INSTANCE OPTIMALITY OF THE ADAPTIVE MAXIMUM STRATEGY

LARS DIENING, CHRISTIAN KREUZER, AND ROB STEVENSON

Abstract. In this paper, we prove that the standard adaptive finite element method with a (modified) *maximum marking strategy* is *instance optimal* for the *total error*, being the square root of the squared energy error plus the squared oscillation. This result will be derived in the model setting of Poisson's equation on a polygon, linear finite elements, and conforming triangulations created by newest vertex bisection.

1. Introduction

Adaptive algorithms for the solution of PDEs that have been proposed since the 70's are nowadays standard tools in science and engineering. In contrast to uniform refinements, adaptive mesh modifications do not guarantee that the maximal mesh size tends to zero. For this reason, even convergence of adaptive finite element methods (AFEM's) was unclear for a long time, though practical experiences often showed optimal convergence rates.

In one dimension, convergence of an AFEM for elliptic problems was proved by Babuška and Vogelius in [BV84] under some heuristic assumptions. Later, Dörfler introduced in [Dör96] a bulk chasing marking strategy thereby proving linear convergence of an AFEM in two space dimensions for a sufficiently fine initial triangulation. This restriction was removed in [MNS00, MNS02] by Morin, Nochetto, and Siebert.

In [BDD04], Binev, Dahmen and DeVore extended the AFEM analysed in [MNS00] by a so-called *coarsening* routine, and showed that the resulting method is *instance optimal*, cf. also [Bin07]. This means that the energy norm of the error in any approximation produced by the algorithm, with underlying triangulation denoted as $\mathcal{T}$, is less than some constant multiple of the error w.r.t. any *admissible* triangulation $\tilde{\mathcal{T}}$ satisfying $\#(\tilde{\mathcal{T}} \setminus \mathcal{T}_\perp) \leq \lambda \#(\mathcal{T} \setminus \mathcal{T}_\perp)$, for some fixed constant $\lambda \in (0, 1)$. Here, an admissible triangulation is a conforming triangulation, which

Date: June 24, 2014.

is created by finitely many newest vertex bisections (NVB) from a fixed initial triangulation, which we denote as $\mathcal{T}_\perp$.

In [Ste07], it was shown that already without the addition of coarsening, the AFEM is *class optimal*: Whenever the solution can be approximated at some asymptotic (algebraic) convergence rate s by finite element approximations, and the right-hand side can be approximated by piecewise polynomials at rate s, then the AFEM produces a sequence of approximations, which converges with precisely this rate s. In [CKNS08], a similar result was presented with a refinement routine that is not required to produce "interior nodes", and with a different treatment of the approximation of the right-hand side, that is assumed to be in L_2. In that paper, the AFEM is considered as a procedure for reducing the *total error*, being the square root of the squared error in the energy norm plus the squared so-called oscillation. This is also the point of view that will be taken in the present work.

In the last few years, in numerous works class optimality results for AFEMs have been derived for arbitrary space dimensions, finite elements of arbitrary orders, the error measured in L^2, right-hand sides in H^{-1}, nonconforming triangulations, discontinuous Galerkin methods, general diffusion tensors, (mildly) non-symmetric problems, nonlinear diffusion equations, and indefinite problems.

In all these works the marking strategy is *bulk chasing*, also called *Dörfler marking*. In [MSV08], Morin, Siebert and Veeser considered also the maximum and equidistribution strategies, without proving any rates though.

In the present work, we consider a standard AFEM, so without coarsening, in the model setting of Poisson's equations with homogeneous Dirichlet boundary conditions on a two-dimensional polygonal domain, the error measured in the energy norm, square integrable right-hand side, linear finite elements, and conforming triangulations created by NVB. The refinement routine is not required to create interior nodes in refined triangles. Our method utilizes a (modified) *maximum marking strategy* for the standard residual error estimator organised by edges.

The maximum strategy marks all edges for bisection whose indicator is greater or equal to a constant $\sqrt{\mu} \in (0, 1]$ times the largest indicator. This strategy is usually preferred by practitioners since, other than with Dörfler marking, it does not require the sorting of the error indicators, and in practise the results turn out to be very insensible to the choice of the marking parameter $\mu \in (0, 1]$.

Roughly speaking, our modification of the maximum marking strategy replaces the role of the error indicator associated with an edge S by the square root of the sum of the squared error indicators over those edges that necessarily have to be bisected together with S in order to retain a conforming triangulation. The precise AFEM is stated in Section 5.

The main result of this paper states that for any $\mu \in (0,1]$, for some constants $C, \tilde{C} \geq 1$ it holds that

$$|u - u_{\mathcal{T}_k}|^2_{H^1(\Omega)} + \operatorname{osc}^2_{\mathcal{T}_k}(\mathcal{T}_k) \leq \tilde{C}\left(|u - u_{\mathcal{T}}|^2_{H^1(\Omega)} + \operatorname{osc}^2_{\mathcal{T}}(\mathcal{T})\right),$$

for all admissible $\mathcal{T}$ with $\#(\mathcal{T} \setminus \mathcal{T}_\perp) \leq \frac{\#(\mathcal{T}_k \setminus \mathcal{T}_\perp)}{C}$. Here $u_{\mathcal{T}}$ is the Galerkin approximation to the exact solution u from the finite element space corresponding to $\mathcal{T}$, $\operatorname{osc}^2_{\mathcal{T}}(\mathcal{T}) := \sum_{T \in \mathcal{T}} |T| \|f - f_T\|^2_{L^2(T)}$, where $f_T := \frac{1}{|T|}\int_T f\,dx$, and $\mathcal{T}_k$ is the triangulation produced in the kth iteration of our AFEM. This result means that our AFEM is *instance optimal* for the *total error*. Clearly, instance optimality implies class optimality for any (algebraic) rate s, but not vice versa.

Our AFEM is driven by the usual residual based a posteriori estimator, that is only equivalent to the *total* error. Consequently, we do not obtain instance optimality for the plain energy error, so without the oscillation term. The oscillation encodes approximability of the right hand side and is in most cases of higher order (e.g. as when $f \in H^s$ for some $s > 0$), and thus asymptotically neglectable.

To prove instance optimality, we will show that the *total energy* associated with any triangulation $\mathcal{T}$ produced by our AFEM is not larger than the total energy associated with any conforming triangulation $\tilde{\mathcal{T}}$ created by NVB with $\#(\tilde{\mathcal{T}} \setminus \mathcal{T}_\perp) \leq \lambda \#(\mathcal{T} \setminus \mathcal{T}_\perp)$, for some fixed constant $\lambda \in (0,1)$. Here the total energy is defined as the Dirichlet energy plus the "squared element residual part of the a posteriori estimator".

The outline of this paper is as follows: Sect. 2 is devoted to the newest vertex bisection refinement procedure. On the set of vertices of the triangulations that can be created by NVB from $\mathcal{T}_\perp$, we introduce a tree structure where nodes generally have multiple parents. Because of the resemblance of this tree structure with that of a family tree, we refer to such a tree as a population. The concept of population is the key for the derivation of some interesting new properties of NVB.

In Sect. 3, we show that the squared norm of the difference of Galerkin solutions on nested triangulations is equivalent to the sum of squared norms of the differences of the Galerkin solution on the fine triangulation and that on some intermediate triangulations. We call this the *lower diamond estimate.*

Sect. 4 is devoted to a posteriori error bounds. It is shown that the difference of total energies associated with a triangulation $\mathcal{T}_*$ and a coarser triangulation $\mathcal{T}$ is equivalent to the sum of the squared error indicators over exactly those edges in $\mathcal{T}$ that are refined in $\mathcal{T}_*$.

Based on the presented refinement framework and error estimator, we precisely specify our AFEM in Section 5.

In Sect. 6, we investigate some fine properties of populations, and thus of conforming triangulations created by NVB. Calling the vertices in such a triangulation "free" when they can be removed while retaining a conforming triangulation, the most striking property says that the number of free nodes cannot be reduced by more than a constant factor in any further conforming NVB refinement.

Finally, in Sect. 7 we combine these tools to prove instance optimality of our AFEM.

Throughout this paper we use the notation $a \lesssim b$ to indicate $a \leq C\, b$, with a generic constant C only possibly depending on fixed quantities like the initial triangulation $\mathcal{T}_\perp$, which will be introduced in the next subsection. Obviously, $a \gtrsim b$ means $b \lesssim a$, and we denote $a \lesssim b \lesssim a$ by $a \eqsim b$.

2. Newest vertex bisection

We recall properties of the newest vertex bisection (NVB) algorithm for creating locally refined triangulations. Moreover, we introduce new concepts related to conforming NVB that allow us to derive some new interesting properties.

2.1. **Triangulations and binary trees.**

We denote by $\mathcal{T}_\perp$ a conforming initial or "bottom" triangulation of a polygonal domain $\Omega \subset \mathbb{R}^2$. We restrict ourselves to mesh adaptation by newest vertex bisection in $2d$; compare with [Bän91, Kos94, Mau95, Tra97, BDD04, Ste08] as well as [NSV09, SS05] and the references therein.

To be more precise, for each $T \in \mathcal{T}_\perp$, we label one of its vertices as its *newest vertex*. Newest vertex bisection splits T into two sub-triangles by connecting the newest vertex to the midpoint of the opposite edge of T, called the *refinement edge* of T. This midpoint is labelled as the newest vertex of both newly created triangles, called *children* of T. A recursive application of this rule uniquely determines all possible NVB refinements of $\mathcal{T}_\perp$.

The triangles of any triangulation of Ω that can be created in this way are the leaves of a subtree of an infinite binary tree $\mathfrak{T}$ of triangles, having as roots the triangles of $\mathcal{T}_\perp$. The newest vertex of any $T \in \mathfrak{T}$ is determined by the labelling of newest vertices in $\mathcal{T}_\perp$. We define the generation $\mathtt{gen}(T)$ of $T \in \mathfrak{T}$ as the number of bisections that are needed to create T starting from $\mathcal{T}_\perp$. In particular, $\texttt{gen}(T) = 0$ for $T \in \mathcal{T}_\perp$. We have *uniform shape regularity* of $\mathfrak{T}$ in the sense that

$$\sup_{T\in\mathfrak{T}} \operatorname{diam}(T)/|T|^{\frac{1}{2}} < \infty,$$

only dependent on $\mathcal{T}_\perp$. We denote by $\mathcal{N}(T)$ the set of *nodes* or vertices of $T \in \mathfrak{T}$.

Among all triangulations that can be created by newest vertex bisection from $\mathcal{T}_\perp$, we are interested in those that are *conforming* and denote the set of these triangulations as $\mathbb{T}$. Note that $\mathcal{T}_\perp \in \mathbb{T}$ by assumption.

In the following we shall *always* assume that in $\mathcal{T}_\perp$ the labelling of the newest vertices is such that $\mathcal{T} = \mathcal{T}_\perp$ satisfies the *matching condition*:

(2.1) *If, for $T, T' \in \mathcal{T}$, $T \cap T'$ is the refinement edge of T, then it is the refinement edge of T'.*

It is shown in [BDD04], that such a labelling can be found for any conforming $\mathcal{T}_\perp$.

By induction one shows that for any $k \in \mathbb{N}_0$, the uniform refinement of the initial triangulation $\{T \in \mathfrak{T} \colon \texttt{gen}(T) = k\}$ is in $\mathbb{T}$, and satisfies the matching condition. Moreover, the following result is valid:

Proposition 2.1 ([Ste08, Corollary 4.6])**.** *Let $\mathcal{T} \in \mathbb{T}$ and $T, T' \in \mathcal{T}$ be such that $S = T \cap T'$ is the refinement edge of T. Then,*

- *either* $\texttt{gen}(T') = \texttt{gen}(T)$ *and S is the refinement edge of T', or*
- $\texttt{gen}(T') = \texttt{gen}(T) - 1$ *and S is the refinement edge of one of the two children of T'.*

We denote by $\mathcal{S}(\mathcal{T})$ ($\mathcal{S}_0(\mathcal{T})$) the set of (interior) *sides* or edges, and by $\mathcal{N}(\mathcal{T})$ ($\mathcal{N}_0(\mathcal{T})$) the set of (interior) *nodes* or vertices of a triangulation $\mathcal{T} \in \mathbb{T}$.

Finally, we note that if, for $\mathcal{T} \in \mathbb{T}$, we replace each $T \in \mathcal{T}$ by its grandchildren, i.e., the children of its children, then we obtain a conforming triangulation, that will be denoted as $\mathcal{T}^{++}$; compare with Figure 1.

2.2. **Populations.** A triangulation $\mathcal{T} \in \mathbb{T}$ can alternatively be described in terms of *populations*, which we shall introduce now. To this end, we denote the elements of

$$\mathcal{P}^\top := \bigcup_{T\in\mathfrak{T}} \mathcal{N}(T),$$

i.e., the union of the vertices of all $T \in \mathfrak{T}$, as *persons*.

We call a collection of persons a *population* when it is equal to $\mathcal{N}(\mathcal{T})$ for some $\mathcal{T} \in \mathbb{T}$, and denote with $\mathbb{P}$ the collection of all populations.

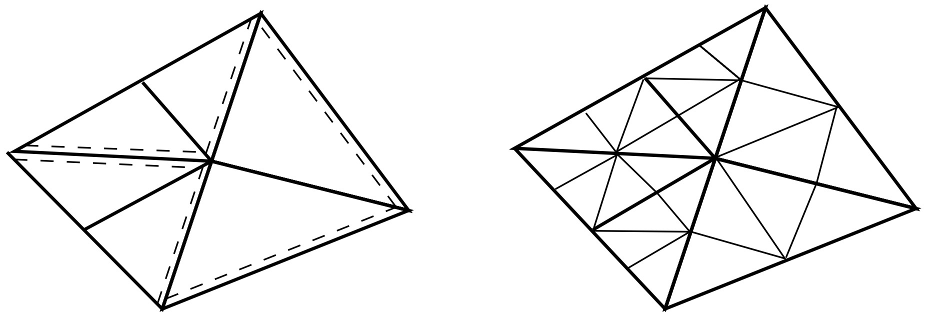

FIGURE 1. $\mathcal{T}$ with dashed refinement edges and the resulting $\mathcal{T}^{++}$.

Since a triangulation $\mathcal{T} \in \mathbb{T}$ is uniquely defined by its nodes $\mathcal{N}(\mathcal{T})$, we have a one-to-one correspondence between populations $\mathcal{P} \in \mathbb{P}$ and triangulations $\mathcal{T} \in \mathbb{T}$. When $\mathcal{P} = \mathcal{N}(\mathcal{T})$, we write $\mathcal{P} = \mathcal{P}(\mathcal{T})$ respectively $\mathcal{T} = \mathcal{T}(\mathcal{P})$ and set

$$\mathcal{P}_\perp := \mathcal{P}(\mathcal{T}_\perp)$$

for the initial or bottom population.

The set $\mathcal{P}^\top$ can be equipped with a family tree structure. Let $P \in \mathcal{P}^\top \setminus \mathcal{P}_\perp$, then there exists a $T \in \mathfrak{T}$ such that P is the midpoint of the refinement edge of T. We call the newest vertex of T a parent of P, respectively P its child. If $P \in \partial\Omega$ then P has one parent. Otherwise, when $P \in \Omega$, it has two parents. The *generation* of P is defined by $\texttt{gen}(P) = \texttt{gen}(T) + 1$. Since any uniform refinement of the initial partition is conforming and satisfies the matching condition, this definition is unique. Indeed, if P is the midpoint of a refinement edge of another element in $\tilde{T} \in \mathfrak{T}$, then $\texttt{gen}(T) = \texttt{gen}(\tilde{T})$. Defining $\texttt{gen}(P) = 0$ when $P \in \mathcal{P}_\perp$, we infer that the generation of a child is one plus the generation of its parent(s), which in particular are of equal generation.

Since an equivalent definition of $\texttt{gen}(P)$ is given by $\min\{\texttt{gen}(T) : T \in \mathfrak{T}, P \in \mathcal{N}(T)\}$, no two vertices of a $T \in \mathfrak{T}$ can have the same generation, unless they have generation zero.

Thanks to the uniform shape regularity of $\mathfrak{T}$, the number of children a single person can have is uniformly bounded. It is easy to see, that a person $P \in \mathcal{P}^\top \setminus \mathcal{P}_\perp$ has either two (when it is on the boundary) or four children in $\mathcal{P}^\top$; cf. Figure 2. For $P \in \mathcal{P}^\top$, we denote by $\texttt{child}(P)$ the set of the children of P, and by $\texttt{parent}(P)$ the set of its parents.

Any $\mathcal{T} \in \mathbb{T}$ is obtained from $\mathcal{T}_\perp$ by a sequence of simultaneous bisections of pairs of triangles that share their refinement edge (or by individual bisections of triangles that have their refinement edge on

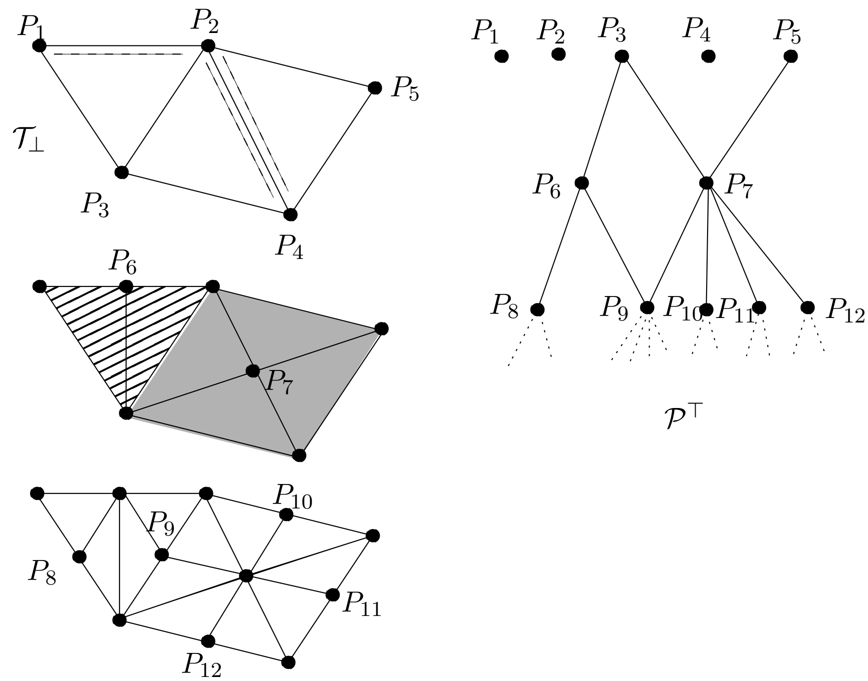

FIGURE 2. Parents–children relations, and $\Omega(P_6)$ and $\Omega(P_7)$.

the boundary). Each of such (simultaneous) bisections corresponds to the addition of a person to the population whose both its parents (or its single parent when the person is on the boundary) are already in the population. We conclude the following result.

Proposition 2.2. *A collection $\mathcal{U} \subset \mathcal{P}^\top$ is a population if and only if $\mathcal{P}_\perp \subset \mathcal{U}$ and, for each $P \in \mathcal{U}$, we have that* all *parents of P are contained in $\mathcal{U}$.*

This intrinsic characterization of a population as a family tree will be the key to derive many interesting properties of populations, and so of triangulations in $\mathbb{T}$.

As we have seen above, a person $P \in \mathcal{P}^\top \setminus \mathcal{P}_\perp$ is the (newest) vertex of four, or, when $P \in \partial\Omega$, two triangles from $\mathfrak{T}$, each of them having the same generation as P. For $P \in \mathcal{P}^\top$, we set

$$\Omega(P) := \text{interior} \bigcup \big\{T \in \mathfrak{T} \colon P \in T \text{ and } \mathtt{gen}(T) = \mathtt{gen}(P)\big\},$$

cf. Figure 2. This definition extends to subsets $\mathcal{U} \subset \mathcal{P}^\top$ setting

$$\Omega(\mathcal{U}) := \text{interior} \bigcup_{P \in \mathcal{U}} \overline{\Omega(P)}. \tag{2.2}$$

One easily verifies the following result:

Proposition 2.3.

(a) *Let* $P_1, P_2 \in \mathcal{P}^\top \setminus \mathcal{P}_\perp$ *with* $P_1 \neq P_2$ *and* $\mathtt{gen}(P_1) = \mathtt{gen}(P_2)$. *Then* $\Omega(P_1) \cap \Omega(P_2) = \emptyset$.

(b) *Let* $P \in \mathcal{P}^\top \setminus \mathcal{P}_\perp$. *Then* $\Omega(P) \subset \Omega(\mathtt{parent}(P))$.

2.3. **Refinements and coarsenings.** For $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$ we write $\mathcal{T} \leq \mathcal{T}_*$ or $\mathcal{T}_* \geq \mathcal{T}$, when $\mathcal{T}_*$ is a *refinement* of $\mathcal{T}$ or, equivalently, $\mathcal{T}$ is a *coarsening* of $\mathcal{T}_*$, i.e., when the tree of $\mathcal{T}$ is a subtree of that of $\mathcal{T}_*$. This defines a partial ordering on $\mathbb{T}$. On $\mathbb{P}$, we define a *partial ordering* by $\mathcal{P} \leq \mathcal{P}_*$ when $\mathcal{P} \subset \mathcal{P}_*$. We call $\mathcal{P}_*$ a *refinement* of $\mathcal{P}$ or, equivalently, $\mathcal{P}$ a *coarsening* of $\mathcal{P}_*$. These orderings are equivalent:

Proposition 2.4. *For* $\mathcal{P}, \mathcal{P}_* \in \mathbb{P}$, *we have*

$$\mathcal{P} \leq \mathcal{P}_* \quad \Longleftrightarrow \quad \mathcal{T}(\mathcal{P}) \leq \mathcal{T}(\mathcal{P}_*).$$

The partially ordered set $(\mathbb{P}, \leq)$ is a *lattice*, since for any $\mathcal{P}_1, \mathcal{P}_2 \in \mathbb{P}$, the *lowest upper bound* $\mathcal{P}_1 \vee \mathcal{P}_2$ and the *greatest lower bound* $\mathcal{P}_1 \wedge \mathcal{P}_2$ exist in $\mathbb{P}$, and are given by

$$\mathcal{P}_1 \vee \mathcal{P}_2 = \mathcal{P}_1 \cup \mathcal{P}_2 \quad \text{and} \quad \mathcal{P}_1 \wedge \mathcal{P}_2 = \mathcal{P}_1 \cap \mathcal{P}_2, \tag{2.3}$$

respectively. We call $\mathcal{P}_1 \wedge \mathcal{P}_2$ the *largest common coarsening*, and $\mathcal{P}_1 \vee \mathcal{P}_2$ the *smallest common refinement* of $\mathcal{P}_1$ and $\mathcal{P}_2$.

Since $\mathcal{P}_\perp \leq \mathcal{P}$ for all $\mathcal{P} \in \mathbb{P}$, we have that $\mathcal{P}_\perp$ is the *bottom* of $(\mathbb{P}, \leq)$. Moreover, if we define $\widehat{\mathbb{P}} := \mathbb{P} \cup \{\mathcal{P}^\top\}$ and set $\mathcal{P}^\top \geq \mathcal{P}$ for all $\mathcal{P} \in \widehat{\mathbb{P}}$, then $\mathcal{P}^\top$ is the *top* of $\widehat{\mathbb{P}}$ and whence $\widehat{\mathbb{P}}$ is a bounded lattice.

These notions can be transferred to triangulations $\mathcal{T}_1, \mathcal{T}_2 \in \mathbb{T}$ via

$$\begin{aligned} \mathcal{T}_1 \vee \mathcal{T}_2 &:= \mathcal{T}\big(\mathcal{P}(\mathcal{T}_1) \vee \mathcal{P}(\mathcal{T}_2)\big), \\ \mathcal{T}_1 \wedge \mathcal{T}_2 &:= \mathcal{T}\big(\mathcal{P}(\mathcal{T}_1) \wedge \mathcal{P}(\mathcal{T}_2)\big). \end{aligned}$$

Consequently, $(\mathbb{T}, \leq)$ is a lattice with bottom $\mathcal{T}_\perp$. Moreover, we can add a largest element $\mathcal{T}^\top = \mathcal{T}(\mathcal{P}^\top)$ to $\mathbb{T}$ and define $\widehat{\mathbb{T}} := \mathbb{T} \cup \{\mathcal{T}^\top\}$ and $\mathcal{T}^\top \geq \mathcal{T}$ for all $\mathcal{T} \in \mathbb{T}$. Then $\mathcal{T}^\top$ is the top of the bounded lattice $\widehat{\mathbb{T}}$.

Remark 2.5. *An interpretation of* $\mathcal{T}_1 \vee \mathcal{T}_2$ *and* $\mathcal{T}_1 \wedge \mathcal{T}_2$ *is given in the following (cf.* [NSV09, Lemma 4.3]*). For* $\mathcal{T}_1, \mathcal{T}_2 \in \mathbb{T}$, *let* $T_1 \in \mathcal{T}_1$, $T_2 \in \mathcal{T}_2$ *with* $|T_1 \cap T_2| > 0$, *so that either* $T_1 \subset T_2$ *or* $T_2 \subset T_1$. *W.l.o.g., we assume* $T_1 \subset T_2$. *Then* $T_1 \in \mathcal{T}_1 \vee \mathcal{T}_2$ *and* $T_2 \in \mathcal{T}_1 \wedge \mathcal{T}_2$.

For $\mathcal{T} \in \mathbb{T}$ and $\mathcal{U} \subset \mathcal{T}$, we define

$$\Omega(\mathcal{U}) := \text{interior} \bigcup \{T \,:\, T \in \mathcal{U}\}.$$

For $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$ with $\mathcal{T} \leq \mathcal{T}_*$, we call $\Omega(\mathcal{T} \setminus \mathcal{T}_*) = \Omega(\mathcal{T}_* \setminus \mathcal{T})$ the *area of coarsening*. It is the union of all triangles that are coarsened when

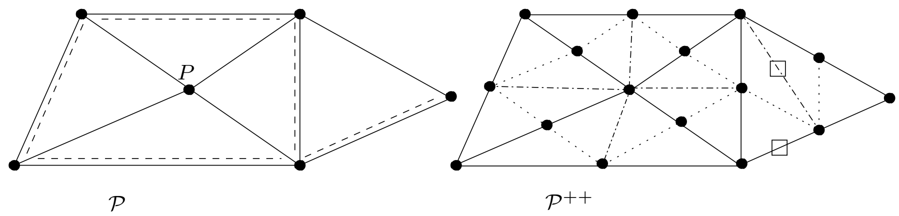

FIGURE 3. $\mathcal{P}$, $\mathcal{P}^{++}$, and grandchildren ($\square$) of $P \in \mathcal{P}$ that are not in $\mathcal{P}^{++}$

passing from $\mathcal{T}_*$ to $\mathcal{T}$, or, equivalently, the union all triangles that are refined when passing from $\mathcal{T}$ to $\mathcal{T}_*$. The coarsening point of view, however, will often turn out to be more relevant, in particular in Sect. 3.

Recalling the definition $\mathcal{T}^{++}$ for $\mathcal{T} \in \mathbb{T}$, we set $\mathcal{P}^{++} := \mathcal{P}((\mathcal{T}(\mathcal{P})^{++})$. Then $\mathcal{P}^{++} \setminus \mathcal{P} \subset \big(\cup_{P\in\mathcal{P}} \texttt{child}(P) \cup \texttt{child}(\texttt{child}(P))\big) \setminus \mathcal{P}$, with equality only when all $T \in \mathcal{T}(\mathcal{P})$ have the same generation, cf. Figure 3. There is a one-to-one correspondence of $\mathcal{S}(\mathcal{T}(\mathcal{P}))$ and $\mathcal{P}^{++} \setminus \mathcal{P}$. Indeed, denote the midpoint of a side $S \in \mathcal{S}(\mathcal{T})$ by $\mathrm{midpt}(S)$ and set $\mathrm{midpt}(\mathcal{S}) := \{\mathrm{midpt}(S) : S \in \mathcal{S}\}$ for a collection $\mathcal{S}$ of sides, then

$$\mathcal{P}^{++} \setminus \mathcal{P} = \mathrm{midpt}(\mathcal{S}(\mathcal{T}(\mathcal{P}))). \tag{2.4}$$

More general, if $\mathcal{P}, \mathcal{P}_* \in \mathbb{P}$ with $\mathcal{P} \le \mathcal{P}_*$, then

$$\mathcal{P}_* \cap (\mathcal{P}^{++} \setminus \mathcal{P}) = \mathrm{midpt}(\mathcal{S}(\mathcal{T}(\mathcal{P})) \setminus \mathcal{S}(\mathcal{T}(\mathcal{P}_*))). \tag{2.5}$$

For $\mathcal{T} \in \mathbb{T}$, we define

$$\begin{aligned} \mathbb{V}_0(\mathcal{T}) &:= \{v \in H^1_0(\Omega) : v|_T \in P_1(T)\,(T \in \mathcal{T})\}, \\ \mathbb{V}(\mathcal{T}) &:= \{v \in H^1(\Omega) : v|_T \in P_1(T)\,(T \in \mathcal{T})\}, \end{aligned}$$

Thanks to the nodal Lagrange basis representation of any finite element function, the degrees of freedom (DOFs) of $\mathbb{V}_0(\mathcal{T})$ or $\mathbb{V}(\mathcal{T})$ can be identified with $\mathcal{N}_0(\mathcal{T})$ or $\mathcal{N}(\mathcal{T})$, respectively. We set $\mathbb{V}_0(\mathcal{T}^\top) := H^1_0(\Omega)$ and $\mathbb{V}(\mathcal{T}^\top) := H^1(\Omega)$.

The proof of the next proposition is left to the reader.

Proposition 2.6. *The mapping $\mathcal{T} \mapsto \mathbb{V}_0(\mathcal{T})$ from $\widehat{\mathbb{T}}$ to the lattice of vector spaces is compatible with the lattice structure, i.e.,*

$$\begin{aligned} \mathcal{T} \le \mathcal{T}_* \quad &\Rightarrow \quad \mathbb{V}_0(\mathcal{T}) \subset \mathbb{V}_0(\mathcal{T}_*), \\ \mathbb{V}_0(\mathcal{T} \wedge \mathcal{T}_*) &= \mathbb{V}_0(\mathcal{T}) \cap \mathbb{V}_0(\mathcal{T}_*), \\ \mathbb{V}_0(\mathcal{T} \vee \mathcal{T}_*) &= \mathbb{V}_0(\mathcal{T}) + \mathbb{V}_0(\mathcal{T}_*), \end{aligned}$$

The same holds true when we replace $\mathbb{V}_0(\mathcal{T})$ by $\mathbb{V}(\mathcal{T})$.

2.4. **The refinement routine.** For $\mathcal{P} \in \mathbb{P}$ and a finite set $\mathcal{C} \subset \mathcal{P}^\top$, we denote by $\mathcal{P} \oplus \mathcal{C}$ the *smallest refinement of* $\mathcal{P}$ *in* $\mathbb{P}$ *that contains* $\mathcal{C}$, i.e.,

$$\mathcal{P} \oplus \mathcal{C} := \bigwedge \{\mathcal{P}' \in \mathbb{P} \,:\, \mathcal{P}' \geq \mathcal{P}, \mathcal{C} \subset \mathcal{P}'\}.$$

This is well-defined. To see this, recall that, thanks to the matching condition, we have that $\{P \in \mathcal{P}^\top \,:\, \mathtt{gen}(P) \leq k\} \in \mathbb{P}$ for all $k \in \mathbb{N}_0$, so that the largest common coarsening can be taken over the finitely many $\mathcal{P}' \in \mathbb{P}$ with $\max_{P \in \mathcal{P}'} \mathtt{gen}(P) \leq \max_{P \in \mathcal{P} \cup \mathcal{C}} \mathtt{gen}(P)$.

For $P \in \mathcal{P}^\top$, we also write $\mathcal{P} \oplus P$ instead of $\mathcal{P} \oplus \{P\}$. Note that

$$\mathcal{P}_* \oplus \mathcal{P} = \mathcal{P}_* \oplus (\mathcal{P} \setminus \mathcal{P}_*) = \mathcal{P}_* \vee \mathcal{P} \qquad \text{for all } \mathcal{P}, \mathcal{P}_* \in \mathbb{P}.$$

For $\mathcal{P} \in \mathbb{P}$ and $\mathcal{C} \subset \mathcal{P}^\top \setminus \mathcal{P}_\perp$ we denote by $\mathcal{P} \ominus \mathcal{C}$ the *greatest coarsening of* $\mathcal{P}$ *in* $\mathbb{P}$ *that does not contain* $\mathcal{C}$, i.e.,

$$\mathcal{P} \ominus \mathcal{C} := \bigvee \{\mathcal{P}' \in \mathbb{P} \,:\, \mathcal{P}' \leq \mathcal{P},\, \mathcal{C} \cap \mathcal{P}' = \emptyset\}.$$

For $P \in \mathcal{P}^\top \setminus \mathcal{P}_\perp$, we also write $\mathcal{P} \ominus P$ for $\mathcal{P} \ominus \{P\}$.

For $\mathcal{T} \in \mathbb{T}$ and $\mathcal{U} \subset \mathcal{T}$, we denote by $\mathcal{T}_* = \mathrm{Ref}(\mathcal{T};\mathcal{U})$ the *smallest refinement of* $\mathcal{T}$ *in* $\mathbb{T}$ *with* $\mathcal{U} \cap \mathcal{T}_* = \emptyset$, i.e.,

$$\mathrm{Ref}(\mathcal{T};\mathcal{U}) \;= \bigwedge \{\mathcal{T}' \in \mathbb{T} \,:\, \mathcal{T}' \geq \mathcal{T},\, \mathcal{U} \cap \mathcal{T}' = \emptyset\}.$$

In this definition $\mathcal{T}' \in \mathbb{T}$ can be restricted to $\mathcal{T}' \leq \mathcal{T}^{++}$. The set $\mathcal{U}$ is commonly referred to as the set of triangles that are marked for refinement.

Although no uniform bound for $\#(\mathcal{T}_* \setminus \mathcal{T})/\#\mathcal{U}$ can be shown, the following important result is valid:

Theorem 2.7 ([BDD04])**.** *For any sequence* $(\mathcal{T}_k)_k \subset \mathbb{T}$ *defined by* $\mathcal{T}_0 = \mathcal{T}_\perp$ *and* $\mathcal{T}_{k+1} = \mathrm{Ref}(\mathcal{T}_k;\mathcal{U}_k)$ *for some* $\mathcal{U}_k \subset \mathcal{T}_k$, $k = 0, 1, \ldots,$ *we have that*

$$\#(\mathcal{T}_k \setminus \mathcal{T}_\perp) \lesssim \sum_{i=0}^{k-1} \#\mathcal{U}_i.$$

Our adaptive finite element routine will be driven by the marking of *edges* for refinement. Therefore, for $\mathcal{T} \in \mathbb{T}$ and $\mathcal{M} \subset \mathcal{S}(\mathcal{T})$, let $\mathcal{T}_* = \mathrm{Refine}(\mathcal{T};\mathcal{M})$ denote the *smallest refinement of* $\mathcal{T}$ *in* $\mathbb{T}$ *with* $\mathcal{M} \cap \mathcal{S}(\mathcal{T}_*) = \emptyset$, i.e.,

$$\mathrm{Refine}(\mathcal{T};\mathcal{M}) \;= \bigwedge \{\mathcal{T}' \in \mathbb{T} \,:\, \mathcal{T}' \geq \mathcal{T},\, \mathcal{M} \cap \mathcal{S}(\mathcal{T}') = \emptyset\}.$$

Note that $\mathrm{Refine}(\mathcal{T};\mathcal{M}) = \mathcal{T}(\mathcal{P}(\mathcal{T}) \oplus \mathrm{midpt}(\mathcal{M}))$.

Setting

$$\mathcal{U}_1 = \{T \in \mathcal{T} : \mathcal{M} \text{ contains an edge of } T\},$$
$$\mathcal{U}_2 = \{T' \in \mathrm{Ref}(\mathcal{T};\mathcal{U}_1) : \mathcal{M} \text{ contains an edge of } T'\},$$

we have that

$$\mathrm{Refine}(\mathcal{T};\mathcal{M}) = \mathrm{Ref}(\mathrm{Ref}(\mathcal{T};\mathcal{U}_1);\mathcal{U}_2).$$

Since moreover $\#\mathcal{U}_1 + \#\mathcal{U}_2 \leq 4 \cdot \#\mathcal{M}$, we conclude the following result.

Corollary 2.8. *For any sequence $(\mathcal{T}_k)_k \subset \mathbb{T}$ defined by $\mathcal{T}_0 = \mathcal{T}_\perp$ and $\mathcal{T}_{k+1} = \mathrm{Refine}(\mathcal{T}_k;\mathcal{M}_k)$ for some $\mathcal{M}_k \subset \mathcal{S}(\mathcal{T}_k)$, $k = 0, 1, \ldots$, we have that*

$$\#(\mathcal{T}_k \setminus \mathcal{T}_\perp) \lesssim \sum_{i=0}^{k-1} \#\mathcal{M}_i.$$

Since every simultaneous bisection of a pair of triangles that share their refinement edge increases the population by one, and the number of triangles by two, and every bisection of a triangle that has its refinement edge on the boundary increases both the population and the number of triangles by one, we observe that for $\mathcal{P}, \mathcal{P}_* \in \mathbb{P}$ with $\mathcal{P}_* \geq \mathcal{P}$,

$$\#(\mathcal{P}_* \setminus \mathcal{P}) \leq \#(\mathcal{T}(\mathcal{P}_*) \setminus \mathcal{T}(\mathcal{P})) \leq 2\,\#(\mathcal{P}_* \setminus \mathcal{P}). \tag{2.6}$$

This result will allow us to transfer Corollary 2.8 in terms of populations.

3. Continuous problem, its discretisation, and the lower diamond estimate

In this section, we shall introduce the model problem. Moreover, we shall investigate a splitting of the difference of energies related to nested spaces. To the best of our knowledge, this so-called lower diamond estimate is new, and it plays a crucial role in the proof of the instance optimality of the AFEM in Section 7.

3.1. Continuous and discrete problem.

We consider the model setting of Poisson's equation

$$\begin{aligned} -\Delta u &= f \qquad \text{on } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{3.1}$$

where, in view of the application of an a posteriori error estimator, we assume that $f \in L^2(\Omega)$. In weak form, it reads as finding $u := u_{\mathcal{T}^\top} \in H^1_0(\Omega) = \mathbb{V}_0(\mathcal{T}^\top)$ such that

$$\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x \qquad (v \in H^1_0(\Omega)).$$

For $\mathcal{T} \in \mathbb{T}$, the Galerkin approximation $u_{\mathcal{T}} \in \mathbb{V}_0(\mathcal{T})$ of u is uniquely defined by

$$\int_\Omega \nabla u_{\mathcal{T}} \cdot \nabla v_{\mathcal{T}} \,\mathrm{d}x = \int_\Omega f v_{\mathcal{T}} \,\mathrm{d}x \qquad (v_{\mathcal{T}} \in \mathbb{V}_0(\mathcal{T})). \tag{3.2}$$

It is well known, that for $\mathcal{T} \in \hat{\mathbb{T}}$, $u_{\mathcal{T}}$ is the unique minimiser of the (Dirichlet) energy

$$\mathcal{J}(v) := \int_\Omega \frac{1}{2}|\nabla v|^2 - f v \,\mathrm{d}x \qquad (v \in \mathbb{V}_0(\mathcal{T})).$$

Setting

$$\mathcal{J}(\mathcal{T}) := \mathcal{J}(u_{\mathcal{T}}),$$

Proposition 2.6 shows that $\mathcal{J}$ is non-increasing with respect to $(\widehat{\mathbb{T}}, \leq)$, i.e., for $\mathcal{T}, \mathcal{T}_* \in \widehat{\mathbb{T}}$, we have

$$\mathcal{T} \leq \mathcal{T}_* \qquad \Rightarrow \qquad \mathcal{J}(\mathcal{T}) \geq \mathcal{J}(\mathcal{T}_*). \tag{3.3}$$

Moreover, from basic calculations we observe that for $\mathcal{T} \in \mathbb{T}$, we have

$$\mathcal{J}(\mathcal{T}) - \mathcal{J}(\mathcal{T}_*) = \frac{1}{2}|u_{\mathcal{T}} - u_{\mathcal{T}_*}|^2_{H^1(\Omega)} \tag{3.4}$$

for all $\mathcal{T}_* \in \hat{\mathbb{T}}$ with $\mathcal{T} \leq \mathcal{T}_*$.

3.2. **The lower diamond estimate.** To formulate the main result from this subsection, we have to start with a definition.

Definition 3.1. *For $\{\mathcal{T}_1, \ldots, \mathcal{T}_m\} \subset \mathbb{T}$, we call $(\mathcal{T}^\wedge, \mathcal{T}_\vee; \mathcal{T}_1, \ldots, \mathcal{T}_m)$ a* lower diamond *in $\mathbb{T}$ of size m, when $\mathcal{T}^\wedge = \bigwedge_{j=1}^m \mathcal{T}_j$, $\mathcal{T}_\vee = \bigvee_{j=1}^m \mathcal{T}_j$, and the areas of* coarsening $\Omega(\mathcal{T}_j \setminus \mathcal{T}_\vee)$ *are pairwise disjoint, cf. Figure 4 for an illustration.*

It is called an upper diamond *in $\mathbb{T}$ of size m, when the last condition reads as the areas of* refinement $\Omega(\mathcal{T}^\wedge \setminus \mathcal{T}_j)$ *being pairwise disjoint.*

Obviously, for any $\mathcal{T} \in \mathbb{T}$, $(\mathcal{T}, \mathcal{T}; \mathcal{T})$ is a lower (and upper) diamond in $\mathbb{T}$ of size 1. More interesting is the following result:

Lemma 3.2. *For any $\mathcal{T}_1 \neq \mathcal{T}_2 \in \mathbb{T}$, $(\mathcal{T}_1 \wedge \mathcal{T}_2, \mathcal{T}_1 \vee \mathcal{T}_2; \mathcal{T}_1, \mathcal{T}_2)$ is a lower (and upper) diamond in $\mathbb{T}$ of size 2.*

Proof. Setting $\mathcal{T}^\wedge := \mathcal{T}_1 \wedge \mathcal{T}_2$, $\mathcal{T}_\vee := \mathcal{T}_1 \vee \mathcal{T}_2$, assume that $\Omega(\mathcal{T}_1 \setminus \mathcal{T}_\vee)$ and $\Omega(\mathcal{T}_2 \setminus \mathcal{T}_\vee)$ are not disjoint. Recalling that $\Omega(\mathcal{T}_j \setminus \mathcal{T}_\vee) = \Omega(\mathcal{T}_\vee \setminus \mathcal{T}_j)$, then there exists a $T \in (\mathcal{T}_\vee \setminus \mathcal{T}_1) \cap (\mathcal{T}_\vee \setminus \mathcal{T}_2) = \mathcal{T}_\vee \setminus (\mathcal{T}_1 \cup \mathcal{T}_2)$. This contradicts $\mathcal{T}_\vee = \mathcal{T}_1 \vee \mathcal{T}_2$; compare also with Remark 2.5, and thus $(\mathcal{T}^\wedge, \mathcal{T}_\vee; \mathcal{T}_1, \mathcal{T}_2)$ is a lower diamond.

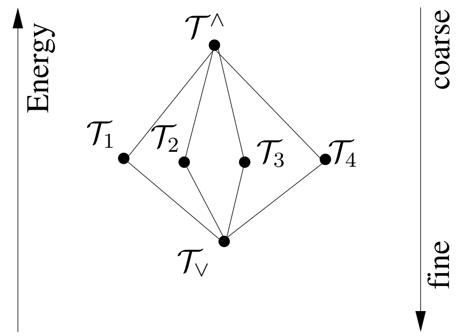

FIGURE 4. Lower (or upper) diamond of size 4.

Similarly, one finds that $(\mathcal{T}_1 \setminus \mathcal{T}^\wedge) \cap (\mathcal{T}_2 \setminus \mathcal{T}^\wedge) = \emptyset$, i.e., $(\mathcal{T}^\wedge, \mathcal{T}_\vee; \mathcal{T}_1, \mathcal{T}_2)$ is an upper diamond. $\square$

The main goal of this subsection is to prove the following result:

Theorem 3.3. *Let $(\mathcal{T}^\wedge, \mathcal{T}_\vee; \mathcal{T}_1, \dots, \mathcal{T}_m)$ be a lower diamond in $\mathbb{T}$. Then*

$$|u_{\mathcal{T}_\vee} - u_{\mathcal{T}^\wedge}|^2_{H^1(\Omega)} \eqsim \sum_{j=1}^m |u_{\mathcal{T}_\vee} - u_{\mathcal{T}_j}|^2_{H^1(\Omega)}, \tag{3.5}$$

only dependent on $\mathcal{T}_\perp$.

The first ingredient to prove this theorem is the following observation.

Lemma 3.4. *Let $\mathcal{T}, \mathcal{T}_* \in \widehat{\mathbb{T}}$ with $\mathcal{T} \leq \mathcal{T}_*$, and let $\Pi : \mathbb{V}_0(\mathcal{T}_*) \to \mathbb{V}_0(\mathcal{T}_*)$ be a linear projector onto $\mathbb{V}_0(\mathcal{T})$ which is $H^1(\Omega)$-bounded, uniformly in $\mathcal{T}, \mathcal{T}_*$. Then,*

$$|u_{\mathcal{T}_*} - u_{\mathcal{T}}|_{H^1(\Omega)} \eqsim |u_{\mathcal{T}_*} - \Pi u_{\mathcal{T}_*}|_{H^1(\Omega)}.$$

Proof. Use that $u_{\mathcal{T}}$ is the best approximation from $\mathbb{V}_0(\mathcal{T})$ to $u_{\mathcal{T}_*}$ in $|\cdot|_{H^1(\Omega)}$, and $|u_{\mathcal{T}_*} - \Pi u_{\mathcal{T}_*}|_{H^1(\Omega)} \leq |u_{\mathcal{T}_*} - v_{\mathcal{T}}|_{H^1(\Omega)} + |\Pi(v_{\mathcal{T}} - u_{\mathcal{T}_*})|_{H^1(\Omega)} \lesssim |u_{\mathcal{T}_*} - v_{\mathcal{T}}|_{H^1(\Omega)}$ for all $v_{\mathcal{T}} \in \mathbb{V}_0(\mathcal{T})$. $\square$

In order to localize the projection error to the area of coarsening, we shall consider a particular Scott-Zhang type quasi-interpolator [SZ90].

Lemma 3.5. *Let $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$ with $\mathcal{T} \leq \mathcal{T}_*$. Let $\Omega_1 := \Omega(\mathcal{T} \setminus \mathcal{T}_*)$ and $\Omega_2 := \Omega \setminus \overline{\Omega}_1$. There exists a projector $\Pi_{\mathcal{T}_* \to \mathcal{T}} : H^1(\Omega) \to H^1(\Omega)$ onto $\mathbb{V}(\mathcal{T})$ with the following properties*

(Pr1) $|\Pi_{\mathcal{T}_* \to \mathcal{T}} v|_{H^1(\Omega)} \lesssim |v|_{H^1(\Omega)}$ *for all $v \in H^1(\Omega)$.*

(Pr2) *There exist projectors* $\bar{\Pi}_{\mathcal{T},i} : H^1(\Omega_i) \to H^1(\Omega_i)$ *onto* $\mathbb{V}(\mathcal{T})|_{\Omega_i} := \{v|_{\Omega_i} : v \in \mathbb{V}(\mathcal{T})\}$, *with for any* $T \in \mathcal{T}$ *with* $T \subset \bar{\Omega}_i$,

$$|\bar{\Pi}_{\mathcal{T},i} v_i|^2_{H^1(T)} \lesssim \sum_{\{T' \in \mathcal{T} : T' \cap T \neq \emptyset,\, T' \subset \bar{\Omega}_i\}} |v_i|^2_{H^1(T')}$$

for all $v_i \in H^1(\Omega_i)$, $i = 1, 2$, *such that*

$$(\Pi_{\mathcal{T}_* \to \mathcal{T}} v)|_{\Omega_i} = \bar{\Pi}_{\mathcal{T},i}(v|_{\Omega_i}), \qquad i = 1, 2.$$

(Pr3) $v - \Pi_{\mathcal{T}_* \to \mathcal{T}} v$ *vanishes on* $\overline{\Omega}_2$ *for all* $v \in \mathbb{V}(\mathcal{T}_*)$.

(Pr4) $\Pi_{\mathcal{T}_* \to \mathcal{T}}(\mathbb{V}_0(\mathcal{T}_*)) \subset \mathbb{V}_0(\mathcal{T})$.

Proof. For the construction of $\Pi_{\mathcal{T}_* \to \mathcal{T}}$, we assign to each node $z \in \mathcal{N}(\mathcal{T})$ some edge $S_z \in \mathcal{S}(\mathcal{T})$ such that $z \in S_z$ and

$$S_z \subset \begin{cases} \partial\Omega_1, & \text{if } z \in \partial\Omega_1, \\ \partial\Omega_2, & \text{if } z \in \partial\Omega_2. \end{cases} \tag{3.6}$$

These restrictions are well posed, since Ω is a domain, which excludes the case that Ω_1 and Ω_2 touch at some isolated point. We denote by $\Pi := \Pi_{\mathcal{T}_* \to \mathcal{T}}$ the Scott-Zhang projector according to the above assignments (3.6), i.e., for $z \in \mathcal{N}(\mathcal{T})$, the nodal value $(\Pi v)(z)$ is defined by means of $L^2(S_z)$ dual functions of the local nodal basis functions on S_z; compare with [SZ90]. Then $\Pi\colon H^1(\Omega) \to H^1(\Omega)$ is a projector onto $\mathbb{V}(\mathcal{T})$, and (Pr1) follows from [SZ90].

Thanks to (3.6), we may define the Scott-Zhang projectors $\bar{\Pi}_{\mathcal{T},i} : H^1(\Omega_i) \to H^1(\Omega_i)$ onto $\mathbb{V}(\mathcal{T})|_{\Omega_i}$ according to S_z, $z \in \mathcal{N}(\mathcal{T}) \cap \overline{\Omega}_i$, $i = 1, 2$. With these definitions the properties listed in (Pr2) are valid.

Let $v \in \mathbb{V}(\mathcal{T}_*)$. Then $v|_{\Omega_2} \in \mathbb{V}(\mathcal{T})|_{\Omega_2}$ and since $\bar{\Pi}_{\mathcal{T},2}$ is a projector onto $\mathbb{V}(\mathcal{T})|_{\Omega_2}$, we have that $\bar{\Pi}_{\mathcal{T},2} v|_{\Omega_2} = v|_{\Omega_2}$, i.e., $\Pi v = v$ on $\overline{\Omega}_2$. This proves (Pr3).

In order to prove (Pr4) let $v \in \mathbb{V}_0(\mathcal{T}_*)$. Then (Pr3) implies $v = 0$ on $\partial\Omega \cap \partial\Omega_2$. Therefore, let $z \in \mathcal{N}(\mathcal{T})$ with $z \in \partial\Omega \setminus \partial\Omega_2$. Then locally $\partial\Omega_1$ coincides with $\partial\Omega$, and thus $S_z \subset \partial\Omega_1 \cap \partial\Omega$ according to (3.6). Since $v = 0$ on $\partial\Omega$, it follows from properties of the Scott-Zhang projector that $(\Pi v)(z) = 0$. Consequently, we have $(\Pi v)(z) = 0$ for all $z \in \mathcal{N}(\mathcal{T})$, i.e., $\Pi v = 0$ on $\partial\Omega$. □

Remark 3.6. *Note that the projector constructed in Lemma 3.5 does not map* $H^1_0(\Omega)$ *into* $\mathbb{V}_0(\mathcal{T})$ *when* Ω_1 *touches the boundary. In such a situation we might have* $z \in \mathcal{N}(\mathcal{T}) \cap \partial\Omega \cap \partial\Omega_1 \cap \partial\Omega_2$ *but* $\partial\Omega \cap \partial\Omega_1 \cap \partial\Omega_2$ *contains no edge. Hence, in view of* (3.6) *it is not possible to require additionally that* $z \in \partial\Omega$ *implies* $S_z \subset \partial\Omega$.

Theorem 3.7. *Let $(\mathcal{T}^\wedge, \mathcal{T}_\vee; \mathcal{T}_1, \dots, \mathcal{T}_m)$ be a lower diamond in $\mathbb{T}$. Set $\Pi_j := \Pi_{\mathcal{T}_\vee \to \mathcal{T}_j}$, and $\Omega_j := \Omega(\mathcal{T}_j \setminus \mathcal{T}_\vee)$, $j = 1, \dots, m$. Then the projectors Π_j commute as operators from $\mathbb{V}(\mathcal{T}_\vee) \to \mathbb{V}(\mathcal{T}_\vee)$.*

Define $\Pi := \Pi_1 \circ \dots \circ \Pi_m : \mathbb{V}(\mathcal{T}_\vee) \to \mathbb{V}(\mathcal{T}_\vee)$. Then Π is a projector onto $\mathbb{V}(\mathcal{T}^\wedge)$, and $\Pi(\mathbb{V}_0(\mathcal{T}_\vee)) \subset \mathbb{V}_0(\mathcal{T}^\wedge)$. Moreover, for all $v_\vee \in \mathbb{V}(\mathcal{T}_\vee)$ we have

$$\Pi v_\vee = \begin{cases} \Pi_j v_\vee & \text{on } \overline{\Omega}_j, \\ v_\vee & \text{on } \overline{\Omega \setminus \cup_{j=1}^m \Omega_j} \end{cases} \tag{3.7}$$

and $|\Pi v_\vee|_{H^1(\Omega)} \lesssim |v_\vee|_{H^1(\Omega)}$, only dependent on $\mathcal{T}_\perp$.

Proof. Thanks to Lemma 3.5 we have that Π_j is a projector onto $\mathbb{V}(\mathcal{T}_j)$, and $\Pi_j(\mathbb{V}_0(\mathcal{T}_\vee)) \subset \mathbb{V}_0(\mathcal{T}_j)$. We fix $i \neq j$. Since Ω_i and Ω_j are disjoint, we have $\overline{\Omega}_i \subset \overline{\Omega \setminus \overline{\Omega}_j}$. Hence, we conclude from `(Pr3)` that $\Pi_j v_\vee = v_\vee$ on $\overline{\Omega}_i$ and $\Pi_i v_\vee = \Pi_j \Pi_i v_\vee$ on $\overline{\Omega}_i$. Since `(Pr2)` implies that $(\Pi_i w)|_{\Omega_i}$ only depends on $w|_{\Omega_i}$, we conclude that $\Pi_i \Pi_j v_\vee = \Pi_i v_\vee$ on $\overline{\Omega}_i$, and thus $\Pi_i \Pi_j v_\vee = \Pi_i v_\vee = \Pi_j \Pi_i v_\vee$ on $\overline{\Omega}_i$. Analogously, we have $\Pi_i \Pi_j v_\vee = \Pi_j v_\vee = \Pi_j \Pi_i v_\vee$ on $\overline{\Omega}_j$. Moreover, thanks to `(Pr3)`, we have $\Pi_i \Pi_j v_\vee = v_\vee = \Pi_j \Pi_i v_\vee$ on $\overline{\Omega \setminus (\Omega_i \cup \Omega_j)}$ and thus $\Pi_i \Pi_j v_\vee = \Pi_j \Pi_i v_\vee$.

Since the Π_j commute, we conclude that Π is a projector onto $\bigcap_{j=1}^m \mathbb{V}(\mathcal{T}_j) = \mathbb{V}(\mathcal{T}^\wedge)$; compare also with Proposition 2.6. The claim $\Pi(\mathbb{V}_0(\mathcal{T}_\vee)) \subset \mathbb{V}_0(\mathcal{T}^\wedge)$ follows analogously using `(Pr4)`. Again since the Π_j commute we infer (3.7) from `(Pr2)`.

Thanks to (3.7) and `(Pr2)` we conclude

$$\begin{aligned} |\Pi v_\vee|^2_{H^1(\Omega)} &= |v_\vee|^2_{H^1(\Omega \setminus \bigcup_{j=1}^m \Omega_j)} + \sum_{j=1}^m |\Pi_j v_\vee|^2_{H^1(\Omega_j)} \\ &\lesssim |v_\vee|^2_{H^1(\Omega \setminus \bigcup_{j=1}^m \Omega_j)} + \sum_{j=1}^m |v_\vee|^2_{H^1(\Omega_j)} = |v_\vee|^2_{H^1(\Omega)}, \end{aligned}$$

with constants independent of m. □

With the projectors Π_j and Π at hand, we are now ready to prove the main result of this section.

Proof of Theorem 3.3. Thanks to Lemma 3.5, $\Pi, \Pi_j : \mathbb{V}_0(\mathcal{T}_\vee) \to \mathbb{V}_0(\mathcal{T}_\vee)$ are (uniformly) bounded projectors onto $\mathbb{V}_0(\mathcal{T}^\wedge)$ or $\mathbb{V}_0(\mathcal{T}_j)$. From this, together with Lemma 3.4, (3.7), and (Pr3), we infer that

$$|u_{\mathcal{T}_\vee} - u_{\mathcal{T}^\wedge}|^2_{H^1(\Omega)} \eqsim |u_{\mathcal{T}_\vee} - \Pi u_{\mathcal{T}_\vee}|^2_{H^1(\Omega)} = \sum_{j=1}^m |u_{\mathcal{T}_\vee} - \Pi_j u_{\mathcal{T}_\vee}|^2_{H^1(\Omega_j)}$$
$$= \sum_{j=1}^m |u_{\mathcal{T}_\vee} - \Pi_j u_{\mathcal{T}_\vee}|^2_{H^1(\Omega)} \eqsim \sum_{j=1}^m |u_{\mathcal{T}_\vee} - u_{\mathcal{T}_j}|^2_{H^1(\Omega)}. \qquad \square$$

Thanks to (3.4), the latter result directly transfers to energy differences. In fact, under the conditions of Theorem 3.3, we have

$$\mathcal{J}(\mathcal{T}^\wedge) - \mathcal{J}(\mathcal{T}_\vee) \eqsim \sum_{j=1}^m \big(\mathcal{J}(\mathcal{T}_j) - \mathcal{J}(\mathcal{T}_\vee)\big).$$

This estimate is fundamental for our optimality analysis in Section 7. We make the following definition:

Definition 3.8. *An energy* $\widetilde{\mathcal{J}} : \mathbb{T} \to \mathbb{R}$ *is said to satisfy the* lower diamond estimate *when for all lower diamonds* $(\mathcal{T}^\wedge, \mathcal{T}_\vee; \mathcal{T}_1, \dots, \mathcal{T}_m)$ *in* $\mathbb{T}$, *it holds that*

$$\widetilde{\mathcal{J}}(\mathcal{T}^\wedge) - \widetilde{\mathcal{J}}(\mathcal{T}_\vee) \eqsim \sum_{j=1}^m \big(\widetilde{\mathcal{J}}(\mathcal{T}_j) - \widetilde{\mathcal{J}}(\mathcal{T}_\vee)\big),$$

independent of the lower diamond.

Corollary 3.9. *The energy* $\mathcal{J}$ *satisfies the lower diamond estimate.*

4. A posteriori error estimation

In this section we shall present an edge-based variant of the standard residual error estimator and recall some of its properties. To this end, we fix some triangulation $\mathcal{T} \in \mathbb{T}$ of Ω. For $S \in \mathcal{S}(\mathcal{T})$ we define $\Omega_{\mathcal{T}}(S)$ as the interior of the union of the triangles with common edge S, and we define the squared local error indicators by

(4.1)
$$\mathcal{E}^2_{\mathcal{T}}(S) := \begin{cases} \displaystyle\sum_{\{T\in\mathcal{T}: T\subset\overline{\Omega_{\mathcal{T}}(S)}\}} h_T^2 \|f\|^2_{L^2(T)} + h_S \big\|[\![\nabla u_{\mathcal{T}}]\!]\big\|^2_{L^2(S)} & \text{for } S \subset \Omega \\ \displaystyle\sum_{\{T\in\mathcal{T}: T\subset\overline{\Omega_{\mathcal{T}}(S)}\}} h_T^2 \|f\|^2_{L^2(T)} & \text{for } S \subset \partial\Omega. \end{cases}$$

Here $h_S := |S|$, $h_T := |T|^{\frac{1}{2}}$, and

$$[\![\nabla u_{\mathcal{T}}]\!]|_S := \sum_{\{T\in\mathcal{T}: S\subset T\}} \nabla u_{\mathcal{T}}|_T \cdot \boldsymbol{n}_T$$

with $\boldsymbol{n}_T$ being the outward pointing unit normal on ∂T. Note that, thanks to the choice $h_T = |T|^{\frac{1}{2}}$, we have that the local mesh size h_T decreases strictly with the factor $2^{-1/2}$ at every bisection of T.

For $\tilde{\mathcal{S}} \subset \mathcal{S}(\mathcal{T})$ we define the accumulated squared error indicator by

$$\mathcal{E}^2_{\mathcal{T}}(\tilde{\mathcal{S}}) := \sum_{S\in\tilde{\mathcal{S}}} \mathcal{E}^2_{\mathcal{T}}(S). \tag{4.2}$$

For $\mathcal{T} \in \mathbb{T}$ we define the squared oscillation $\operatorname{osc}^2(\mathcal{U})$ on $\mathcal{U} \subset \mathcal{T}$ by

$$\operatorname{osc}^2(\mathcal{U}) := \sum_{T\in\mathcal{U}} h_T^2 \|f - f_T\|^2_{L^2(T)},$$

where $f_T := \frac{1}{|T|}\int_T f\,dx$. It is well known that the estimator, defined in (4.1), is reliable and efficient in the following sense; compare e.g. with [Ver96].

Proposition 4.1. *For $\mathcal{T} \in \mathbb{T}$ we have the bounds*

$$|u - u_{\mathcal{T}}|^2_{H^1(\Omega)} \lesssim \mathcal{E}^2_{\mathcal{T}}(\mathcal{S}(\mathcal{T})) \lesssim |u - u_{\mathcal{T}}|^2_{H^1(\Omega)} + \operatorname{osc}^2(\mathcal{T}).$$

This proposition shows that the error estimator mimics the error $|u - u_{\mathcal{T}}|_{H^1(\Omega)}$ up to oscillation.

We define the *total energy* $\mathcal{G} : \mathbb{T} \to \mathbb{R}$ by

$$\mathcal{G}(\mathcal{T}) := \mathcal{J}(\mathcal{T}) + \mathcal{H}(\mathcal{T}), \quad \text{where} \quad \mathcal{H}(\mathcal{T}) := \sum_{T\in\mathcal{T}} h_T^2 \|f\|^2_{L^2(T)}. \tag{4.3}$$

Note that $\mathcal{G}$, $\mathcal{H}$ and osc^2 are non-increasing with respect to $\mathbb{T}$. Since $\mathcal{H}(\mathcal{T})$, $\operatorname{osc}^2(\mathcal{T}) \to 0$ for $\mathcal{T} \to \mathcal{T}^\top$, it is natural to set $\mathcal{H}(\mathcal{T}^\top) := 0$, $\operatorname{osc}^2(\mathcal{T}^\top) := 0$, and $\mathcal{G}(\mathcal{T}^\top) := \mathcal{J}(\mathcal{T}^\top)$.

From Proposition 4.1 and $\mathcal{E}^2_{\mathcal{T}}(\mathcal{S}(\mathcal{T})) \geq \mathcal{H}(\mathcal{T}) \geq \operatorname{osc}^2(\mathcal{T})$, we obtain

$$\mathcal{E}^2_{\mathcal{T}}(\mathcal{S}(\mathcal{T})) \eqsim |u - u_{\mathcal{T}}|^2_{H^1(\Omega)} + \operatorname{osc}^2(\mathcal{T}) \eqsim |u - u_{\mathcal{T}}|^2_{H^1(\Omega)} + \mathcal{H}(\mathcal{T}). \tag{4.4}$$

The term $(|u - u_{\mathcal{T}}|^2_{H^1(\Omega)} + \operatorname{osc}^2(\mathcal{T}))^{1/2}$ is referred to as the *total error*. Similarly, we have in terms of an energy difference, that

$$\mathcal{E}^2_{\mathcal{T}}(\mathcal{S}(\mathcal{T})) \eqsim (\mathcal{J} + \operatorname{osc}^2)(\mathcal{T}) - (\mathcal{J} + \operatorname{osc}^2)(\mathcal{T}^\top) \eqsim \mathcal{G}(\mathcal{T}) - \mathcal{G}(\mathcal{T}^\top) \tag{4.5}$$

Therefore, in order to prove instance optimality for the total error, it suffices to prove instance optimality of the energy difference of the total energy $\mathcal{G}$.

In order two compare the energies of two discrete solutions, we need the following lemma.

Lemma 4.2. *Let $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$ with $\mathcal{T} \leq \mathcal{T}_*$, then*

$$\mathcal{H}(\mathcal{T}) - \mathcal{H}(\mathcal{T}_*) \eqsim \mathcal{H}(\mathcal{T} \setminus \mathcal{T}_*) \quad := \sum_{T \in \mathcal{T} \setminus \mathcal{T}_*} h_T^2 \|f\|_{L^2(T)}^2.$$

$$\operatorname{osc}^2(\mathcal{T}) - \operatorname{osc}^2(\mathcal{T}_*) \eqsim \operatorname{osc}^2(\mathcal{T} \setminus \mathcal{T}_*) = \sum_{T \in \mathcal{T} \setminus \mathcal{T}_*} h_T^2 \|f - f_T\|_{L^2(T)}^2.$$

Proof. Since every bisection locally reduces the mesh size by a factor of $2^{-1/2}$, we have $\mathcal{H}(\mathcal{T}_* \setminus \mathcal{T}) \leq \frac{1}{2}\mathcal{H}(\mathcal{T} \setminus \mathcal{T}_*)$. This and $\mathcal{H}(\mathcal{T}) - \mathcal{H}(\mathcal{T}_*) = \mathcal{H}(\mathcal{T} \setminus \mathcal{T}_*) - \mathcal{H}(\mathcal{T}_* \setminus \mathcal{T})$ proves the claim for $\mathcal{H}$. The proof of the second claim is similar using also $\inf_{c \in \mathbb{R}} \|f - c\|_{L^2(T)} = \|f - f_T\|_{L^2(T)}$. $\square$

We shall now derive a discrete analogue of Proposition 4.1. To this end, we need the Scott-Zhang type interpolation $\Pi_{\mathcal{T}_* \to \mathcal{T}}$ introduced in Section 3.

Lemma 4.3. *Let $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$, with $\mathcal{T}_* \geq \mathcal{T}$ and denote by $\mathcal{S} = \mathcal{S}(\mathcal{T})$ and $\mathcal{S}_* = \mathcal{S}(\mathcal{T}_*)$ the respective sets of sides. Then we have*

$$|u_{\mathcal{T}} - u_{\mathcal{T}_*}|_{H^1(\Omega)}^2 \lesssim \mathcal{E}_{\mathcal{T}}^2(\mathcal{S} \setminus \mathcal{S}_*) \lesssim |u_{\mathcal{T}} - u_{\mathcal{T}_*}|_{H^1(\Omega)}^2 + \mathcal{H}(\mathcal{T} \setminus \mathcal{T}_*),$$

where $\mathcal{S} \setminus \mathcal{S}_*$ is the set of sides in $\mathcal{S}$ that are refined in $\mathcal{S}_*$.

Proof. Let $e_* := u_{\mathcal{T}_*} - u_{\mathcal{T}}$, then by Lemma 3.5`(Pr4)` we have that

$$\begin{aligned} |e_*|_{H^1(\Omega)}^2 &= \int_\Omega \nabla e_* \cdot \nabla e_* \, \mathrm{d}x = \int_\Omega \nabla e_* \cdot (\nabla e_* - \nabla \Pi_{\mathcal{T}_* \to \mathcal{T}} e_*) \, \mathrm{d}x \\ &= \sum_{T \in \mathcal{T}} \int_T f(e_* - \Pi_{\mathcal{T}_* \to \mathcal{T}} e_*) \, \mathrm{d}x - \sum_{S \in \mathcal{S}} \int_S [\![\nabla u_{\mathcal{T}}]\!](e_* - \Pi_{\mathcal{T}_* \to \mathcal{T}} e_*) \, \mathrm{d}s. \end{aligned}$$

It follows from Lemma 3.5`(Pr3)` that $e_* = \Pi_{\mathcal{T}_* \to \mathcal{T}} e_*$ on $\Omega \setminus \overline{\Omega(\mathcal{T} \setminus \mathcal{T}_*)}$. The first inequality to be shown follows by the trace theorem for the second sum, the Cauchy-Schwarz inequality and Lemma 3.5`(Pr2)`.

In order to prove the second inequality, let $S \in \mathcal{S} \setminus \mathcal{S}_*$, i.e., S is refined in $\mathcal{S}_*$. In other words, we have for the midpoint z of S, that $z \in \mathcal{N}(\mathcal{T}_*)$. If $S \subset \partial\Omega$, then trivially have

$$\mathcal{E}_{\mathcal{T}}^2(S) \lesssim |u_{\mathcal{T}} - u_{\mathcal{T}_*}|_{H^1(\Omega_{\mathcal{T}}(S))}^2 + \sum_{\{T \in \mathcal{T} : T \subseteq \Omega_{\mathcal{T}}(S)\}} h_T^2 \|f\|_{L^2(T)}^2. \tag{4.6}$$

For $S \not\subset \partial\Omega$, let $\mathcal{T}' := \operatorname{Refine}(\mathcal{T}; S)$, and let $\varphi_z \in \mathbb{V}_0(\mathcal{T}')$ be defined by $\varphi_z(z) = 1$, and $\varphi_z(z') = 0$ for $z' \in \mathcal{N}(\mathcal{T}') \setminus \{z\}$. Note that $\varphi_z \in \mathbb{V}_0(\mathcal{T}_*)$, and $\operatorname{supp} \varphi_z \subseteq \Omega_{\mathcal{T}}(S)$. We recall that $[\![\nabla u_{\mathcal{T}}]\!]|_S \in \mathbb{R}$ and deduce from

(3.2), that

$$\begin{aligned}\frac{1}{2}\int_S h_S[\![\nabla u_{\mathcal{T}}]\!]^2\,\mathrm{d}s &= \int_S h_S[\![\nabla u_{\mathcal{T}}]\!]^2\varphi_z\,\mathrm{d}s\\ &= \int_{\Omega_{\mathcal{T}}(S)} f h_S[\![\nabla u_{\mathcal{T}}]\!]\varphi_z\,\mathrm{d}x\\ &\quad - \int_{\Omega_{\mathcal{T}}(S)} (\nabla u_{\mathcal{T}_*} - \nabla u_{\mathcal{T}}) h_S[\![\nabla u_{\mathcal{T}}]\!]\nabla\varphi_z\,\mathrm{d}x\\ &\le h_S\|f\|_{L^2(\Omega_{\mathcal{T}}(S))}\|[\![\nabla u_{\mathcal{T}}]\!]\varphi_z\|_{L^2(\Omega_{\mathcal{T}}(S))}\\ &\quad + \|\nabla u_{\mathcal{T}_*} - \nabla u_{\mathcal{T}}\|_{L^2(\Omega_{\mathcal{T}}(S))}\|[\![\nabla u_{\mathcal{T}}]\!]h_S\nabla\varphi_z\|_{L^2(\Omega_{\mathcal{T}}(S))}.\end{aligned}$$

With standard scaling arguments we obtain that

$$\big\|[\![\nabla u_{\mathcal{T}}]\!]h_S\nabla\varphi_z\big\|^2_{L^2(\Omega_{\mathcal{T}}(S))} \lesssim \big\|[\![\nabla u_{\mathcal{T}}]\!]\varphi_z\big\|^2_{L^2(\Omega_{\mathcal{T}}(S))} \lesssim \int_S h_S[\![\nabla u_{\mathcal{T}}]\!]^2\,\mathrm{d}s.$$

Thus it follows from Young's inequality that

$$\int_S h_S[\![\nabla u_{\mathcal{T}}]\!]^2\,\mathrm{d}s \lesssim \|\nabla u_{\mathcal{T}_*} - \nabla u_{\mathcal{T}}\|^2_{L^2(\Omega_{\mathcal{T}}(S))} + \sum_{\{T\in\mathcal{T}: T\subseteq\Omega_{\mathcal{T}}(S)\}} h_T^2\|f\|^2_{L^2(T)}.$$

and consequently we have (4.6) for all $S \in \mathcal{S}\setminus\mathcal{S}_*$. Since we have at most a triple overlap of the $\Omega_{\mathcal{T}}(S)$, $S\in\mathcal{S}$, the assertion follows by summing over all $S\in\mathcal{S}\setminus\mathcal{S}_*$. □

The next result is the discrete analogue of (4.5) and shows that the total energy $\mathcal{G}$ matches perfectly the squared a posteriori error estimator defined in (4.1).

Proposition 4.4. *Let $\mathcal{T},\mathcal{T}_*\in\mathbb{T}$ with $\mathcal{T}\le\mathcal{T}_*$. Then we have*

$$\mathcal{G}(\mathcal{T}) - \mathcal{G}(\mathcal{T}_*) \eqsim \mathcal{E}^2_{\mathcal{T}}(\mathcal{S}(\mathcal{T})\setminus\mathcal{S}(\mathcal{T}_*)).$$

Proof. By (3.4) and Lemma 4.2, we have $\mathcal{G}(\mathcal{T})-\mathcal{G}(\mathcal{T}_*) \eqsim |u_{\mathcal{T}} - u_{\mathcal{T}_*}|^2_{H^1(\Omega)} + \mathcal{H}(\mathcal{T}\setminus\mathcal{T}_*)$. From

$$\overline{\Omega(\mathcal{T}\setminus\mathcal{T}_*)} = \bigcup_{T\in\mathcal{T}\setminus\mathcal{T}_*} T = \bigcup_{S\in\mathcal{S}(\mathcal{T})\setminus\mathcal{S}(\mathcal{T}_*)} \overline{\Omega_{\mathcal{T}}(S)}, \tag{4.7}$$

with an at most triple overlap of the sets $\overline{\Omega_{\mathcal{T}}(S)}$, it follows that $\mathcal{H}(\mathcal{T}\setminus\mathcal{T}_*) \lesssim \mathcal{E}^2_{\mathcal{T}}(\mathcal{S}(\mathcal{T})\setminus\mathcal{S}(\mathcal{T}_*))$. An application of Lemma 4.3 completes the proof. □

Let us turn to the case of coarsenings in mutual disjoint areas. As a direct consequence of Lemma 4.2 and the fact that $\mathcal{J}$ satisfies the lower diamond estimate (Corollary 3.9), we get the following result.

Corollary 4.5. *The energies $\mathcal{H}$, osc^2 and $\mathcal{G}$ satisfy the lower diamond estimate.*

Remark 4.6. *In this paper we resort to edge based error indicators* (4.1) *for the following reason. In the situation of Proposition 4.4 consider e.g. the element based squared error indicators*

$$\tilde{\mathcal{E}}^2_{\mathcal{T}}(T) := h_T^2 \|f\|^2_{L^2(T)} + \tfrac{1}{2} h_T \|[\![\nabla u_{\mathcal{T}}]\!]\|^2_{L^2(\partial T)}$$

from [Ver96]*. Then we have the estimate*

$$\tilde{\mathcal{E}}^2_{\mathcal{T}}(T) \lesssim \sum_{S \subset T} \|\nabla u_{\mathcal{T}} - \nabla u_{\mathcal{T}_*}\|^2_{L^2(\Omega_{\mathcal{T}}(S))} + h_T^2 \|f\|^2_{L^2(\Omega_{\mathcal{T}}(S))}$$

only if all three edges of T are bisected at least once in $\mathcal{T}_*$; compare e.g. with [Dör96, MNS00, MNS02]*. Since $\mathcal{T}_*$ is conforming, this can only be true for all elements $T \in \mathcal{T} \setminus \mathcal{T}_*$ when all elements of $\mathcal{T}$ are at least refined twice in $\mathcal{T}_*$. Consequently, either we have estimates similar to those in Proposition 4.4 for squared element based error indicators running over different sets of elements for both inequalities respectively, or we need to resort to global refinement. In the latter case we have $\mathcal{T} \setminus \mathcal{T}_* = \mathcal{T}$.*

Our optimality proof later will be based on the language of populations. Naturally, we define $\mathcal{G}(\mathcal{P}) := \mathcal{G}(\mathcal{T}(\mathcal{P}))$ for $\mathcal{P} \in \widehat{\mathbb{P}}$. Now, let us reformulate our error estimator estimates in terms of populations.

Due to the one-to-one correspondence of $\mathcal{S}(\mathcal{T}(\mathcal{P}))$ and $\mathcal{P}^{++} \setminus \mathcal{P}$, see (2.4), we set for $\mathcal{U} \subset \mathcal{P}^{++} \setminus \mathcal{P}$

$$\mathcal{E}^2_{\mathcal{P}}(\mathcal{U}) := \mathcal{E}^2_{\mathcal{T}(\mathcal{P})}(\mathsf{midpts}^{-1}(\mathcal{U})).$$

This allows us to rewrite Proposition 4.4 as follows.

Corollary 4.7. *Let $\mathcal{P}, \mathcal{P}_* \in \mathbb{P}$ with $\mathcal{P} \le \mathcal{P}_*$. Then we have*

$$\mathcal{G}(\mathcal{P}) - \mathcal{G}(\mathcal{P}_*) \eqsim \mathcal{E}^2_{\mathcal{P}}\big(\mathcal{P}_* \cap (\mathcal{P}^{++} \setminus \mathcal{P})\big).$$

Remark 4.8 (Upper diamond estimate)**.** *Let $(\mathcal{T}^\wedge, \mathcal{T}_\vee; \mathcal{T}_1, \ldots, \mathcal{T}_m)$ be an upper diamond in $\mathbb{T}$. Since the requirement of the areas of refinement $\Omega(\mathcal{T}^\wedge \setminus \mathcal{T}_j)$ being mutually disjoint is equivalent to the requirement that the sets $\mathcal{T}^\wedge \setminus \mathcal{T}_j$, or the sets $\mathcal{S}(\mathcal{T}^\wedge) \setminus \mathcal{S}(\mathcal{T}_j)$ being mutually disjoint, from*

Proposition 4.4 we obtain that

$$\mathcal{G}(\mathcal{T}^\wedge) - \mathcal{G}(\mathcal{T}_\vee) \eqsim \mathcal{E}^2_{\mathcal{T}^\wedge}(\mathcal{S}(\mathcal{T}^\wedge) \setminus \mathcal{S}(\mathcal{T}_\vee)) = \sum_{j=1}^m \mathcal{E}^2_{\mathcal{T}^\wedge}(\mathcal{S}(\mathcal{T}^\wedge) \setminus \mathcal{S}(\mathcal{T}_j))$$
$$\eqsim \sum_{j=1}^m \big(\mathcal{G}(\mathcal{T}^\wedge) - \mathcal{G}(\mathcal{T}_j)\big).$$

5. The adaptive finite element method (AFEM)

According to [BR78], the maximum marking strategy, marks sides for refinement that correspond to squared local error indicators that are not less than some constant multiple $\mu \in (0,1]$ of the maximum squared local error indicator.

In view of the fact that, in order to refine a side, generally more sides have to be bisected to retain conformity of the triangulation, we consider the following *modified maximum marking strategy*: First we determine a side S such that the sum $\bar{\mathcal{E}}^2$ of *all* local squared error indicators of the sides that have to be bisected in order to refine S is maximal. Then, in some arbitrary order, running over the sides in the triangulation, we mark those sides $\tilde{S}$ for refinement for which the sum of all squared local error indicators that correspond to the sides that have to be bisected in order to refine $\tilde{S}$, but that do not have to be bisected for the refinement of sides that are marked earlier, is not less than $\mu\bar{\mathcal{E}}^2$.

To give a formal description, for $\mathcal{T} \in \mathbb{T}$ and $S \in \mathcal{S}(\mathcal{T})$, let

$$\text{ref'd}(\mathcal{T}; S) := \mathcal{S}(\mathcal{T}) \setminus \mathcal{S}(\text{Refine}(\mathcal{T}; S)),$$

being the subset of sides in $\mathcal{S}(\mathcal{T})$ that are bisected when passing to the smallest refinement (in $\mathbb{T}$) of $\mathcal{T}$ in which S has been bisected. Then the adaptive finite element method reads as follows:

Algorithm 5.1 (AFEM)**.** *Fix $\mu \in (0,1]$ and set $\mathcal{T}_0 := \mathcal{T}_\perp$ and $k = 0$. The adaptive loop is an iteration of the following steps:*

(1) SOLVE: *compute* $u_{\mathcal{T}_k} \in \mathbb{V}_0(\mathcal{T}_k)$;

(2) ESTIMATE: *compute* $\{\mathcal{E}^2_{\mathcal{T}_k}(S) : S \in \mathcal{S}(\mathcal{T}_k)\}$;

(3) MARK: $\bar{\mathcal{E}}^2_{\mathcal{T}_k} := \max\{\mathcal{E}^2_{\mathcal{T}_k}(\text{ref'd}(\mathcal{T}_k; S)) : S \in \mathcal{S}(\mathcal{T}_k)\}$;

$\mathcal{M}_k := \emptyset$; $\mathcal{C}_k := \mathcal{S}(\mathcal{T}_k)$; $\widetilde{\mathcal{M}}_k := \emptyset$;

while $\mathcal{C}_k \neq \emptyset$ *do*

select $S \in \mathcal{C}_k$;

if $\mathcal{E}^2_{\mathcal{T}_k}(\text{ref'd}(\mathcal{T}_k;S)\setminus\widetilde{\mathcal{M}}_k)\geq\mu\bar{\mathcal{E}}^2_{\mathcal{T}_k}$;

then $\mathcal{M}_k:=\mathcal{M}_k\cup\{S\}$;

$\widetilde{\mathcal{M}}_k:=\widetilde{\mathcal{M}}_k\cup\text{ref'd}(\mathcal{T}_k;S)$;

end if;

$\mathcal{C}_k:=\mathcal{C}_k\setminus\text{ref'd}(\mathcal{T}_k;S)$;

end while;

(4) REFINE: compute $\mathcal{T}_{k+1}=\text{Refine}(\mathcal{T}_k;\mathcal{M}_k)$ and increment k.

If we define $\mathcal{P}_k:=\mathcal{P}(\mathcal{T}_k)$, then we can rewrite our algorithm also in the language of populations:

(1) SOLVE: compute $u_{\mathcal{T}_k}\in\mathbb{V}_0(\mathcal{T}_k)$;

(2) ESTIMATE: compute $\{\mathcal{E}^2_{\mathcal{P}_k}(P):P\in\mathcal{P}_k^{++}\setminus\mathcal{P}_k\}$;

(3) MARK: $\bar{\mathcal{E}}^2_{\mathcal{P}_k}:=\max\{\mathcal{E}^2_{\mathcal{P}_k}((\mathcal{P}_k\oplus P)\setminus\mathcal{P}_k)\ :\ P\in\mathcal{P}_k^{++}\setminus\mathcal{P}_k)\}$;

$\mathcal{M}_k:=\emptyset$; $\mathcal{C}_k:=\mathcal{P}_k^{++}\setminus\mathcal{P}_k$; $\widetilde{\mathcal{M}}_k:=\emptyset$;

while $\mathcal{C}_k\neq\emptyset$ *do*

select $P\in\mathcal{C}_k$;

if $\mathcal{E}^2_{\mathcal{P}_k}((\mathcal{P}_k\oplus P)\setminus(\mathcal{P}_k\cup\widetilde{\mathcal{M}}_k))\geq\mu\bar{\mathcal{E}}^2_{\mathcal{P}_k}$;

then $\mathcal{M}_k:=\mathcal{M}_k\cup\{P\}$;

$\widetilde{\mathcal{M}}_k:=\widetilde{\mathcal{M}}_k\cup((\mathcal{P}_k\oplus P)\setminus\mathcal{P}_k)$;

end if;

$\mathcal{C}_k:=\mathcal{C}_k\setminus\big((\mathcal{P}_k\oplus P)\setminus\mathcal{P}_k\big)$;

end while;

(4) REFINE: $\mathcal{P}_{k+1}:=\mathcal{P}_k\oplus\mathcal{M}_k\ \big[=\mathcal{P}_k\cup\widetilde{\mathcal{M}}_k\big]$ and increment k.

Proposition 5.1. *For the sequences $(\mathcal{P}_k)_{k\in\mathbb{N}_0}$ and $(\mathcal{M}_k)_{k\in\mathbb{N}_0}$ produced by Algorithm 5.1 (second formulation), we have $\mathcal{M}_k\neq\emptyset$ and*

$$\mathcal{E}^2_{\mathcal{P}_k}\big(\mathcal{P}_{k+1}\cap(\mathcal{P}_k^{++}\setminus\mathcal{P}_k)\big)=\mathcal{E}^2_{\mathcal{P}_k}\big((\mathcal{P}_k\oplus\mathcal{M}_k)\setminus\mathcal{P}_k\big)\geq\mu\,\#\mathcal{M}_k\,\bar{\mathcal{E}}^2_{\mathcal{P}_k}.$$

Proof. Consider the while-loop in MARK. As long as $\mathcal{M}_k=\emptyset$, we have $\widetilde{\mathcal{M}}_k=\emptyset$. Thus for every $P\in\mathcal{P}_k^{++}\setminus\mathcal{P}_k$ that has been considered and all $P'\in(\mathcal{P}_k\oplus P)\setminus\mathcal{P}_k$, we conclude $\mathcal{E}^2_{\mathcal{P}_k}((\mathcal{P}_k\oplus P')\setminus\mathcal{P}_k)\leq\mathcal{E}^2_{\mathcal{P}_k}((\mathcal{P}_k\oplus P)\setminus\mathcal{P}_k)<\mu\bar{\mathcal{E}}^2_{\mathcal{P}_k}$. Hence assuming that $\mathcal{M}_k$ remains empty, at some moment

$P \in \mathcal{P}_k^{++} \setminus \mathcal{P}_k$ is encountered with $\mathcal{E}^2_{\mathcal{P}_k}((\mathcal{P}_k \oplus P) \setminus (\mathcal{P}_k \cup \widetilde{\mathcal{M}}_k)) = \bar{\mathcal{E}}^2_{\mathcal{P}_k}$, which yields a contradiction. Therefore, after termination of the while-loop in MARK, we have $\mathcal{M}_k \neq \emptyset$.

Each time a P is added to $\mathcal{M}_k$, the quantity $\mathcal{E}^2_{\mathcal{P}_k}((\mathcal{P}_k \oplus \mathcal{M}_k) \setminus \mathcal{P}_k)$ increased by at least $\mu \bar{\mathcal{E}}^2_{\mathcal{P}_k}$, which shows $\mathcal{E}^2_{\mathcal{P}_k}\big((\mathcal{P}_k \oplus \mathcal{M}_k) \setminus \mathcal{P}_k\big) \geq \mu \, \#\mathcal{M}_k \, \bar{\mathcal{E}}^2_{\mathcal{P}_k}$.

Since $\mathcal{M}_k \subset \mathcal{P}_k^{++}$ and $\mathcal{P}_{k+1} = \mathcal{P}_k \oplus \mathcal{M}_k$, we have $\mathcal{P}_{k+1} \wedge \mathcal{P}_k^{++} = \mathcal{P}_k \oplus \mathcal{M}_k$. Thus

$$\mathcal{P}_{k+1} \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k) = (\mathcal{P}_{k+1} \wedge \mathcal{P}_k^{++}) \setminus \mathcal{P}_k = (\mathcal{P}_k \oplus \mathcal{M}_k) \setminus \mathcal{P}_k = \widetilde{\mathcal{M}}_k,$$

which concludes the proof. □

Remark 5.2. *Generally, the set of marked edges or persons determined in MARK depend on the ordering in which the sets $\mathcal{S}(\mathcal{T}_k)$ or $\mathcal{P}_k^{++} \setminus \mathcal{P}_k$ are traversed. In particular, the marking strategy does not necessarily mark the $S \in \mathcal{S}(\mathcal{T})$ with $\mathcal{E}^2_{\mathcal{T}_k}(\mathrm{ref'd}(\mathcal{T}_k; S)) = \bar{\mathcal{E}}_{\mathcal{T}_k}$ for refinement.*

Remark 5.3. *It is not difficult to see that ESTIMATE can be implemented in $\mathcal{O}(\#\mathcal{P}_k)$ operations. Also REFINE can be implemented in $\mathcal{O}(\#\mathcal{P}_k)$ operations, once $\mathcal{M}_k$ is determined by MARK; recall $\mathcal{P}_{k+1} = \mathcal{P}_k \oplus \mathcal{M}_k = \mathcal{P}_k \cup \widetilde{\mathcal{M}}_k$. In Appendix A it is demonstrated that the same holds true for a slightly modified MARK, which yields a set $\mathcal{M}_k$, which has qualitatively the same properties as given in Proposition 5.1. Consequently, our instance optimality result still holds with this modified marking.*

6. Fine properties of populations

Before we get to our optimality proof, we need some fine properties of populations.

6.1. **Ancestors, descendants and free elements.**

For $P \in \mathcal{P}^\top$ we define its *set of ancestors* $\texttt{anc}(P)$ as follows: If $\texttt{gen}(P) = 0$, then $\texttt{anc}(P) := \emptyset$. For $\texttt{gen}(P) \geq 1$, we define inductively

$$\texttt{anc}(P) := \texttt{parent}(P) \cup \bigcup_{Q \in \texttt{parent}(P)} \texttt{anc}(Q).$$

Moreover, we denote the set of the *descendants* of P by

$$\texttt{desc}(P) := \{P' \in \mathcal{P}^\top \, : \, P \in \texttt{anc}(P')\}.$$

As a shorthand notation, we write $P' \lhd P$ or $P \rhd P'$, when $P \in \texttt{child}(P')$, or equivalently, $P' \in \texttt{parent}(P)$; and $P' \lll P$ or $P \ggg P'$, when $P \in \texttt{desc}(P')$, or equivalently, $P' \in \texttt{anc}(P)$.

For $\mathcal{U} \subset \mathcal{P}^\top$ we define

$$\texttt{child}(\mathcal{U}) := \bigcup_{P\in\mathcal{U}} \texttt{child}(P), \qquad \texttt{parent}(\mathcal{U}) := \bigcup_{P\in\mathcal{U}} \texttt{parent}(P),$$
$$\texttt{anc}(\mathcal{U}) := \bigcup_{P\in\mathcal{U}} \texttt{anc}(P), \qquad \texttt{desc}(\mathcal{U}) := \bigcup_{P\in\mathcal{U}} \texttt{desc}(P).$$

If $P, P' \in \mathcal{P}^\top$, with $P \neq P'$, have a joint child, then we call them *partners*, and write $P \infty P'$.

For $k \in \mathbb{N}_0$, we define

$$\texttt{gen}^{-1}(k) := \{P \in \mathcal{P}^\top : \texttt{gen}(P) = k\}.$$

The following lemma summarises some apparent basic properties without proof.

Lemma 6.1. *For $\mathcal{P}, \mathcal{P}_* \in \mathbb{P}$, we have*

(a) $\texttt{anc}(\mathcal{P}) \subset \mathcal{P}$;
(b) *if* $\mathcal{P} \leq \mathcal{P}_*$, *then* $\texttt{anc}(\mathcal{P}) \cap (\mathcal{P}_* \setminus \mathcal{P}) = \emptyset$;
(c) *if* $\mathcal{P} \leq \mathcal{P}_*$, *then* $\texttt{desc}(\mathcal{P}_* \setminus \mathcal{P}) \cap \mathcal{P} = \emptyset$;
(d) *if* $\mathcal{C} \subset \mathcal{P}^\top$, *then* $\mathcal{P} \oplus \mathcal{C} = \mathcal{P} \cup \texttt{anc}(\mathcal{C}) \cup \mathcal{C}$.

So far we have stated only very general properties of populations. However, populations correspond to conforming triangulations created by newest vertex bisection of the initial triangulation $\mathcal{T}_\bot$. In the following we shall exploit the structures inherited by this fact in order to prove much stronger results.

The following lemma shows that the number of ancestors of the same generation is for every person bounded by a uniform constant. For apparent reasons we call this property *limited genetic diversity (LGD).*

Proposition 6.2. *We have*

$$\sup_{P\in\mathcal{P}^\top} \sup_{k\in\mathbb{N}} \#\big(\texttt{anc}(P) \cap \texttt{gen}^{-1}(k)\big) =: c_{\mathrm{GD}} < \infty.$$

The scalar c_{GD} is called the genetic diversity constant.

Proof. Thanks to the refinement by bisection, for $T \in \mathfrak{T}$, we have $|T| \eqsim 2^{-\texttt{gen}(T)}$. Consequently, by the uniform shape regularity of $\mathfrak{T}$, for $P' \in \mathcal{P}^\top$ and $P \in \texttt{child}(P')$, we have that $\operatorname{dist}(P', P) \eqsim 2^{-\texttt{gen}(P')/2}$. Applying a geometrical series argument, we thus infer that for $P' \in \texttt{anc}(P) \cap \texttt{gen}^{-1}(k)$, $\operatorname{dist}(P', P) \eqsim 2^{-k/2}$.

Again by the uniform shape regularity of $\mathfrak{T}$, any ball of radius $2^{-k/2}$ contains at most an uniformly bounded number of vertices of the uniform refinement of $\mathcal{T}_\bot$ with triangles of generation k. This completes the proof. ∎

The following property of the newest vertex bisection is even more peculiar:

Proposition 6.3. *Let $P_1, P_2 \in \mathcal{P}^\top$ be partners with* $\texttt{gen}(P_1) \geq 2$*. Then P_1 and P_2 have a joint parent.*

Proof. Let $k := \text{gen}(P_1)$ $(= \text{gen}(P_2))$, and let P be a child of P_1 and P_2. A patch of the coarsest triangulation in $\mathbb{T}$ that contains P looks as indicated in Figure 5. Here, and in the following figures, the arrows

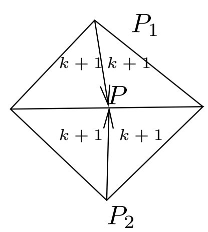

FIGURE 5. Coarsest triangulation that contains P

indicate the parent-child relationships, and the numbers indicate the generations of the triangles.

The two possible patches (up to symmetries) of coarsest triangulations in $\mathbb{T}$ that contain P_1 and P_2 look as indicated in Figure 6. In the

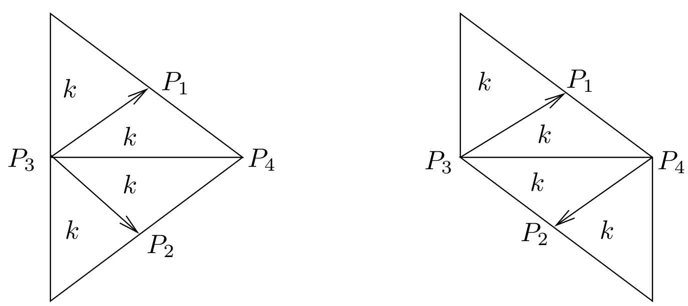

FIGURE 6. Two possible coarsest triangulations that contain P_1 and P_2

left picture, P_1 and P_2 have a joint parent P_3.

In the right picture, P_1 has parent P_3, and P_2 has parent P_4. Since P_3 and P_4 are vertices of a joint $T \in \mathfrak{T}$ and have the same generation, their generation must be zero, i.e., $\text{gen}(P_1) = 1$. □

The next lemma shows that any two ancestors of some person that have the same (non-zero) generation are linked via a sequence of partners.

Lemma 6.4. *Let $P \in \mathcal{P}^\top$ and $\bar{P}, \tilde{P} \in \mathtt{anc}(P)$, $\tilde{P} \neq \bar{P}$, with $\mathtt{gen}(\tilde{P}) = \mathtt{gen}(\bar{P}) \geq 1$. Then, for some $1 \leq m \leq \mathtt{gen}(P) - \mathtt{gen}(\bar{P})$, there exist $P_0, \dots, P_m \in \mathtt{anc}(P)$ such that $P_0 \infty \cdots \infty P_m$ and $P_0 = \bar{P}$ and $P_m = \tilde{P}$. In particular, $\mathtt{child}(P_{i-1}) \cap \mathtt{child}(P_i) \cap (\mathtt{anc}(P) \cup \{P\}) \neq \emptyset$, $i = 1, \dots, m$.*

Proof. Fix $P \in \mathcal{P}^\top$. We prove the claim by induction over $k := \mathtt{gen}(P) - \mathtt{gen}(\bar{P})$. If $k = 1$, then $\bar{P} \neq \tilde{P}$ have the joint child P, and hence $\mathtt{child}(\bar{P}) \cap \mathtt{child}(\tilde{P}) \cap \{P\} \neq \emptyset$.

Now let $k \geq 2$, and assume that the claim is already true for $k-1$. Let $\bar{P}' \in \mathtt{child}(\bar{P}) \cap \mathtt{anc}(P)$ and $\tilde{P}' \in \mathtt{child}(\tilde{P}) \cap \mathtt{anc}(P)$. If $\bar{P}' = \tilde{P}'$, then $\mathtt{child}(\bar{P}') \cap \mathtt{child}(\tilde{P}') \cap \mathtt{anc}(P)) \neq \emptyset$.

Otherwise, by induction for some $1 \leq m-1 \leq k-1$, there exist $\bar{P}_0, \dots, \bar{P}_{m-1} \in \mathtt{anc}(P)$ such that $\bar{P}_0 \infty \cdots \infty \bar{P}_{m-1}$ and $\bar{P}_0 = \bar{P}'$ and $\bar{P}_{m-1} = \tilde{P}'$.

Because of $\mathtt{gen}(\bar{P}_i) = \mathtt{gen}(\bar{P}) + 1 \geq 2$, by Proposition 6.3 there exist $P_1, \dots, P_{m-1}$ such that for $i = 1, \dots, m-1$, P_i is a parent of $\bar{P}_{i-1}$ and $\bar{P}_i$; in particular $P_i \in \mathtt{anc}(P)$.

Setting, $P_0 := \bar{P}$ and $P_m := \tilde{P}$, we have found a sequence in $\mathtt{anc}(P)$ such that subsequent persons have a joint child in $\mathtt{anc}(P)$. By removing possibly subsequent equal persons, we have found a sequence with the required properties. ☐

Definition 6.5. *We say that a set $\mathcal{U} \subset \mathcal{P}^\top$ is* descendant-free *when $\mathtt{desc}(\mathcal{U}) \cap \mathcal{U} = \emptyset$.*

The next proposition generalizes upon Proposition 6.2.

Proposition 6.6. *Let $P \in \mathcal{P}^\top$ and $\mathcal{U} \subset \mathtt{anc}(P) \setminus \mathcal{P}_\perp$. If $\mathcal{U}$ is descendant-free, then $\#\mathcal{U} \leq c_{\mathrm{GD}}$.*

Proof. Since $\#\mathtt{anc}(P) < \infty$, there are at most finitely many descendant-free subsets of $\mathtt{anc}(P) \setminus \mathcal{P}_\perp$. Among them, let $\mathcal{U}$ denote the one that first maximizes $\#\mathcal{U}$ and then $\sum_{Q \in \mathcal{U}} \mathtt{gen}(Q)$. We shall show that this implies that all persons in $\mathcal{U}$ are of the same generation. Therefore, by Proposition 6.2, we conclude the claim $\#\mathcal{U} \leq c_{\mathrm{GD}}$.

Let $\mathtt{gen}(P) \geq 2$, so that $\mathcal{U} \neq \emptyset$. Define $k := \min\{\mathtt{gen}(P') : P' \in \mathcal{U}\}$. In order to show that $\mathcal{U} \subset \mathtt{gen}^{-1}(k)$, we proceed by contradiction and assume that $\mathcal{U} \not\subset \mathtt{gen}^{-1}(k)$, i.e., there exists a $Q \in \mathcal{U}$ with $\mathtt{gen}(Q) > k$, and so $\mathtt{gen}(P) > k + 1$. Since $\mathcal{U}$ is descendent-free, there exists a $\tilde{P} \in (\mathtt{anc}(Q) \cap \mathtt{anc}(P) \cap \mathtt{gen}^{-1}(k)) \setminus \mathcal{U}$.

By definition of k there exists $\bar{P} \in \mathcal{U}$ with $\mathtt{gen}(\bar{P}) = \mathtt{gen}(\tilde{P}) = k$. Due to Lemma 6.4 we find a finite sequence of partners in $\mathtt{anc}(P)$, starting with $\bar{P}$ and ending with $\tilde{P}$, where each couple has a common child in $\texttt{anc}(P)$. Since $\tilde{P} \notin \mathcal{U}$, we can select from this sequence a couple $\bar{P}'\infty\tilde{P}'$, with $\bar{P}' \in \mathcal{U}$, $\tilde{P}' \notin \mathcal{U}$, $\texttt{gen}(\tilde{P}') = \texttt{gen}(\bar{P}') = k$, and that has a common child $\hat{P}' \in \texttt{anc}(P)$.

On the one hand, since $\mathcal{U}$ is descendant-free, $\bar{P}' \in \mathcal{U}$ and $\hat{P}' \rhd \bar{P}'$, we conclude that $\texttt{desc}(\hat{P}') \cap \mathcal{U} \subset \texttt{desc}(\bar{P}') \cap \mathcal{U} = \emptyset$. On the other hand, thanks to the definition of k and $\tilde{P}' \notin \mathcal{U}$, we have that $\bar{P}'$ is the only ancestor of $\hat{P}'$ in $\mathcal{U}$. In other words, $\mathcal{U}' := (\mathcal{U} \setminus \{\bar{P}'\}) \cup \{\hat{P}'\}$ is a descendent-free subset of $\texttt{anc}(P)$. Since $\#\mathcal{U}' = \#\mathcal{U}$ and $\sum_{Q\in\mathcal{U}'} \texttt{gen}(Q) = 1 + \sum_{Q\in\mathcal{U}} \texttt{gen}(Q)$, this is the desired contradiction. □

Definition 6.7. *Let $\mathcal{U} \subset \mathcal{P}^\top \setminus \mathcal{P}_\perp$. We call the subset*

$$\texttt{free}(\mathcal{U}) := \{P \in \mathcal{U} \colon \texttt{desc}(P) \cap \mathcal{U} = \emptyset\}.$$

the set of free persons *in $\mathcal{U}$.*

The following lemma collects some basic properties of free subsets.

Lemma 6.8. *Let $\mathcal{U} \subset \mathcal{P}^\top \setminus \mathcal{P}_\perp$.*

(a) *The set $\texttt{free}(\mathcal{U})$ is descendant-free.*
(b) *If $\mathcal{U}$ is descendant-free, then $\texttt{free}(\mathcal{U}) = \mathcal{U}$.*
(c) *If $\#\mathcal{U} < \infty$, then $\mathcal{U} \cup \texttt{anc}(\mathcal{U}) = \texttt{free}(\mathcal{U}) \cup \texttt{anc}(\texttt{free}(\mathcal{U}))$.*
(d) *If $\#\mathcal{U} < \infty$ and $\mathcal{P} \in \mathbb{P}$, then $\mathcal{P} \oplus \mathcal{U} = \mathcal{P} \oplus \texttt{free}(\mathcal{U})$.*
(e) *If $\#\mathcal{U} < \infty$ and $\mathcal{U} \neq \emptyset$, then $\texttt{free}(\mathcal{U}) \neq \emptyset$.*

Proof. (a): Let $P \in \texttt{free}(\mathcal{U})$, then by definition we have $\texttt{desc}(P) \cap \mathcal{U} = \emptyset$. Since $\texttt{free}(\mathcal{U}) \subset \mathcal{U}$, we conclude that $\texttt{desc}(P) \cap \texttt{free}(\mathcal{U}) = \emptyset$, i.e., $\texttt{free}(\mathcal{U})$ is descendant-free.

(b): The claim follows directly from the assumption $\texttt{desc}(\mathcal{U}) \cap \mathcal{U} = \emptyset$.

(c): Obviously, it is sufficient to prove $\mathcal{U} \cup \texttt{anc}(\mathcal{U}) \subset \texttt{free}(\mathcal{U}) \cup \texttt{anc}(\texttt{free}(\mathcal{U}))$. Let $P \in \mathcal{U}$. If $P \in \texttt{free}(\mathcal{U})$, then $P \cup \texttt{anc}(P) \subset \texttt{free}(\mathcal{U}) \cup \texttt{anc}(\texttt{free}(\mathcal{U}))$. Otherwise, if $P \notin \texttt{free}(\mathcal{U})$, then pick a $P' \in \mathcal{U} \cap \texttt{desc}(P)$. If $P' \notin \texttt{free}(\mathcal{U})$, then, because $\#\mathcal{U} < \infty$, by continuing this process, after finitely many steps a descendant P'' of P', and thus of P, is found, which is in $\texttt{free}(\mathcal{U})$. We conclude that $P \cup \texttt{anc}(P) \subset \texttt{anc}(P'') \subset \texttt{anc}(\texttt{free}(\mathcal{U}))$, which finishes the proof.

(d): By (c) and Lemma 6.1 (d), we have $\mathcal{P} \oplus \mathcal{U} = \mathcal{P} \cup \mathcal{U} \cup \texttt{anc}(\mathcal{U}) = \mathcal{P} \cup \texttt{free}(\mathcal{U}) \cup \texttt{anc}(\texttt{free}(\mathcal{U})) = \mathcal{P} \oplus \texttt{free}(\mathcal{U})$.

(e): Let $\#\mathcal{U} < \infty$ with $\mathcal{U} \neq \emptyset$. Then $\texttt{free}(\mathcal{U}) = \emptyset$ together with (c) implies $\mathcal{U} = \emptyset$, which contradicts $\mathcal{U} \neq \emptyset$. □

The following lemma states that removing free persons from a population results in a (smaller) population.

Lemma 6.9. *Let $\mathcal{P}_*, \mathcal{P} \in \mathbb{P}$ Then,*

(a) *for* $\mathcal{C} \subset \mathtt{free}(\mathcal{P}_* \setminus \mathcal{P}_\perp)$, *we have* $\mathcal{P}_* \ominus \mathcal{C} = \mathcal{P}_* \setminus \mathcal{C}$;
(b) $\mathtt{free}(\mathcal{P}_* \setminus \mathcal{P}) \subset \mathtt{free}(\mathcal{P}_* \setminus \mathcal{P}_\perp)$.

Proof. (a) By assumption, $\mathcal{P}_* \setminus \mathcal{C} \in \mathbb{P}$ and thus $\mathcal{P}_* \ominus \mathcal{C} = \mathcal{P}_* \setminus \mathcal{C}$.

(b) By $\mathtt{free}(\mathcal{P}_* \setminus \mathcal{P}) \subset \mathcal{P}_* \setminus \mathcal{P}$, we have $\mathtt{free}(\mathcal{P}_* \setminus \mathcal{P}) \cap \mathcal{P} = \emptyset$. Since $\mathcal{P} \in \mathbb{P}$, the latter shows that $\mathtt{desc}(\mathtt{free}(\mathcal{P}_* \setminus \mathcal{P})) \cap \mathcal{P} = \emptyset$. Finally, from from $\mathtt{desc}(\mathtt{free}(\mathcal{P}_* \setminus \mathcal{P})) \cap (\mathcal{P}_* \setminus \mathcal{P}) = \emptyset$, we conclude that $\mathtt{desc}(\mathtt{free}(\mathcal{P}_* \setminus \mathcal{P})) \cap \mathcal{P}_* = \emptyset$. □

For $\mathcal{P} \in \mathbb{P}$, the set $\mathtt{free}(\mathcal{P} \setminus \mathcal{P}_\perp)$ are the nodes of $\mathcal{T}(\mathcal{P})$ that are "free" in the sense that they can be removed while retaining a conforming triangulation, i.e., a triangulation in $\mathbb{T}$. Remarkably, as follows from the following proposition, the number of free nodes in any triangulation in $\mathbb{T}$ cannot be reduced by more than some constant factor by whatever further refinement in $\mathbb{T}$. This proposition plays a crucial role in the optimality proof in Section 7.

Theorem 6.10. *Let* $\mathcal{U} \subset \mathcal{V} \subset \mathcal{P}^\top \setminus \mathcal{P}_\perp$ *with* $\#\mathcal{V} < \infty$. *Then*

$$\#\mathtt{free}(\mathcal{U}) \le c_{\mathrm{GD}}\, \#\mathtt{free}(\mathcal{V}).$$

Proof. It follows from $\mathtt{free}(\mathcal{U}) \subset \mathcal{U} \subset \mathcal{V}$ and Lemma 6.8 (c), applied to $\mathcal{V}$, that

$$\mathtt{free}(\mathcal{U}) \subset \mathtt{free}(\mathcal{V}) \cup \mathtt{anc}(\mathtt{free}(\mathcal{V})).$$

Thus we can write $\mathtt{free}(\mathcal{U})$ as

$$\mathtt{free}(\mathcal{U}) = \bigcup_{P \in \mathtt{free}(\mathcal{V})} \underbrace{\Big(\big(\{P\} \cup \mathtt{anc}(P)\big) \cap \mathtt{free}(\mathcal{U})\Big)}_{=:\mathcal{V}_P}.$$

Now the claim follows, when $\#\mathcal{V}_P \le c_{\mathrm{GD}}$ for all $P \in \mathtt{free}(\mathcal{V})$. To this end, let $P \in \mathtt{free}(\mathcal{V})$.

Assume first that $P \in \mathtt{free}(\mathcal{U})$. Since $\mathtt{free}(\mathcal{U})$ is descendent-free we have $\mathtt{anc}(P) \cap \mathtt{free}(\mathcal{U}) = \emptyset$. Thus $\mathcal{V}_P = \{P\}$ and $\#\mathcal{V}_P = 1 \le c_{\mathrm{GD}}$.

Now assume $P \notin \mathtt{free}(\mathcal{U})$. Then $\mathcal{V}_P = \mathtt{anc}(P) \cap \mathtt{free}(\mathcal{U})$. Since $\mathtt{free}(\mathcal{U})$ is descendant-free, the subset $\mathcal{V}_P$ of $\mathtt{anc}(P)$ is descendant-free, and it follows by Proposition 6.6 that $\#\mathcal{V}_P \le c_{\mathrm{GD}}$. □

6.2. **Populations and the lower diamond estimate.** In this subsection we shall translate the lower diamond estimate of Section 3 to the setting of populations. We start with the definition of a lower diamond.

Definition 6.11. *For* $\{\mathcal{P}_1, \ldots, \mathcal{P}_m\} \subset \mathbb{P}$, *we call* $(\mathcal{P}^\wedge, \mathcal{P}_\vee; \mathcal{P}_1, \ldots, \mathcal{P}_m)$ *a* lower diamond *in* $\mathbb{P}$ *of size* m, *when* $\mathcal{P}^\wedge = \bigwedge_{j=1}^m \mathcal{P}_j$, $\mathcal{P}_\vee = \bigvee_{j=1}^m \mathcal{P}_j$, *and the sets* $\mathcal{P}_\vee \setminus \mathcal{P}_j$ *are mutually disjoint.*

As we shall see in Corollary 6.14 below, Definition 6.11 in terms of populations is consistent with Definition 3.1 in terms of triangulations. In particular, all results of Section 3 dealing with lower diamonds transfer to populations.

Recall from Subsection 2.2 the definition of $\Omega(P)$ for $P \in \mathcal{P}^\top \setminus \mathcal{P}_\bot$.

Lemma 6.12. *Let $P_1, P_2 \in \mathcal{P}^\top \setminus \mathcal{P}_\bot$ with $\Omega(P_1) \cap \Omega(P_2) \neq \emptyset$. Then either $P_1 = P_2$ or $P_1 \lll P_2$ or $P_2 \lll P_1$.*

Proof. If $\mathtt{gen}(P_1) = \mathtt{gen}(P_2)$, then the claim follows by Proposition 2.3 (a). W.l.o.g. assume now that $\mathtt{gen}(P_1) < \mathtt{gen}(P_2)$. By (a repeated) application of Proposition 2.3 (b), we have

$$\overline{\Omega(P_2)} \subset \overline{\Omega(\mathtt{anc}(P_2) \cap \mathtt{gen}^{-1}(\mathtt{gen}(P_1)))}.$$

Thus $\Omega(P_1) \cap \Omega(\mathtt{anc}(P_2) \cap \mathtt{gen}^{-1}(\mathtt{gen}(P_1)) \neq \emptyset$, and therefore there exists a $P_3 \lll P_2$ with $\mathtt{gen}(P_3) = \mathtt{gen}(P_1)$ and $\Omega(P_1) \cap \Omega(P_3) \neq \emptyset$. Finally, Proposition 2.3 (a) implies $P_1 = P_3$, i.e., $P_1 \lll P_2$. □

Lemma 6.13. *Let $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_* \in \mathbb{P}$ with $\mathcal{P}_1, \mathcal{P}_2 \leq \mathcal{P}_*$. Define $\mathcal{R}_j := \mathcal{P}_* \setminus \mathcal{P}_j$ for $j = 1, 2$. Then $\mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset$ if and only if $\Omega(\mathcal{R}_1) \cap \Omega(\mathcal{R}_2) = \emptyset$.*

Proof. Assume that $\mathcal{R}_1 \cap \mathcal{R}_2 \neq \emptyset$. Then for $P \in \mathcal{R}_1 \cap \mathcal{R}_2$ we obviously have $\emptyset \neq \Omega(P) \subset \Omega(\mathcal{R}_1) \cap \Omega(\mathcal{R}_2)$.

Assume now that $\Omega(\mathcal{R}_1) \cap \Omega(\mathcal{R}_2) \neq \emptyset$. Therefore, there exists $P_1 \in \mathcal{R}_1$ and $P_2 \in \mathcal{R}_2$ with $\Omega(P_1) \cap \Omega(P_2) \neq \emptyset$. It follows from Lemma 6.12 that either $P_1 = P_2$, and thus $\mathcal{R}_1 \cap \mathcal{R}_2 \neq \emptyset$, or $P_1 \lll P_2$ or $P_2 \lll P_1$. In the case $P_1 \lll P_2$ we obtain from $P_1 \in \mathcal{R}_1$ that $P_1 \notin \mathcal{P}_1$, and thus $P_2 \notin \mathcal{P}_1$ since $\mathcal{P}_1 \in \mathbb{P}$. Consequently, we have $P_2 \in \mathcal{R}_1 \cap \mathcal{R}_2$. The same argument shows that $\mathcal{R}_1 \cap \mathcal{R}_2 \neq \emptyset$ when $P_2 \lll P_1$. □

For $\mathcal{P}, \mathcal{P}_* \in \mathbb{P}$, we have $\Omega(\mathcal{P}_* \setminus \mathcal{P}) = \Omega(\mathcal{T}(\mathcal{P}_*) \setminus \mathcal{T}(\mathcal{P}))$. Hence, as an immediate consequence we obtain the following result.

Corollary 6.14. *$(\mathcal{P}^\wedge, \mathcal{P}_\vee; \mathcal{P}_1, \dots, \mathcal{P}_m)$ is a lower diamond in $\mathbb{P}$ if and only if $(\mathcal{T}(\mathcal{P}^\wedge), \mathcal{T}(\mathcal{P}_\vee); \mathcal{T}(\mathcal{P}_1), \dots, \mathcal{T}(\mathcal{P}_m))$ is a lower diamond in $\mathbb{T}$.*

This allows us to reformulate the lower diamond estimate in terms of populations. In particular, Corollary 4.5 reads as:

Corollary 6.15 (Lower diamond estimate)**.** *Let $(\mathcal{P}^\wedge, \mathcal{P}_\vee; \mathcal{P}_1, \dots, \mathcal{P}_m)$ be a lower diamond in $\mathbb{P}$, then*

$$\mathcal{G}(\mathcal{P}^\wedge) - \mathcal{G}(\mathcal{P}_\vee) \eqsim \sum_{j=1}^{m} \big(\mathcal{G}(\mathcal{P}_j) - \mathcal{G}(\mathcal{P}_\vee)\big).$$

7. Energy optimality and instance optimality

For each $m \in \mathbb{N}_0$, we define the minimal energy level of populations with not more than $\#P_\perp + m$ persons by

$$\mathcal{G}_m^{\text{opt}} = \min\{\mathcal{G}(\mathcal{P}) \,:\, \mathcal{P} \in \mathbb{P},\, \#(\mathcal{P} \setminus \mathcal{P}_\perp) \leq m\}.$$

Since the set on the right-hand side is finite, the minimum is attained and there exists a population $\mathcal{P}_m^{\text{opt}} \in \mathbb{P}$ such that $\mathcal{G}_m^{\text{opt}} = \mathcal{G}(\mathcal{P}_m^{\text{opt}})$. Our analysis does not rely on the particular choice of $\mathcal{P}_m^{\text{opt}}$ and therefore we may ignore the fact that the choice of $\mathcal{P}_m^{\text{opt}}$ may not be unique.

The AFEM, algorithm 5.1, produces a monotone increasing sequence $\mathcal{P}_k$ of populations with $\mathcal{P}_0 = \mathcal{P}_\perp$. We say that the AFEM algorithm is *energy optimal*, when there exists a constant $C > 0$ such that $\mathcal{G}(\mathcal{P}_k) \leq \mathcal{G}_m^{\text{opt}}$, whenever $\#(P_k \setminus \mathcal{P}_\perp) \geq C\, m$.

Lemma 7.1. *Consider the sequences $(\mathcal{P}_k)_{k\in\mathbb{N}_0}$ and $(\mathcal{M}_k)_{k\in\mathbb{N}_0}$ produced by Algorithm 5.1 (second formulation). Then there exists a constant $\gamma > 0$, only possibly depending on $\mathcal{T}_\perp$ and on $\mu \in (0,1]$ when it tends to zero (i.e., $\forall\varepsilon > 0$, $\inf_{\mu\in[\varepsilon,1]} \gamma(\mu) > 0$), such that: If, for some $k, m \in \mathbb{N}_0$, we have $\mathcal{G}_m^{\text{opt}} \geq \mathcal{G}(\mathcal{P}_k) > \mathcal{G}_{m+1}^{\text{opt}}$, then*

$$\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_{k+1}) \geq \gamma\, \#\mathcal{M}_k \left(\mathcal{G}_m^{\text{opt}} - \mathcal{G}_{m+1}^{\text{opt}}\right).$$

Proof. Let $k, m \in \mathbb{N}_0$ be such that $\mathcal{G}_m^{\text{opt}} \geq \mathcal{G}(\mathcal{P}_k) > \mathcal{G}_{m+1}^{\text{opt}}$. From Corollary 4.7 and Proposition 5.1, we obtain

$$\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_{k+1}) \gtrsim \mathcal{E}^2_{\mathcal{P}_k}(\mathcal{P}_{k+1} \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k)) \geq \mu\, \#\mathcal{M}_k\, \bar{\mathcal{E}}^2_{\mathcal{P}_k}. \tag{7.1}$$

Setting $\mathcal{P}_\vee := \mathcal{P}_k \vee \mathcal{P}_{m+1}^{\text{opt}}$ and $\mathcal{P}^\wedge := \mathcal{P}_k \wedge \mathcal{P}_{m+1}^{\text{opt}}$, we choose

$$\mathcal{U} := \mathtt{free}(\mathcal{P}_\vee \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k));$$

see Figure 7. Since $\mathcal{G}(\mathcal{P}_{m+1}^{\text{opt}}) < \mathcal{G}(\mathcal{P}_k)$, we have $\mathcal{U} \neq \emptyset$; compare with Lemma 6.8(e). Thanks to the definition of $\bar{\mathcal{E}}^2_{\mathcal{P}_k}$ (see Algorithm 5.1) and $\mathcal{U} \subset \mathcal{P}_k^{++} \setminus \mathcal{P}_k$ we obtain that

$$\bar{\mathcal{E}}^2_{\mathcal{P}_k} \geq \frac{1}{\#\mathcal{U}} \sum_{P\in\mathcal{U}} \mathcal{E}^2_{\mathcal{P}_k}((\mathcal{P}_k \oplus P) \setminus \mathcal{P}_k) \geq \frac{1}{\#\mathcal{U}}\, \mathcal{E}^2_{\mathcal{P}_k}\Big(\bigcup_{P\in\mathcal{U}} (\mathcal{P}_k \oplus P) \setminus \mathcal{P}_k\Big).$$

By Lemma 6.8 (c) we have $\mathcal{U} \cup \mathtt{anc}(\mathcal{U}) \supset \mathcal{P}_\vee \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k)$, and thus

$$\begin{aligned}
\bigcup_{P\in\mathcal{U}} (\mathcal{P}_k \oplus P) \setminus \mathcal{P}_k &= \bigcup_{P\in\mathcal{U}} \big(\{P\} \cup \mathtt{anc}(P)\big) \setminus \mathcal{P}_k \\
&= \big(\mathcal{U} \cup \mathtt{anc}(\mathcal{U})\big) \setminus \mathcal{P}_k \\
&\supset \big(\mathcal{P}_\vee \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k)\big) \setminus \mathcal{P}_k \\
&= \mathcal{P}_\vee \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k).
\end{aligned}$$

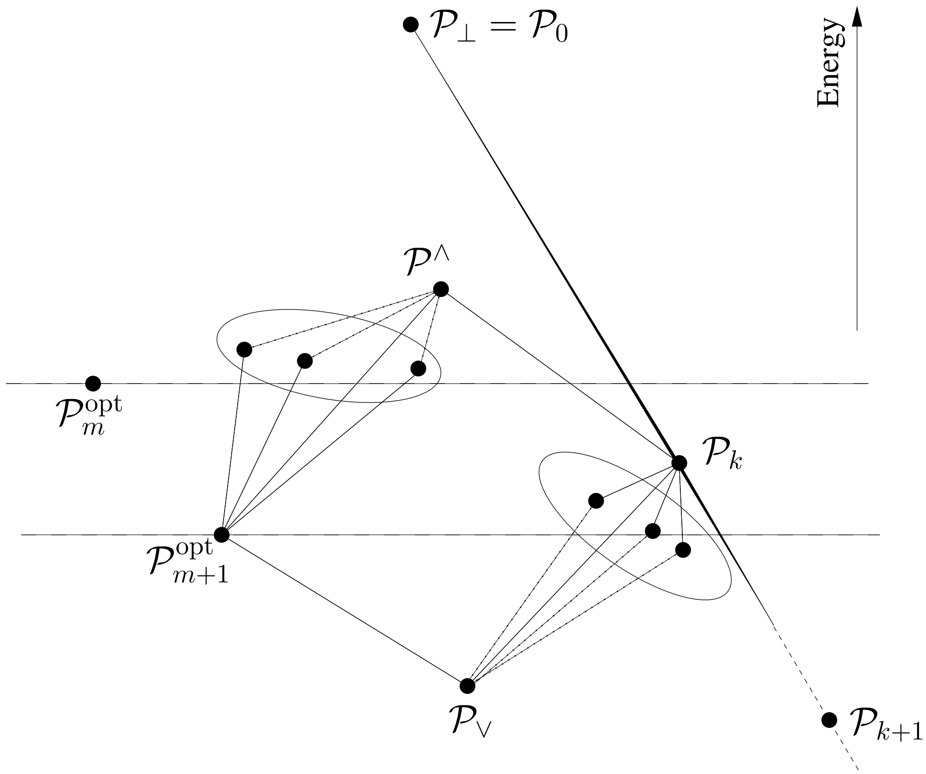

FIGURE 7. Illustration with the proof of Lemma 7.1. The populations represented by the dots in the left and right ellipses are $\{\mathcal{P}_{m+1}^{\mathrm{opt}} \ominus P : P \in \mathcal{C}\}$ and $\{\mathcal{P}_k \oplus P : P \in \mathcal{U}\}$, respectively.

This and the previous estimate prove

$$\bar{\mathcal{E}}_{\mathcal{P}_k}^2 \geq \frac{1}{\#\mathcal{U}}\, \mathcal{E}_{\mathcal{P}_k}^2\big(\mathcal{P}_\vee \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k)\big).$$

An application of Corollary 4.7 then shows that

$$\mathcal{E}_{\mathcal{P}_k}^2(\mathcal{P}_\vee \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k)) \gtrsim \mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_\vee).$$

Since $\mathcal{P}_k \neq \mathcal{P}_{m+1}^{\mathrm{opt}}$, by Lemma 3.2 and Corollary 6.14 we have that $(\mathcal{P}^\wedge, \mathcal{P}_\vee; \mathcal{P}_k, \mathcal{P}_{m+1}^{\mathrm{opt}})$ is a lower diamond in $\mathbb{P}$. By the lower diamond estimate Corollary 6.15 together with $\mathcal{G}(\mathcal{P}_k) \geq \mathcal{G}(\mathcal{P}_{m+1}^{\mathrm{opt}}) \geq \mathcal{G}(\mathcal{P}_\vee)$, this implies

$$\begin{aligned}\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_\vee) &\geq \frac{1}{2}\Big(\big(\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_\vee)\big) + \big(\mathcal{G}(\mathcal{P}_{m+1}^{\mathrm{opt}}) - \mathcal{G}(\mathcal{P}_\vee)\big)\Big) \\ &\eqsim \mathcal{G}(\mathcal{P}^\wedge) - \mathcal{G}(\mathcal{P}_\vee) \geq \mathcal{G}(\mathcal{P}^\wedge) - \mathcal{G}(\mathcal{P}_{m+1}^{\mathrm{opt}}).\end{aligned}$$

Combining the above observations with (7.1), yields

$$\begin{aligned}\mathcal{G}(\mathcal{P}_k)-\mathcal{G}(\mathcal{P}_{k+1}) &\gtrsim \mu\,\frac{\#\mathcal{M}_k}{\#\mathcal{U}}\,\mathcal{E}^2_{\mathcal{P}_k}(\mathcal{P}_\vee\cap(\mathcal{P}_k^{++}\setminus\mathcal{P}_k))\\ &\gtrsim \mu\,\frac{\#\mathcal{M}_k}{\#\mathcal{U}}\,\big(\mathcal{G}(\mathcal{P}^\wedge)-\mathcal{G}(\mathcal{P}^{\rm opt}_{m+1})\big).\end{aligned}\tag{7.2}$$

For every $P\in\mathcal{C}:=\texttt{free}(\mathcal{P}^{\rm opt}_{m+1}\setminus\mathcal{P}^\wedge)\neq\emptyset$, we set $\mathcal{P}'_P:=\mathcal{P}^{\rm opt}_{m+1}\ominus P$; see Figure 7. Thanks to Lemma 6.9, we know $\mathcal{P}'_P=\mathcal{P}^{\rm opt}_{m+1}\setminus\{P\}$, and thus $\mathcal{P}^\wedge\leq\bigwedge_{P\in\mathcal{C}}\mathcal{P}'_P$. If $\#\mathcal{C}>1$, then the sets $\mathcal{P}^{\rm opt}_{m+1}\setminus\mathcal{P}'_P$ for $P\in\mathcal{C}$ are mutually disjoint, and $\bigvee_{P\in\mathcal{C}}\mathcal{P}'_P=\mathcal{P}^{\rm opt}_{m+1}$. Applying the lower diamond estimate to the lower diamond $(\bigwedge_{P\in\mathcal{C}}\mathcal{P}'_P,\mathcal{P}^{\rm opt}_{m+1};(\mathcal{P}'_P)_{P\in\mathcal{C}})$ in $\mathbb{P}$, we obtain that

$$\mathcal{G}(\mathcal{P}^\wedge)-\mathcal{G}(\mathcal{P}^{\rm opt}_{m+1})\geq\mathcal{G}(\bigwedge_{P'\in\mathcal{C}}\mathcal{P}'_{P'})-\mathcal{G}(\mathcal{P}^{\rm opt}_{m+1})\eqsim\sum_{P\in\mathcal{C}}\big(\mathcal{G}(\mathcal{P}'_P)-\mathcal{G}(\mathcal{P}^{\rm opt}_{m+1})\big).$$

If $\#\mathcal{C}=1$, then the last step is obvious, whence the estimate is true in any case. Since $\#(\mathcal{P}'_P\setminus\mathcal{P}_\perp)=\#(\mathcal{P}^{\rm opt}_{m+1}\setminus\mathcal{P}_\perp)-1\leq m$, we have

$$\mathcal{G}(\mathcal{P}'_P)\geq\mathcal{G}(\mathcal{P}^{\rm opt}_m).$$

Therefore, we conclude from (7.2), that

$$\mathcal{G}(\mathcal{P}_k)-\mathcal{G}(\mathcal{P}_{k+1})\gtrsim\mu\,\frac{\#\mathcal{M}_k\#\mathcal{C}}{\#\mathcal{U}}\,\big(\mathcal{G}(\mathcal{P}^{\rm opt}_m)-\mathcal{G}(\mathcal{P}^{\rm opt}_{m+1})\big).\tag{7.3}$$

It remains to prove that $\#\mathcal{C}\gtrsim\#\mathcal{U}$. Define $\mathcal{V}:=\mathcal{P}_\vee\cap(\mathcal{P}_k^{++}\setminus\mathcal{P}_k)$, then $\mathcal{U}=\texttt{free}(\mathcal{V})$ and $\mathcal{V}\subset\mathcal{P}_\vee\setminus\mathcal{P}_k$. Since $\mathcal{P}_\vee\setminus\mathcal{P}_k=(\mathcal{P}_k\cup\mathcal{P}^{\rm opt}_{m+1})\setminus\mathcal{P}_k=\mathcal{P}^{\rm opt}_{m+1}\setminus(\mathcal{P}_k\cap\mathcal{P}^{\rm opt}_{m+1})=\mathcal{P}^{\rm opt}_{m+1}\setminus\mathcal{P}^\wedge$, we have $\mathcal{V}\subset\mathcal{P}^{\rm opt}_{m+1}\setminus\mathcal{P}^\wedge$. Thus Theorem 6.10 and $\mathcal{C}=\texttt{free}(\mathcal{P}^{\rm opt}_{m+1}\setminus\mathcal{P}^\wedge)$ yield

$$\#\mathcal{U}=\#\texttt{free}(\mathcal{V})\leq c_{\rm GD}\,\#\texttt{free}(\mathcal{P}^{\rm opt}_{m+1}\setminus\mathcal{P}^\wedge)=c_{\rm GD}\,\#\mathcal{C}.\tag{7.4}$$

Therefore, (7.3) and (7.4) imply the desired estimate

$$\mathcal{G}(\mathcal{P}_k)-\mathcal{G}(\mathcal{P}_{k+1})\gtrsim\mu\,\#\mathcal{M}_k\,\big(\mathcal{G}(\mathcal{P}^{\rm opt}_m)-\mathcal{G}(\mathcal{P}^{\rm opt}_{m+1})\big).\qquad\square$$

Instance optimality of our AFEM would follow from Lemma 7.1 when there would be a uniform bound on the cardinalities of the sets $\mathcal{M}_k$ of marked edges. Such a bound, however, does not exist, and the case of having 'many' marked edges will be covered by the following lemma.

Lemma 7.2. *Consider the sequences* $(\mathcal{P}_k)_{k\in\mathbb{N}_0}$ *and* $(\mathcal{M}_k)_{k\in\mathbb{N}_0}$ *produced by Algorithm 5.1 (second formulation). Then there exists a constant* $K\geq 1$, *only possibly depending on* $\mathcal{T}_\perp$, *and on* $\mu\in(0,1]$ *when it tends to zero, such that:*

If, for some $k, m \in \mathbb{N}_0$, $\mathcal{G}_m^{\mathrm{opt}} \geq \mathcal{G}(\mathcal{P}_k) > \mathcal{G}_{m+1}^{\mathrm{opt}}$, *then*

$$\mathcal{G}(\mathcal{P}_{k+1}) \leq \mathcal{G}^{\mathrm{opt}}_{m+\lfloor \frac{\#\mathcal{M}_k}{K} \rfloor}.$$

Proof. For some $k, m \in \mathbb{N}_0$, let $\mathcal{G}_m^{\mathrm{opt}} \geq \mathcal{G}(\mathcal{P}_k) > \mathcal{G}_{m+1}^{\mathrm{opt}}$. For $\#\mathcal{M}_k < K$, we have that $\lfloor \frac{\#\mathcal{M}_k}{K} \rfloor = 0$, and the claim is a direct consequence of the monotonicity of the total energy. Therefore, assume that $\#\mathcal{M}_k \geq K$. We set $\alpha := \lfloor \frac{\#\mathcal{M}_k}{K} \rfloor$. Setting $\mathcal{P}_\vee := \mathcal{P}_k \vee \mathcal{P}^{\mathrm{opt}}_{m+\alpha}$ and $\mathcal{P}^\wedge := \mathcal{P}_k \wedge \mathcal{P}^{\mathrm{opt}}_{m+\alpha}$. We repeat the steps in the proof of Lemma 7.1 up to (7.2), now using the diamond $(\mathcal{P}^\wedge, \mathcal{P}_\vee; \mathcal{P}_k, \mathcal{P}^{\mathrm{opt}}_{m+\alpha})$ in $\mathbb{P}$. Then, for $\mathcal{U} := \mathtt{free}(\mathcal{P}_\vee \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k))$ we get

$$\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_{k+1}) \gtrsim \frac{\mu \,\#\mathcal{M}_k}{\#\mathcal{U}} \big(\mathcal{G}(\mathcal{P}^\wedge) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha})\big). \tag{7.5}$$

We define $\mathcal{C} := \mathtt{free}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha} \setminus \mathcal{P}^\wedge)$, and set $N := \lfloor \frac{\#\mathcal{C}}{\alpha} \rfloor$. Exactly as in Theorem 6.10 equation (7.4), we have $\#\mathcal{U} \leq c_{\mathrm{GD}} \#\mathcal{C}$.

If $N = 0$, then $\#\mathcal{C} < \frac{\#\mathcal{M}_k}{K}$, and hence $\frac{\#\mathcal{M}_k}{\#\mathcal{U}} > \frac{K}{c_{\mathrm{GD}}}$. By taking the constant K to be sufficiently large, depending on μ when it tends to zero, we conclude that $\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_{k+1}) \geq \mathcal{G}(\mathcal{P}^\wedge) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha})$, and thus by $\mathcal{G}(\mathcal{P}^\wedge) \geq \mathcal{G}(\mathcal{P}_k)$ we arrive at $\mathcal{G}(\mathcal{P}_{k+1}) \leq \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha})$.

Else, if $N \geq 1$, then $N \geq \frac{\#\mathcal{C}}{2\alpha}$. For $\mathcal{C}_1, \ldots, \mathcal{C}_N$ being mutually disjoint subsets of $\mathcal{C}$, each having α elements, we set $\mathcal{P}'_j := \mathcal{P}^{\mathrm{opt}}_{m+\alpha} \ominus \mathcal{C}_j$, $j = 1, \ldots, N$. It follows from Lemma 6.9, that $\mathcal{P}'_j = \mathcal{P}^{\mathrm{opt}}_{m+\alpha} \setminus \mathcal{C}_j$, and hence we have $\mathcal{P}^\wedge \leq \bigwedge_{j=1}^N \mathcal{P}'_j$ and $\#(\mathcal{P}'_j \setminus \mathcal{P}_\perp) \leq m$. The last inequality implies that

$$\mathcal{G}(\mathcal{P}'_j) \geq \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_m), \qquad j = 1, \ldots, N.$$

If $N > 1$, then the sets $\mathcal{P}^{\mathrm{opt}}_{m+\alpha} \setminus \mathcal{P}'_j = \mathcal{C}_j$ for $1 \leq j \leq N$ are mutually disjoint, and $\bigvee_{j=1}^N \mathcal{P}'_j = \mathcal{P}^{\mathrm{opt}}_{m+\alpha}$. By applying the lower diamond estimate to the lower diamond $(\bigwedge_{j=1}^N \mathcal{P}'_j, \mathcal{P}^{\mathrm{opt}}_{m+\alpha}; (\mathcal{P}'_j)_{1 \leq j \leq N})$ in $\mathbb{P}$, we obtain that

$$\mathcal{G}(\mathcal{P}^\wedge) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha}) \geq \mathcal{G}\Big(\bigwedge_{j=1}^N \mathcal{P}'_j\Big) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha}) \eqsim \sum_{j=1}^N \big(\mathcal{G}(\mathcal{P}'_j) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha})\big).$$

If $N = 1$, then this estimate is obvious, whence it is true for $N \geq 1$.

Therefore, we can further estimate (7.5) by

$$\begin{aligned}\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_{k+1}) &\gtrsim \frac{\mu N \#\mathcal{M}_k}{\#\mathcal{U}} \big(\mathcal{G}(\mathcal{P}^{\mathrm{opt}}_m) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha})\big) \\ &\geq \frac{\mu K}{2 c_{\mathrm{GD}}} \big(\mathcal{G}(\mathcal{P}^{\mathrm{opt}}_m) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+\alpha})\big).\end{aligned}$$

In the last estimate, we have used that $\frac{N\#\mathcal{M}_k}{\#\mathcal{U}} \geq \frac{K}{2c_{\mathrm{GD}}}$, since $N \geq \frac{\#\mathcal{C}}{2\alpha}$, $\frac{1}{\alpha} \geq \frac{K}{\#\mathcal{M}_k}$, and $\frac{1}{\#\mathcal{U}} \geq \frac{1}{c_{\mathrm{GD}}\#\mathcal{C}}$. By taking the constant K to be sufficiently large, depending on μ when it tends to zero, we conclude that $\mathcal{G}(\mathcal{P}_k) - \mathcal{G}(\mathcal{P}_{k+1}) \geq \mathcal{G}(\mathcal{P}_m^{\mathrm{opt}}) - \mathcal{G}(\mathcal{P}_{m+\alpha}^{\mathrm{opt}})$. Thus from $\mathcal{G}(\mathcal{P}_m^{\mathrm{opt}}) \geq \mathcal{G}(\mathcal{P}_k)$, we obtain that $\mathcal{G}(\mathcal{P}_{k+1}) \leq \mathcal{G}(\mathcal{P}_{m+\alpha}^{\mathrm{opt}})$. This proves the claim in the case $N \geq 1$. □

Theorem 7.3. *Consider the sequences $(\mathcal{P}_k)_{k\in\mathbb{N}_0}$ and $(\mathcal{M}_k)_{k\in\mathbb{N}_0}$ produced by Algorithm 5.1 (second formulation). Then, there exists a constant $C \geq 1$, only possibly depending on $\mathcal{T}_\perp$ and on $\mu \in (0,1]$ when it tends to zero, such that $\#(\mathcal{P}_k \setminus \mathcal{P}_\perp) \geq Cm$ implies $\mathcal{G}(\mathcal{P}_k) \leq \mathcal{G}_m^{\mathrm{opt}}$, i.e., the algorithm is* energy optimal *with respect to the total energy $\mathcal{G}$.*

Proof. Let γ and K be the constants from Lemmas 7.1 and 7.2. Setting

$$C_k := \sum_{\ell=0}^{k-1} \#\mathcal{M}_\ell, \quad R := \left\lceil \frac{1}{\gamma} \right\rceil, \quad L := 2(R-1)(K-1) + 2K,$$

the claim follows from

$$\mathcal{G}(\mathcal{P}_k) \leq \mathcal{G}(\mathcal{P}_{\lfloor C_k/L \rfloor}^{\mathrm{opt}}). \tag{7.6}$$

Indeed, Corollary 2.8 and (2.6) imply that there exists some constant $D > 0$, depending solely on $\mathcal{T}_\perp$, such that $\#(\mathcal{P}_k \setminus \mathcal{P}_\perp) \leq D\,C_k$. Taking $C := 2DL$, we conclude that if $\#(\mathcal{P}_k \setminus \mathcal{P}_\perp) \geq Cm$, then $\frac{C_k}{2L} \geq m$, and thus $\lfloor C_k/L \rfloor \geq m$, which shows $\mathcal{G}(\mathcal{P}_k) \leq \mathcal{G}(\mathcal{P}_m^{\mathrm{opt}})$ by (7.6).

We shall prove (7.6) by induction. Obviously, (7.6) is valid for $k = 0$. Fixing an arbitrary $k \in \mathbb{N}$, assume that (7.6) is valid for $\mathbb{N}_0 \ni k' < k$. Since (7.6) is obviously true when $\mathcal{G}(\mathcal{P}_k) = \mathcal{G}(\mathcal{P}^\top)$, we assume that $\mathcal{G}(\mathcal{P}_k) > \mathcal{G}(\mathcal{P}^\top)$.

First, assume that the set

$$\big\{\ell' \in \{\max(k-R,0), \max(k-R+1,0), \ldots, k-1\} : \#\mathcal{M}_{\ell'} \geq K\big\} \tag{7.7}$$

is non-empty and set ℓ to be its maximal element. By the induction hypothesis, there exists an $m \geq \lfloor C_\ell/L \rfloor$ such that $\mathcal{G}(\mathcal{P}_{m+1}^{\mathrm{opt}}) < \mathcal{G}(\mathcal{P}_\ell) \leq \mathcal{G}(\mathcal{P}_m^{\mathrm{opt}})$. By Lemma 7.2 we obtain that $\mathcal{G}(\mathcal{P}_k) \leq \mathcal{G}(\mathcal{P}_{\ell+1}) \leq \mathcal{G}(\mathcal{P}_{m+\lfloor \#\mathcal{M}_\ell/K \rfloor}^{\mathrm{opt}})$. Using that $\lfloor a \rfloor + \lfloor b \rfloor \geq \lfloor a + b/2 \rfloor$ for $b \geq 1$, $\#\mathcal{M}_\ell \geq K$, the definition of L, and $\#\mathcal{M}_{\ell'} \leq K-1$ for the at most $R-1$ integers $\ell < \ell' \leq k-1$, we arrive at

$$\begin{aligned} m + \lfloor \#\mathcal{M}_\ell/K \rfloor &\geq \lfloor C_\ell/L \rfloor + \lfloor \#\mathcal{M}_\ell/K \rfloor \geq \lfloor C_\ell/L + \#\mathcal{M}_\ell/(2K) \rfloor \\ &\geq \lfloor (C_\ell + \#\mathcal{M}_\ell + (R-1)(K-1))/L \rfloor \geq \lfloor C_k/L \rfloor. \end{aligned}$$

This completes the proof of (7.6) in the case that the set in (7.7) is non-empty.

Suppose now, that the set in (7.7) is empty. If $k < R$, then $C_k \leq (R-1)(K-1) < L$, and we have (7.6). Now let $k \geq R$. By the induction hypothesis, there exists an $m \geq \lfloor C_{k-R}/L \rfloor$ such that $\mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+1}) < \mathcal{G}(\mathcal{P}_{k-R}) \leq \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m})$. By a repeated application of Lemma 7.1 with k reading as $k-R, k-R+1, \ldots, \ell$, as long as $\mathcal{G}(\mathcal{P}_\ell) > \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+1})$, we find that

$$\begin{aligned}\mathcal{G}(\mathcal{P}_{k-R}) - \mathcal{G}(\mathcal{P}_{\ell+1}) &= \mathcal{G}(\mathcal{P}_{k-R}) - \mathcal{G}(\mathcal{P}_{k-R+1}) + \cdots + \mathcal{G}(\mathcal{P}_\ell) - \mathcal{G}(\mathcal{P}_{\ell+1}) \\ &\geq \gamma(\ell - k + R + 1)(\mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m}) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+1})) \\ &\geq \gamma(\ell - k + R + 1)(\mathcal{G}(\mathcal{P}_{k-R}) - \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+1}).\end{aligned}$$

Therefore, by definition of R, for $\ell = k-1$ at the latest it holds that $\mathcal{G}(\mathcal{P}_{\ell+1}) \leq \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+1})$, and thus $\mathcal{G}(\mathcal{P}_k) \leq \mathcal{G}(\mathcal{P}^{\mathrm{opt}}_{m+1})$. Since $L \geq R(K-1)$ and $\#\mathcal{M}_{\ell'} \leq K-1$ for $\ell' \in \{k-R, \ldots, k-1\}$, we have

$$m+1 \geq \lfloor 1 + C_{k-R}/L \rfloor \geq \lfloor (R(K-1) + C_{k-R})/L \rfloor \geq \lfloor C_k/L \rfloor.$$

This proves (7.6). □

We are now ready to prove *instance optimality* of our AFEM as was announced in the introduction:

Theorem 7.4. *There exist constants $C, \tilde{C} \geq 1$ such that for $(\mathcal{T}_k)_{k\in\mathbb{N}_0} \subset \mathbb{T}$ being the sequence of triangulations produced by Algorithm 5.1 (first formulation), it holds that*

$$|u - u_{\mathcal{T}_k}|^2_{H^1(\Omega)} + \operatorname{osc}^2_{\mathcal{T}_k}(\mathcal{T}_k) \leq \tilde{C}\left(|u - u_{\mathcal{T}}|^2_{H^1(\Omega)} + \operatorname{osc}^2_{\mathcal{T}}(\mathcal{T})\right)$$

for all $\mathcal{T} \in \mathbb{T}$ with $\#(\mathcal{T} \setminus \mathcal{T}_\perp) \leq \frac{\#(\mathcal{T}_k \setminus \mathcal{T}_\perp)}{C}$. The constant $\tilde{C}$ depends only possibly on $\mathcal{T}_\perp$. The constant C may additionally depend on $\mu \in (0,1]$ when it tends to zero.

Proof. We know from (4.4) that the total energy $\mathcal{G}$ satisfies

$$\mathcal{G}(\mathcal{T}) - \mathcal{G}(\mathcal{T}^\top) \eqsim |u - u_{\mathcal{T}}|^2_{H^1(\Omega)} + \operatorname{osc}^2_{\mathcal{T}}(\mathcal{T})$$

for all $\mathcal{T} \in \mathbb{T}$. Hence the assertion follows from Theorem 7.3. □

Appendix A. A slightly modified marking

In this section we propose a routine MARK, resorting to slightly modified accumulated indicators, that can be implemented in $\mathcal{O}(\#\mathcal{P}_k)$ operations. The important fact is that Proposition 5.1 remains valid, which ensures the instance optimality in Section 7 also for this modified marking step.

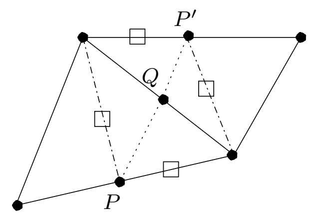

FIGURE 8. $\mathcal{T}(\mathcal{P})$ (solid lines), $P, P', Q \in \mathcal{P}^{++} \setminus \mathcal{P} = \operatorname{midpt}(\mathcal{S}(\mathcal{T}(\mathcal{P})))$ with $\{P, P'\} = \texttt{parent}(Q)$, and $\{\square\} = \texttt{child}(Q)$.

To this end, we first compute a modified maximal accumulated indicator $\bar{E}^2_{\mathcal{P}_k}$. This value can be determined with the help of the following recursive routine.

Algorithm A.1 (Maximal Indicator)**.** *Set* $\bar{E}^2_{\mathcal{P}_k} := 0$ *and call* max-ind$(P, 0)$ *for all* $P \in (\mathcal{P}_k^{++} \setminus \mathcal{P}_k)$ *with no parents in* $\mathcal{P}_k^{++} \setminus \mathcal{P}_k$*.*

procedure max-ind$(P,$ value-parent$)$

$\bar{E}^2_{\mathcal{P}_k} := \max\{\bar{E}^2_{\mathcal{P}_k}, \mathcal{E}^2_{\mathcal{P}_k}(P) + \text{value-parent}\}$;

for each child $C \in \texttt{child}(P) \cap (\mathcal{P}_k^{++} \setminus \mathcal{P}_k)$ do

max-ind$(C, \mathcal{E}^2_{\mathcal{P}_k}(P) + \text{value-parent})$;

end for

end procedure max-ind

In general we have $\bar{E}^2_{\mathcal{P}_k} \neq \bar{\mathcal{E}}^2_{\mathcal{P}_k}$, since $C \in \mathcal{P}_k^{++} \setminus \mathcal{P}_k$ may have *two* parents $P, P' \in \mathcal{P}_k^{++} \setminus \mathcal{P}_k$. However, such a C cannot have children in $\mathcal{P}^{++} \setminus \mathcal{P}$ as is illustrated in Figure 8, and so we conclude that

$$\bar{E}^2_{\mathcal{P}_k} \geq \frac{1}{2}\bar{\mathcal{E}}^2_{\mathcal{P}_k}.$$

Next, the sets $\mathcal{M}_k$ and $\widetilde{\mathcal{M}}_k$ are determined by running the following routine.

Algorithm A.2 (Marking)**.** *Set* $\mathcal{M}_k := \widetilde{\mathcal{M}}_k := \emptyset$ *and call* accum-est$(P, 0)$ *for all* $P \in \mathcal{P}_k^{++} \setminus \mathcal{P}_k$ *with no parents in* $\mathcal{P}_k^{++} \setminus \mathcal{P}_k$*.*

boolean function accum-est$(P,$ value-parent$)$

$E^2_P := \mathcal{E}^2_{\mathcal{P}_k}(P) + \text{value-parent}$;

```
    is_marked := false;
    if E_P^2 ≥ μ Ē^2_{P_k} then
      M_k := M_k ∪ {P}; M̃_k := M̃_k ∪ {P};
      E_P^2 := 0;
      is_marked := true;
    end if;
    for each child C ∈ child(P) ∩ (P_k^{++} \ P_k) do
      if accum-est(C, E_P^2) then    % child is marked, so mark the parent
        M̃_k := M̃_k ∪ {P};
        E_P^2 := 0;
        is_marked := true;
      end if
    end for
    return is_marked;
  end function accum-est
```

One verifies that $\mathcal{M}_k, \widetilde{\mathcal{M}}_k \subset \mathcal{P}_k^{++} \setminus \mathcal{P}_k$, $\mathcal{P}_k \oplus \mathcal{M}_k = \mathcal{P}_k \cup \widetilde{\mathcal{M}}_k$, and

$$\mathcal{E}^2_{\mathcal{P}_k}\big((\mathcal{P}_k \oplus \mathcal{M}_k) \setminus \mathcal{P}_k\big) \geq \mu\, \#\mathcal{M}_k\, \bar{E}^2_{\mathcal{P}_k} \geq \frac{1}{2}\mu\, \#\mathcal{M}_k\, \bar{\mathcal{E}}^2_{\mathcal{P}_k};$$

i.e., Proposition 5.1 is still valid.

Finally, the work needed for this evaluation of MARK scales linearly with $\#\mathcal{P}_k$. Indeed, the number of times that a $P \in \mathcal{P}_k^{++} \setminus \mathcal{P}_k$ is accessed by the flow of computation is proportional to the number of calls accum-est$(P, \cdot)$ (being one), plus the number of its children in $\mathcal{P}_k^{++} \setminus \mathcal{P}_k$ (being uniformly bounded).

References

[Bän91] E. Bänsch, *Local mesh refinement in 2 and 3 dimensions*, IMPACT Comput. Sci. Engrg. **3** (1991), 181–191.

[BDD04] P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math. **97** (2004), no. 2, 219–268.

[Bin07] P. Binev, *Adaptive methods and near-best tree approximation*, Oberwolfach Reports, vol. 29, 2007, pp. 1669–1673.

[BR78] I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal. **15** (1978), no. 4, 736–754.

[BV84] I. Babuška and M. Vogelius, *Feedback and adaptive finite element solution of one-dimensional boundary value problems*, Numer. Math. **44** (1984), no. 1, 75–102.

[CKNS08] J. M. Cascón, C. Kreuzer, R. H. Nochetto, and K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal. **46** (2008), no. 5, 2524–2550.

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

[Kos94] I. Kossaczký, *A recursive approach to local mesh refinement in two and three dimensions*, J. Comput. Appl. Math. **55** (1994), 275–288.

[Mau95] J. M. Maubach, *Local bisection refinement for n-simplicial grids generated by reflection*, SIAM J. Sci. Comput. **16** (1995), 210–227.

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

[MNS02] ______, *Convergence of adaptive finite element methods*, SIAM Review **44** (2002), 631–658.

[MSV08] P. Morin, K. G. Siebert, and A. Veeser, *A basic convergence result for conforming adaptive finite elements*, Math. Models Methods Appl. Sci. **18** (2008), no. 5, 707–737.

[NSV09] R. H. Nochetto, K. G. Siebert, and A. Veeser, *Theory of adaptive finite element methods: An introduction*, Multiscale, Nonlinear and Adaptive Approximation (Ronald A. DeVore and Angela Kunoth, eds.), Springer, 2009, pp. 409–542.

[SS05] A. Schmidt and K. G. Siebert, *Design of adaptive finite element software. The finite element toolbox ALBERTA*, Lecture Notes in Computational Science and Engineering 42. Springer, 2005 (English).

[Ste07] R.P. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. **7** (2007), no. 2, 245–269.

[Ste08] ______, *The completion of locally refined simplicial partitions created by bisection*, Math. Comp. **77** (2008), no. 261, 227–241 (electronic).

[SZ90] L. R. Scott and S. Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp. **54** (1990), no. 190, 483–493.

[Tra97] C. T. Traxler, *An algorithm for adaptive mesh refinement in n dimensions*, Computing **59** (1997), 115–137.

[Ver96] R. Verfürth, *A review of a posteriori error estimation and adaptive mesh-refinement techniques*, Adv. Numer. Math., John Wiley, Chichester, UK, 1996.

Lars Diening, Mathematisches Institut der Universität München, Theresienstrasse 39, D-80333 München, Germany

URL: http://www.mathematik.uni-muenchen.de/~diening

E-mail address: diening@math.lmu.de

Christian Kreuzer, Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstrasse 150, D-44801 Bochum, Germany

URL: http://www.ruhr-uni-bochum.de/ffm/Lehrstuehle/Kreuzer/index.html

E-mail address: christan.kreuzer@rub.de

Rob Stevenson, Korteweg-de Vries (KdV) Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

URL: http://staff.science.uva.nl/~rstevens/

E-mail address: r.p.stevenson@uva.nl