case $f(w)$ should be interpreted as $\int_\Omega fw$.

Aiming at results that hold uniformly in the size of variations that $\rho := \rho(\mathbf{A})$ may have, we introduce a weighted $L_2(\Omega)$-scalar product

$$\langle u,w\rangle_0 = \int_\Omega \rho u w,$$

and define the weighted norms

$$\|w\|_0 = \langle w,w\rangle_0^{\frac{1}{2}}, \quad |w|_1 = a(w,w)^{\frac{1}{2}}, \quad \|g\|_{-1} = \sup_{0\neq w\in H_0^1(\Omega)} \frac{|g(w)|}{|w|_1}$$

on $L_2(\Omega)$, $H_0^1(\Omega)$ and $H^{-1}(\Omega)$ respectively. Equipped with these norms, L is an isomorphism between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$.

For $\Sigma \subset \Omega$, we define $|w|_{1,\Sigma} := (\int_\Sigma \mathbf{A}|\nabla w|^2)^{\frac{1}{2}}$.

3. Partitions of Ω

We are going to approximate the solution of (2.1) by continuous piecewise linear functions subordinate to a partition of Ω into triangles. In this subsection we precisely describe the type of partitions we will consider.

We will use P to denote a *partition* of Ω, defined as a collection of closed triangles $\triangle$ such that $\Omega = \cup_{\triangle \in P} \triangle$ and $\text{meas}(\triangle \cap \tilde{\triangle}) = 0$ for any two different $\triangle, \tilde{\triangle} \in P$. When $\triangle \cap \tilde{\triangle} \neq \emptyset$, such triangles will be called *neighbours*. A partition $\tilde{P}$ is called a *refinement* of P, when $\tilde{P}$ can be constructed by, for zero or more $\triangle \in P$, replacing $\triangle$ by the four subtriangles created by connecting the midpoints of edges of $\triangle$, or by a recursive application of this elementary 'red' refinement step. The above $\triangle$ will be referred to as being the *parent* of its four subtriangles, called *children* of $\triangle$. As expected, children of children of $\triangle$ are called *grandchildren* of $\triangle$.

Throughout this paper we consider only partitions P that are refinements of some *fixed initial partition* P_0 of Ω. Many statements we are going to derive will involve constants that actually depend on P_0. However since P_0 is assumed to be fixed, for ease of presentation we ignore these dependencies. Clearly, any $\triangle \in P$ is *similar* to a triangle from P_0. For $\triangle \in P$, $\text{gen}(\triangle)$ will denote the number of elementary refinement steps needed to create $\triangle$ starting from some $\tilde{\triangle} \in P_0$, where $\text{gen}(\tilde{\triangle}) := 0$.

We call v a *vertex of* P, when there exists a $\triangle \in P$ such that v is a vertex of $\triangle$. A vertex v of P is called *non-hanging* when it is a vertex of all $\triangle \in P$ that contain v, otherwise it is called a *hanging* vertex of P. With $\bar{V}_P$ or V_P we will denote the set of all *non-hanging* vertices of P or all *non-hanging, interior* vertices of P respectively. We assume that P_0 is *conforming*, i.e., all its vertices are non-hanging.

A vertex v of P is called *regular* when for all $\triangle \in P$ that contain v, $\text{gen}(\triangle)$ has the same value. Note that a regular vertex is non-hanging.

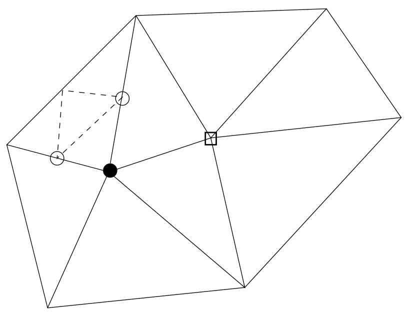

FIGURE 1. Regular ($\square$), non-hanging but non-regular ($\bullet$), and hanging vertices ($\circ$).

For a vertex v of P, the number of $\triangle \in P$ that contain v is called the *valence* of v in P. The valence of any v of P is less or equal to the maximum of 6 and the maximum valence of all vertices of P_0. If for a $\triangle \in P$, $\text{gen}(\triangle) = \max_{\tilde{\triangle} \in P} \text{gen}(\tilde{\triangle})$, then its edges cannot contain hanging nodes. As a consequence, for such $\triangle$, the number of its neighbours in P is given by the sum of the valences of its vertices minus 6 plus the number of edges of $\triangle$ on $\partial\Omega$, in particular showing that this number is uniformly bounded.

Proposition 3.1. *For any partition P of the type we consider, there exists a unique sequence of partitions*

$$P_0, P_1, \ldots, P_n$$

with $\max_{\triangle \in P_i} \operatorname{gen}(\triangle) = i$, $P_n = P$, *and where* P_{i+1} *is created from* P_i *by refining some* $\triangle \in P_i$ *with* $\operatorname{gen}(\triangle) = i$. *For convenience, we put* $P_{-1} = \emptyset$ *and so* $V_{P_{-1}} = \emptyset$. *The following properties are valid:*

(i) $\sum_{i=0}^{n} \#P_i \leq \frac{4}{3}\#P$,

(ii) $V_{P_{i-1}} \subset V_{P_i}$ *and so* $V_P = \cup_{i=0}^{n} V_{P_i} \backslash V_{P_{i-1}}$ *with empty mutual intersections,*

(iii) *a* $v \in V_{P_i} \backslash V_{P_{i-1}}$ *is not a vertex of* P_{i-1}, *and so it is a regular vertex of* P_i.

Proof. The existence and uniqueness of the sequence $(P_i)_i$ and also (ii) are obvious. By each elementary refinement step the left-hand side of (i) increases by 4, whereas the right-hand side increases by 3, which by induction on the number of these steps shows (i).

Suppose for some $1 \leq i \leq n$, $v \in V_{P_i} \backslash V_{P_{i-1}}$ is a vertex of P_{i-1}, then it is a hanging vertex of P_{i-1}. So there is a $\triangle \in P_{i-1}$ with $\operatorname{gen}(\triangle) < i-1$ that contains v. However by definition of the sequence $(P_i)_i$, such a $\triangle$ is not refined when going to P_i, meaning that v is also a hanging vertex of P_i which gives a contradiction. $\square$

Notation 3.2. Throughout this paper, for any partition P (or $\hat{P}$, $\tilde{P}$, etc.), with $(P_i)_i$ (or $(\hat{P}_i)_i$, $(\tilde{P}_i)_i$, etc.) we will always mean the corresponding sequence as in Proposition 3.1. When we write $P = P_n$, we mean that P is a partition with $\max_{\triangle \in P} \operatorname{gen}(\triangle) = n$.

In view of the forthcoming discussion on adaptively refined partitions, we emphasize here that given any partition P, the definition of the corresponding sequence $(P_i)_i$ is independent of the way P has been constructed.

Definition 3.3. A partition P^a is called *admissible* when for all neighbours $\triangle, \tilde{\triangle} \in P^a$, $|\operatorname{gen}(\triangle) - \operatorname{gen}(\tilde{\triangle})| \leq 1$.

As will turn out later, the reason to consider this restricted class of admissible partitions is given by the following proposition.

Proposition 3.4. *Let* P^a *be admissible. For any* $\triangle \in P^a$ *with* $i := \operatorname{gen}(\triangle) > 0$ *the vertices of the parent* $\tilde{\triangle} \in P^a_{i-1}$ *of* $\triangle$ *are regular vertices of* P^a_{i-1}.

Proof. For $i = 1$ the statement is true. Now let $i > 1$. Suppose that some vertex v of $\tilde{\triangle}$ is not a regular vertex of P^a_{i-1}. Then there exists a $\hat{\triangle} \in P^a_{i-1}$ with $v \in \hat{\triangle}$ and $\operatorname{gen}(\hat{\triangle}) < i-1$. Since by definition of the sequence $(P^a_i)_i$, this $\hat{\triangle}$ will never be refined, we get a contradiction with the fact that P^a is admissible. $\square$

Proposition 3.5. *If* $P^a = P^a_n$ *is admissible, then for any* $0 \leq i \leq n$, P^a_i *is also admissible.*

Proof. Suppose P^a_i is not admissible, then it contains neighbours $\hat{\triangle}, \triangle$ with $\operatorname{gen}(\hat{\triangle}) < \operatorname{gen}(\triangle) - 1$. Since $\operatorname{gen}(\hat{\triangle}) < i$, it will never be refined and so P^a cannot be admissible. $\square$

Although not any partition is admissible, it has an admissible refinement with a number of triangles which is at most a constant factor larger. Indeed, given a partition $P = P_n$, consider the following algorithm to compute the partition $P^a = P^a_n$:

Algorithm 3.6.

$P_0^a := P_0$
`for` $i = 0, \ldots, n$ `do`
define P_{i+1}^a *as the union of* P_{i+1} *and, when* $i \leq n-2$*, the collection of children of those* $\triangle \in P_i^a$ *that have a neighbour in* P_i *with grandchildren in* P_{i+2}
`od`

Proposition 3.7. *The partition* P^a *yielded by Algorithm 3.6 is an admissable refinement of* P *with* $\#P^a \lesssim \#P$.

Proof. The criterion to add children of a $\triangle \in P_i^a$ to P_{i+1}^a can only be fulfilled when $\triangle$ is neighbour of a $\tilde{\triangle} \in P_i$ which was refined when going to P_{i+1}, and thus with $\mathrm{gen}(\tilde{\triangle}) = i$. As we have seen, since $\max_{\triangle \in P_i^a} \mathrm{gen}(\triangle) = i$, the number of neighbours in P_i^a of such a $\tilde{\triangle}$ is uniformly bounded. So defining λ_i or λ_i^a as the number of triangles that were refined when going from P_i to P_{i+1} or from P_i^a to P_{i+1}^a respectively, we have $\lambda_i^a \lesssim \lambda_i$.

Note that $(P_i^a)_{0 \leq i \leq n}$ corresponds to P^a in the sense of Proposition 3.1. Since each time a triangle in a partition is refined the number of triangles is increased by 3, we conclude that

$$\#P^a = \#P_0 + 3\sum_{i=0}^{n-1} \lambda_i^a \lesssim \#P_0 + 3\sum_{i=0}^{n-1} \lambda_i = \#P.$$

What is left to show is that P^a is admissable. Obviously the partitions P_0^a and P_1^a are admissable. Suppose that there exists an $1 \leq i \leq n-1$ such that P_i^a is admissable, whereas P_{i+1}^a is not. Then there exist neighbours $\triangle, \tilde{\triangle} \in P_{i+1}^a$ with $\mathrm{gen}(\triangle) - \mathrm{gen}(\tilde{\triangle}) > 1$. Since P_i^a is admissable, necessarily $\mathrm{gen}(\triangle) = i+1$ and $\mathrm{gen}(\tilde{\triangle}) = i-1$.

If $\triangle \in P_{i+1}$, then $\tilde{\triangle} \in P_{i-1}^a$ has a neighbour in P_{i-1} with grandchildren in P_{i+1}. So by construction $\tilde{\triangle}$ would have been refined when going to P_i^a which gives a contradiction with the assumption that $\tilde{\triangle} \in P_{i-1}^a$.

If $\triangle \in P_{i+1}^a \backslash P_{i+1}$, then by construction its parent $\triangle_f \in P_i^a$ has a neighbour $\hat{\triangle} \in P_i$ with grandchildren in P_{i+2}, whereas obviously also $\triangle_f \in P_i^a$ and $\tilde{\triangle}$ are neighbours. Let $\triangle_{ff} \in P_{i-1}^a$ and $\hat{\triangle}_f \in P_{i-1}$ denote the parents of $\triangle_f$ and $\hat{\triangle}$ respectively. We are going to show that $\hat{\triangle}_f$ and $\tilde{\triangle}$ are neighbours in P_{i-1}^a, meaning, because $\hat{\triangle}_f$ has grandchildren in P_{i+1}, that $\tilde{\triangle}$ must have been refined when going to P_i^a which gives a contradiction with the assumption that $\tilde{\triangle} \in P_{i+1}^a$. We have to distinguish between two cases: If $\triangle_f$ is the central subtriangle of $\triangle_{ff}$, then both $\triangle_{ff}$ and $\hat{\triangle}_f$, and $\triangle_{ff}$ and $\tilde{\triangle}$ share an edge, and so $\hat{\triangle}_f$ and $\tilde{\triangle}$ are neighbours (cf. left picture in Figure 3). If $\triangle_f$ is a corner subtriangle of $\triangle_{ff}$, i.e, $\triangle_f$ and $\triangle_{ff}$ share a vertex v, then v is also a vertex of both $\hat{\triangle}_f$ and $\tilde{\triangle}$, again showing that they are neighbours (cf. right picture in Figure 3). $\square$

With $P_0^* = P_0$, by induction on i we construct the partition P_i^* from P_{i-1}^* by applying a red refinement step to *all* $\triangle \in P_{i-1}^*$, i.e., P_i^* is the result of applying recursively i *uniform refinement steps* to P_0. Note that these definitions are in accordance with Notation 3.2. We define

$$V_* = \cup_{i \geq 0} V_{P_i^*} \backslash V_{P_{i-1}^*},$$

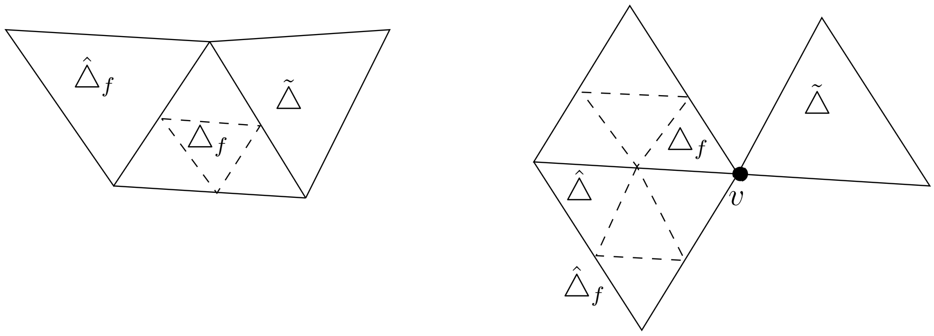

Figure 2. Illustration with the proof of Proposition 3.7.

which set contains V_P for any partition $P = P_n$. Obviously, for $0 \leq i \leq n$, $V_{P_i} \subset V_{P_i^*}$, and Proposition 3.1(iii) shows that $V_{P_i} \backslash V_{P_{i-1}} \subset V_{P_i^*} \backslash V_{P_{i-1}^*}$.

Finally in this subsection, having defined the initial partition P_0, we are able to formulate all assumptions on the coefficient matrix $\mathbf{A}$ that we will need:

Assumption 3.8. In addition to assuming $\mathbf{A} \in L_\infty(\Omega)$ with $\mathbf{A}^*(x) = \mathbf{A}(x) > 0$ a.e., we assume that $\mathbf{A} \eqsim \rho(\mathbf{A})\mathbf{id}$, uniformly over the domain (*isotropic diffusion*), and that $\mathbf{A}$ is *piecewise constant with respect to* P_0. Further, following [DSW96], we assume that $\rho = \rho(\mathbf{A})$ is *quasi-monotone* with respect to P_0. That is, defining for $v \in V_{P_0}$, $P_0(v) = \{\triangle \in P_0 : v \in \triangle\}$ and $\triangle(v) = \operatorname{argmax}\{\rho|_\triangle : \triangle \in P_0(v)\}$, we assume that for some absolute constant $c > 0$, for all $v \in V_{P_0}$ and $\triangle \in P_0(v)$ there exist $\triangle = \triangle_1, \ldots, \triangle_m = \triangle(v)$ such that $\triangle_i$ shares an edge with $\triangle_{i+1}$ and $\rho|_{\triangle_i} \leq c\rho|_{\triangle_{i+1}}$. Under these assumptions, all results we are going to derive that depend on $\mathbf{A}$ should be interpreted as to hold *uniformly in* $(\rho|_\triangle)_{\triangle \in P_0}$.

4. Finite element spaces and bases, and Galerkin approximations

For a given partition P, let $\mathcal{S}_P \subset H_0^1(\Omega)$ denote the space of continuous, piecewise linear functions subordinate to P which vanish at $\partial\Omega$. The solution $u_P \in \mathcal{S}_P$ of

$$a(u_P, w_P) = f(w_P), \quad (w_P \in \mathcal{S}_P), \tag{4.1}$$

is called the *Galerkin approximation* of the solution u of (2.1). Defining $L_P : \mathcal{S}_P \to (\mathcal{S}_P)' \supset H^{-1}(\Omega)$ by $(L_P u_P)(w_P) = a(u_P, w_P)$, the solution of (4.1) is $L_P^{-1} f$.

On some places we will replace the right-hand side f by some approximation from $\mathcal{S}_P^{(0)}$, being defined as the space of functions that are piecewise constant with respect to P.

If $\tilde{P}$ is a refinement of P, then $\mathcal{S}_P \subset \mathcal{S}_{\tilde{P}}$ and $\mathcal{S}_P^{(0)} \subset \mathcal{S}_{\tilde{P}}^{(0)}$. Each $u \in \mathcal{S}_P$ is uniquely determined by its values on V_P, and so in particular $\#V_P = \dim \mathcal{S}_P$. Defining, for $v \in V_P$, $\phi_P^v \in \mathcal{S}_P$ by

$$\phi_P^v(\tilde{v}) = \begin{cases} 1 & v = \tilde{v}, \\ 0 & v \neq \tilde{v} \in V_P, \end{cases}$$

the set

$$\{\phi_P^v : v \in V_P\}$$

is a basis for $\mathcal{S}_P$, called *nodal basis*.

One easily verifies the following

Lemma 4.1. *Let v be a regular vertex of a partition P. Then with $i := \mathrm{gen}(\triangle)$ for any (and thus all) $\triangle \in P$ that contain v, it holds that $\phi^v_{P_i^*} \in \mathcal{S}_P$.*

Proposition 4.2. *For any partition $P = P_n$, $\cup_{i=0}^n \{\phi^v_{P_i^*} : v \in V_{P_i} \backslash V_{P_{i-1}}\}$ is a basis for $\mathcal{S}_P$, called* hierarchical basis*.*

Proof. Proposition 3.1(iii) and Lemma 4.1 show that for $v \in V_{P_i} \backslash V_{P_{i-1}}$, $\phi^v_{P_i^*} \in \mathcal{S}_{P_i} \subset \mathcal{S}_P$. Since it vanishes on $V_{P_{i-1}}$, by induction on i we conclude that for given scalars $(d_v)_{v \in V_P}$ the interpolation problem of finding scalars $(c_v)_{v \in V_P}$ with $\sum_{i=0}^n \sum_{v \in V_{P_i} \backslash V_{P_{i-1}}} c_v \phi^v_{P_i^*}(\tilde{v}) = d_{\tilde{v}}$ ($\tilde{v} \in V_P$) has a unique solution. Since $\dim \mathcal{S}_P = \# V_P = \sum_{i=0}^n \#(V_{P_i} \backslash V_{P_{i-1}})$ by Proposition 3.1(ii), the proof is completed. ☐

Besides the nodal and hierarchical bases, for admissible partitions P^a we introduce another basis for $\mathcal{S}_{P^a}$, that as we will see, is appropriately called a *wavelet basis*.

Let $v \in V_*$, then there exist a unique $i \in I\!N$ such that $v \in V_{P_i^*} \backslash V_{P_{i-1}^*}$. When $i > 0$, v is the midpoint of the common edge of two triangles $\triangle_1, \triangle_2 \in P_{i-1}^*$. Let us denote with $v_1(v), \ldots, v_4(v)$ the vertices of these $\triangle_1, \triangle_2$, with $v_2(v), v_3(v)$ being the vertices on the edge containing v, see Figure 4 For some scalars $\mu_{v,j}$ that we will specify later on, with

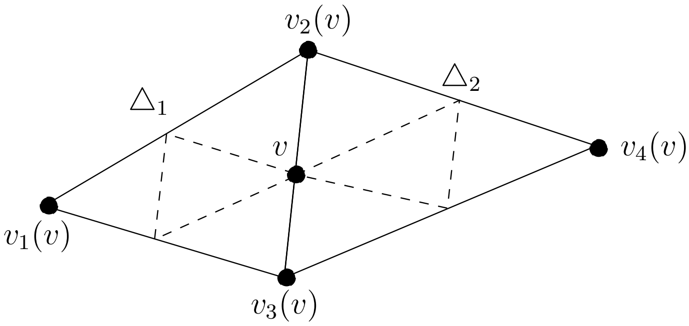

FIGURE 3. Definition of $v_j(v)$.

$\mu_{v,j} := 0$ when $v_j(v) \in \partial\Omega$, we define

$$\psi^v := \phi^v_{P_i^*} - \sum_{j=1}^4 \mu_{v,j} \phi^{v_j(v)}_{P_{i-1}^*}, \tag{4.2}$$

and for convenience, for $v \in V_{P_0^*}$ we put $\psi^v = \phi^v_{P_0^*}$.

Proposition 4.3. *If P^a is* admissible *then*

$$\{\psi^v : v \in V_{P^a}\} \tag{4.3}$$

is a basis for $\mathcal{S}_{P^a}$.

Proof. Obviously $\{\psi^v : v \in V_{P_0}\}$ is a basis for $\mathcal{S}_{P_0}$. Assuming that $\{\psi^v : v \in V_{P^a_{n-1}}\}$ is a basis for $\mathcal{S}_{P^a_{n-1}}$, the same argument as was applied in the proof of Proposition 4.2 shows that $\{\psi^v : v \in V_{P^a_{n-1}}\} \cup \{\phi^v_{P^*_n} : v \in V_{P^a_n}\backslash V_{P^a_{n-1}}\}$ is basis for $\mathcal{S}_{P^a_n}$. Proposition 3.1(iii) shows that each $v \in V_{P^a_n}\backslash V_{P^a_{n-1}}$ is the midpoint of $\triangle_1, \triangle_2 \in P^a_{n-1}$ with $\mathrm{gen}(\triangle_1) = \mathrm{gen}(\triangle_2) = n-1$, both which are refined in the transition to P^a_n. Proposition 3.4 shows that each of $v_1(v), \ldots, v_4(v)$, which are the vertices of $\triangle_1$ and $\triangle_2$, is a regular vertex of P^a_{n-1}, obviously with $\mathrm{gen}(\hat{\triangle}) = n-1$ for any, and thus all $\hat{\triangle} \in P^a_{n-1}$ which contain this vertex. From Lemma 4.1 we now conclude that $\sum_{j=1}^4 \mu_{v,j}\phi^{v_j(v)}_{P^*_{i-1}} \in \mathcal{S}_{P^a_{n-1}}$, from which it follows that also $\{\psi^v : v \in V_{P^a_n}\}$ is a basis for $\mathcal{S}_{P^a_n}$. $\square$

One may verify that for $P^a = P^a_n$ and $w_{P^a} \in \mathcal{S}_{P^a}$ given by its values $(w_{P^a}(v))_{v\in V_{P^a}}$, i.e., the coefficients of its representation with respect to the nodal basis, an application of the following routine yields the coefficients $(c_v)_{v\in V_{P^a}}$ of its representation with respect to the wavelet basis (4.3):

Algorithm 4.4.

$\lambda^{(n)}_v := w_{P^a}(v)$ *($v \in \bar{V}_{P^a}$)*
`for` $i = n, \ldots, 1$ `do`
  $\lambda^{(i-1)}_v := \lambda^{(i)}_v$ *($v \in \bar{V}_{P^a_{i-1}}$)*
  $c_v := \lambda^{(i)}_v - \frac{1}{2}(\lambda^{(i-1)}_{v_2(v)} + \lambda^{(i-1)}_{v_3(v)})$ *($v \in V_{P^a_i}\backslash V_{P^a_{i-1}}$)*
  $\lambda^{(i-1)}_{v_j(v)} := \lambda^{(i-1)}_{v_j(v)} + c_v\mu_{v,j}$ *($v \in V_{P^a_i}\backslash V_{P^a_{i-1}}$, $1 \le j \le 4$)*
`od`
$c_v := \lambda^{(0)}_v$ *($v \in V_{P^a_0}$)*

Conversely, if $w_{P^a} \in \mathcal{S}_{P^a}$ is given by its coefficients $(c_v)_{v\in V_{P^a}}$ with respect to the wavelet basis (4.3), then the values $(w_{P^a}(v))_{v\in V_{P^a}}$ are yielded by the following routine:

Algorithm 4.5.

$\lambda^{(0)}_v := c_v$ *($v \in V_{P^a_0}$)*
`for` $i = 1, \ldots, n$ `do`
  $\lambda^{(i-1)}_v := 0$ *($v \in \bar{V}_{P^a_{i-1}}\backslash V_{P^a_{i-1}}$)*
  $\lambda^{(i-1)}_{v_j(v)} := \lambda^{(i-1)}_{v_j(v)} - c_v\mu_{v,j}$ *($v \in V_{P^a_i}\backslash V_{P^a_{i-1}}$, $1 \le j \le 4$)*
  $\lambda^{(i)}_v := \lambda^{(i-1)}_v$ *($v \in V_{P^a_{i-1}}$)*
  $\lambda^{(i)}_v := c_v + \frac{1}{2}(\lambda^{(i-1)}_{v_2(v)} + \lambda^{(i-1)}_{v_3(v)})$ *($v \in V_{P^a_i}\backslash V_{P^a_{i-1}}$)*
`od`
$w_{P^a}(v) := \lambda^{(n)}_v$ *($v \in V_{P^a}$)*

From Proposition 3.1(i), we infer

Proposition 4.6. *Both the above algorithms to switch between the representation of a $w \in \mathcal{S}_{P^a}$ with respect to the nodal basis to its representation with respect to (4.3) and vice versa take not more than $\mathcal{O}(\dim \mathcal{S}_{P^a})$ operations.*

Now we come to the specification of the coefficients $\mu_{v,j}$ from (4.2): We take

$$\mu_{v,j} = \frac{3(\rho|_{\triangle_1} \cdot \text{meas}(\triangle_1) + \rho|_{\triangle_2} \cdot \text{meas}(\triangle_2))}{8\sum_{\{\triangle \in P^*_{i-1}, v_j(v) \in \triangle\}} \rho|_\triangle \cdot \text{meas}(\triangle)} \text{ when } j \in \{2,3\} \text{ and } v_j(v) \notin \partial\Omega, \tag{4.4}$$

and $\mu_{v,j} = 0$ otherwise. An alternative choice of the coefficients will later be discussed in Remark 4.9. When both $v_2(v), v_3(v) \notin \partial\Omega$, a simple calculation reveals that, for $i \geq 1$ and $v \in V_{P^*_i} \backslash V_{P^*_{i-1}}$, $\int_\Omega \rho\psi^v = 0$, so that it is appropriate to call ψ^v a *wavelet*.

For coefficient matrices $\mathbf{A}$ that satisfy Assumption 3.8, a combination of results from [Ste98b, Ste98a, DSW96] shows the following

Theorem 4.7. *With* $\bar{\psi}^v := \psi^v/|\psi^v|_1$,

$$\{\bar{\psi}^v : v \in V_*\}, \tag{4.5}$$

is a Riesz basis for $H^1_0(\Omega)$ *equipped with* $|\cdot|_1$*, where here for completeness we emphasize that this result is valid uniformly in* $(\rho|_\triangle)_{\triangle \in P_0}$.

Defining $|||w|||_1 = (\sum_{v \in V_*} c_v^2)^{\frac{1}{2}}$, where $w = \sum_{v \in V_*} c_v \bar{\psi}^v$ is the unique expansion of $w \in H^1_0(\Omega)$, $\lambda_{\bar{\Psi}}, \Lambda_{\bar{\Psi}} > 0$ will denote the largest or smallest constant such that

$$\lambda_{\bar{\Psi}} |||\cdot|||_1^2 \leq |\cdot|_1^2 \leq \Lambda_{\bar{\Psi}} |||\cdot|||_1^2, \tag{4.6}$$

and $\kappa_{\bar{\Psi}} := \frac{\Lambda_{\bar{\Psi}}}{\lambda_{\bar{\Psi}}}$. Here for completeness we emphasize that this equivalence between the $|\cdot|_1$-norm of a function in $H^1_0(\Omega)$ and the ℓ_2-norm of its, generally infinite, coefficient vector in particular holds for functions from $\mathcal{S}_{P^a}$ for admissible P^a, which by Proposition 4.3 have a finite wavelet expansion.

Remark 4.8. Since the proof of Theorem 4.7 can only be deduced by combining results from different papers, we briefly comment on its derivation. For any $\triangle \in \cup_{i \geq 0} P^*_i$, let $I_\triangle : C(\triangle) \to P_1(\triangle)$ be the nodal value interpolant, let the bilinear form $\langle\langle u, w \rangle\rangle_\triangle := \frac{1}{3}\text{meas}(\triangle) \cdot \sum_{v \text{ vertex of } \triangle} u(v)w(v)$, and

$$\langle u, w \rangle_\triangle := \langle\langle I_\triangle u, I_\triangle w \rangle\rangle_\triangle + \sum_{k=1}^{4} \left[\langle\langle u, w \rangle\rangle_{\triangle_k} - \langle\langle I_\triangle u, I_\triangle w \rangle\rangle_{\triangle_k}\right],$$

where $\triangle_1, \ldots, \triangle_4$ are the children of $\triangle$. Note that for $u, w \in P_1(\triangle)$, $\langle u, w \rangle_\triangle = \langle\langle u, w \rangle\rangle_\triangle$ which is a scalar product on $P_1(\triangle)$. Let $Y_\triangle : C(\triangle) \cap \prod_{k=1}^4 P_1(\triangle_k) \to P_1(\triangle)$ be the orthogonal projector with respect to $\langle\, ,\, \rangle_\triangle$. Finally, for $i \geq 1$, on $\mathcal{S}_{P^*_i} \times \mathcal{S}_{P^*_i}$ let

$$\langle u, w \rangle_{\mathcal{S}_{P^*_i}} := \sum_{\triangle \in P^*_{i-1}} \rho|_\triangle \langle u, w \rangle_\triangle.$$

Using that $\{\phi^v_{P^*_{i-1}} : v \in V_{P^*_{i-1}}\}$ is an orthogonal set with respect to $\langle\, ,\, \rangle_{\mathcal{S}_{P^*_i}}$, in [Ste98a, §6] it was shown that for $i \geq 1$ the sets

$$\{\psi^v/\|\psi^v\|_0 : v \in V_{P^*_i} \backslash V_{P^*_{i-1}}\},$$

defined by (4.2), (4.4), are uniform Riesz bases for $\mathcal{S}_{P_i^*} \cap \mathcal{S}_{P_{i-1}^*}^{\perp_{\langle\,,\,\rangle_{\mathcal{S}_{P_i^*}}}}$ equipped with $\|\cdot\|_0$, where 'uniform' refers to both the parameter i and to $(\rho|_\triangle)_{\triangle\in P_0}$.

With

$$t := \sup_{0\neq w\in C(\triangle)\cap\prod_{k=1}^4 P_1(\triangle_k)} \frac{\langle w,w\rangle_\triangle - \langle (I-Y_\triangle)w,(I-Y_\triangle)w\rangle_\triangle}{\sum_{k=1}^4 \langle\!\langle w,w\rangle\!\rangle_{\triangle_k}}$$

which value is independent of the triangle $\triangle$, from [Ste98b, Th. 3.1 and (4.7)] it follows that in case of *constant* $\mathbf{A} = \mathbf{id}$, for $\frac{3}{2} > s > \log_2 \sqrt{t}$ the infinite collection

$$\cup_{i\geq 0}\{2^{(s-1)i}\psi^v : v \in V_{P_i^*}\backslash V_{P_{i-1}^*}\}$$

is a Riesz basis for $H_0^s(\Omega)$. Some calculations show that $t = \frac{181}{64}$, and so $\sqrt{t} = .74992\ldots$, meaning in particular that (4.5) is a Riesz basis for $H_0^1(\Omega)$.

With $Q_i : L_2(\Omega) \to \mathcal{S}_{P_i^*}$ being the $\langle\,,\,\rangle_0$-orthogonal projector onto $\mathcal{S}_{P_i^*}$, and $Q_{-1} := 0$, for coefficient matrices $\mathbf{A}$ that satisfy Assumption 3.8, in [DSW96] it was shown that

$$\|w\|_1^2 \eqsim \sum_{i=0}^\infty 4^i \|(Q_i - Q_{i-1})w\|_0^2 \quad (w \in H_0^1(\Omega)),$$

uniformly in $(\rho|_\triangle)_{\triangle\in P_0}$. As shown in [Ste98b], from this result and the fact that $t < 1$ it even follows that (4.5) is a Riesz basis for $H_0^1(\Omega)$ equipped with $|\cdot|_1$, uniformly in $(\rho|_\triangle)_{\triangle\in P_0}$.

Remark 4.9. For *constant* $\mathbf{A} = \mathbf{id}$, in [CES00] other values for the coefficients $\mu_{v,1},\ldots,\mu_{v,4}$ from (4.2) were proposed, which generally all four are non-zero. Although uniform refinements of an arbitrary initial partition are considered, just as outlined above, for admissible $P^a = P_n^a$ a subset of the wavelet basis for $\mathcal{S}_{P_n^*}$ spans $\mathcal{S}_{P^a}$. For some $\tilde{s} > 0$ and $s \in (-\tilde{s}, \frac{3}{2})$, the infinite collection of properly scaled wavelets from [CES00] is shown to generate a Riesz basis for $\mathcal{H}^s(\Omega) = \begin{cases} H_0^s(\Omega) & \text{when } s \geq 0, \\ (H_0^{-s}(\Omega))' & \text{when } s < 0. \end{cases}$

On regular "type-I triangulations" of the whole of $\mathbb{R}^2$, the wavelet proposals from [Ste98a, §6] or [CES00] reduce to the so-called coarse-grid stabilized HB-systems from [LO96] with parameters $a = \frac{1}{8}$ or $a = -\frac{3}{16}$ respectively. There it is shown that for these uniform triangulations the exact $H^s(\mathbb{R}^2)$ stability ranges are $s \in (0.022818, \frac{3}{2})$ and $s \in (-0.440765, \frac{3}{2})$ respectively (cf. also [CS93]).

Finally, numerical results ([LO96, Table 1.2]) show that both wavelet bases are also quantitatively well-conditioned ($\kappa_{\bar{\Psi}}$ approximately 16 or 10 for $\mathbf{A} = \mathbf{id}$ and P_0 being the standard regular partition of the unit square into 8 triangles, so that $\#V_{P^0} = 1$).

5. A coarsening algorithm

Although they often perform well in practical computations, the adaptive finite element methods that can be found in the literature usually are not proven even to convergence, let alone that they are shown to outperform the standard non-adaptive methods. Only quite recently in the works of Dörfler ([Dör96]) and Morin, Nochetto and Siebert ([MNS00]), adaptive methods were constructed for which convergence was demonstrated. In [BDD02], Binev, Dahmen and DeVore added a coarsening step to the method from [MNS00], with which they obtained a method for which they could show convergence with an optimal rate.

Note that in contrast to this paper, in [MNS00, BDD02] conforming partitions were considered created by the newest vertex bisection procedure. Another difference is that the results from these papers are either restricted to the Laplace operator ([BDD02]), or not valid uniformly in the size of possible jumps of the coefficient matrix $\mathbf{A}$.

In [CDD01], a coarsening step was introduced into an adaptive algorithm in the framework of a wavelet method. The idea of such a step, that has to be applied after each fixed number of iterations that produce increasingly more accurate approximations, is to remove a possible large number of small terms in the current approximation that hardly contribute to its quality, but which because of their number spoil the complexity. With wavelet methods such 'small terms' stand for terms in a wavelet expansion with small coefficients, and in our finite element setting they correspond to a representation of the approximation as a piecewise linear function subordinate to a locally fine partition, whereas it is close to being linear on the union of these triangles.

Given some current approximation defined on some partition, in order to find a more efficient representation without increasing the error too much, one cannot simply join arbitrary collections of triangles since generally their union will not be a triangle. Instead one can only join groups of all siblings of one parent, that is, one has to respect the underlying tree structure. In view of this, in [BDD02], for each triangle in the tree associated to the partition an error functional was defined. It was shown that for any subtree, the squared sum over its leaves of these error functionals is bounded by some multiple of the squared error of the best continuous piecewise linear approximation subordinate to the partition defined by this subtree. Giving a tolerance that one allows to be added to the current error, a tree-coarsening algorithm from [BD02] was run, that, modulo some constant factor, yields the smallest subtree for which the above squared sum is less than this squared tolerance. The squared sum could not be shown being equivalent to the squared $|\cdot|_1$-norm of the error in the best approximation. Therefore, afterwards a fixed number of uniform refinement steps was needed to guarantee that a partition was obtained with respect to which there exists a continuous piecewise linear function with $|\cdot|_1$-distance to the approximation before applying this coarsening step that is less or equal to the squared tolerance.

In this paper, for the different type of partitions we consider, based on ideas from [BDD02, BD02] an alternative coarsening procedure is developed, which we hope is more attractive for practical computations. Given a current approximation from $\mathcal{S}_P$, in case P is not admissible, we will first embed it into $\mathcal{S}_{P^a}$, where P^a is constructed using Algorithm 3.6. Next, we determine its finite set of wavelets coefficients. Now using the norm equivalence (4.6), an obvious coarsening procedure would be just to order these coefficients by their modulus, and then to remove coefficients, starting with the smallest one, until the tolerance is met. Yet, the task is not to find an approximation with a minimum number of wavelet coefficients, but to find an approximation from a finite element space subordinate to a partition that has, modulo some constant factor, a minimum number of triangles, and the suggested procedure will generally fail to do this. Therefore we will equip the infinite index set V_* of all wavelets with a tree structure, and run the algorithm from [BD02] to find a subtree approximation on distance less or equal to the tolerance that, modulo some constant factor, has a minimum number of terms. Since the tree structure will be designed such that both the enlargement of an index set V_{P^a} of wavelets spanning a space $\mathcal{S}_{P^a}$ to a subtree, and conversely the enlargement of a subtree to such an index set V_{P^a} will increase the cardinality of the sets by at most a constant factor, we will be able to conclude that we found an approximation subordinate to a partition that, modulo some constant factor, has a minimum number of triangles. An additional advantage of our coarsening procedure will be that it not only gives a (quasi-) optimal partition, but at the same time that it also yields a (quasi-) optimal continuous piecewise linear approximation subordinate to this partition.

The tree structure with which we equip V_* is defined as follows: The vertices from $V_{P_0^*}$ are the roots of the tree. For $i \geq 1$ and $v \in V_{P_i^*} \backslash V_{P_{i-1}^*}$, either one or both vertices $v_2(v)$, $v_3(v)$ are in $V_{P_{i-1}^*} \backslash V_{P_{i-2}^*}$, which are the obvious candidates for being a parent of v. Yet, to apply results from [BD02], we need a tree structure in which each child has exactly one parent. When $i = 1$, both $v_2(v)$, $v_3(v) \in V_{P_0^*}$, and we just pick one as being the parent of v. When $i > 1$, and only one of $v_2(v)$, $v_3(v)$ is in $V_{P_{i-1}^*} \backslash V_{P_{i-2}^*}$, then this one is the parent of v. If both $v_2(v)$, $v_3(v) \in V_{P_{i-1}^*} \backslash V_{P_{i-2}^*}$, then there exists a unique $\triangle \in P_{i-1}^*$, that has vertices $v_2(v), v_3(v), \hat{v}$, where $\hat{v} \in V_{P_{i-2}^*}$. After ordering these vertices in clockwise direction starting with $\hat{v}$, we select the second one as being the parent of v, see Figure 5. Since the

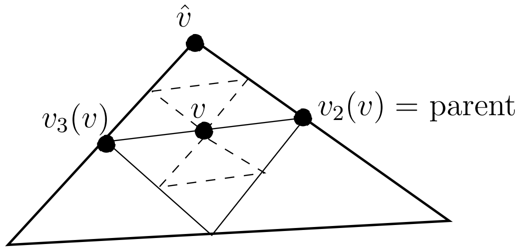

FIGURE 4. Parent of v.

valence of all vertices is uniformly bounded, so is the number of children of any parent in this tree.

For $w \in H_0^1(\Omega)$, let $w = \sum_{v \in V_*} c_v \bar{\psi}^v$ be its expansion with respect to (4.5). Obviously, for any $\tilde{V} \subset V_*$, its best approximation with respect to $|||\,|||_1$ from $\operatorname{span}\{\bar{\psi}^v : v \in \tilde{V}\}$ is $\sum_{v \in \tilde{V}} c_v \bar{\psi}^v$. The squared error $E(\tilde{V})$ of this approximation with respect to $|||\,|||_1$ is $E(\tilde{V}) = \sum_{v \in V_* \backslash \tilde{V}} |c_v|^2$.

We will call $\tilde{V} \subset V_*$ a *subtree* when it contains all roots $V_{P_0^*}$, and when for any $v \in \tilde{V}$, all its ancestors and all its siblings, i.e., those $w \in V_*$ that have the same parent as v, are also in $\tilde{V}$. The set of leaves $\mathcal{L}(\tilde{V})$ is defined as the set of those $v \in \tilde{V}$ which have no children in $\tilde{V}$. Defining for $v \in V_*$ the error functional $e(v) := \sum_{\bar{v} \text{ a descendant of } v} |c_{\bar{v}}|^2$, we have $E(\tilde{V}) = \sum_{v \in \mathcal{L}(\tilde{V})} e(v)$.

Following [BD02], we define a modified error functional $\tilde{e}(v)$ for $v \in V_*$ as follows: For the roots $v \in V_{P_0^*}$, $\tilde{e}(v) := e(v)$. Assuming $\tilde{e}(v)$ has been defined, then for all its children $v_1, \ldots, v_m$,

$$\tilde{e}(v_j) := \frac{\sum_{i=1}^m e(v_i)}{e(v) + \tilde{e}(v)} \tilde{e}(v).$$

Now given a $w \in H_0^1(\Omega)$ and a tolerance $\varepsilon > 0$, the *thresholding second algorithm* from [BD02] for determining a quasi-optimal *subtree approximation* runs as follows:

Algorithm 5.1.

$\tilde{V} := V_{P_0^*}$
`While` $E(\tilde{V}) > \varepsilon^2$ `do`
  compute $\rho = \max_{v \in \mathcal{L}(\tilde{V})} \tilde{e}(v)$
  `Forall` $v \in \mathcal{L}(\tilde{V})$ *with* $\tilde{e}(v) = \rho$ `do` *add all children of* v *to* $\tilde{V}$ `od`
`od`

Remark 5.2. During the evaluation of Algorithm 5.1, the values $\tilde{e}(v)$ for the current leaves should be stored as an ordered list. As a consequence, with $\tilde{V}$ being the subtree at termination, the operation count of Algorithm 5.1 will contain a term $\mathcal{O}(\#\tilde{V} \log((\#\tilde{V}))$ due the insertions of $\tilde{e}(v)$ for newly created leaves in this list. Since the other costs of the algorithm are $\mathcal{O}(\#\tilde{V})$, asymptotically the costs of these insertions will dominate. Although it seems unlikely that this will happen with practical problem sizes, for mathematical completeness we sketch here a modification with which the log-factor is avoided (see [Bar03, Met02, Ste02] for solutions of a similar problem in a wavelet context).

Noting that $\tilde{e}(v_j) \leq \sum_{i=1}^m e(v_i) \leq e(v) \leq |\!|\!| w |\!|\!|_1^2$, we may store the current leaves v in binary bins $V_0, \ldots V_q$, where for $0 \leq i \leq q-1$, V_i contains those v with $\tilde{e}(v) \in (2^{-(i+1)} |\!|\!| w |\!|\!|_1^2, 2^{-i} |\!|\!| w |\!|\!|_1^2]$, and remaining v, thus with $\tilde{e}(v) \leq 2^{-q} |\!|\!| w |\!|\!|_1^2$, are put into V_q. Instead of replacing all $v \in \mathcal{L}(\tilde{V})$ with maximal $\tilde{e}(v)$ by their children, in each iteration of the while-loop we replace just one v taken from the first non-empty bin by its children. Again from $\sum_{i=1}^m e(v_i) \leq e(v)$, we have $\tilde{e}(v_j) \leq \tilde{e}(v)$ meaning that for any i, once $V_0, \ldots, V_i$ got empty they will remain empty.

If q is chosen such that during the iteration we only extract v from bins V_i with $i < q$, then the corresponding $\tilde{e}(v)$ will be at most a factor 2 smaller that the current maximal value of $\tilde{e}$. As a consequence, one may verify that, with the exception of the operations counts, all results proven in [BD02] about Algorithm 5.1 are also valid for this modified version (making use of the property $\sum_{i=1}^m e(v_i) \leq e(v)$, which is stronger than the assumption made in [BD02], only (5.13) has to be adapted).

With $\tilde{V}$ being the subtree at termination, the number of operations required by this modified Algorithm 5.1 is $\lesssim \#\tilde{V} + q$, where here q enters as the maximum number of bins that have to be generated or inspected for containing leaves. Thinking of the situation that in the course of the iteration the maximum value of $\tilde{e}$ over the leaves varies largely in size, note that generally q cannot be bounded in terms of $\#\tilde{V}$.

We will apply the modified Algorithm 5.1 only in the situation that there exists a finite subtree $\bar{V} \subset V_*$ such that

$$e(v) = 0 \qquad (v \in \mathcal{L}(\bar{V})), \tag{5.1}$$

which allows us to make a suitable choice for q. Note that the subtree $\tilde{V}$ at termination satisfies $\tilde{V} \subset \bar{V}$. The definition of $\tilde{e}(v)$ shows that $\sum_{j=1}^{m} \frac{e(v_j)}{\tilde{e}(v_j)} = 1 + \frac{e(v)}{\tilde{e}(v)}$. A recursive application of this formula gives $\sum_{v\in\mathcal{L}(\tilde{V})} \frac{e(v)}{\tilde{e}(v)} = \#(\tilde{V}\backslash\mathcal{L}(\tilde{V})) + \#V_{P_0^*} \leq \#\tilde{V} \leq \#\bar{V}$, and so $E(\tilde{V}) \leq \#\bar{V} \max_{v\in\mathcal{L}(\tilde{V})} \tilde{e}(v)$. We conclude that the modified Algorithm 5.1 terminates before any leaf v with $\tilde{e}(v) \leq \varepsilon^2/\#\bar{V}$ is replaced by its children. Solving the smallest $q \in \mathbb{N}_0$ with $2^{-q}|||w|||_1^2 = \varepsilon^2/\#\bar{V}$ yields $q = \max\{0, \lceil \log_2(\varepsilon^{-2}|||w|||_1^2\#\bar{V})\rceil\}$, which q thus satisfies the assumption we made earlier.

The analysis of Algorithm 5.1 from [BD02], together with the additions from the above remark concerning our slight modification yields the following

Proposition 5.3 ([BD02, Corollary 5.3])**.** *The subtree $\tilde{V}$ yielded by (the modified) Algorithm 5.1 satisfies $E(\tilde{V}) \leq \varepsilon^2$. Moreover, there exists absolute constants $t_1, T_2 > 0$, necessarily with $t_1 \leq 1 \leq T_2$, such that if $\hat{V}$ is a subtree with $E(\hat{V}) \leq t_1\varepsilon^2$, then $\#\tilde{V} \leq T_2\#\hat{V}$. The number of evaluations of e and the number of additional arithmetic operations required by the modified Algorithm 5.1 are $\lesssim \#\tilde{V} + \max\{0, \log(\varepsilon^{-2}|||w|||_1^2\#\bar{V})\}$, with $\bar{V}$ being any subtree satisfying* (5.1).

We are now ready to define our coarsening routine. For P^a being an admissible partition, note that by Proposition 3.4 the set V_{P^a} is almost a subtree in the sense that it contains the roots V_{P_0} as well as all ancestors of any $v \in V_{P^a}$.

COARSE$[P, w_P, \varepsilon] \to [\tilde{P}^a, w_{\tilde{P}^a}]$:
% $P = P_n$ is some admissible partition, $w_P \in \mathcal{S}_P$ is given by its values $(w_P(v))_{v\in V_P}$ and
% $\varepsilon > 0$. The output $\tilde{P}^a$ is an admissible partition, and $w_{\tilde{P}^a} \in \mathcal{S}_{\tilde{P}^a}$ is given by its values
% $(w_{\tilde{P}^a}(v))_{v\in V_{\tilde{P}^a}}$

(i) *When P is admissible, let $P^a = P$, otherwise compute an admissible refinement P^a of P by applying Algorithm 3.6. Compute the values $(w_{P^a}(v))_{v\in V_{P^a}}$ of $w_{P^a} := w_P$.*
(ii) *Compute the wavelet coefficients $(c_v)_{v\in V_{P^a}}$ of w_{P^a} using Algorithm 4.4.*
(iii) *Compute the error functional $(e(v))_{v\in V_*}$ as follows:*

$e(v) := 0$ ($v \in V_*$).
`for` $i = n, \ldots, 1$ `do`
  `Forall` $v \in V_{P_i^a}\backslash V_{P_{i-1}^a}$ `do`
    $e(\hat{v}) := e(\hat{v}) + e(v) + c_v^2$, where $\hat{v}$ is the parent of v
  `od`
`od`

(iv) *Apply the modified Algorithm 5.1 yielding a subtree $\tilde{V}$ and the approximation $\sum_{v\in\tilde{V}} c_v\bar{\psi}^v$ for w_{P^a}.*
(v) *Determine a partition $\tilde{P}$ as follows:*

$\tilde{P}_0 := P_0$
`for` $i = 1, \ldots, n$ `do`
construct $\tilde{P}_i$ *from* $\tilde{P}_{i-1}$ *by refining those* $\triangle \in \tilde{P}_{i-1}$ *that contain a* $v \in \tilde{V} \cap V_{P_i^*} \backslash V_{P_{i-1}^*}$
`od`

(vi) *Apply Algorithm 3.6 to determine an admissable refinement* $\tilde{P}^a$ *of* $\tilde{P}$. *Note that* $\tilde{V} \subset V_{\tilde{P}} \subset V_{\tilde{P}^a}$, *and thus that* $w_{\tilde{P}^a} := \sum_{v \in \tilde{V}} c_v \bar{\psi}^v \in \mathcal{S}_{\tilde{P}^a}$.

(vii) *Apply Algorithm 4.5 to compute* $(w_{\tilde{P}^a}(v))_{v \in V_{\tilde{P}^a}}$.

Theorem 5.4. *(a).* $[\tilde{P}^a, w_{\tilde{P}^a}] :=$ **COARSE**$[P, w_P, \varepsilon]$ *satisfies* $|\!|\!| w_P - w_{\tilde{P}^a} |\!|\!|_1 \leq \varepsilon$. *There exist an absolute constant* $D > 0$, *such that for any partition* $\hat{P}$ *for which there exists a* $w_{\hat{P}} \in \mathcal{S}_{\hat{P}}$ *with* $|\!|\!| w_P - w_{\hat{P}} |\!|\!|_1 \leq t_1^{\frac{1}{2}} \varepsilon$, *it holds that* $\#\tilde{P}^a \leq D \# \hat{P}$.
(b). The number of arithmetic operations required by the call is $\lesssim \#P + \max\{0, \log(\varepsilon^{-1} |\!|\!| w_P |\!|\!|_1)\}$.

Proof. (a). The first statement follows by construction. Let $(c_v)_{v \in V_*}$ be the wavelet coefficients in the expansion $w_P = \sum_{v \in V_*} c_v \bar{\psi}_v$. Let $\hat{P}$, $w_{\hat{P}} \in \mathcal{S}_{\hat{P}}$ with $|\!|\!| w - w_{\hat{P}} |\!|\!|_1 \leq t_1^{\frac{1}{2}} \varepsilon$. Let $\hat{P}^a$ be the admissible refinement of $\hat{P}$ constructed by applying Algorithm 3.6, and let $(\hat{c}_v)_{v \in V_{\hat{P}^a}}$ be the wavelet coefficients of $w_{\hat{P}} \in \mathcal{S}_{\hat{P}} \subset \mathcal{S}_{\hat{P}^a}$. Let $\hat{V}$ be the enlargement of $V_{\hat{P}^a}$ by adding all siblings of all $v \in V_{\hat{P}^a}$ to this set. Then $\hat{V}$ is a subtree with $\#\hat{V} \lesssim \# V_{\hat{P}^a} \lesssim \#\hat{P}^a \lesssim \#\hat{P}$. Because of

$$E(\hat{V}) = \sum_{v \in V_* \backslash \hat{V}} |c_v|^2 \leq \sum_{v \in V_*} |c_v - \hat{c}_v|^2 = |\!|\!| w - \hat{w} |\!|\!|_1^2 \leq t_1 \varepsilon^2,$$

an application of Proposition 5.3 shows that the subtree constructed in (iv) satisfies $\#\tilde{V} \leq T_2 \#\hat{V}$. The proof is completed by noting that the partitions $\tilde{P}$ and $\tilde{P}^a$ constructed in (v) and (vi) satisfy $\#\tilde{P}^a \lesssim \#\tilde{P} \lesssim \#\tilde{V}$.
(b). The number of arithmetic operations required by (i)-(iii) is $\lesssim \#P^a \lesssim \#P$.

Let $\bar{V}$ be the enlargement of V_{P^a} by adding all siblings of all $v \in V_{P^a}$ to this set. Then $\bar{V}$ is a subtree with $\#\tilde{V} \leq \#\bar{V} \lesssim \#V_{P^a} \lesssim \#P^a \lesssim \#P$, and $e(v) = 0$ for all $v \in \mathcal{L}(\bar{V})$. Choosing q in the modified Algorithm 5.1 equal to $q = \max\{0, \lceil \log_2(\varepsilon^{-2} |\!|\!| w_P |\!|\!|_1^2 \#\bar{V}) \rceil\}$, Proposition 5.3 shows that the number of arithmetic operations required by (iv) is $\lesssim \#\tilde{V} + \max\{0, \log(\varepsilon^{-2} |\!|\!| w_P |\!|\!|_1^2 \#\bar{V})\} \lesssim \#P + \max\{0, \log(\varepsilon^{-1} |\!|\!| w_P |\!|\!|_1)\}$.

The number of arithmetic operations required by (v)-(vii) is $\lesssim \#\tilde{V} \lesssim \#P$. □

When the unmodified Algorithm 5.1 would have been applied inside **COARSE**, the required number of arithmetic operations would have been $\lesssim \#P \log(\#P)$. In contrast to such a log-factor, for our application it will turn out that the log-*term* from Theorem 5.4(b) is fully harmless. The next corollary demonstrates the strength of applying a coarsening procedure. It shows that if for any $u \in H_0^1(\Omega)$ and $s > 0$ the errors of the best approximations from any $\mathcal{S}_P$ with $\#P \leq n$ are $\mathcal{O}(n^{-s})$, then given any $\varepsilon > 0$, a partition P and a $w_P \in \mathcal{S}_P$ with $|u - w_P|_1 \leq \varepsilon$, by allowing the tolerance to increase by some suitable, sufficiently large constant factor, the coarsening procedure yields an (admissible) partition $\tilde{P}^a$ and a $\tilde{w} \in \mathcal{S}_{\tilde{P}^a}$ with $|u - w_{\tilde{P}^a}|_1 \lesssim \varepsilon$ and $\#\tilde{P}^a \lesssim \varepsilon^{-1/s}$, which in view of the assumption is the smallest size, modulo some constant factor, one generally can expect for an approximation with this accuracy. The short proof of this corollary is based on an argument taken from [BDD02, proof of Theorem 4.9].

Corollary 5.5. *Let $\gamma > t_1^{-\frac{1}{2}}$. Then for any $\varepsilon > 0$, $u \in H_0^1(\Omega)$, a partition P, $w_P \in \mathcal{S}_P$ with $|||u - w_P|||_1 \leq \varepsilon$, for $[\tilde{P}^a, w_{\tilde{P}^a}] := \mathbf{COARSE}[P, w_P, \gamma\varepsilon]$ it holds that $|||u - w_{\tilde{P}^a}|||_1 \leq (1+\gamma)\varepsilon$ and*

$$\tilde{P}^a \leq D\#\hat{P},$$

for any partition $\hat{P}$ with $\inf_{w_{\hat{P}} \in \mathcal{S}_{\hat{P}}} |||u - w_{\hat{P}}|||_1 \leq (t_1^{\frac{1}{2}}\gamma - 1)\varepsilon$.

Proof. The first statement is an obvious consequence of Theorem 5.4. The second one also follows from this theorem using that

$$\inf_{w_{\hat{P}} \in \mathcal{S}_{\hat{P}}} |||w_P - w_{\hat{P}}|||_1 \leq |||u - w_P|||_1 + \inf_{w_{\hat{P}} \in \mathcal{S}_{\hat{P}}} |||w_{\hat{P}} - u|||_1 \leq \varepsilon + (t_1^{\frac{1}{2}}\gamma - 1)\varepsilon = t_1^{\frac{1}{2}}\gamma\varepsilon.$$

□

Remark 5.6. The argument, introduced in [BDD02], that was used to conclude the above corollary can also be applied to give a much simpler proof of the earlier similar result [CDD01, Corollary 5.2] in the wavelet context. Moreover, it shows that in that case, which corresponds to $t_1 = 1$, it is sufficient that the coarsening increases the tolerance with any factor larger than 2 instead of a factor larger than 5.

6. A convergent adaptive refinement strategy

We derive an a posteriori estimate of the error in the Galerkin approximation (4.1) of the solution of the boundary value problem (2.1), where we, temporarily, assume that the right-hand side $f \in L_2(\Omega)$. Secondly, under the assumption that f is even piecewise constant with respect to the current partition, we derive a refinement strategy that guarantees that the difference between the Galerkin solutions on the new and old partition is greater than some fixed multiple of the error estimate for the solution on the old partition. Exploiting Galerkin orthogonality we can therefore conclude that the error in the new solution is less than some absolute constant times the error in the old solution.

This section is largely based on ideas from [MNS00] by Morin, Nochetto and Siebert on the construction of an adaptive finite element method that can be proven to converge. Differences are that

- we consider a different type of partitions (non-conforming ones generated by red-refinements, vs. conforming ones generated by newest vertex bisection).
- Under Assumption 3.8 our results are valid uniformly in the size of jumps of $\rho = \rho(\mathbf{A})$.
- We analyze the adaptive method as it is performed in practice. That is, we allow that the discrete systems are set up with an approximation of the right-hand

side, with which the application of quadrature can be modeled, and that they are solved inexactly. We consider the a posteriori error estimator as it depends on the approximation of the right-hand side and the inexact discrete solution.

- We allow right-hand sides $f \in H^{-1}(\Omega)$ instead of $f \in L_2(\Omega)$,

We start by introducing some notation. We call e *an edge of a partition* P, when e is an edge of some $\triangle \in P$ and e connects two vertices from $\bar{V}_P$. Note that since we allow non-conforming partitions, not all edges of $\triangle \in P$ are edges of P. With $\bar{E}_P$ or E_P respectively, we denote the set of all edges of P or all edges of P which are not part of $\partial\Omega$. Note that for an admissible partition P^a, any edge $e \in E_{P^a}$ is either the common edge of $\triangle_1, \triangle_2 \in P^a$, or it is the common edge of $\triangle_1, \hat{\triangle}$ where $\triangle_1 \in P^a$ and $\hat{\triangle}$ is the parent of four triangles in P^a.

For $e \in E_P$ and $u \in H^1(\Omega)$, we put

$$\eta_e(u) := \frac{\operatorname{diam}(e)}{\max\{\rho|_{e^-}, \rho|_{e^+}\}} \|[\mathbf{A}\nabla u]_e \cdot \mathbf{n}_e\|^2_{L_2(e)},$$

where $\mathbf{n}_e$ is an unit vector orthogonal to e, $[\mathbf{A}\nabla u]_e$ denotes the jump of $\mathbf{A}\nabla u$ in the direction of $\mathbf{n}_e$, and $\rho|_{e^\pm} = \rho(x \pm \delta\mathbf{n}_e)$ for arbitrary $x \in e$ and $\delta > 0$ small enough. For $\triangle \in \cup_{i\geq 0} P_i^*$ and $f \in L_2(\Omega)$, we put

$$\zeta_\triangle(f) := \frac{\operatorname{diam}(\triangle)^2}{\rho|_\triangle} \|f\|^2_{L_2(\triangle)}.$$

Finally, for a partition P, $f \in L_2(\Omega)$ and $u \in H^1(\Omega)$, we put

$$\mathcal{E}(P, f, u) := [\sum_{\triangle \in P} \zeta_\triangle(f) + \sum_{e \in E_P} \eta_e(u)]^{\frac{1}{2}}$$

For $\mathbf{A} = \mathbf{id}$, and thus $\rho \equiv 1$, and conforming partitions the a posteriori estimate given by the following theorem can already be found in [BM87, Ver96].

Theorem 6.1. *There exists an absolute constant C_1, such that for any $f \in L_2(\Omega)$ and admissible partition P^a, with $u := L^{-1}f$ and $u_{P^a} := L_{P^a}^{-1} f$, it holds that*

$$|u - u_{P^a}|_1 \leq C_1 \mathcal{E}(P^a, f, u_{P^a}).$$

Proof. For any $w \in H_0^1(\Omega)$, $w_{P^a} \in \mathcal{S}_{P^a}$, by the Galerkin orthogonality and because $\mathbf{A}$ is piecewise constant, integration by parts shows that

$$\begin{aligned} a(u - u_{P^a}, w) = a(u - u_{P^a}, w - w_{P^a}) &= \int_\Omega f(w - w_{P^a}) - a(u_{P^a}, w - w_{P^a}) \\ &= \int_\Omega f(w - w_{P^a}) - \sum_{\triangle \in P^a} \int_{\partial\triangle} (\mathbf{A}\nabla u_{P^a} \cdot \mathbf{n}_\triangle)(w - w_{P^a}) \\ &= \sum_{\triangle \in P^a} \int_\triangle f(w - w_{P^a}) - \sum_{e \in E_{P^a}} \int_e ([\mathbf{A}\nabla u_{P^a}]_e \cdot \mathbf{n}_e)(w - w_{P^a}), \end{aligned} \tag{6.1}$$

where $\mathbf{n}_\triangle$ denotes the unit exterior normal to $\triangle$.

Instead of the Clément interpolator used in [Ver96], we apply a quasi-interpolator to construct a suitable w_{P^a}: As shown in [Osw94, p. 17-18], for any triangle $\triangle$ and any of its vertices v, there exist a $\varphi(\triangle, v) \in L_\infty(\triangle)$ such that $\int_\triangle \varphi(\triangle, v)p = p(v)$ for all polynomials p of degree 1, whereas $\|\varphi(\triangle, v)\|_{L_\infty} \lesssim \operatorname{meas}(\triangle)^{-1}$ independent of $\triangle$. For each $v \in V_{P^a}$ we now select $\triangle_v = \operatorname{argmax}\{\rho|_\triangle : \triangle \in P^a, v \in \triangle\}$, and define $w_{P^a} \in \mathcal{S}_{P^a}$ by $w_{P^a}(v) = \int_{\triangle_v} \varphi(\triangle_v, v)w$.

We start with estimating the first sum from (6.1). Let $\triangle \in P^a$, say $\triangle \in P^a_i$. The triangle $\triangle$ has between 0 and 3 hanging vertices. Yet, by Proposition 3.4, since P^a is admissible, each of these hanging vertices is the midpoint of an edge connecting two vertices from $\bar{V}_{P^a_{i-1}}$. In the following we consider the case that $\triangle$ has one hanging vertex, and so in particular $i > 0$, but the other cases can treated similarly.

Let $\tilde{\triangle} \in P^a_{i-1}$ be the parent of $\triangle$, v_1, v_2 the vertices of $\tilde{\triangle}$ such that the hanging vertex of $\triangle$ is the midpoint of the edge connecting v_1 and v_2, and let v_3 and v_4 be the non-hanging vertices of $\triangle$, see Figure 6. The linear function $w_{P^a}|_\triangle$ is the solution of the elementary

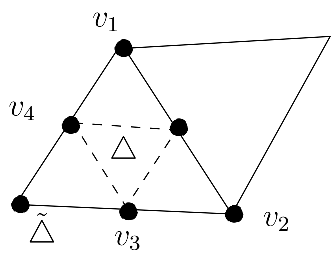

FIGURE 5. Illustration with the proof of Theorem 6.1.

interpolation problem with data $w_{P^a}(v_i)$ $(1 \leq i \leq 4)$, and it is easily seen that $\|w_{P^a}\|_{L_2(\triangle)} \lesssim \operatorname{meas}(\triangle)^{\frac{1}{2}} \max_{1 \leq i \leq 4} |w_{P^a}(v_i)|$. By construction $|w_{P^a}(v_i)| \lesssim \operatorname{meas}(\triangle_{v_i})^{-\frac{1}{2}} \|w\|_{L_2(\triangle_{v_i})}$, and so $\|w_{P^a}\|_{L_2(\triangle)} \lesssim \|w\|_{L_2(\Omega_\triangle)}$, where $\Omega_\triangle = \cup_{i=1}^4 \triangle_{v_i}$.

Since ρ is piecewise constant with respect to P^a_{i-1}, it holds that $\rho|_\triangle = \rho|_{\tilde{\triangle}} \leq \inf_{x \in \Omega_\triangle} \rho(x)$ by definition of the mapping $v \mapsto \triangle_v$. Furthermore, we may assume that, if not inside $\tilde{\triangle}$, both $\triangle_{v_3}$ and $\triangle_{v_4}$ have an edge on $\partial\tilde{\triangle}$.

For $i = 1$ or 2, if $\triangle_{v_i} \neq \tilde{\triangle}$ and these triangles also do not share an edge, then by definition of ρ being quasi-monotone (see Assumption 3.8), there exist $\tilde{\triangle} = \triangle_1^{(i)}, \ldots, \triangle_{m_i}^{(i)} = \triangle_{v_i} \subset P^a$ such that $v_i \in \cap_j \triangle_j^{(i)}$, $\triangle_j^{(i)}$ shares an edge with $\triangle_{j+1}^{(i)}$, and $\rho|_{\tilde{\triangle}} \lesssim \inf_j \rho|_{\triangle_j^{(i)}}$.

We conclude that the set $\tilde{\Omega}_\triangle := \Omega_\triangle \cup \tilde{\triangle} \cup \{\triangle_j^{(i)} : i \in \{1,2\}, 1 \leq j \leq m_i\}$ is a simply connected uniformly Lipschitz domain that consists of a uniformly bounded number of triangles from P^a, with $\rho|_\triangle \lesssim \inf_{x \in \tilde{\Omega}_\triangle} \rho(x)$ and $\|w - w_{P^a}\|_{L_2(\triangle)} \lesssim \|w\|_{L_2(\tilde{\Omega}_\triangle)} \leq \|w\|_{H^1(\tilde{\Omega}_\triangle)}$. Since our interpolator reproduces all polynomials of first order degree, and so in particular any constant, from an application of the Bramble-Hilbert lemma and a homogeneity argument we infer that

$$\|w - w_{P^a}\|^2_{L_2(\triangle)} \lesssim \operatorname{diam}(\tilde{\Omega}_\triangle)^2 |w|^2_{H^1(\tilde{\Omega}_\triangle)} \lesssim \frac{\operatorname{diam}(\triangle)^2}{\rho|_\triangle} \sum_{\{\hat{\triangle} \in P^a : \hat{\triangle} \subset \tilde{\Omega}_\triangle\}} \rho|_{\hat{\triangle}} |w|^2_{H^1(\hat{\triangle})}.$$

From an application of the Cauchy-Schwarz inequality, we conclude that

$$|\sum_{\triangle \in P^a} \int_\triangle f(w - w_{P^a})|^2 \lesssim \sum_{\triangle \in P^a} \frac{\operatorname{diam}(\triangle)^2}{\rho|_\triangle} \|f\|^2_{L_2(\triangle)} \sum_{\triangle \in P^a} \rho|_\triangle |w|^2_{H^1(\triangle)}. \tag{6.2}$$

Now we estimate the second sum from (6.1). Let $P^a = P^a_n$ and $e \in E_{P^a}$, then $e = \triangle_1 \cap \triangle_2$ for some $\triangle_1, \triangle_2 \in \cup_{i=0}^n P^a_i$. Let us assume that $\rho|_{\triangle_1} \geq \rho|_{\triangle_2}$. Note that either $\triangle_1 \in P^a$ or it is the parent of four triangles from P^a.

From the trace theorem we have $\|w\|_{L_2(e)} \lesssim \|w\|_{H^1(\triangle_1)}$. With $v_1, v_2 \in V_{P^a}$ being the endpoints of e, we have $\|w_{P^a}\|_{L_2(e)} \leq \operatorname{meas}(e)^{\frac{1}{2}} \max\{|w_{P^a}(v_1)|, |w_{P^a}(v_2)|\}$. From $|w_{P^a}(v_i)| \lesssim \operatorname{meas}(\triangle_{v_i})^{-\frac{1}{2}} \|w\|_{L_2(\triangle_{v_i})}$, we find that $\|w - w_{P^a}\|_{L_2(e)} \lesssim \operatorname{meas}(e)^{-\frac{1}{2}} \|w\|_{H^1(\triangle_1 \cup \triangle_{v_1} \cup \triangle_{v_2})}$.

If $\triangle_1 \cup \triangle_{v_1} \cup \triangle_{v_2}$ is not already a simply connected uniformly Lipschitz domain, then since ρ is quasi-monotone, analogous arguments as used above show that it can be extended to such a domain Ω_e consisting of a uniformly bounded number of triangles from P^a with $\max\{\rho|_{e^-}, \rho|_{e^+}\} \lesssim \inf_{x \in \Omega_e} \rho(x)$.

Since our interpolator reproduces all polynomials of first order degree, and so in particular any constant, the Bramble-Hilbert lemma and a homogeneity argument show that

$$\|w - w_{P^a}\|^2_{L_2(e)} \lesssim \operatorname{diam}(e) |w|^2_{H^1(\Omega_e)} \lesssim \frac{\operatorname{diam}(e)}{\max\{\rho|_{e^-}, \rho|_{e^+}\}} \sum_{\{\triangle \in P^a : \triangle \subset \Omega_e\}} \rho|_\triangle |w|^2_{H^1(\triangle)}.$$

From an application of the Cauchy-Schwarz inequality, we conclude that

$$\begin{aligned} &|\sum_{e \in E_{P^a}} \int_e ([\mathbf{A}\nabla u_{P^a}]_e \cdot \mathbf{n}_e)(w - w_{P^a})|^2 \\ &\qquad \lesssim \sum_{e \in E_{P^a}} \frac{\operatorname{diam}(e)}{\max\{\rho|_{e^-}, \rho|_{e^+}\}} \|[\mathbf{A}\nabla u_{P^a}]_e \cdot \mathbf{n}_e\|^2_{L_2(e)} \sum_{\triangle \in P^a} \rho|_\triangle |w|^2_{H^1(\triangle)}. \end{aligned} \tag{6.3}$$

Upon using that $\sum_{\triangle \in P^a} \rho|_\triangle |w|^2_{H^1(\triangle)} \eqsim |w|^2_1$ and by substituting $w = u - u_{P^a}$, the proof follows from (6.1), (6.2), (6.3). □

The next lemma, which is based on [MNS00, Lemma 4.2], gives local lower bounds of the difference between the Galerkin solutions on some partition and a refinement of this partition. Differences with [MNS00] are that we consider a different class of partitions, and that our results hold uniformly in the size of jumps of ρ. Furthermore, we simply assume that the right-hand side f is piecewise constant with respect to the first partition and postpone the analysis of having an f that does not satisfy this.

Lemma 6.2. *Let P^a be an admissible partition and $\tilde{P}$ a refinement of P^a. Let $f \in L_2(\Omega)$ be* piecewise constant with respect to *P^a, i.e., $f \in \mathcal{S}^0_{P^a}$, and let $u_{P^a} = L^{-1}_{P^a} f, u_{\tilde{P}} = L^{-1}_{\tilde{P}} f$ be the corresponding Galerkin solutions.*

(a). Let $\triangle_1, \triangle_2 \in P^a$ such that $e := \triangle_1 \cap \triangle_2 \in E_{P^a}$. Assume that $V_{\tilde{P}}$ contains points interior to $\triangle_1$, $\triangle_2$ and e, see Figure 6.2. Then

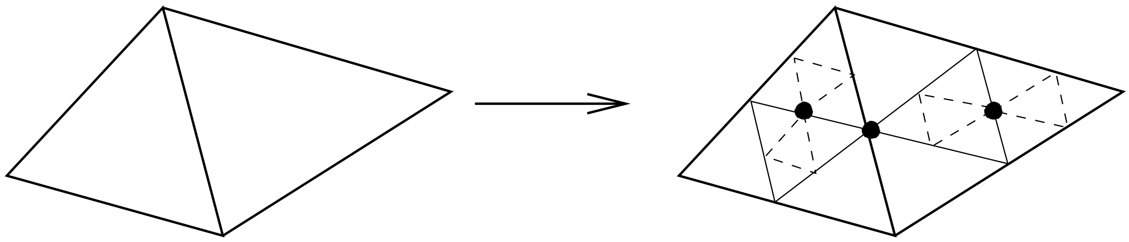

FIGURE 6. Illustration with Lemma 6.2(a).

$$|u_{\tilde{P}} - u_{P^a}|^2_{1,\triangle_1 \cup \triangle_2} \gtrsim \eta_e(u_{P^a}) + \sum_{i=1}^{2} \zeta_{\triangle_i}(f).$$

(b). Let $\triangle_1, \hat{\triangle}$ such that $e := \triangle_1 \cap \hat{\triangle} \in E_{P^a}$, $\triangle_1 \in P^a$ and $\hat{\triangle}$ is the parent of four triangles $\triangle_2, \ldots, \triangle_5 \in P^a$, numbered such that $\triangle_2$, $\triangle_3$ have an edge e_2, e_3 on e. Assume that $V_{\tilde{P}}$ contains points interior to $\triangle_1$, $\triangle_2$, $\triangle_3$, e_2 and e_3, see Figure 6.2. Then

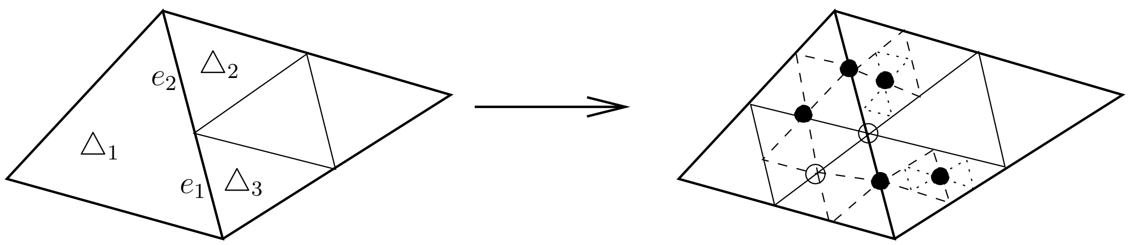

FIGURE 7. Illustration with Lemma 6.2(b). Open circles correspond to degrees of freedom that are not used.

$$|u_{\tilde{P}} - u_{P^a}|^2_{1,\triangle_1 \cup \triangle_2 \cup \triangle_3} \gtrsim \eta_e(u_{P^a}) + \sum_{i=1}^{3} \zeta_{\triangle_i}(f).$$

(c). Assume that $V_{\tilde{P}}$ contains a point interior to $\triangle \in P^a$. Then

$$|u_{\tilde{P}} - u_{P^a}|^2_{1,\triangle} \gtrsim \zeta_{\triangle}(f).$$

Proof. (a). By assumption there exist $\varphi_1, \varphi_2, \varphi_3 \in H^1_0(\triangle_1 \cup \triangle_2) \cap \mathcal{S}_{\tilde{P}}$ with $|\varphi_i|_1 = 1$, and for $i \in \{1, 2\}$

$$\operatorname{supp}(\varphi_i) \subset \triangle_i, \quad \int_{\triangle_i} \varphi_i \eqsim \frac{\operatorname{meas}(\triangle_i)}{\rho\big|_{\triangle_i}^{\frac12}}, \quad \int_{\triangle_i} \varphi_3 \eqsim \frac{\operatorname{meas}(\triangle_i)}{\max\{\rho\big|_{e^-}, \rho\big|_{e^+}\}^{\frac12}}, \quad \int_e \varphi_3 \eqsim \frac{\operatorname{meas}(e)}{\max\{\rho\big|_{e^-}, \rho\big|_{e^+}\}^{\frac12}}.$$

For any $\varphi = \sum_{i=1}^{3} \mathbf{c}_i \varphi_i$, integration by parts shows that

$$\begin{aligned}(6.4)\qquad \int_{\triangle_1\cup\triangle_2} f\varphi - \int_e ([\mathbf{A}\nabla u_{P^a}]_e\cdot\mathbf{n}_e)\varphi &= a(u_{\tilde{P}} - u_{P^a}, \varphi) \leq \\ |u_{\tilde{P}} - u_{P^a}|_{1,\triangle_1\cup\triangle_2} |\varphi|_1 &\lesssim |u_{\tilde{P}} - u_{P^a}|_{1,\triangle_1\cup\triangle_2} \|\mathbf{c}\|.\end{aligned}$$

Let $g_j \in H_0^1(\triangle_1\cup\triangle_2)'$ be defined by $g_j(\varphi) = \frac{\rho\big|_{\triangle_j}^{\frac{1}{2}}}{\operatorname{diam}(\triangle_j)\operatorname{meas}(\triangle_j)^{\frac{1}{2}}}\int_{\triangle_j}\varphi$ when $j = 1$ or 2, and $g_3(\varphi) = \frac{\max\{\rho|_{e^-},\rho|_{e^+}\}^{\frac{1}{2}}}{(\operatorname{diam}(e)\operatorname{meas}(e))^{\frac{1}{2}}}\int_e \varphi$, and let $\mathbf{B}_{ij} := g_j(\varphi_i)$. Then with $\mathbf{d}_j = \frac{\operatorname{diam}(\triangle_j)}{\rho\big|_{\triangle_j}^{\frac{1}{2}}}\|f\|_{L_2(\triangle)}$ when $j = 1$ or 2, and $\mathbf{d}_3 = \frac{\operatorname{diam}(e)^{\frac{1}{2}}}{\max\{\rho|_{e^-},\rho|_{e^+}\}^{\frac{1}{2}}}\|[\mathbf{A}\nabla u]_e\cdot\mathbf{n}_e\|_{L_2(e)}$, the left-hand side of (6.4) reads as $\sum_{i,j=1}^{3}\mathbf{d}_j\mathbf{c}_i\mathbf{B}_{ij}$. Using (6.4), the statement of the lemma reduces to the question whether $\sup_{\mathbf{c}}\frac{\langle\mathbf{Bd},\mathbf{c}\rangle}{\|\mathbf{c}\|} \gtrsim \|\mathbf{d}\|$ or $\|\mathbf{B}^{-1}\| \lesssim 1$. Since $\mathbf{B}_{ii} \eqsim 1$, and $\mathbf{B}_{31}, \mathbf{B}_{32} \lesssim 1$, whereas the other coefficients of $\mathbf{B}$ are zero, the proof of (a) is completed.

Part (b) can be proven similar to (a). Note that generally $[\mathbf{A}\nabla u_{P^a}]_e$ has different values on e_2 and e_3. The proof of (c) poses no additional difficulties. $\square$

As an immediate consequence we have

Corollary 6.3. *Let P^a be an admissible partition, $f \in \mathcal{S}^0_{P^a}$, and let $\tilde{P}$ be a refinement of P^a, such that for some $G \subset P^a$, $F \subset E_{P^a}$, for all $\triangle \in G$, $V_{\tilde{P}}$ satisfies the conditions from Lemma 6.2(c), and for all $e \in F$, $V_{\tilde{P}}$ satisfies the conditions from either Lemma 6.2(a) or Lemma 6.2(b) (sufficient is when $\tilde{P}$ contains all grandchildren of all $\triangle \in P^a$ which either are in G or have an edge on an $e \in F$). Then for $u_{P^a} = L_{P^a}^{-1}f, u_{\tilde{P}} = L_{\tilde{P}}^{-1}f$ being the corresponding Galerkin solutions, it holds that*

$$|u_{\tilde{P}} - u_{P^a}|_1^2 \geq c_2^2\{\sum_{\triangle\in F}\zeta_\triangle(f) + \sum_{e\in G}\eta_e(u_{P^a})\},$$

for some absolute constant $c_2 > 0$.

Since Lemma 6.2 also applies when $u_{\tilde{P}} = L_{\tilde{P}}^{-1}f$ is replaced by $u = L^{-1}f$, for later use we state the following

Corollary 6.4. *Let P^a be an admissible partition and let $f \in \mathcal{S}^0_{P^a}$. With $u := L^{-1}f$, $u_{P^a} := L_{P^a}^{-1}f$ and c_2 the constant from Corollary 6.3, we have*

$$|u - u_{P^a}|_1 \geq c_2\mathcal{E}(P^a, f, u_{P^a}).$$

The idea to obtain a convergent adaptive refinement strategy is to select the sets F and G from Corollary 6.3 such that $\sum_{\triangle\in F}\zeta_\triangle(f) + \sum_{e\in G}\eta_e(u_{P^a})$ is bounded from below by some multiple of $\sum_{\triangle\in P^a}\zeta_\triangle(f) + \sum_{e\in E_{P^a}}\eta_e(u_{P^a}) = \mathcal{E}(P^a, f, u_{P^a})^2$. Convergence then follows from a combination of Theorem 6.1 and this Corollary 6.3. In Corollary 6.3 it was assumed that the right-hand side f is piecewise constant with respect to the current partition P^a, and that the resulting exact Galerkin solution is available. The following two lemmas will be used to relax both these two unrealistic assumptions.

Lemma 6.5. *There exists an absolute constant $C_3 > 0$ such that for any partition P, $f, \tilde{f} \in L_2(\Omega)$, $u, \tilde{u} \in H^1(\Omega)$,*

$$|\mathcal{E}(P,f,u) - \mathcal{E}(P,\tilde{f},\tilde{u})| \leq [\sum_{\triangle\in P} \zeta_\triangle(f-\tilde{f})]^{\frac{1}{2}} + C_3|u-\tilde{u}|_1$$

Proof. It holds that

$$\begin{aligned}|\mathcal{E}(P,f,u) - \mathcal{E}(P,\tilde{f},\tilde{u})| &= [\sum_{\triangle\in P} \zeta_\triangle(f) + \sum_{e\in E_P} \eta_e(u)]^{\frac{1}{2}} - [\sum_{\triangle\in P} \zeta_\triangle(\tilde{f}) + \sum_{e\in E_P} \eta_e(\tilde{u})]^{\frac{1}{2}} \\ &\leq [\sum_{\triangle\in P} (\zeta_\triangle(f)^{\frac{1}{2}} - \zeta_\triangle(\tilde{f})^{\frac{1}{2}})^2 + \sum_{e\in E_P} (\eta_e(u)^{\frac{1}{2}} - \eta_e(\tilde{u})^{\frac{1}{2}})^2]^{\frac{1}{2}} \\ &\leq [\sum_{\triangle\in P} (\zeta_\triangle(f)^{\frac{1}{2}} - \zeta_\triangle(\tilde{f})^{\frac{1}{2}})^2]^{\frac{1}{2}} + [\sum_{e\in E_P} (\eta_e(u)^{\frac{1}{2}} - \eta_e(\tilde{u})^{\frac{1}{2}})^2]^{\frac{1}{2}},\end{aligned}$$

and $(\zeta_\triangle(f)^{\frac{1}{2}} - \zeta_\triangle(\tilde{f})^{\frac{1}{2}})^2 \leq \zeta_\triangle(f-\tilde{f})$. For any $e \in E_P$,

$$\begin{aligned}|\eta_e(u)^{\frac{1}{2}} - \eta_e(\tilde{u})^{\frac{1}{2}}| &= \frac{\operatorname{diam}(e)^{\frac{1}{2}}}{\max\{\rho|_{e^-}, \rho|_{e^+}\}^{\frac{1}{2}}} \big|\|[\mathbf{A}\nabla u]_e \cdot \mathbf{n}_e\|_{L_2(e)} - \|[\mathbf{A}\nabla \tilde{u}]_e \cdot \mathbf{n}_e\|_{L_2(e)}\big| \\ &\leq \frac{\operatorname{diam}(e)^{\frac{1}{2}}}{\max\{\rho|_{e^-}, \rho|_{e^+}\}^{\frac{1}{2}}} \|[\mathbf{A}\nabla(u-\tilde{u})]_e \cdot \mathbf{n}_e\|_{L_2(e)}.\end{aligned}$$

The proof is completed by the observation that for any edge e_i of a $\triangle \in P$, and any $w \in P_1(\triangle)$ and unit vector $\mathbf{n}$, from the trace theorem and a homogeneity argument it follows that

$$\operatorname{diam}(e_i)^{\frac{1}{2}} \|\mathbf{A}\nabla w \cdot \mathbf{n}\|_{L_2(e_i)} \lesssim \rho|_\triangle |w|_{H^1(\triangle)} \eqsim \rho|_\triangle^{\frac{1}{2}} |w|_{1,\triangle}.$$

□

For a partition P, let $Q_P^{(0)} : L_2(\Omega) \to \mathcal{S}_P^{(0)}$ be defined by $(Q_P^{(0)} g)|_\triangle = \operatorname{meas}(\triangle)^{-1} \int_\triangle g$ ($\triangle \in P$), and for $g \in L_2(\Omega)$, let

$$\operatorname{osc}(g,P) := \inf_{g_p \in \mathcal{S}_P^{(0)}} [\sum_{\triangle\in P} \zeta_\triangle(g-g_P)]^{\frac{1}{2}} = [\sum_{\triangle\in P} \zeta_\triangle(g - Q_P^{(0)} g)]^{\frac{1}{2}}.$$

Lemma 6.6. *There exists an absolute constant $C_4 > 0$, such that for any partition and $g \in L_2(\Omega)$,*

$$\|g - Q_P^{(0)} g\|_{-1} \leq C_4 \operatorname{osc}(g,P).$$

Proof. The proof follows from $\|g - Q_p^0 g\|_{-1} = \sup_{0\neq w\in H_0^1(\Omega)} \frac{|\int_\Omega (g-Q_P^{(0)}g)w|}{|w|_1}$ and

$$
\begin{aligned}
&|\int_\Omega (g - Q_P^{(0)} g)w| = \int_\Omega (g - Q_P^{(0)} g)(w - Q_P^{(0)} w)| \\
&\leq \sum_{\triangle\in P} \|g - Q_P^{(0)} g\|_{L_2(\triangle)} \|w - Q_P^{(0)} w\|_{L_2(\triangle)} \lesssim \sum_{\triangle\in P} \|g - Q_P^{(0)} g\|_{L_2(\triangle)} \mathrm{diam}(\triangle) \|\nabla w\|_{L_2(\triangle)} \\
&\leq [\sum_{\triangle\in P} \frac{\mathrm{diam}(\triangle)^2}{\rho|_\triangle} \|g - Q_P^{(0)} g\|^2_{L_2(\triangle)}]^{\frac{1}{2}} [\sum_{\triangle\in P} \rho|_\triangle \|\nabla w\|^2_{L_2(\triangle)}]^{\frac{1}{2}} \\
&\lesssim \mathrm{osc}(g, P)|w|_1.
\end{aligned}
$$

□

For some *fixed constant* $\theta \in (0, 1]$, we consider the following refinement procedure:

REFINE$[P^a, f, w_{P^a}] \to \tilde{P}$
% P^a is an admissible partition, $f \in L_2(\Omega)$ with $(\|f\|_{L_2(\triangle)})_{\triangle\in P^a}$ being computationally
% available, and $w_{P^a} \in \mathcal{S}_{P^a}$ is given by its values $(w(v))_{v\in V_{P^a}}$.
Select $F \subset P^a$, $G \subset E_{P^a}$ such that

$$\sum_{\triangle\in F} \zeta_\triangle(f) + \sum_{e\in G} \eta_e(w_{P^a}) \gtrsim \theta^2\, \mathcal{E}(P^a, f, w_{P^a})^2.$$

Determine a refinement $\tilde{P}$ of P^a such that for all $\triangle \in F$, $V_{\tilde{P}}$ satisfies the conditions from Lemma 6.2(c), and for all $e \in G$, $V_{\tilde{P}}$ satisfies the conditions from either Lemma 6.2(a) or Lemma 6.2(b), where at the same time each $\tilde{\triangle} \in \tilde{P}$ is either in P^a or it is a child or a grandchild of a $\triangle \in P^a$.

As long as the selection of F and G is organized such that it does not involve the exact ordering of all $\zeta_\triangle(f)$ and $\eta_e(w_{P^a})$ by their modulus (cf. discussion from Remark 5.2), we have

Proposition 6.7. *The call* **REFINE**$[P^a, f, w_{P^a}]$ *requires a number of arithmetic operations $\lesssim \#P^a$.*

As stated before, we will consider right-hand sides $f \in H^{-1}(\Omega)$, but at the same time we will allow an approximate right-hand side to be used for setting up a discrete system. In fact, since the terms $\zeta_\triangle(f)$ are only defined for $f \in L_2(\Omega)$, generally we will need approximate right-hand sides different from the exact one. The following theorem shows that if the current approximate right-hand side is sufficiently close to both the exact one and to some piecewise constant function subordinate to the current partition, and **REFINE** is called with this approximate right-hand side and a sufficiently accurate approximation of the discrete solution, then for the discrete solution on the new partition with again an approximate right-hand side that is sufficiently close to the exact one, it holds that the error is less than some constant multiple less than 1 of the error on the previous partition. Together with the coarsening routine from §5, in the next two subsections this theorem will be the basis to construct adaptive finite element methods that converge with optimal rates.

Theorem 6.8. *Let $f \in H^{-1}(\Omega)$, $u = L^{-1}f$, P^a be an admissible partition, $f_{P^a} \in L_2(\Omega)$, $u_{P^a} = L_{P^a}^{-1} f_{P^a}$, $\bar{u}_{P^a} \in \mathcal{S}_{P^a}$, $\tilde{P} = \mathbf{REFINE}[P^a, f_{P^a}, \bar{u}_{P^a}]$ or a refinement of it, $f_{\tilde{P}} \in L_2(\Omega)$ and $u_{\tilde{P}} = L_{\tilde{P}}^{-1} f_{\tilde{P}}$. Then*

$$(6.5)\qquad \begin{aligned}|u - u_{\tilde{P}}|_1 \leq & [1 - \tfrac{1}{2}(\tfrac{c_2\theta}{C_1})^2]^{\frac{1}{2}} |u - u_{P^a}|_1 + 2c_2 C_3 |u_{P^a} - \bar{u}_{P^a}|_1 \\ & + C_5 \|f - f_{P^a}\|_{-1} + C_6 \|f - f_{\tilde{P}}\|_{-1} + C_7 \mathrm{osc}(f_{P^a}, P^a),\end{aligned}$$

where $C_5, C_6, C_7 > 0$ are some absolute constants.

Proof. To be able to apply Theorem 6.1 or Corollary 6.3, we need a right-hand side in $L_2(\Omega)$ or even in $\mathcal{S}^0_{P^a}$ respectively. Let $\hat{f}_{P^a} = Q^0_{P^a} f_{P^a} \in \mathcal{S}^0_{P^a}$, and let $\hat{u} = L^{-1}\hat{f}_{P^a}$, $\hat{u}_{P^a} = L_{P^a}^{-1}\hat{f}_{P^a}$ and $\hat{u}_{\tilde{P}} = L_{\tilde{P}}^{-1}\hat{f}_{P^a}$. With $F \subset P^a$, $G \subset E_{P^a}$ as determined in the call $\tilde{P} := \mathbf{REFINE}[P^a, f_{P^a}, \bar{u}_{P^a}]$, Corollary 6.3, two applications of Lemma 6.5, and Theorem 6.1 show that

$$\begin{aligned}|\hat{u}_{\tilde{P}} - \hat{u}_{P^a}|_1 &\geq c_2[\{\sum_{\triangle \in F} \zeta_\triangle(\hat{f}_{P^a}) + \sum_{e \in G} \eta_e(\hat{u}_{P^a})\}]^{\frac{1}{2}} \\ &\geq c_2[\{\sum_{\triangle \in F} \zeta_\triangle(f_{P^a}) + \sum_{e \in G} \eta_e(\bar{u}_{P^a})\}^{\frac{1}{2}} - \{\mathrm{osc}(f_{P^a}, P^a) + C_3|\hat{u}_{P^a} - \bar{u}_{P^a}|_1\}] \\ &\geq c_2[\theta\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) - \{\mathrm{osc}(f_{P^a}, P^a) + C_3|\hat{u}_{P^a} - \bar{u}_{P^a}|_1\}] \\ &\geq c_2[\theta\mathcal{E}(P^a, \hat{f}_{P^a}, \hat{u}_{P^a}) - 2\{\mathrm{osc}(f_{P^a}, P^a) + C_3|\hat{u}_{P^a} - \bar{u}_{P^a}|_1\}] \\ &\geq c_2[\tfrac{\theta}{C_1}|\hat{u} - \hat{u}_{P^a}|_1 - 2\{\mathrm{osc}(f_{P^a}, P^a) + C_3|\hat{u}_{P^a} - \bar{u}_{P^a}|_1\}].\end{aligned}$$

Since for any scalars a, b, $(a-b)^2 \geq \frac{1}{2}a^2 - b^2$, we infer that

$$|\hat{u}_{\tilde{P}} - \hat{u}_{P^a}|_1^2 \geq \tfrac{1}{2}(\tfrac{c_2\theta}{C_1})^2 |\hat{u} - \hat{u}_{P^a}|_1^2 - 4c_2^2\{\mathrm{osc}(f_{P^a}, P^a) + C_3|\hat{u}_{P^a} - \bar{u}_{P^a}|_1\}^2.$$

Since $\hat{u}_{\tilde{P}} \in \mathcal{S}_{\tilde{P}}$ is the Galerkin approximation of $\hat{u}$ on $\mathcal{S}_{\tilde{P}}$ and $\hat{u}_{P^a} - \hat{u}_{\tilde{P}} \in \mathcal{S}_{\tilde{P}}$, we have

$$\begin{aligned}|\hat{u} - \hat{u}_{\tilde{P}}|_1^2 &= |\hat{u} - \hat{u}_{P^a}|_1^2 - |\hat{u}_{P^a} - \hat{u}_{\tilde{P}}|_1^2 \\ &\leq [1 - \tfrac{1}{2}(\tfrac{c_2\theta}{C_1})^2]|\hat{u} - \hat{u}_{P^a}|_1^2 + 4c_2^2\{\mathrm{osc}(f_{P^a}, P^a) + C_3|\hat{u}_{P^a} - \bar{u}_{P^a}|_1\}^2 \\ &\leq \{[1 - \tfrac{1}{2}(\tfrac{c_2\theta}{C_1})^2]^{\frac{1}{2}}|\hat{u} - \hat{u}_{P^a}|_1 + 2c_2\mathrm{osc}(f_{P^a}, P^a) + 2c_2C_3|\hat{u}_{P^a} - \bar{u}_{P^a}|_1\}^2,\end{aligned}$$

where we have used that $c_2 \leq C_1$ and thus that $1 - \frac{1}{2}(\frac{c_2\theta}{C_1})^2 > 0$.

The proof is completed by observing that

$$\begin{aligned}|u - \hat{u}|_1 &\leq \|f - \hat{f}_{P^a}\|_{-1} \leq \|f - f_{P^a}\|_{-1} + \|f_{P^a} - \hat{f}_{P^a}\|_{-1}, \\ |u_{P^a} - \hat{u}_{P^a}|_1 &\leq \|f_{P^a} - \hat{f}_{P^a}\|_{-1}, \\ |u_{\tilde{P}} - \hat{u}_{\tilde{P}}|_1 &\leq \|f_{\tilde{P}} - \hat{f}_{P^a}\|_{-1} \leq \|f_{\tilde{P}} - f\|_{-1} + \|f - f_{P^a}\|_{-1} + \|f_{P^a} - \hat{f}_{P^a}\|_{-1},\end{aligned}$$

and $\|f_{P^a} - \hat{f}_{P^a}\|_{-1} \leq C_4 \mathrm{osc}(f_{P^a}, P^a)$ by Lemma 6.6. □

7. A first optimal adaptive finite element method

We start with a corollary that is an easy consequence of Theorem 6.8. It extends the reduction under some conditions of the error in the discrete solutions when moving to the next partition created by **REFINE** to such a reduction of the error in sufficiently accurate approximations of the discrete solutions.

Corollary 7.1. *For any $\mu \in ([1-\frac{1}{2}(\frac{c_2\theta}{C_1})^2]^{\frac{1}{2}}, 1)$, there exists a constant $\delta > 0$ being small enough, such that if for $f \in H^{-1}(\Omega)$, an admissible partition P^a, $\bar{u}_{P^a} \in \mathcal{S}_{P^a}$, $f_{P^a} \in L_2(\Omega)$, $\tilde{P} = \textbf{REFINE}[P^a, f_{P^a}, \bar{u}_{P^a}]$ or a refinement of it, $\bar{u}_{\tilde{P}} \in \mathcal{S}_{\tilde{P}}$, $f_{\tilde{P}} \in L_2(\Omega)$ and $\varepsilon > 0$, with $u = L^{-1}f$, $u_{P^a} = L^{-1}_{P^a} f_{P^a}$ and $u_{\tilde{P}} = L^{-1}_{\tilde{P}} f_{\tilde{P}}$ it holds that $|u - \bar{u}_{P^a}|_1 \leq \varepsilon$ and*

$$|u_{P^a} - \bar{u}_{P^a}|_1 + \|f - f_{P^a}\|_{-1} + \operatorname{osc}(f_{P^a}, P^a) + |u_{\tilde{P}} - \bar{u}_{\tilde{P}}|_1 + \|f - f_{\tilde{P}}\|_{-1} \leq 2(1+\mu)\delta\varepsilon$$

then $|u - \bar{u}_{\tilde{P}}|_1 \leq \mu\varepsilon$.

Proof. The proof is an easy consequence of Theorem 6.8, $|u - \bar{u}_{\tilde{P}}|_1 \leq |u - u_{\tilde{P}}|_1 + |u_{\tilde{P}} - \bar{u}_{\tilde{P}}|_1$ and $|u - u_{P^a}|_1 \leq |u - \bar{u}_{P^a}|_1 + |u_{P^a} - \bar{u}_{P^a}|_1$. ☐

We assume the availability of the following routine

$\textbf{GALSOLVE}[P^a, f_{P^a}, u^{(0)}_{P^a}, \varepsilon] \to \bar{u}_{P^a}$
% P^a is an admissible partition, $f_{P^a} \in (\mathcal{S}_{P^a})'$ is given by $(f_{P^a}(\phi^v_{P^a}))_{v \in V_{P^a}}$, and $u^{(0)}_{P^a} \in \mathcal{S}_{P^a}$
% is given by $(u^{(0)}_{P^a}(v))_{v \in V_{P^a}}$. The output $\bar{u}_{P^a} \in \mathcal{S}_{P^a}$ is given by $(\bar{u}_{P^a}(v))_{v \in V_{P^a}}$. With
% $u_{P^a} := L^{-1}_{P^a} f_{P^a}$, it satisfies

$$|u_{P^a} - \bar{u}_{P^a}|_1 \leq \varepsilon.$$

% The call requires $\lesssim \max\{1, \log(\varepsilon^{-1}|u_{P^a} - u^{(0)}_{P^a}|_1|)\}\#P^a$ arithmetic operations.

So we do not only assume that we have an iterative solver at our disposal that converges with a rate independent of the problem size, but in accordance with the idea of an adaptive solver, additionally we assume that we have an efficient and reliable control of the algebraic error. As a consequence the number of iterations to be applied has not to depend on a possibly pessimistic a priori bound of the initial error. Two possible realizations of **GALSOLVE** are discussed in the next remark.

Remark 7.2. One can apply Conjugate Gradients starting with $u^{(0)}_{P^a}$ to the representation of $L_{P^a} u_{P^a} = f_{P^a}$ with respect to $\{\bar{\psi}^v : v \in V_{P^a}\}$. In each iteration, that takes $\eqsim \#P^a$ operations, the $|\cdot|_1$-norm of the error is multiplied with a factor less or equal to some constant $\tau < 1$ only dependent on $\kappa_{\bar{\Psi}}$, meaning that after $\lceil \log_\tau(\varepsilon |u_{P^a} - u^{(0)}_{P^a}|_1^{-1}) \rceil$ iterations this norm is $\leq \varepsilon$. The $|\cdot|_1$-norm of the error in an approximation for u_{P^a} from $\mathcal{S}_{P^a}$ is less (greater) or equal to $\lambda_{\bar{\Psi}}^{-\frac{1}{2}}$ ($\Lambda_{\bar{\Psi}}^{-\frac{1}{2}}$) times the Euclidean norm of the corresponding residual vector. So if one stops the iteration as soon as the latter norm is $\leq \lambda_{\bar{\Psi}}^{\frac{1}{2}}\varepsilon$, then the $|\cdot|_1$-norm of the error is $\leq \varepsilon$, whereas the number of iterations is bounded by

$$\lceil \log_\tau(\kappa_{\bar{\Psi}}^{-\frac{1}{2}} \varepsilon |u_{P^a} - u^{(0)}_{P^a}|_1^{-1}) \rceil \lesssim \max\{1, \log(\varepsilon^{-1}|u_{P^a} - u^{(0)}_{P^a}|_1)\},$$

showing that this approach results in a a valid routine **GALSOLVE**.

Alternatively one may apply Conjugate Gradients to the representation of $L_{P^a}u_{P^a} = f_{P^a}$ with respect to the nodal basis $\{\phi^v_{P^a} : v \in V_{P^a}\}$ using a BPX preconditioner, where similarly as above the Euclidean norm of the residual of the preconditioned system may serve to develop a stop criterium. Indeed, when $P^a = P^a_n$, for $0 \leq i \leq n$ one can select $\tilde{V}_{P^a_i} \subset V_{P^a_i}$ such that both $\mathrm{span}\{\bar{\psi}^v : v \in V_{P^a_i} \backslash V_{P^a_{i-1}}\} \subset \mathrm{span}\{\phi^v_{P^a_i} : v \in \tilde{V}_{P^a_i}\}$ and $\#\tilde{V}_{P^a_i} \lesssim \#(P^a_i \backslash P^a_{i-1})$. Using (4.6), then it can be proven that on $\mathcal{S}_{P^a}$, $\inf_{u=\sum_{i=0}^n \sum_{v\in\tilde{V}_{P^a_i}} c^v_i \phi^v_{P^a_i}} |c^v_i|^2 |\phi^v_{P^a_i}|^2_1 \eqsim |u|^2_1$ showing that the resulting BPX preconditioner, or in view of possible jumps of ρ, more precisely the MDS preconditioner (cf. [Osw94]) gives rise to uniformly well-conditioned systems, whereas it can be implemented in $\lesssim \#P^a$ operations.

Before continuing let us first explain what we mean with an optimal method for solving the boundary value problem (2.1). We consider a method being optimal, if whenever the solution u is such that for some $s > 0$ the errors of the best approximation from any $\mathcal{S}_P$ with $\#P \leq n$ are $\lesssim n^{-s}$, then for any $\varepsilon > 0$, the method yields a partition P and an $w_P \in \mathcal{S}_P$ with $|u - w_P|_1 \leq \varepsilon$ taking only $\lesssim \#P$ operations where $\#P \lesssim \varepsilon^{-1/s}$. Indeed, note that in view of the assumption on u, the smallest partition P for which there exists such a $w_P \in \mathcal{S}_P$ generally has cardinality $\eqsim \varepsilon^{-1/s}$. A definition of the class of functions $u \in H^1_0(\Omega)$ for which for some $s > 0$ the errors of the best approximations decay as indicated above is given by

$$\mathcal{A}^s = \{u \in H^1_0(\Omega) : |u|_{\mathcal{A}^s} := |u|_1 + \sup_{n\geq 0} n^s \inf_{\#P \leq \#P_0+n} \inf_{u_P\in\mathcal{S}_P} |u - u_P|_1 < \infty\},$$

where P is any partition of the type we consider.

It is well-known that for $s \leq \frac{1}{2}$, $H^1_0(\Omega) \cap H^{1+2s}(\Omega) \subset \mathcal{A}^s$. Indeed, for functions in $H^1_0(\Omega) \cap H^{1+2s}(\Omega)$, the errors of the best continuous piecewise linear approximations subordinate to the uniform refinements P^*_i of P_0 already exhibit a decay of $\lesssim (\#P^*_i)^{-s}$. Obviously, the class $\mathcal{A}^s$ contains many more functions than only those, which is the reason to consider adaptive methods anyway. For partitions generated by the so-called newest vertex bisection, a characterization of $\mathcal{A}^s$ in terms of Besov spaces can be found in [BDDP02]. We expect the same results to be valid for the type of partitions that we consider. The characterization via Besov spaces together with regularity results as can be found in [Dah99] allows to obtain a priori knowledge about in which class $\mathcal{A}^s$ the solution u of the boundary value problem (2.1) is contained. Although for any $s > 0$ the class $\mathcal{A}^s$ is non-trivial, as it contains all $u \in \mathcal{S}_P$ for any partition P, because we are approximating with piecewise linears, only for $s \leq \frac{1}{2}$ membership of $\mathcal{A}^s$ can be guaranteed by only imposing suitable smoothness conditions.

Since u is only implicitly given as the solution of (2.1), we will need an assumption about how well the right-hand side can be approximated by finite expansions, that additionally, in view of our refinement strategy, should be close to being piecewise constants. We assume the availability of the following routine:

RHS$[P, f, \varepsilon] \to [\tilde{P}^a, f_{\tilde{P}^a}]$
% P is a partition, $f \in H^{-1}(\Omega)$ and $\varepsilon > 0$. The output consists of an admissible refinement

% *$\tilde{P}^a$ of P, and an $f_{\tilde{P}^a} \in L_2(\Omega)$ with $\|f - f_{\tilde{P}^a}\|_{-1} + \operatorname{osc}(f_{\tilde{P}^a}, \tilde{P}^a) \leq \varepsilon$. This $f_{\tilde{P}^a}$ should be*
% *computationally available via $(f_{\tilde{P}^a}(\phi^v_{\tilde{P}^a}))_{v \in V_{\tilde{P}^a}}$ and $(\|f_{\tilde{P}^a}\|_{L_2(\triangle)})_{\triangle \in \tilde{P}^a}$.*

Although **RHS** allows more general approximations, one may think of $f_{\tilde{P}^a} \in \mathcal{S}^{(0)}_{P^a}$ in which case $\operatorname{osc}(f_{\tilde{P}^a}, \tilde{P}^a) = 0$.

Assuming that the solution $u \in \mathcal{A}^s$ for some $s > 0$, the costs of approximating the right-hand side f using a routine **RHS** will generally not dominate the other costs of our adaptive method only if there is some constant c_f such that for any $\varepsilon > 0$ and any partition P, for $[P^a, f_{P^a}] := \mathbf{RHS}[P, f, \varepsilon]$ both $\#P^a$ and the number of arithmetic operations required by this call are $\lesssim \#P + c_f^{1/s}\varepsilon^{-1/s}$. We will call such a pair $(f, \mathbf{RHS})$ to be *s-optimal*.

Generally the realization of a suitable routine **RHS** depends on the function $f \in H^{-1}(\Omega)$ at hand. Yet, in case $f \in L_2(\Omega)$ with $\|\rho^{-\frac{1}{2}} f\|_{L_2(\Omega)} \lesssim 1$, a simple uniform refinement procedure suffices: For some integer i to be determined below, let $\tilde{P}$ denote the smallest common refinement of the given P and P_i^*, let $\tilde{P}^a$ be its admissible refinement as a result of applying Algorithm 3.6, and let $f_{\tilde{P}^a} = Q^{(0)}_{\tilde{P}^a} f$ so that $\operatorname{osc}(f_{\tilde{P}^a}, \tilde{P}^a) = 0$. Similarly as in the proof of Lemma 6.6, for any $w \in H_0^1(\Omega)$ we have

$$|\int_\Omega (f - f_{\tilde{P}^a})w| = |\int_\Omega (f - f_{\tilde{P}^a})(w - Q^{(0)}_{\tilde{P}^a} w)| \leq C 2^{-i} \|\rho^{-\frac{1}{2}} f\|_{L_2(\Omega)} |w|_1,$$

where $C > 0$ is some absolute constant. By taking i to be the smallest integer such that $2^{-i} C \|\rho^{-\frac{1}{2}} f\|_{L_2(\Omega)} \leq \varepsilon$, we have $\|f - f_{\tilde{P}^a}\|_{-1} \leq \varepsilon$, and

$$\#\tilde{P}^a \lesssim \#P + \#P_i^* \lesssim \#P + (2^i)^2 \lesssim \#P + \varepsilon^{-2} \|\rho^{-\frac{1}{2}} f\|^2_{L_2(\Omega)}.$$

So in case for any $\triangle \in \cup_{j \geq 0} P_j^*$ the evaluation of $\int_\triangle f$ takes $\mathcal{O}(1)$ operations, we may conclude that for **RHS** based on this procedure, $(f, \mathbf{RHS})$ is s-optimal for any $s \leq \frac{1}{2}$, which as we have seen is the range of main interest. Alternatively, instead of assuming the exact evaluation of $\int_\triangle f$, one easily infers that it also suffices to approximate it with an error $\lesssim \rho^{\frac{1}{2}} \operatorname{diam}(\triangle)$, which in any case is possible in $\mathcal{O}(1)$ operations when $\rho^{-\frac{1}{2}} f$ has some piecewise smoothness with respect to P_0.

assuming additional smoothness of $\rho^{-\frac{1}{2}} f$.

Finally, although as we have seen, for $f \in L_2(\Omega)$ a simple uniform refinement strategy yields the right asymptotics, obviously depending on f an adaptive procedure may give quantitative advantages.

Remark 7.3. In [MNS00], and as a consequence in [BDD02], the exact right-hand side f is assumed to be used for setting up the discrete systems. As a consequence, to ensure linear convergence of the adaptive method, the sequence of partitions has to be selected such that in any case the term $\operatorname{osc}(f, P^a)$ decreases linearly (cf. (6.5) with $f_{P^a} = f_{\tilde{P}} = f$). A necessary condition is that $f \in L_2(\Omega)$, since otherwise $\operatorname{osc}(f, P^a)$ is even not defined. Assuming $u \in \mathcal{A}^s$, for the adaptive method developed in [BDD02], the costs of additional refinements to control $\operatorname{osc}(f, P^a)$ were shown generally not to dominate the other costs only if for any partition P and $\varepsilon > 0$, a refinement $\tilde{P}$ could be constructed with $\operatorname{osc}(f, \tilde{P}) \leq$

ε, where both the costs and $\#\tilde{P}$ are $\lesssim \#P + \varepsilon^{-1/s}$. In [BDD02, §4.4], an auxiliary adaptive procedure was developed which meets this requirement whenever f allows this, i.e., whenever $\sup_{n\geq 0} n^s \inf_{\#P\leq \#P_0+n} \operatorname{osc}(f,P) < \infty$.

By allowing inexact right-hand sides to be applied, we ended up with the condition that $\sup_{n\geq 0} n^s \inf_{\#P\leq \#P_0+n} \|f - Q_P^{(0)} f\|_{-1} < \infty$, that because of $\|f - Q_P^{(0)} f\|_{-1} \lesssim \operatorname{osc}(f,P)$ (see Lemma 6.6) is milder, and that for $s \leq \frac{1}{2}$ is satisfied for any $f \in L_2(\Omega)$, where moreover suitable partitions can simply be constructed by uniform refinements.

We are ready to give our first adaptive finite element method for solving (2.1). As expected, it is based on the repeated application of the triple **REFINE**, **RHS** and **GALSOLVE**, the latter two with suitable tolerances, which by Corollary 7.1 give rise to linearly convergent approximations. To obtain an optimal work-accuracy balance, **COARSE** is applied after any M iterations of the above triple, with M being some fixed constant.

SOLVE1$[f,\varepsilon,\bar{u}_{P_0},\varepsilon_0] \to [P^a,\bar{u}_{P^a}]$:
% *The following constants are fixed:* $\delta < \frac{1}{3}$ *and it is small enough so that it correspond to*
% *a* $\mu < 1$ *as in Corollary 7.1;* $\gamma > t_1^{-\frac{1}{2}}$ *with* t_1 *as in Proposition 5.3;* $M \in \mathbb{N}$ *such that*
% $\frac{(1+\gamma)\kappa_\Psi^{\frac{1}{2}}\mu^M}{1-3\delta} < 1$.
% *The input must satisfy* $f \in H^{-1}(\Omega)$, $\varepsilon > 0$, $\bar{u}_{P_0} \in \mathcal{S}_{P_0}$ *and* $\varepsilon_0 \geq |u - \bar{u}_{P_0}|_1$.

$P^a := P_0$, $f_{P^a} := f_{P_0}$, $\bar{u}_{P^a} := \bar{u}_{P_0}$
`if` $\varepsilon \geq \varepsilon_0$ `then` $N := 0$
`else` $N \geq 1$ *is the smallest integer with* $(\frac{\mu^M}{1-3\delta})^N((1+\gamma)\kappa_\Psi^{\frac{1}{2}})^{N-1}\varepsilon_0 \leq \varepsilon$ `fi`
`for` $i = 1,\ldots,N$ `do`
  `if` $i = 1$ `then` $\varepsilon_1 := \frac{\varepsilon_0}{1-3\delta}$
  `else` $[P^a,\bar{u}_{P^a}] :=$ **COARSE**$[P^a,\bar{u}_{P^a},\gamma\lambda_\Psi^{-\frac{1}{2}}\mu^M\varepsilon_{i-1}]$, $\varepsilon_i := \frac{(1+\gamma)\kappa_\Psi^{\frac{1}{2}}\mu^M}{1-3\delta}\varepsilon_{i-1}$ `fi`
  $[P^a, f_{P^a}] :=$ **RHS**$[P^a,f,\delta\varepsilon_i]$
  $\bar{u}_{P^a} :=$ **GALSOLVE**$[P^a,f_{P^a},\bar{u}_{P^a},\delta\varepsilon_i]$
  `for` $j = 1,\ldots,M$ `do`
    $P :=$ **REFINE**$[P^a,f_{P^a},\bar{u}_{P^a}]$
    $[P^a,f_{P^a}] :=$ **RHS**$[P,f,\delta\mu^j\varepsilon_i]$
    $\bar{u}_{P^a} :=$ **GALSOLVE**$[P^a,f_{P^a},\bar{u}_{P^a},\delta\mu^j\varepsilon_i]$
  `od`
`od`

The next theorem shows that **SOLVE1** is an optimal method, whenever this is allowed by the $(f,\textbf{RHS})$ pair.

Theorem 7.4. $[P^a,\bar{u}_{P^a}] :=$ **SOLVE1**$[f,\varepsilon,\bar{u}_{P_0},\varepsilon_0]$ *satisfies* $|u-\bar{u}_{P^a}|_1 \leq \varepsilon$. *Assuming* $\varepsilon_0 \lesssim |u|_1$, *if for some* $s > 0$, $u \in \mathcal{A}^s$ *and* $(f,\textbf{RHS})$ *is* s*-optimal, then both* $\#P^a$ *and the number of arithmetic operations required by this call are* $\lesssim \max\{1,\varepsilon^{-1/s}(c_f^{1/s} + |u|_{\mathcal{A}^s}^{1/s})\}$.

Proof. For $\varepsilon \geq \varepsilon_0$ there is nothing to prove, so let us assume that $\varepsilon < \varepsilon_0$. By induction on i we prove that at termination of the if-then-else-fi clause inside the loop over i,

$$|u - \bar{u}_{P^a}|_1 \leq (1-3\delta)\varepsilon_i. \tag{7.1}$$

For $i = 1$, this follows from the input condition on ε_0. Let us now assume (7.1) for some $i \geq 1$. Then after the call $[P^a, f_{P^a}] := \mathbf{RHS}[P^a, f, \delta\varepsilon_i]$, by definition

$$\|f - f_{P^a}\|_{-1} + \operatorname{osc}(f_{P^a}, P^a) \leq \delta\varepsilon_i. \tag{7.2}$$

So for $u_{P^a} := L_{P^a}^{-1} f_{P^a}$ we have

$$\begin{aligned}
|u - u_{P^a}|_1 &\leq |u - L^{-1} f_{P^a}|_1 + |L^{-1} f_{P^a} - u_{P^a}|_1 \leq |u - L^{-1} f_{P^a}|_1 + |L^{-1} f_{P^a} - \bar{u}_{P^a}|_1 \\
&\leq 2|u - L^{-1} f_{P^a}|_1 + |u - \bar{u}_{P^a}|_1 = 2\|f - f_{P^a}\|_{-1} + |u - \bar{u}_{P^a}|_1 \\
&\leq 2\delta\varepsilon_i + (1-3\delta)\varepsilon_i = (1-\delta)\varepsilon_i,
\end{aligned} \tag{7.3}$$

where for the second inequality we have used that $\bar{u}_{P^a} \in \mathcal{S}_{P^a}$ and that u_{P^a} is the best approximation with respect to $|\cdot|_1$ of $L^{-1} f_{P^a}$ from $\mathcal{S}_{P^a}$. We conclude that after the update of $\bar{u}_{P^a}$ by the call of **GALSOLVE**,

$$|u_{P^a} - \bar{u}_{P^a}|_1 \leq \delta\varepsilon_i \text{ and so } |u - \bar{u}_{P^a}|_1 \leq |u - u_{P^a}|_1 + |u_{P^a} - \bar{u}_{P^a}|_1 \leq \varepsilon_i. \tag{7.4}$$

After the first calls of **REFINE**, **RHS** and **GALSOLVE** in the inner loop, i.e., when $j = 1$, for the new P^a, f_{P^a}, $\bar{u}_{P^a}$, $u_{P^a} := L_{P^a}^{-1} f_{P^a}$ we have $\|f - f_{P^a}\|_{-1} + \operatorname{osc}(f_{P^a}, P^a) \leq \delta\mu\varepsilon_i$ and $|u_{P^a} - \bar{u}_{P^a}|_1 \leq \delta\mu\varepsilon_i$, and so by (7.2), (7.4), Corollary 7.1 shows that $|u - \bar{u}_{P^a}|_1 \leq \mu\varepsilon_i$. Repeating this argument for $j = 2, \ldots, M$ shows that at termination of the inner loop over j, it holds that $|u - \bar{u}_{P^a}|_1 \leq \mu^M \varepsilon_i$.

In particular, when $i = N$, we have that $|u - \bar{u}_{P^a}|_1 \leq \mu^M \varepsilon_N = (\frac{\mu^M}{1-3\delta})^N ((1+\gamma)\kappa_{\bar{\Psi}}^{\frac{1}{2}})^{N-1} \varepsilon_0 \leq \varepsilon$ by definition of N.

Otherwise, if $i < N$, then in the next iteration, thus after increasing i by one, just before the call of **COARSE**, it holds that $|||u - \bar{u}_{P^a}|||_1 \leq \lambda_{\bar{\Psi}}^{-\frac{1}{2}} |u - \bar{u}_{P^a}|_1 \leq \lambda_{\bar{\Psi}}^{-\frac{1}{2}} \mu^M \varepsilon_{i-1}$. By Corollary 5.5, after this call we have

$$|u - \bar{u}_{P^a}|_1 \leq \Lambda_{\bar{\Psi}}^{\frac{1}{2}} |||u - \bar{u}_{P^a}|||_1 \leq \Lambda_{\bar{\Psi}}^{\frac{1}{2}} (1+\gamma) \lambda_{\bar{\Psi}}^{-\frac{1}{2}} \mu^M \varepsilon_{i-1} = (1-3\delta)\varepsilon_i, \tag{7.5}$$

which completes the proof of (7.1), and thus that of $|u - \bar{u}_{P^a}|_1 \leq \varepsilon$ at termination of **SOLVE1**.

Now we will prove that for any $i = 1, \ldots, N$, both $\#P^a$ at the end of the outer cycle for this i and the costs of this cycle excluding, for $i > 1$, the costs of the **COARSE**, but including, for $i < N$, the costs of the **COARSE** in the next cycle, are $\lesssim \varepsilon_i^{-1/s}(c_f^{1/s} + |u|_{\mathcal{A}^s}^{1/s})$. Because of $\sum_{i=1}^N \varepsilon_i^{-1/s} \lesssim \varepsilon_N^{-1/s} \leq \varepsilon^{-1/s}$ this will prove the statement about the complexity.

At the start of the outer cycle for $i = 1$, we have $\#P^a \lesssim \varepsilon_i^{-1/s} |u|_{\mathcal{A}^s}^{1/s}$, which follows from $\#P^a = \#P_0 \lesssim 1$ and the assumption that $\varepsilon_0 \lesssim |u|_1$. Since, as we have seen, for $i > 1$ before the call of **COARSE**, it holds that $|||u - \bar{u}_{P^a}|||_1 \leq \lambda_{\bar{\Psi}}^{-\frac{1}{2}} \mu^M \varepsilon_{i-1}$, Corollary 5.5 shows that after this call, $\#P^a \leq D\#\hat{P}$ for any partition $\hat{P}$ with $\inf_{u_{\hat{P}}\in\mathcal{S}_{\hat{P}}} |||u-u_{\hat{P}}|||_1 \leq (t_1^{\frac{1}{2}}\gamma-1)\lambda_{\breve{\Psi}}^{-\frac{1}{2}}\mu^M\varepsilon_{i-1}$. From $|||\cdot|||_1 \leq \lambda_{\breve{\Psi}}^{-\frac{1}{2}}|\cdot|_1$ and $u\in\mathcal{A}^s$, we find that

$$\#P^a \leq D(\#P_0 + [(t_1^{\frac{1}{2}}\gamma-1)\mu^M\varepsilon_{i-1}]^{-1/s}|u|_{\mathcal{A}^s}^{1/s}) \lesssim \varepsilon_i^{-1/s}|u|_{\mathcal{A}^s}^{1/s}.$$

Since $(f, \mathbf{RHS})$ is s-optimal and M is a fixed constant, from the properties of **RHS** and **REFINE** we conclude that for any i, at the end of the outer cycle

$$\#P^a \lesssim \varepsilon_i^{-1/s}(c_f^{1/s} + |u|_{\mathcal{A}^s}^{1/s}), \tag{7.6}$$

whereas the costs of all calls of **RHS** and **REFINE** inside this cycle are also $\lesssim \varepsilon_i^{-1/s}(c_f^{1/s} + |u|_{\mathcal{A}^s}^{1/s})$. Furthermore, for $i < N-1$, Theorem 5.4(b) shows that the costs of **COARSE** in the next iteration are $\lesssim \#P^a + \max\{0, \log(\varepsilon_i^{-1}|||\bar{u}_{P^a}|||_1)\}$. From $\log(\varepsilon_i^{-1}|||\bar{u}_{P^a}|||_1) \leq \varepsilon_i^{-1/s}|||\bar{u}_{P^a}|||_1^{1/s}$, $|||\bar{u}_{P^a}|||_1 \lesssim |\bar{u}_{P^a}|_1 \leq |u|_1 + \varepsilon_0 \lesssim |u|_1 \leq |u|_{\mathcal{A}^s}$ and (7.6), we conclude that also these costs are $\lesssim \varepsilon_i^{-1/s}(c_f^{1/s} + |u|_{\mathcal{A}^s}^{1/s})$.

What is left is to bound the costs of the applications of **GALSOLVE**. As we have seen, just before the call **GALSOLVE**$[P^a, f_{P^a}, \bar{u}_{P^a}, \delta\varepsilon_i]$ outside the inner loop over j it holds that $|u-\bar{u}_{P^a}|_1 \leq (1-3\delta)\varepsilon_i$, and with $u_{P^a} := L_{P^a}^{-1}f_{P^a}$, $|u-u_{P^a}|_1 \leq (1-\delta)\varepsilon_i$, and so $|u_{P^a}-\bar{u}_{P^a}|_1 \leq 2(1-2\delta)\varepsilon_i$. Since $\frac{2(1-2\delta)\varepsilon_i}{\delta\varepsilon_i}$ is a constant, we conclude that the costs of this call are $\lesssim \#P^a$.

Let us now consider a call **GALSOLVE**$[P^a, f_{P^a}, \bar{u}_{P^a}, \delta\mu^j\varepsilon_i]$ inside the loop over j. Just before this call it holds that $|u-\bar{u}_{P^a}|_1 \leq \mu^{j-1}\varepsilon_i$ and $\|f-f_{P^a}\|_{-1} \leq \delta\mu^j\varepsilon_i$. As in (7.3), for $u_{P^a} := L_{P^a}^{-1}f_{P^a}$ we have

$$|u-u_{P^a}|_1 \leq 2\|f-f_{P^a}\|_{-1} + |u-\bar{u}_{P^a}|_1 \leq (2\delta\mu^j + \mu^{j-1})\varepsilon_i,$$

and so $|u_{P^a}-\bar{u}_{P^a}|_1 \leq 2(\delta\mu^j+\mu^{j-1})\varepsilon_i$. Since $\frac{(2\delta\mu^j+\mu^{j-1})\varepsilon_i}{\delta\mu^j\varepsilon_i}$ is a constant, we conclude that the costs of this call are $\lesssim \#P^a$, with which the proof is completed. □

8. An optimal adaptive finite element method with a posteriori error control

As follows from the proof of Theorem 7.4, the approximations $\bar{u}_{P^a}$ on the sequence of partitions produced by **SOLVE1** converge with an asymptotic rate $\leq (\frac{(1+\gamma)\kappa_{\breve{\Psi}}^{\frac{1}{2}}\mu^M}{1-3\delta})^{1/(M+1)}$, that is close to μ when M is not too small. Because of the application of a coarsening, the asymptotic rate is generally even equal to the above number. Indeed, after the evaluation of $[P^a, \bar{u}_{P^a}] := \mathbf{COARSE}[P^a, `\bar{u}_{P^a}^{\mathrm{old}}\text{'}, \gamma\lambda_{\breve{\Psi}}^{-\frac{1}{2}}\mu^M\varepsilon_{i-1}]$, it holds that

$$|u-\bar{u}_{P^a}|_1 \geq \lambda_{\breve{\Psi}}^{\frac{1}{2}}|||u-\bar{u}_{P^a}|||_1 \geq \lambda_{\breve{\Psi}}^{\frac{1}{2}}\left||||\bar{u}_{P^a}-\bar{u}_{P^a}^{\mathrm{old}}|||_1 - |||u-\bar{u}_{P^a}^{\mathrm{old}}|||_1\right| \geq ((\gamma-\eta)-1)\mu^M\varepsilon_{i-1}$$

where η can be arbitrary small, so that this lower bound is only a constant factor smaller than the upper bound for $|u-\bar{u}_{P^a}|_1$ from (7.5). The value μ has to be supplied by the user. It should be large enough to ensure that indeed $\lambda_{\breve{\Psi}}^{-\frac{1}{2}}\mu^M\varepsilon_{i-1}$ is an upper bound for $|||u-\bar{u}_{P^a}^{old}|||_1$, so that the quasi-optimality of the partition after **COARSE** is guaranteed by Corollary 5.5. Yet, a save choice of μ will be the result of a worst case analysis, and so likely it will be unnecessarily close to 1, resulting in a quantitatively less attractive algorithm. All adaptive finite element or wavelet methods based on coarsening introduced so far share this drawback that a judicious choice of such a parameter μ has to be made.

In this final subsection, we develop a modified routine **SOLVE2** in which the tolerances used in the routines **COARSE**, **RHS** and **GALSOLVE** will depend on an a posteriori estimate of the error, instead of on an a priori one.

For $f \in L_2(\Omega)$, in Theorem 6.1 we showed that with $u = L^{-1}f$ and $u_{P^a} = L_{P^a}^{-1}f$, the error $|u - u_{P^a}|_1$ is less or equal to the a posteriori estimate $C_1\mathcal{E}(P^a, f, u_{P^a})$. Below in Proposition 8.1, we develop such an estimator under the relaxed assumptions that $f \in H^{-1}(\Omega)$, an approximation $f_{P^a} \in L_2(\Omega)$ has been used for setting up the discrete system, and that this system is solved inexactly. Necessarily, such an estimator involves (upper bounds for) $\|f - f_{P^a}\|_{-1}$ and the algebraic error in the inexact solution.

Proposition 8.1. *Let $C_8 := 1 + C_1C_3$, where $C_1, C_3 > 0$ are constants from Theorem 6.1 and Lemma 6.5. For $f \in H^{-1}(\Omega)$, an admissible partition P^a, $\bar{u}_{P^a} \in \mathcal{S}_{P^a}$, $f_{P^a} \in L_2(\Omega)$, with $u = L^{-1}f$, $u_{P^a} = L_{P^a}^{-1}f_{P^a}$ we have*

$$|u - \bar{u}_{P^a}|_1 \leq C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + \|f - f_{P^a}\|_{-1} + C_8|u_{P^a} - \bar{u}_{P^a}|_1.$$

Proof. With $\breve{u} := L^{-1}f_{P^a}$, the proof follows from Theorem 6.1 and Lemma 6.5 by

$$\begin{aligned}
|u - \bar{u}_{P^a}|_1 &\leq |u - \breve{u}|_1 + |\breve{u} - u_{P^a}|_1 + |u_{P^a} - \bar{u}_{P^a}|_1 \\
&\leq \|f - f_{P^a}\|_{-1} + C_1\mathcal{E}(P^a, f_{P^a}, u_{P^a}) + |u_{P^a} - \bar{u}_{P^a}|_1 \\
&\leq \|f - f_{P^a}\|_{-1} + C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + (1 + C_1C_3)|u_{P^a} - \bar{u}_{P^a}|_1.
\end{aligned}$$

□

Since constant multiples of the a posteriori error estimate from Proposition 8.1 will be used as tolerances in **COARSE**, **RHS**, and **GALSOLVE**, convergence of the adaptive method can only be shown when a repeated application of the triple **REFINE**, **RHS**, and **GALSOLVE** reduces this error estimate. A combination of both statements from the following corollary of Theorem 6.8 implies such a reduction. The key to this result is that, as basically has been shown in Corollary 6.4, the estimator is not only 'reliable' but also 'efficient'. Note that in contrast to the reduction of $|u - \bar{u}_{P^a}|_1$ shown in the earlier Corollary 7.1, the reduction of the a posteriori estimate will not necessarily be monotone.

Corollary 8.2. *For any $f \in H^{-1}(\Omega)$, an admissible partition P^a, $\bar{u}_{P^a} \in \mathcal{S}_{P^a}$, and $f_{P^a} \in L_2(\Omega)$, let ζ_{P^a} denote an upper bound for $\|f - f_{P^a}\|_{-1} + \operatorname{osc}(f_{P^a}, P^a) + C_8|u_{P^a} - \bar{u}_{P^a}|_1$, where $u_{P^a} = L_{P^a}^{-1}f_{P^a}$. Then for any absolute constant $C > 0$, we have*

$$C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + \zeta_{P^a} \eqsim |u - \bar{u}_{P^a}|_1 + C\zeta_{P^a}. \tag{8.1}$$

*Let $\tilde{P}^a$ be an admissible refinement of $\tilde{P} =$ **REFINE**$[P^a, f_{P^a}, \bar{u}_{P^a}]$, $\bar{u}_{\tilde{P}^a} \in \mathcal{S}_{\tilde{P}^a}$, $f_{\tilde{P}^a} \in L_2(\Omega)$. Then for any $\mu \in ([1 - \frac{1}{2}(\frac{c_2\theta}{C_1})^2]^{\frac{1}{2}}, 1)$ there exist a $\bar{\delta} > 0$ being small enough and a $C_9 > 0$ being large enough, such that if*

$$\zeta_{\tilde{P}^a} \leq \bar{\delta}(1 + C_8)[C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + \zeta_{P^a}],$$

where thus $\zeta_{\tilde{P}^a}$ *denotes an upperbound for* $\|f - f_{\tilde{P}^a}\|_{-1} + \operatorname{osc}(f_{\tilde{P}^a}, \tilde{P}^a) + C_8|u_{\tilde{P}^a} - \bar{u}_{\tilde{P}^a}|_1$ *where* $u_{\tilde{P}^a} = L_{\tilde{P}^a}^{-1} f_{\tilde{P}^a}$*, then*

$$|u - \bar{u}_{\tilde{P}^a}|_1 + C_9\zeta_{\tilde{P}^a} \leq \mu[|u - \bar{u}_{P^a}|_1 + C_9\zeta_{P^a}]. \tag{8.2}$$

Proof. Let $\hat{f}_{P^a} = Q_{P^a}^{(0)} f_{P^a}$, $\hat{u} = L^{-1}\hat{f}_{P^a}$, $\hat{u}_{P^a} = L_{P^a}^{-1}\hat{f}_{P^a}$. By Lemma 6.5, Corollary 6.4, and $\|f_{P^a} - \hat{f}_{P^a}\|_{-1} \leq C_4 \operatorname{osc}(f_{P^a}, P^a)$ by Lemma 6.6, we have

$$\begin{aligned} |\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) - \mathcal{E}(P^a, \hat{f}_{P^a}, \hat{u}_{P^a})| &\leq \|f_{P^a} - \hat{f}_{P^a}\|_{-1} + C_3|\bar{u}_{P^a} - \hat{u}_{P^a}|_1 \\ &\leq (1 + C_3)C_4 \operatorname{osc}(f_{P^a}, P^a) + C_3|\bar{u}_{P^a} - u_{P^a}|_1, \end{aligned}$$

$$\begin{aligned} c_2\mathcal{E}(P^a, \hat{f}_{P^a}, \hat{u}_{P^a}) &\leq |\hat{u} - \hat{u}_{P^a}|_1 \\ &\leq |\hat{u} - u|_1 + |u - \bar{u}_{P^a}|_1 + |\bar{u}_{P^a} - u_{P^a}|_1 + |u_{P^a} - \hat{u}_{P^a}|_1 \\ &\leq \|f - f_{P^a}\|_{-1} + 2C_4 \operatorname{osc}(f_{P^a}, P^a) + |u - \bar{u}_{P^a}|_1 + |\bar{u}_{P^a} - u_{P^a}|_1, \end{aligned}$$

which shows $C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) \lesssim |u - \bar{u}_{P^a}|_1 + \zeta_{P^a}$, and so together with Proposition 8.1 proves (8.1).

Theorem 6.8 shows that with $C_{10} := \max\{\frac{1+2c_2C_3}{C_8}, C_5, C_6, C_7\}$,

$$|u - \bar{u}_{\tilde{P}^a}|_1 + C_9\zeta_{\tilde{P}^a} \leq [1 - \tfrac{1}{2}(\tfrac{c_2\theta}{C_1})^2]^{\frac{1}{2}}|u - \bar{u}_{P^a}|_1 + C_{10}(\zeta_{P^a} + \zeta_{\tilde{P}^a}) + C_9\zeta_{\tilde{P}^a},$$

which, given a $\mu \in ([1 - \frac{1}{2}(\frac{c_2\theta}{C_1})^2]^{\frac{1}{2}}, 1)$, is less or equal to $\mu[|u - \bar{u}_{P^a}|_1 + C_9\zeta_{P^a}]$ if and only if

$$(C_{10} + C_9)\zeta_{\tilde{P}^a} \leq (\mu - [1 - \tfrac{1}{2}(\tfrac{c_2\theta}{C_1})^2]^{\frac{1}{2}})|u - \bar{u}_{P^a}|_1 + (\mu C_9 - C_{10})\zeta_{P^a}.$$

So by selecting the constant $C_9 > \frac{C_{10}}{\mu}$, the proof of (8.2) is completed by observing that for $\bar{\delta}$ small enough,

$$\bar{\delta}(1 + C_8)[C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + \zeta_{P^a}] \leq \frac{(\mu - [1 - \frac{1}{2}(\frac{c_2\theta}{C_1})^2]^{\frac{1}{2}})|u - \bar{u}_{P^a}|_1 + (\mu C_9 - C_{10})\zeta_{P^a}}{C_{10} + C_9},$$

which is a consequence of (8.1). □

We are ready to formulate the adaptive finite element method **SOLVE2** in which the tolerances are controlled by the a posteriori error estimator. A convergence of the approximations produced by the **REFINE**, **RHS**, **GALSOLVE** triple that is faster than it appears from a priori estimates can be expected to lead to better quantitative properties of **SOLVE2** compared to that of **SOLVE1**.

SOLVE2$[f, \varepsilon, \bar{u}_{P_0}, \varepsilon_0] \to [P^a, \bar{u}_{P^a}]$:
% *The following constants are fixed:* $\bar{\delta}$ *being small enough so that it correspond to a*
% $\mu < 1$ *as in Corollary 8.2;* $\gamma > t_1^{-\frac{1}{2}}$ *with* t_1 *as in Proposition 5.3;* $\xi > 0$*, e.g.,* $\xi = \bar{\delta}$*;*
% *and* $\sigma \in (0, 1)$*.*
% *The input must satisfy* $f \in H^{-1}(\Omega)$*,* $\varepsilon > 0$*,* $\bar{u}_{P_0} \in \mathcal{S}_{P_0}$ *and* $\varepsilon_0 \geq |u - \bar{u}_{P_0}|_1$*.*

$P^a := P_0$, $f_{P^a} := f_{P_0}$, $\bar{u}_{P^a} := \bar{u}_{P_0}$, $\bar{e} := \tilde{e} := \varepsilon_0$
while $\bar{e} > \varepsilon$ do
    if *not first iteration* then
        $[P^a, \bar{u}_{P^a}] := \mathbf{COARSE}[P^a, \bar{u}_{P^a}, \gamma\lambda_{\Psi}^{-\frac{1}{2}}\bar{e}]$, $\tilde{e} := (1+\gamma)\kappa_{\Psi}^{\frac{1}{2}}\bar{e}$
    fi
    $[P^a, f_{P^a}] := \mathbf{RHS}[P^a, f, \xi\tilde{e}]$
    $\bar{u}_{P^a} := \mathbf{GALSOLVE}[P^a, f_{P^a}, \bar{u}_{P^a}, \xi\tilde{e}]$
    $\tilde{e} := C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + (1+C_8)\xi\tilde{e}$
    while $\tilde{e} > \sigma\bar{e}$ do
        $P := \mathbf{REFINE}[P^a, f_{P^a}, \bar{u}_{P^a}]$
        $[P^a, f_{P^a}] := \mathbf{RHS}[P, f, \bar{\delta}\tilde{e}]$
        $\bar{u}_{P^a} := \mathbf{GALSOLVE}[P^a, f_{P^a}, \bar{u}_{P^a}, \bar{\delta}\tilde{e}]$
        $\tilde{e} := C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + (1+C_8)\bar{\delta}\tilde{e}$
    od
    $\bar{e} := \tilde{e}$
od

The next theorem shows that **SOLVE2** is an optimal method, whenever this is allowed by the $(f, \mathbf{RHS})$ pair.

Theorem 8.3. $[P^a, \bar{u}_{P^a}] := \mathbf{SOLVE1}[f, \varepsilon, \bar{u}_{P_0}, \varepsilon_0]$ *satisfies* $|u - \bar{u}_{P^a}|_1 \leq \varepsilon$. *Assuming* $\varepsilon_0 \gtrsim |u|_1$, *if for some* $s > 0$, $u \in \mathcal{A}^s$ *and* $(f, \mathbf{RHS})$ *is s-optimal, then both* $\#P^a$ *and the number of arithmetic operations required by this call are* $\lesssim \max\{1, \varepsilon^{-1/s}(c_f^{1/s} + |u|_{\mathcal{A}^s}^{1/s})\}$.

Proof. At the beginning of a cycle of the outer while-loop, it holds that $|u - \bar{u}_{P^a}|_1 \leq \bar{e}$, which for a cycle other than the first one is a consequence of Proposition 8.1. In particular, when the outer loop terminates, we have $|u - \bar{u}_{P^a}|_1 \leq \varepsilon$.

After the if-then-fi clause, it holds that $|u - \bar{u}_{P^a}|_1 \leq \tilde{e}$, where $\tilde{e} = \bar{e}$ for the first iteration, and $\tilde{e} = (1+\gamma)\kappa_{\Psi}^{\frac{1}{2}}\bar{e}$ otherwise (apply Corollary 5.5 and (4.6)). After the call of **RHS**, by definition we have $\|f - f_{P^a}\|_{-1} + \operatorname{osc}(f_{P^a}, P^a) \leq \xi\tilde{e}$, and so as in (7.3), for $u_{P^a} := L_{P^a}^{-1} f_{P^a}$ we have $|u - u_{P^a}|_1 \leq 2\|f - f_{P^a}\|_{-1} + |u - \bar{u}_{P^a}|_1 \leq (2\xi + 1)\tilde{e}$. After the call of **GALSOLVE**, by definition we have $|u_{P^a} - \bar{u}_{P^a}|_1 \leq \xi\tilde{e}$, and so $|u - \bar{u}_{P^a}|_1 \leq (3\xi + 1)\tilde{e}$. By applying (8.1), these estimates show that just before starting the inner while-loop, the new $\tilde{e}$ satisfies

$$\tilde{e} \leq C\bar{e}, \tag{8.3}$$

for some absolute constant $C > 0$.

Let us now consider any newly computed $\bar{u}_{P^a}$ in the inner while-loop, and let us denote with τ_1, τ_2 the tolerances that were used in the corresponding calls of **RHS** and **GALSOLVE** and let $\zeta_{P^a} = \tau_1 + C_8\tau_2$. It holds that ζ_{P^a} is equal to $(1+C_8)\bar{\delta}[C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + \zeta_{P^a}]$ where in the latter expression P^a, f_{P^a}, $\bar{u}_{P^a}$ and ζ_{P^a} refer to the previous partition, right-hand side, approximate solution and ζ_{P^a}. Since by assumption $\bar{\delta}$ corresponds to a $\mu < 1$ as in Corollary 8.2, formula (8.2) shows that in each iteration of the inner loop $|u - \bar{u}_{P^a}|_1 + C_9\zeta_{P^a}$ is multiplied with a factor $\leq \mu$. Since by (8.1), $C_1\mathcal{E}(P^a, f_{P^a}, \bar{u}_{P^a}) + \zeta_{P^a} \eqsim$

$|u-\bar{u}_{P^a}|_1+C_9\zeta_{P^a}$, the geometric decrease of $|u-\bar{u}_{P^a}|_1+C_9\zeta_{P^a}$ together with (8.3) shows that the inner while-loop terminates within an absolute constant number of iterations. After termination of the inner while-loop, the new $\bar{e}$ will be less or equal to $\sigma<1$ times the previous $\bar{e}$ showing that **SOLVE2** terminates, with, as we have seen, $|u-\bar{u}_{P^a}|_1\leq\varepsilon$.

Let us consider any cycle of the outer loop with $\bar{e}>\varepsilon$ being the value at the beginning of this cycle. After the if-then-fi clause, it holds that $\#P^a\lesssim\bar{e}^{-1/s}|u|_{\mathcal{A}^s}^{1/s}$, which for the first iteration follows from $\#P_0\lesssim 1$ and $\varepsilon_0\lesssim|u|_1$ by assumption, and which for any other cycle follows from Corollary 5.5 analogously as in the proof of Theorem 7.4. The properties of **RHS** and **REFINE** and the fact that the inner while-loop terminates within a fixed number of iterations show that at termination of this outer cycle, $\#P^a\lesssim\bar{e}^{-1/s}(c_f^{1/s}+|u|_{\mathcal{A}^s}^{1/s})$, which in particular proves the statement about $\#P^a$ at termination of **SOLVE1**. Furthermore, the costs of all calls of **RHS** and **REFINE** as well as the costs of **COARSE** in the possibly next cycle are $\lesssim\bar{e}^{-1/s}(c_f^{1/s}+|u|_{\mathcal{A}^s}^{1/s})$. Assuming for a while that the costs of any application of **GALSOLVE** in **SOLVE2** on a partition P^a are $\lesssim\#P^a$, by the geometric decrease of the values of $\bar{e}$ at the beginning of the outer while-loop we conclude that the total costs of **SOLVE2** are $\lesssim\varepsilon^{-1/s}(c_f^{1/s}+|u|_{\mathcal{A}^s}^{1/s})$.

As we have seen, just before the evaluation of **GALSOLVE**$[P^a,f_{P^a},\bar{u}_{P^a},\xi\tilde{e}]$ outside the inner while-loop, we have $|u-\bar{u}_{P^a}|_1\leq\tilde{e}$, and with $u_{P^a}:=L_{P^a}^{-1}f_{P^a}$, $|u-u_{P^a}|_1\leq(2\xi+1)\tilde{e}$ and so $|u_{P^a}-\bar{u}_{P^a}|_1\leq 2(\xi+1)\tilde{e}$. Since $\xi>0$ is some fixed constant, we conclude that the costs of this call are $\lesssim\#P^a$.

Analogously, just before an evaluation of **GALSOLVE**$[P^a,f_{P^a},\bar{u}_{P^a},\bar{\delta}\tilde{e}]$ inside the inner while-loop, we have $|u-\bar{u}_{P^a}|_1\leq\tilde{e}$, $\|f-f_{P^a}\|_{-1}\leq\bar{\delta}\tilde{e}$, and so, as in (7.3), with $u_{P^a}:=L_{P^a}^{-1}f_{P^a}$, $|u-u_{P^a}|_1\leq 2\bar{\delta}\tilde{e}+\tilde{e}$ and so $|u_{P^a}-\bar{u}_{P^a}|_1\leq 2(\bar{\delta}\tilde{e}+\tilde{e})$. Since $\bar{\delta}>0$ is a fixed constant, we conclude that also the costs of such a call are $\lesssim\#P^a$.

□

Acknowledgments

The author would like to thank Peter Binev (Univ. of South Carolina) for his explanations concerning adaptive tree approximations.

References

[Bar03] A. Barinka. *Fast Evaluation Tools for Adaptive Wavelet Schemes.* PhD thesis, RTWH Aachen, 2003. To appear.

[BD02] P. Binev and R. DeVore. Fast computation in adaptive tree approximation. IMI Research Reports 02:11, Univ. of South Carolina, 2002.

[BDD02] P. Binev, W. Dahmen, and R. DeVore. Adaptive finite element methods with convergence rates. IGPM report, RWTH Aachen, June 2002.

[BDDP02] P. Binev, W. Dahmen, R. DeVore, and P. Petruchev. Approximation classes for adaptive methods. *Serdica Math. J.*, 28:1001–1026, 2002.

[BM87] I. Babuška and A. Miller. A feedback finite element method with a posteriori error estimation. I. The finite element method and some basic properties of the a posteriori error estimator. *Comput. Methods Appl. Mech. Engrg.*, 61(1):1–40, 1987.

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

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

[CS93] A. Cohen and J.M. Schlenker. Compactly supported wavelets with hexagonal symmetry. *Constr. Approx.*, 9:209–236, 1993.

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

[Dör96] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33:1106–1124, 1996.

[DSW96] M. Drya, M. V. Sarkis, and O. B. Widlund. Multilevel Schwarz methods for elliptic problems with discontinuous coefficients in three dimensions. *Numer. Math.*, 72(3):313–348, 1996.

[LO96] R. Lorentz and P. Oswald. Multilevel finite element Riesz bases in Sobolev spaces. In P. Bjørstad, M. Espedal, and Keyes D., editors, *Proc. 9th. Symp. on Domain Decomposition Methods*. John Wiley & Sons Ltd., 1996.

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

[MNS00] P. Morin, R. Nochetto, and K Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488, 2000.

[Osw94] P. Oswald. *Multilevel finite element approximation: Theory and applications*. B.G. Teubner, Stuttgart, 1994.

[Ste98a] R.P. Stevenson. Piecewise linear (pre-)wavelets on non-uniform meshes. In Hackbusch W. and G. Wittum, editors, *Multigrid Methods V,* Proceedings of the Fifth European Multigrid Conference held in Stuttgart, Germany, October 1-4, 1996, pages 306–319, Heidelberg, 1998. Lecture Notes in Computational Science and Engineering, Volume 3, Springer-Verlag.

[Ste98b] R.P. Stevenson. Stable three-point wavelet bases on general meshes. *Numer. Math.*, 80:131–158, 1998.

[Ste02] R.P. Stevenson. Adaptive solution of operator equations using wavelet frames. Technical Report 1239, University of Utrecht, May 2002. To appear in *SIAM J. Numer. Anal.*

[Ver96] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley-Teubner, Chichester, 1996.

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

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