HIERARCHICAL ERROR ESTIMATES FOR THE ENERGY FUNCTIONAL IN OBSTACLE PROBLEMS

QINGSONG ZOU, ANDREAS VEESER, RALF KORNHUBER, AND CARSTEN GRÄSER

Abstract. We present a hierarchical a posteriori error analysis for the minimum value of the energy functional in symmetric obstacle problems. The main result is that the energy of the exact solution is, up to oscillation terms, equivalent to an appropriate hierarchical estimator. The proof does not invoke any saturation assumption. Moreover, we prove an a posteriori error estimate indicating that the estimator from [12] is asymptotically reliable and we give sufficient conditions for the validity of a saturation assumption. Finally, we corroborate and complement our theoretical results with a numerical example.

1. Introduction and main results

A posteriori error estimates are an important tool for the numerical solution of boundary value problems. For example, they can be used to quantify the error of a given approximate solution in a computable manner. Moreover, they often split into local contributions, so-called indicators, and then these indicators may be used to direct the mesh modifications in an adaptive algorithm.

The hierarchical approach to a posteriori error estimates (see [9, 23] or the monographs [1, 22]) is based upon a finite-dimensional extension of the given finite element space $\mathcal{S}$ by a suitable incremental space $\mathcal{V}$. The indicators are obtained from local defect problems associated with low-dimensional subspaces of $\mathcal{V}$, e.g., the one-dimensional subspaces spanned by the nodal basis functions. Usually, these local defect problems are solved explicitely, providing explicit a posteriori error estimates.

An attractive feature of the hierarchical approach is that lower bounds typically come without unknown constants. On the other hand, as $\mathcal{V}$ has only finite dimension, upper bounds must involve additional terms (see [4, Proposition 2.2]) that measure oscillations beyond $\mathcal{V}$. Ideally these terms are of higher order and computable. For linear elliptic problems, upper bounds have been shown by local equivalence to standard residual indicators [4] or with the help of the so-called saturation assumption [2, 9]. The saturation assumption holds, if (data) oscillation is relatively small [10]. A direct approach to upper bounds has been presented in [20], where a suitable quasi-interpolation operator onto $\mathcal{V}$ is used. All these proofs rely on the Galerkin orthogonality, or, equivalently, on the fact that the residual of the finite element approximation is vanishing on $\mathcal{S}$.

Hierarchical concepts have been applied successfully in numerical computations for various non-smooth nonlinear problems [14], in particular for obstacle problems

Date: May 18, 2009.

The first author is partially supported by NSFC under the grant 10601070 and by an Alexander von Humboldt fellowship hosted by Freie Universität Berlin.

[12, 13], and two-body contact problems in linear elasticity [17]. In all these applications, hierarchical indicators provided satisfying effectivity rates and quasioptimal convergence rates without any extra scaling of the various estimator contributions. On the other hand, the theory of hierarchical error estimates for nonlinear problems still seems to be in its infancy. Only recently, lower and upper bounds for the discretization error have been established in [15], based on a saturation assumption and suitable regularity requirements on the mesh. The a posteriori analysis in [18] avoids the saturation assumption. However, if no obstacle is present, then the upper bound does not reduce to well-known results in the unconstrained case. This is because [18], whose main concern is the convergence of the adaptive algorithm, does not fully exploit the cancellations of the linear finite element solution.

In this paper, we derive and analyze hierarchical error estimates for the following symmetric, elliptic obstacle problem. Let $\Omega \subset \mathbb{R}^2$ be a bounded polygonal Lipschitz domain, $\psi \in C(\overline{\Omega})$ a lower obstacle satisfying $\psi \leq 0$ on the boundary $\partial\Omega$ and $f \in L^2(\Omega)$ a load term. Find

$$u \in K : \qquad \mathcal{J}(u) \leq \mathcal{J}(v) \qquad \forall v \in K, \tag{1.1}$$

where

$$K = \{v \in H_0^1(\Omega) \mid v \geq \psi \text{ a.e. in } \Omega\}$$

and

$$\mathcal{J}(v) = \tfrac{1}{2}a(v,v) - (f,v) \tag{1.2}$$

is the quadratic functional induced by the symmetric bilinear form with associated energy norm

$$a(v,w) = (\nabla v, \nabla w), \qquad \|v\| = a(v,v)^{1/2}$$

and $(\cdot,\cdot)$ denotes the $L^2(\Omega)$-scalar product. Since K is a nonempty, closed, and convex set, and $a(\cdot,\cdot)$ is $H_0^1(\Omega)$-coercive, (1.1) has a unique solution u.

A key feature of obstacle problems is that (1.1) is equivalent to the variational inequality

$$a(u, v-u) \geq (f, v-u) \qquad \forall v \in K \tag{1.3}$$

and not to an equality. This is related to the property that a perturbation of the load f not necessarily affects the solution u. These two features imply that, in general, the residual is no longer orthogonal to $\mathcal{S}$. This complicates the a posteriori error analysis. In particular, the sharpness of an upper bound cannot be verified through the continuous dependence of the usual dual residual norm on the approximate solution; cf. [3, 5, 6, 16, 19], where averaging or residual techniques are considered. Insensitivity of estimators with respect to certain perturbations of the load f has been established by means of the notion of full-contact introduced in [11].

In what follows, we consider a linear finite element solution $u_{\mathcal{S}}$ to (1.1), take $\mathcal{V}$ as the span of the quadratic edge bubbles and assume that the obstacle ψ is continuous and piecewise affine. Our main result is the equivalence

$$\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) \approx -\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}) \tag{1.4}$$

up to constants depending only on the shape regularity of the mesh and oscillation terms that are formally of higher order. Here $\mathcal{I}_{\mathcal{Q}}$ is a quadratic functional of the form (1.2). The load is given by the residual of $u_{\mathcal{S}}$, and the bilinear form is a hierarchical preconditioner of $a(\cdot,\cdot)$, such that the minimum $\varepsilon_{\mathcal{V}}$ of $\mathcal{I}_{\mathcal{Q}}$ on certain localized defect constraints is explicitely known. See Section 2 for precise definitions. The hidden oscillation terms in (1.4) are computable once $u_{\mathcal{S}}$ is known. Moreover, as a corollary, we bound the discretization error $\|u_{\mathcal{S}} - u\|$ in terms of a hierarchical estimator similar to the one proposed in [12]. The equivalence (1.4) seems to be the first theoretical validation of hierarchical a posteriori error estimates for variational inequalities that reduces to well-known estimates in the unconstrained case [4] and that does not not rely on a saturation assumption. Moreover, in Section 4, we even show that the upper bound in (1.4) implies the following saturation assumption

$$\mathcal{J}(u_{\mathcal{Q}}) - \mathcal{J}(u) \leq \alpha\big(\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)\big),$$

where $\alpha \in (0,1)$ and $u_{\mathcal{Q}}$ is the quadratic finite element solution of (1.1). In this way, we generalize well-known results on the relation of error estimates, oscillation and saturation assumptions from the linear, unconstrained case [4, 10] to obstacle problems.

In order to prove (1.4) in Section 3, we apply techniques from [18, 20]. In particular, we handle the possible non-orthogonality of the residual as in [18]. Notice that we improve and partially simplify arguments. Important novelties are Lemma 3.1 and a suitable generalization of the data oscillation in the linear case. Our proof makes use of representation formulas involving local residuals on triangles and jumps of the normal fluxes across their boundary. Such representations do not extend to piecewise quadratic extensions in three space dimensions, because they are associated with edges rather than faces. Face-oriented increments in three space dimensions, like cubic bubble functions, would be covered by our theory, cf. [18]. However, in view of the results on the saturation assumption, we do not consider the three-dimensional case here.

The paper concludes with a numerical example in Section 5 that corroborates and complements the aforementioned theoretical results.

2. Discretization and hierarchical error estimate

In this section, we introduce a finite element approximation of (1.1) and then derive the hierarchical a posteriori error estimate $\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$.

Suppose $\mathcal{T}$ is a conforming triangulation of Ω. Then $\mathcal{S}$ denotes the space of continuous functions that are piecewise affine over $\mathcal{T}$ and vanish on $\partial\Omega$. The space $\mathcal{S}$ is spanned by the nodal basis $\{\phi_P \mid P \in \mathcal{N} \cap \Omega\}$, where $\mathcal{N}$ stands for the set of vertices of $T \in \mathcal{T}$, and the continuous piecewise affine functions ϕ_P associated with $P \in \mathcal{N}$ are characterized by $\phi_P(P') = \delta_{P,P'}$ (Kronecker-δ). The resulting finite element approximation of (1.1) is given by

$$u_{\mathcal{S}} \in K_{\mathcal{S}} : \qquad \mathcal{J}(u_{\mathcal{S}}) \leq \mathcal{J}(v) \quad \forall v \in K_{\mathcal{S}}$$

or, equivalently,

$$u_{\mathcal{S}} \in K_{\mathcal{S}} : \qquad a(u_{\mathcal{S}}, v - u_{\mathcal{S}}) \geq (f, v - u_{\mathcal{S}}) \quad \forall v \in K_{\mathcal{S}}. \tag{2.1}$$

The closed, convex, and non-empty set

$$K_{\mathcal{S}} = \{v \in \mathcal{S} \mid v(P) \geq \psi(P) \ \ \forall P \in \mathcal{N} \cap \Omega\}$$

is the discrete counterpart of K. As in the continuous case, existence and uniqueness follow from the coercivity of of $a(\cdot,\cdot)$ on $\mathcal{S} \subset H_0^1(\Omega)$. We assume that

$$\psi \text{ is continuous and piecewise affine over } \mathcal{T}. \tag{2.2}$$

As a consequence, $K_{\mathcal{S}} \subset K$ so that (2.1) is a conforming method. It is worth mentioning that the continuity in (2.2) may be dropped and is assumed here only for simplicity [18]. The errors arising from the approximation of non-conforming obstacles are not considered here.

We are interested in the a posteriori control of the error

$$\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) \geq 0 \tag{2.3}$$

between the exact and the approximate energy minimum. If no obstacle is present or, formally, if $\psi = -\infty$, then the energy error (2.3) corresponds to the discretization error measured in the energy norm, i.e., $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) = \frac{1}{2}\|u_{\mathcal{S}} - u\|^2$. In presence of an obstacle, however, we only have '$\geq$' instead of '$=$', in general.

We start to derive a hierarchical a posteriori error estimate for the energy error (2.3), by introducing the error function $e = u - u_{\mathcal{S}}$. Let

$$\mathcal{I}(v) = \frac{1}{2}a(v,v) - \rho_{\mathcal{S}}(v), \quad \rho_{\mathcal{S}}(v) = (f,v) - a(u_{\mathcal{S}}, v), \qquad v \in H_0^1(\Omega), \tag{2.4}$$

and

$$\mathcal{A} = \{v \in H_0^1(\Omega) \mid v \geq \psi - u_{\mathcal{S}}\} = -u_{\mathcal{S}} + K.$$

Note that (2.2) implies $0 \in \mathcal{A}$. Then, the error function e solves the defect problem

$$e \in \mathcal{A}: \qquad \mathcal{I}(e) \leq \mathcal{I}(v) \quad \forall v \in \mathcal{A}$$

or, equivalently,

$$e \in \mathcal{A}: \qquad a(e, v-e) \geq \rho_{\mathcal{S}}(v-e) \quad \forall v \in \mathcal{A} \tag{2.5}$$

and there holds

$$\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) = -\mathcal{I}(e). \tag{2.6}$$

Note that the right hand side $\rho_{\mathcal{S}}$ is a key quantity to determine e. It depends only on the load f and on the approximate solution $u_{\mathcal{S}}$. In the context of variational equations, $\rho_{\mathcal{S}}$ is called the residual of $u_{\mathcal{S}}$. In view of the relationship (2.6), we will derive the a posteriori estimate $\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ occurring in (1.4) in two steps. First, the defect problem (2.5) is discretized with respect to an extension of $\mathcal{S}$ that is rich enough to extract enough information from $\rho_{\mathcal{S}}$ or, equivalently, to provide a sufficiently accurate approximation of e. In the second step, the resulting discrete problem is decomposed into local defect problems that can be solved explicitely.

Piecewise quadratic finite elements provide approximations of higher order [8], and, therefore, are natural candidates for the discretization of (2.5) (see [4, 9, 10] for the unconstrained case). Let $\mathcal{Q}$ denote the space of continuous functions that are piecewise quadratic over $\mathcal{T}$ and vanish on $\partial\Omega$. Each function $v \in \mathcal{Q}$ is uniquely determined by its values in $\mathcal{N}_{\mathcal{Q}} = \mathcal{N} \cup \{x_E \mid E \in \mathcal{E}\}$. Here, $\mathcal{E}$ stands for the set of interior edges of $T \in \mathcal{T}$ and x_E denotes the midpoint of $E \in \mathcal{E}$. The approximation $e_{\mathcal{Q}}$ of e in $\mathcal{Q}$ is the unique solution of the discrete defect problem

$$e_{\mathcal{Q}} \in \mathcal{A}_{\mathcal{Q}}: \qquad a(e_{\mathcal{Q}}, v - e_{\mathcal{Q}}) \geq \rho_{\mathcal{S}}(v - e_{\mathcal{Q}}) \quad \forall v \in \mathcal{A}_{\mathcal{Q}}, \tag{2.7}$$

where the closed, convex, and non-empty set

$$\mathcal{A}_{\mathcal{Q}} = \{v \in \mathcal{Q} \mid v(P) \geq \psi(P) - u_{\mathcal{S}}(P)\; \forall P \in \mathcal{N}_{\mathcal{Q}} \cap \Omega\}$$

is the discrete counterpart of the defect constraints $\mathcal{A}$. Note that, in general, $\mathcal{A}_{\mathcal{Q}} \not\subset \mathcal{A}$.

In order to localize (2.7), we modify the bilinear form $a(\cdot,\cdot)$ and the constraints $\mathcal{A}_\mathcal{Q}$. To this end, we introduce the hierarchical splitting

$$\mathcal{Q} = \mathcal{S} + \mathcal{V}, \qquad \mathcal{V} = \text{span}\,\{\phi_E \mid E \in \mathcal{E}\},$$

involving the quadratic bubble functions $\phi_E \in \mathcal{Q}$ characterized by $\phi_E(P) = \delta_{x_E,P}$ for all $P \in \mathcal{N}_\mathcal{Q}$ (Kronecker-δ). Since $\mathcal{S} \cap \mathcal{V} = \{0\}$, the decomposition $v = v_\mathcal{S} + v_\mathcal{V}$ of each $v \in \mathcal{Q}$ into the contributions $v_\mathcal{S} \in \mathcal{S}$ and $v_\mathcal{V} \in \mathcal{V}$ is uniquely determined. Therefore, using this notation, the bilinear form

$$a_\mathcal{Q}(v,w) = a(v_\mathcal{S}, w_\mathcal{S}) + \sum_{E\in\mathcal{E}} v_\mathcal{V}(x_E) w_\mathcal{V}(x_E) a(\phi_E,\phi_E), \qquad v,w \in \mathcal{Q},$$

is well-defined. Note that $a_\mathcal{Q}(\cdot,\cdot)$ is resulting by decoupling of $\mathcal{S}$ and $\mathcal{V}$ and subsequent diagonalization of $a(\cdot,\cdot)$ on the incremental space $\mathcal{V}$. It provides an optimal preconditioner of $a(\cdot,\cdot)$ in the sense that the associated energy norm

$$\|v\|_\mathcal{Q} = a_\mathcal{Q}(v,v)^{1/2}, \qquad v \in \mathcal{Q},$$

is equivalent to $\|\cdot\|$ with constants depending only on the shape regularity of $\mathcal{T}$ [4, 9]. The bilinear form $a_\mathcal{Q}(\cdot,\cdot)$ gives rise to the approximate energy

$$\mathcal{I}_\mathcal{Q}(v) = \frac{1}{2} a_\mathcal{Q}(v,v) - \rho_\mathcal{S}(v), \qquad v \in \mathcal{Q}.$$

However, in contrast to the unconstrained case, the minimization problem

$$\varepsilon_\mathcal{Q} \in \mathcal{A}_\mathcal{Q}: \qquad \mathcal{I}_\mathcal{Q}(\varepsilon_\mathcal{Q}) \le \mathcal{I}_\mathcal{Q}(v) \quad \forall v \in \mathcal{A}_\mathcal{Q}$$

or, equivalently, the preconditioned defect problem

$$\varepsilon_\mathcal{Q} \in \mathcal{A}_\mathcal{Q}: \qquad a_\mathcal{Q}(\varepsilon_\mathcal{Q}, v - \varepsilon_\mathcal{Q}) \ge \rho_\mathcal{S}(v - \varepsilon_\mathcal{Q}) \quad \forall v \in \mathcal{A}_\mathcal{Q} \tag{2.8}$$

cannot be solved explicitely, because the contributions from $\mathcal{S}$ and $\mathcal{V}$ are still coupled through the constraints $\mathcal{A}_\mathcal{Q}$. As a remedy, we suppress the contributions from $\mathcal{S}$ by introducing

$$\mathcal{A}_\mathcal{V} = \{v \in \mathcal{V} \mid v(x_E) \ge \psi(x_E) - u_\mathcal{S}(x_E)\ \forall E \in \mathcal{E}\},$$

which is a proper subset of $\mathcal{A}_\mathcal{Q}$. Note that (2.2) implies $0 \in \mathcal{A}_\mathcal{V}$. We finally arrive at the localized discrete defect problem

$$\varepsilon_\mathcal{V} \in \mathcal{A}_\mathcal{V}: \qquad \mathcal{I}_\mathcal{Q}(\varepsilon_\mathcal{V}) \le \mathcal{I}_\mathcal{Q}(v) \quad \forall v \in \mathcal{A}_\mathcal{V}$$

or, equivalently,

$$\varepsilon_\mathcal{V} \in \mathcal{A}_\mathcal{V}: \qquad a_\mathcal{Q}(\varepsilon_\mathcal{V}, v - \varepsilon_\mathcal{V}) \ge \rho_\mathcal{S}(v - \varepsilon_\mathcal{V}) \quad \forall v \in \mathcal{A}_\mathcal{V}. \tag{2.9}$$

Observe that (2.9) is completely decoupled into local defect problems associated with the edges $E \in \mathcal{E}$. Their solution is explicitly given by

$$\varepsilon_\mathcal{V}(x_E) = \frac{\max\{-d_E, \rho_E\}}{\|\phi_E\|} \tag{2.10}$$

where

$$d_E = \big(u_\mathcal{S}(x_E) - \psi(x_E)\big)\|\phi_E\| \ge 0, \qquad \rho_E = \frac{\rho_\mathcal{S}(\phi_E)}{\|\phi_E\|}. \tag{2.11}$$

The quantity

$$\|\varepsilon_\mathcal{V}\|_\mathcal{Q} = \left(\sum_{E\in\mathcal{E}} \eta_E^2\right)^{1/2}, \qquad \eta_E = |\varepsilon_\mathcal{V}(x_E)|\|\phi_E\|, \tag{2.12}$$

has been proposed in [12] as an a posteriori error estimator for $\|u - u_{\mathcal{S}}\|$ (see also [13]). Here we propose

$$-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}) = -\tfrac{1}{2}a_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}, \varepsilon_{\mathcal{V}}) + \rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}}) \tag{2.13}$$

as an a posteriori error estimator for (2.3). Corresponding local indicators will be derived in Section 3.1. We will show in Section 3 that $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ is equivalent to $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)$ up to oscillation terms that are formally of higher order, in spite of the above discretization and localization.

3. A posteriori error analysis

In this section we prove our main result (1.4) stating that $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ defined in (2.13) is a reliable and efficient a posteriori error estimator for the energy error $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)$. In particular, we specify and discuss the hidden terms in (1.4). In what follows, we write '$\lesssim$' instead of '$\leq C$' where the constant C depends only on the shape regularity of $\mathcal{T}$. The notation '$A \approx B$' stands for $A \lesssim B$ and $B \lesssim A$.

3.1. Error, residual and local indicators.

As the starting point of our a posteriori error analysis, we collect some basic properties of the error $e = u - u_{\mathcal{S}}$ and its various approximations.

Lemma 3.1. *The error $e = u - u_{\mathcal{S}}$ satisfies the inequalities*

$$\tfrac{1}{2}\|e\|^2 \leq \tfrac{1}{2}\rho_{\mathcal{S}}(e) \leq -\mathcal{I}(e) \leq \rho_{\mathcal{S}}(e) \tag{3.1}$$

Proof. As a consequence of (2.2), the function $v = 0$ is contained in $\mathcal{A}$. Inserting $v = 0$ into (2.5), we obtain the first inequality of (3.1). From this inequality, we immediately get

$$\tfrac{1}{2}\rho_{\mathcal{S}}(e) \leq -\tfrac{1}{2}\|e\|^2 + \rho_{\mathcal{S}}(e) \leq \rho_{\mathcal{S}}(e).$$

In view of the definition (2.4) of $\mathcal{I}$, this concludes the proof. □

As the approximations $e_{\mathcal{Q}}$, $\varepsilon_{\mathcal{Q}}$, and $\varepsilon_{\mathcal{V}}$ of e solve the variational inequalities (2.7), (2.8), and (2.9), respectively, the arguments in the proof of Lemma 3.1 can be literally repeated to show the related estimates

$$\tfrac{1}{2}\|e_{\mathcal{Q}}\|^2 \leq \tfrac{1}{2}\rho_{\mathcal{S}}(e_{\mathcal{Q}}) \leq -\mathcal{I}(e_{\mathcal{Q}}) \leq \rho_{\mathcal{S}}(e_{\mathcal{Q}}), \tag{3.2}$$

$$\tfrac{1}{2}\|\varepsilon_{\mathcal{Q}}\|_{\mathcal{Q}}^2 \leq \tfrac{1}{2}\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}}) \leq -\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{Q}}) \leq \rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}}), \tag{3.3}$$

$$\tfrac{1}{2}\|\varepsilon_{\mathcal{V}}\|_{\mathcal{Q}}^2 \leq \tfrac{1}{2}\rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}}) \leq -\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}) \leq \rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}}). \tag{3.4}$$

As a first application, we derive local indicators for our approximate energy error $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$. As $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ and $\rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}})$ are equivalent up to the constant 1/2, local contributions to $\rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}})$ can be used as indicators for $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$. Utilizing the definitions (2.10), (2.11), and (2.12), $\rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}})$ can be decomposed according to

$$\rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}}) = \sum_{E \in \mathcal{E}} \varepsilon_{\mathcal{V}}(x_E)\rho_{\mathcal{S}}(\phi_E) = \sum_{E \in \mathcal{E}} \eta_E|\rho_E|. \tag{3.5}$$

This suggests the local indicators $\eta_E|\rho_E|$, $E \in \mathcal{E}$, for $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$. These indicators have been already used in [18], but they were collected in a different way.

In view of the identity (2.6) and Lemma 3.1, the energy error $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)$ is equivalent to the quantity $\rho_{\mathcal{S}}(e)$. Therefore, it is useful to provide some further properties of $\rho_{\mathcal{S}}$. As in the unconstrained case, $\rho_{\mathcal{S}}$ depends only on the load f

and on the discrete solution $u_\mathcal{S}$. More precisely, after integration by parts on each $T \in \mathcal{T}$, the identity $\Delta u_\mathcal{S} = 0$ on each T yields the representation

$$\rho_\mathcal{S}(v) = \int_\Omega fv + \sum_{E\in\mathcal{E}} \int_E j_E v, \qquad j_E = \partial_\mathbf{n} u_\mathcal{S}|_{T_2} - \partial_\mathbf{n} u_\mathcal{S}|_{T_1}.$$

Here, $\mathbf{n}$ denotes the unit normal vector on the common edge $E = T_1 \cap T_2$ of two triangles $T_1, T_2 \in \mathcal{T}$ pointing from T_1 to T_2, and $j_E \in \mathbb{R}$ represents the jump of the normal flux associated with $u_\mathcal{S}$ across E.

In contrast to the unconstrained case, the equivalence $\rho_\mathcal{S} = 0 \iff e = 0$ in general does not hold for variational inequalities. However, Lemma 3.1 implies

$$\big(\rho_\mathcal{S}(v) \le 0 \ \ \forall v \in \mathcal{A}\big) \implies e = 0 \tag{3.6}$$

which is an extension of $\rho_\mathcal{S} = 0 \implies e = 0$. As $u_\mathcal{S}$ solves the discrete problem (2.1), we can insert $v = u_\mathcal{S} + \phi_P \in K_\mathcal{S}$ to obtain $\rho_\mathcal{S}(\phi_P) \le 0$ for all $P \in \mathcal{N} \cap \Omega$, which means that the approximation of e in $\mathcal{S}$ is just zero. If $u_\mathcal{S}(P) > \psi(P)$, we can even insert $v_\alpha = u_\mathcal{S} - \alpha\phi_P \in K_\mathcal{S}$ with sufficiently small $\alpha > 0$ to obtain $\rho_\mathcal{S}(\phi_P) = 0$. Combining our observations, we obtain the discrete complementarity properties

$$\rho_\mathcal{S}(\phi_P) \le 0, \quad \psi(P) - u_\mathcal{S}(P) \le 0, \quad \rho_\mathcal{S}(\phi_P)\big(\psi(P) - u_\mathcal{S}(P)\big) = 0 \tag{3.7}$$

for all $P \in \mathcal{N} \cap \Omega$. Note that $\rho_\mathcal{S}(\phi_P) < 0$ might occur for certain P so that, in contrast to the unconstrained case, $\rho_\mathcal{S}$ in general does not vanish on $\mathcal{S}$. The discrete complementarity properties (3.7) are compatible with the following localization of $\rho_\mathcal{S}$. Invoking the partition of unity

$$\sum_{P\in\mathcal{N}} \phi_P = 1 \qquad \text{in } \Omega, \tag{3.8}$$

we decompose

$$\rho_\mathcal{S} = \sum_{P\in\mathcal{N}} \rho_P \tag{3.9}$$

into the local contributions

$$\rho_P(v) = \rho_\mathcal{S}(v\phi_P) = \int_{\omega_P} fv\phi_P + \sum_{E\in\mathcal{E}_P} \int_E j_E v\phi_P, \quad v \in H^1(\Omega),$$

where

$$\omega_P = \operatorname{supp} \phi_P, \qquad \mathcal{E}_P = \{E \in \mathcal{E} \mid E \ni P\},$$

denote the support of ϕ_P and the internal edges emanating from P, respectively. Then, (3.7) takes the form

$$\rho_P(1) \le 0, \tag{3.10a}$$

$$u_\mathcal{S}(P) > \psi(P) \implies \rho_P(1) = 0, \tag{3.10b}$$

for all $P \in \mathcal{N} \cap \Omega$. These complementarity properties suggest to introduce the sets

$$\mathcal{N}^0 = \{P \in \mathcal{N} \cap \Omega \mid u_\mathcal{S}(P) = \psi(P)\}, \quad \mathcal{N}^+ = \{P \in \mathcal{N} \cap \Omega \mid u_\mathcal{S}(P) > \psi(P)\}$$

of interior contact and non-contact nodes, respectively.

3.2. **Oscillation terms.** As the extension $\mathcal{V}$ has finite dimension, the corresponding approximation $-\mathcal{I}_Q(\varepsilon_\mathcal{V})$ can only provide upper bounds of $-\mathcal{I}(e)$ up to additional terms (see [4, Proposition 2.2]) measuring oscillation beyond $\mathcal{V}$. In this subsection, we introduce the additional terms that appear in the upper bound of $-\mathcal{I}(e)$ in Section 3.3 below and formally explain their oscillation and higher order character. For the the latter, we consider global refinements of a triangulation with (2.2) and with the expected order h of $\|u-u_\mathcal{S}\|$ or $\big(\mathcal{J}(u_\mathcal{S})-\mathcal{J}(u)\big)^{\frac{1}{2}}$. Rigorous proofs would go beyond the scope of this article. Our heuristic reasoning is supported by the numerical example in Section 5.

Oscillation $\operatorname{osc}(u_\mathcal{S},\psi,f)$ consists of two different terms

$$\operatorname{osc}(u_\mathcal{S},\psi,f)=\operatorname{osc}_1(u_\mathcal{S},\psi)+\operatorname{osc}_2(u_\mathcal{S},\psi,f), \tag{3.11}$$

depending on the data of the given problem (1.3) and its discretization (2.1). Note the dependence on the discrete solution $u_\mathcal{S}$, which is a novelty with respect to the unconstrained case and the reason why we use only the term 'oscillation' instead of 'data oscillation'.

The first term $\operatorname{osc}_1(u_\mathcal{S},\psi)$ measures a kind of obstacle oscillation. It is given by

$$\operatorname{osc}_1(u_\mathcal{S},\psi)=\left(\sum_{P\in\mathcal{N}^{0+}}\|\nabla(\psi-u_\mathcal{S})\|^2_{0,\omega_P}\right)^{1/2}, \tag{3.12}$$

where

$$\mathcal{N}^{0+}=\big\{P\in\mathcal{N}^0 \mid u_\mathcal{S}>\psi \text{ in } \omega_P\setminus\{P\}\big\}$$

denotes the set of isolated contact nodes and $\|\cdot\|_{0,\omega_P}$ stands for the norm of $L^2(\omega_P)$. Isolated contact nodes are discrete counterparts of isolated contact points, which are strict minima $x\in\Omega$ of $u-\psi$ with $u(x)=\psi(x)$. If isolated contact nodes persist under refinement, then, under certain regularity conditions, the exact solution u should have corresponding isolated contact points. In this case, the set $\cup_{P\in\mathcal{N}^{0+}}\omega_P$ shrinks towards these isolated contact points of u. This entails that $\operatorname{osc}_1(u_\mathcal{S},\psi)$ has at least the order of the error. Higher order should arise if ψ is smooth in the isolated contact points. In fact, assuming also that u is also smooth enough, we get $\big(\nabla u-\nabla\psi\big)(x)=0$ for all isolated contact points x. One thus expects that $\operatorname{osc}_1(u_\mathcal{S},\psi)$ is vanishing with higher order. A similar argument applies if ψ is the nodal interpolation of some smooth obstacle ψ_0 to $\mathcal{S}$, which however is outside the conforming framework considered here. It is worth mentioning that the set $\mathcal{N}^{0+}$ can be made smaller, at the expense of a slightly more complicated notion of isolated contact nodes [18].

To define $\operatorname{osc}_2(u_\mathcal{S},\psi,f)$, we introduce the set

$$\mathcal{N}^{++}=\{P\in\mathcal{N}^+ \mid \rho_E\geq -d_E\ \forall E\in\mathcal{E}_P\} \tag{3.13}$$

of non-contact nodes where the approximate error $\varepsilon_\mathcal{V}$ solving (2.9) is not in contact and the set

$$\mathcal{N}^{0-}=\{P\in\mathcal{N}^0 \mid u_\mathcal{S}=\psi, f\leq 0 \text{ in } \omega_P \text{ and } j_E\leq 0\ \forall E\in\mathcal{E}_P\}$$

of full-contact nodes with certain monotonicity properties. The latter are equivalent to $f+\Delta\psi\leq 0$ in the interior of ω_P (in distributional sense), which in turn is

necessary for $u = \psi$ in ω_P. Then, we define
(3.14)

$$\operatorname{osc}_2(u_{\mathcal{S}}, \psi, f) = \left(\sum_{P \in \mathcal{N}^{++}} h_P^2 \|f - \bar{f}_P\|_{0,\omega_P}^2 + \sum_{P \in \mathcal{N} \setminus (\mathcal{N}^{0-} \cup \mathcal{N}^{++})} h_P^2 \|f\|_{0,\omega_P}^2 \right)^{1/2},$$

where, for any $P \in \mathcal{N}$, $h_P = \max_{E \in \mathcal{E}_P} |E|$ is a measure for the diameter of ω_P, and

$$\bar{f}_P = \frac{1}{|\omega_p|} \int_{\omega_p} f$$

is the mean value of f on ω_P. The term $\operatorname{osc}_2(u_{\mathcal{S}}, \psi, f)$ is a generalization of well-known data oscillation from the unconstrained case [1, 22] to obstacle problems. In fact, if no obstacle is present, then the definition (3.14) reduces to

$$\operatorname{osc}_2(u_{\mathcal{S}}, \psi, f) = \left(\sum_{P \in \mathcal{N} \cap \Omega} h_P^2 \|f - \bar{f}_P\|_{0,\omega_P}^2 + \sum_{P \in \mathcal{N} \cap \partial\Omega} h_P^2 \|f\|_{0,\omega_P}^2 \right)^{1/2}. \tag{3.15}$$

Observe that unconstrained data oscillation (3.15) has two types of indicators: the indicators associated with non-Dirichlet nodes involve local means, while the indicators associated with Dirichlet nodes do not (because the corresponding hat functions are not in $\mathcal{S}$). In view of the approximation properties of local means and since $\cup_{P \in \mathcal{N} \cap \partial\Omega} \omega_P$ shrinks to the Dirichlet boundary under refinement, data oscillation (3.15) is of higher order for sufficiently smooth loads f.

The oscillation term $\operatorname{osc}_2(u_{\mathcal{S}}, \psi, f)$ has only contributions from outside of the full-contact region. These contributions have a similar structure as data oscillation for unconstrained problems: indicators that are sufficiently far away from the discrete free boundary involve local means, while indicators in the vicinity of the discrete free boundary do not. If the discrete free boundary converges under refinement (see, e.g., [7]), then the set $\cup_{P \in \mathcal{N} \setminus (\mathcal{N}^{0-} \cup \mathcal{N}^{++})} \omega_P$ shrinks towards the exact free boundary, in addition to $\partial\Omega$. Hence, $\operatorname{osc}_2(u_{\mathcal{S}}, \psi, f)$ is expected to be of higher order. The analogy between generalized oscillation $\operatorname{osc}_2(u_{\mathcal{S}}, \psi, f)$ defined in (3.14) and its unconstrained counterpart (3.15) reflects that the obstacle problem (1.3) reduces to an unconstrained Dirichlet problem on a reduced computational domain, once the exact free boundary is known.

3.3. **Reliability.** In this subsection, we derive an upper bound for the energy error $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) = -\mathcal{I}(e)$ consisting of the hierarchical estimate $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$, introduced in (2.13), and an additional oscillation term $\operatorname{osc}(u_{\mathcal{S}}, \psi, f)$, defined in (3.11).

The reduction of the continuous error $e = u - u_{\mathcal{S}} \in H_0^1(\Omega)$, solving the infinite-dimensional defect problem (2.5), to its approximation $\varepsilon_{\mathcal{V}} \in \mathcal{V}$, obtained from the localized discrete defect problem (2.9), will be performed by local projections

$$\Pi_P : \; H^1(\Omega) \to \mathcal{Q}_P = \operatorname{span}\{\phi_P\} \cup \mathcal{V}_P, \quad \mathcal{V}_P = \operatorname{span}\,\{\phi_E \mid E \in \mathcal{E}_P\}, \qquad P \in \mathcal{N}.$$

For given $v \in H^1(\Omega)$, the value $\Pi_P v \in \mathcal{Q}_P$ is uniquely defined by the conditions

$$\int_E \Pi_P v = \int_E v \quad \forall E \in \mathcal{E}_P \quad \text{and} \quad \begin{cases} \int_{\omega_P} \Pi_P v = \int_{\omega_P} v & \text{if } P \in \mathcal{N}^{++}, \\ \Pi_P v \in \mathcal{V}_P & \text{otherwise.} \end{cases} \tag{3.16}$$

In contrast to similar projections [18], Π_P also preserves the mean value in ω_P for all $P \in \mathcal{N}^{++}$. This property prepares the ground for an upper bound with oscillation term $\operatorname{osc}(u_{\mathcal{S}}, \psi, f)$ defined in (3.11). It can be verified by straightforward calculations that the coefficients in the hierarchical basis representation

$$\Pi_P v = \alpha_P(v)\phi_P + \sum_{E\in\mathcal{E}_P} \alpha_E(v)\phi_E \tag{3.17}$$

are given by

$$\alpha_P(v) = \begin{cases} \frac{c_P(v)}{c_P(\phi_P)} & \text{if } P\in\mathcal{N}^{++}, \\ 0 & \text{otherwise,}\end{cases} \qquad \alpha_E(v) = \frac{\int_E v - \alpha_P(v)\int_E \phi_P}{\int_E \phi_E}, \tag{3.18}$$

where

$$c_P(v) = \int_{\omega_P} v - \sum_{E\in\mathcal{E}_P} \Big(\int_E v\Big)\Big(\int_{\omega_P}\phi_E\Big)\Big(\int_E \phi_E\Big)^{-1}.$$

In particular, $c_P(\phi_P) = -\frac{1}{6}|\omega_P|$. The following lemma collects some essential properties of the projections Π_P.

Lemma 3.2. *The coefficients in* (3.17) *satisfy*

$$\max_{Q\in\{P\}\cup\mathcal{E}_P} |\alpha_Q(v)| \lesssim h_P^{-1}\big(\|v\|_{0,\omega_P} + h_P\|\nabla v\|_{0,\omega_P}\big) \tag{3.19}$$

and Π_P *is stable in the sense that*

$$\|\Pi_P v\|_{0,\omega_P} \lesssim \|v\|_{0,\omega_P} + h_P\|\nabla v\|_{0,\omega_P}. \tag{3.20}$$

Moreover, if $P\notin\mathcal{N}^{++}$, *then the coefficients* $\alpha_E(v) = (\int_E v)(\int_E \phi_E)^{-1}$ *have the property*

$$\int_E v \geq \int_E (\psi - u_{\mathcal{S}}) \implies \alpha_E(v) \gtrsim \psi(x_E) - u_{\mathcal{S}}(x_E) \qquad \forall E\in\mathcal{E}_P. \tag{3.21}$$

Proof. In order to show (3.19) and (3.20), we start with

$$\Big|\int_{\omega_P} v\Big| \lesssim h_P\|v\|_{0,\omega_P}, \qquad \Big|\int_E v\Big| \leq h_E^{\frac{1}{2}}\|v\|_{0,E} \lesssim h_P(h_P^{-1}\|v\|_{0,\omega_P} + \|\nabla v\|_{0,\omega_P}),$$

where we have used the Cauchy-Schwarz inequality, the 'scaled' trace theorem, and $h_E = |E| \leq h_P$ for $E\in\mathcal{E}_P$. Inserting these estimates and straightforward bounds of the integrals of ϕ_E and ϕ_P in terms of h_P into (3.18), we obtain (3.19). Then (3.20) follows from the triangle inequality, $\|\phi_P\|_{0,\omega_P}\approx h_P$, and $\|\phi_E\|_{0,\omega_P}\approx h_P$.

If $P\notin\mathcal{N}^{++}$, then $\alpha_P(v) = 0$ so that (3.18) provides the coefficients $\alpha_E(v) = (\int_E v)(\int_E \phi_E)^{-1}$. Moreover, we have $\int_E(\psi - u_{\mathcal{S}}) = |E|\big(\psi(x_E) - u_{\mathcal{S}}(x_E)\big)$ by condition (2.2) and the midpoint rule. Thus (3.21) follows from the identity $\int_E \phi_E = \frac{1}{6}|E|$. □

Utilizing the projections Π_P, we derive an upper bound for $\rho_{\mathcal{S}}(e)$. In light of Lemma 3.1, this is the crucial step towards an upper bound for the energy error $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) = -\mathcal{I}(e)$.

Proposition 3.3. *Assume that* (2.2) *holds. Then*

$$\rho_{\mathcal{S}}(e) \lesssim \sum_{E\in\mathcal{E}} |\rho_E|\eta_E + \operatorname{osc}(u_{\mathcal{S}},\psi,f)^2 \tag{3.22}$$

with ρ_E *defined in* (2.11), η_E *defined in* (2.12), *and* $\operatorname{osc}(u_{\mathcal{S}},\psi,f)$ *defined in* (3.11).

Proof. Using the decomposition (3.9) we write $\rho_{\mathcal{S}}(e) = \sum_{P\in\mathcal{N}} \rho_P(e)$. To derive upper bounds for the local contributions $\rho_P(e)$, we distinguish six cases corresponding to the splitting

$$\mathcal{N} = \mathcal{N}^{++} \cup \big(\mathcal{N}^{+} \setminus \mathcal{N}^{++}\big) \cup (\mathcal{N} \cap \partial\Omega) \cup \big(\mathcal{N}^{0} \setminus (\mathcal{N}^{0+} \cup \mathcal{N}^{0-}\big) \cup \mathcal{N}^{0+} \cup \mathcal{N}^{0-},$$

which will be addressed in the given order.

Case 1: $P \in \mathcal{N}^{++}$. We claim that

$$\rho_P(e) \lesssim \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P\|f - \bar{f}_P\|_{0,\omega_P}\Big)\|\nabla e\|_{0,\omega_P} \tag{3.23}$$

with $\mathcal{E}_P^+ = \{E \in \mathcal{E}_P \mid \rho_E \geq -d_E\}$. Note that $\mathcal{E}_P^+ = \mathcal{E}_P$, because $P \in \mathcal{N}^{++}$. In order to prove (3.23), we set

$$w = (e - c)\phi_P, \qquad c = \frac{1}{|\omega_P|}\int_{\omega_P} e.$$

Then, we derive

$$\begin{aligned}
\rho_P(e) = \rho_P(e - c) &= \int_{\omega_P} fw + \sum_{E\in\mathcal{E}_P}\int_E j_E w \\
&= \int_{\omega_P} f\Pi_P w + \sum_{E\in\mathcal{E}_P}\int_E j_E \Pi_P w + \int_{\omega_P} f(w - \Pi_P w) \\
&= \rho_{\mathcal{S}}(\Pi_P w) + \int_{\omega_P}(f - \bar{f}_P)(w - \Pi_P w) \\
&\leq \sum_{E\in\mathcal{E}_P} \alpha_E(w)\rho_E\|\phi_E\| + \|f - \bar{f}_P\|_{0,\omega_P}\|w - \Pi_P w\|_{0,\omega_P}
\end{aligned} \tag{3.24}$$

from $\mathcal{N}^{++} \subset \mathcal{N}^{+} \cap \Omega$, (3.10b) and thus $\rho_P(1) = \rho_{\mathcal{S}}(\phi_P) = 0$, the definition (3.16) of Π_P, the fact that $j_E \in \mathbb{R}$ is constant, the definition (2.11) of ρ_E, and the Cauchy-Schwarz inequality. Notice that, thanks to the choice of c in the definition of w and $P \in \Omega$, we have

$$\|w\|_{0,\omega_P} \leq \|e - c\|_{0,\omega_P} \lesssim h_P\|\nabla e\|_{0,\omega_P} \tag{3.25}$$

by a Poincaré inequality, cf., e.g., [21]. Utilizing (3.19), $\|\phi_P\|_{\infty,\omega_P} \leq 1$, $\|\nabla\phi_P\|_{\infty,\omega_P} \lesssim h_P^{-1}$, and (3.25), we obtain

$$\begin{aligned}
|\alpha_E(w)| &\lesssim h_P^{-1}\big\{\|w\|_{0,\omega_P} + h_P\|\nabla w\|_{0,\omega_P}\big\} \\
&\lesssim h_P^{-1}\Big\{\|(e - c)\phi_P\|_{0,\omega_P} + h_P\|\nabla\big((e - c)\phi_P\big)\|_{0,\omega_P}\Big\} \\
&\lesssim \|\nabla e\|_{0,\omega_P} \approx \|\phi_E\|^{-1}\|\nabla e\|_{0,\omega_P}
\end{aligned} \tag{3.26}$$

for all $E \in \mathcal{E}_P$. In a similar way, we get

$$\|w - \Pi_P w\|_{0,\omega_P} \lesssim \|w\|_{0,\omega_P} + h_P\|\nabla w\|_{0,\omega_P} \lesssim h_P\|\nabla e\|_{0,\omega_P} \tag{3.27}$$

using (3.20). The desired estimate (3.23) follows by inserting these two inequalities and $\mathcal{E}_P = \mathcal{E}_P^+$ into (3.24).

Case 2: $P \in \mathcal{N}^{+} \setminus \mathcal{N}^{++}$. We claim

$$\rho_P(e) \lesssim \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P\|f\|_{0,\omega_P}\Big)\|\nabla e\|_{0,\omega_P}. \tag{3.28}$$

To show (3.28), we first proceed as in the proof of (3.24), to derive the inequality

$$\rho_P(e) \leq \sum_{E\in\mathcal{E}_P} \alpha_E(w)\rho_E\|\phi_E\| + \|f\|_{0,\omega_P}\|w-\Pi_P w\|_{0,\omega_P}, \tag{3.29}$$

where

$$w = (e-c)\phi_P, \qquad c = \min\Big\{ \Big(\int_E e\phi_P\Big)\Big(\int_E \phi_P\Big)^{-1} \mid E \in \mathcal{E}_P \Big\}.$$

This particular choice of c implies that we have $\alpha_E(w) = (\int_E w)(\int_E \phi_E)^{-1} \geq 0$ for all $E \in \mathcal{E}_P$ and that the inequalities (3.26) and (3.27) follow from the generalized Poincaré-Friedrichs inequality in [18, Lemma 3.4]. Hence, inserting (3.26) and (3.27) into (3.29), we can exploit $\alpha_E(w)\rho_E \leq 0$ whenever $\rho_E < -d_E \leq 0$, to obtain the desired inequality (3.28).

Case 3: $P \in \mathcal{N} \cap \partial\Omega$. We claim

$$\rho_P(e) \lesssim \sum_{E\in\mathcal{E}_P^0} |\rho_E| d_E + \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P\|f\|_{0,\omega_P}\Big)\|\nabla e\|_{0,\omega_P} \tag{3.30}$$

with $\mathcal{E}_P^0 = \mathcal{E}_P \setminus \mathcal{E}_P^+ = \{E \in \mathcal{E}_P \mid \rho_E < -d_E\}$. To prove (3.30), we again start from the inequality

$$\rho_P(e) \leq \sum_{E\in\mathcal{E}_P} \alpha_E(w)\rho_E\|\phi_E\| + \|f\|_{0,\omega_P}\|w-\Pi_P w\|_{0,\omega_P}, \tag{3.31}$$

where this time we set

$$w = (e-c)\phi_P, \qquad c = 0.$$

There is no freedom in the choice of c, since $P \notin \Omega$ so that we cannot invoke (3.10). However, e vanishes at least on one edge of $\partial\omega_P$, because $P \in \partial\Omega$. Hence, the generalized Poincaré-Friedrichs inequality [18, Lemma 3.4] can be applied again to obtain (3.26) and (3.27). Inserting these inequalities into (3.31), we get the desired bound for the contributions from $E \in \mathcal{E}_P^+$. In view of $u_\mathcal{S}+w = (1-\phi_P)u_\mathcal{S}+\phi_P u \geq \psi$, (3.21) implies $\alpha_E(w) \gtrsim \psi(x_E) - u_\mathcal{S}(x_E) = -d_E\|\phi_E\|^{-1}$. Using this inequality, we get the desired bound for the remaining contributions from $E \in \mathcal{E}_P^0$.

Case 4: $P \in \mathcal{N}^0 \setminus \big(\mathcal{N}^{0-} \cup \mathcal{N}^{0+}\big)$. We claim

$$\rho_P(e) \lesssim \sum_{E\in\mathcal{E}_P^0} |\rho_E| d_E + \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P\|f\|_{0,\omega_P}\Big)\|\nabla e\|_{0,\omega_P}. \tag{3.32}$$

In order to show (3.32), we write $\rho_P(e) = \rho_P(e^+) + \rho_P(e^-)$ with $e^+ = \max(e,0)$, $e^- = \min(e,0)$ and prove the desired bound seperately for $\rho_P(e^+)$ and $\rho_P(e^-)$.

Let us start with $\rho_P(e^+)$. Utilizing (3.10a), we proceed as in the Case 2, to derive the usual upper bound

$$\rho_P(e^+) \leq \rho_P(e^+ - c) \leq \sum_{E\in\mathcal{E}_P} \alpha_E(w)\rho_E\|\phi_E\| + \|f\|_{0,\omega_P}\|w-\Pi_P w\|_{0,\omega_P},$$

where

$$w = (e^+ - c)\phi_P, \qquad c = \min\Big\{ \Big(\int_E e^+\phi_P\Big)\Big(\int_E \phi_P\Big)^{-1} \mid E \in \mathcal{E}_P \Big\} \geq 0.$$

Then, we continue literally as in the Case 2 and use $|\nabla e^+| \le |\nabla e|$, to obtain the desired bound

$$\rho_P(e^+) \lesssim \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P\|f\|_{0,\omega_P}\Big)\|\nabla e\|_{0,\omega_P}. \tag{3.33}$$

Next we consider $\rho_P(e^-)$. As in the Case 3 we get the upper bound

$$\rho_P(e^-) \le \sum_{E\in\mathcal{E}_P} \alpha_E(w)\rho_E\|\phi_E\| + \|f\|_{0,\omega_P}\|w-\Pi_P w\|_{0,\omega_P} \tag{3.34}$$

where

$$w = (e^- - c)\phi_P, \qquad c = 0.$$

Since $w = e^-\phi_P \ge e^- \ge \psi - u_\mathcal{S}$, (3.21) provides for all $E \in \mathcal{E}_P$,

$$0 \ge \alpha_E(w) \gtrsim \psi(x_E) - u_\mathcal{S}(x_E) = -d_E\|\phi_E\|^{-1} \tag{3.35}$$

and therefore

$$\alpha_E(w)\rho_E\|\phi_E\| \lesssim |\rho_E| d_E \qquad \forall E \in \mathcal{E}_P^0. \tag{3.36}$$

It remains to bound $|\alpha_E(w)|$ for $E \in \mathcal{E}_P^+$ and $\|w-\Pi_P w\|_{0,\omega_P}$ appropriately. To this end, we exploit that $P \notin \mathcal{N}^{0+}$ providing that there is at least one edge $E \in \mathcal{E}_P$ such that $e^- = 0$ on E. As in Case 3, we can therefore apply the generalized Poincaré-Friedrichs inequality [18, Lemma 3.4] and $|\nabla e^-| \le |\nabla e|$ to show (3.26) and (3.27). In combination with (3.36), this leads to

$$\rho_P(e^-) \lesssim \sum_{E\in\mathcal{E}_P^0} |\rho_E| d_E + \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P\|f\|_{0,\omega_P}\Big)\|\nabla e\|_{\omega_P}. \tag{3.37}$$

We sum the two estimates (3.33) and (3.37), to obtain the desired bound (3.32).

Case 5: $P \in \mathcal{N}^{0+}$. We claim

$$\begin{aligned}\rho_P(e) \lesssim &\sum_{E\in\mathcal{E}_P^0} |\rho_E| d_E \\ &+ \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P\|f\|_{0,\omega_P}\Big)\Big(\|\nabla e\|_{0,\omega_P} + \|\nabla(\psi - u_\mathcal{S})\|_{0,\omega_P}\Big).\end{aligned} \tag{3.38}$$

As in Case 4, we use the splitting $\rho_P(e) = \rho_P(e^+) + \rho_P(e^-)$ and proceed literally as above, to show that $\rho_P(e^+)$ satisfies an inequality of the form (3.33), and that $\rho_P(e^-)$ satisfies

$$\rho_P(e^-) \le \sum_{E\in\mathcal{E}_P} \alpha_E(w)\rho_E\|\phi_E\| + \|f\|_{0,\omega_P}\|w-\Pi_P w\|_{0,\omega_P} \tag{3.39}$$

where

$$w = (e^- - c)\phi_P, \qquad c = 0.$$

As the estimate

$$\alpha_E(w)\rho_E\|\phi_E\| \le |\rho_E| d_E \qquad \forall E \in \mathcal{E}_P^0 \tag{3.40}$$

can be also shown as in Case 4, it remains to bound $|\alpha_E(w)|$ for $E \in \mathcal{E}_P^+$ and $\|w - \Pi_P w\|_{0,\omega_P}$ appropriately. As a first step, we derive the following substitute for (3.26)

$$|\alpha_E(w)|\|\phi_E\| \le |\alpha_E(\psi - u_\mathcal{S})|\|\phi_E\| \lesssim h_P^{-\frac{1}{2}}\|\psi - u_\mathcal{S}\|_{0,E} \lesssim \|\nabla(\psi - u_\mathcal{S})\|_{0,\omega_P} \tag{3.41}$$

by using $\psi - u_{\mathcal{S}} \le e^- \le w \le 0$, the monotonicity of the integral in $\alpha_E(w) = (\int_E w)(\int_E \phi_E)^{-1}$ and a 'scaled' Poincaré-Friedrichs inequality based on $(\psi - u_{\mathcal{S}})(P) = 0$ and $\psi - u_{\mathcal{S}} \in \mathcal{S}$, which in turn holds thanks to (2.2). Similarly, $\psi - u_{\mathcal{S}} \le e^- \le w \le 0$ implies $\|w\|_{0,\omega_P} \le \|\psi - u_{\mathcal{S}}\|_{0,\omega_P}$, and, in view of $(\psi - u_{\mathcal{S}})(P) = 0$, a 'scaled' Poincaré-Friedrichs inequality provides

$$\|w\|_{0,\omega_P} \le \|\psi - u_{\mathcal{S}}\|_{0,\omega_P} \lesssim h_P \|\nabla(\psi - u_{\mathcal{S}})\|_{0,\omega_P}.$$

As consequence of (3.41), we immediately get

$$\|\Pi_P w\|_{0,\omega_P} \le \sum_{E\in\mathcal{E}_P} |\alpha_E(w)| \|\phi_E\|_{0,\omega_P} \lesssim h_P \|\nabla(\psi - u_{\mathcal{S}})\|_{0,\omega_P}.$$

Combining these two estimates we obtain the following substitute for (3.27)

$$\|w - \Pi_P w\|_{0,\omega_P} \lesssim h_P \|\nabla(\psi - u_{\mathcal{S}})\|_{0,\omega_P}. \tag{3.42}$$

We insert the bounds (3.40) for $E \in \mathcal{E}_P^0$, (3.41) for $E \in \mathcal{E}_P^+$ and (3.42) into (3.39), to obtain

$$\rho_P(e^-) \lesssim \sum_{E\in\mathcal{E}_P^0} |\rho_E| d_E + \Big(\sum_{E\in\mathcal{E}_P^+} |\rho_E| + h_P \|f\|_{0,\omega_P} \Big) \|\nabla(\psi - u_{\mathcal{S}})\|_{0,\omega_P}. \tag{3.43}$$

Finally, we sum up the two estimates (3.33) and (3.43), to obtain the desired bound (3.38).

Case 6: $P \in \mathcal{N}^{0-}$. In this case, we have by definition that $e = u - \psi \ge 0$, $f \le 0$ in ω_P, and $j_E \le 0$ for all $E \in \mathcal{E}_P$. Hence

$$\rho_P(e) = \int_{\omega_P} f e \phi_P + \sum_{E\in\mathcal{E}_P} \int_E j_E e \phi_P \le 0.$$

To conclude the proof, we sum the estimates for the six cases, invoke the definition of the oscillation term, and apply the Cauchy-Schwarz inequality, to obtain

$$C\rho_{\mathcal{S}}(e) \le \sum_{E\in\mathcal{E}^0} |\rho_E| d_E + \Big(\sum_{E\in\mathcal{E}^+} |\rho_E|^2 + \mathrm{osc}_2(u_{\mathcal{S}}, \psi, f)^2 \Big)^{\frac{1}{2}} \Big(\|\nabla e\|_{0,\Omega}^2 + \mathrm{osc}_1(u_{\mathcal{S}}, \psi)^2 \Big)^{\frac{1}{2}}.$$

The constant $C > 0$ depends only on the shape regularity of $\mathcal{T}$ and we have set $\mathcal{E}^+ = \cup_{P\in\mathcal{N}} \mathcal{E}_P^+$, $\mathcal{E}^0 = \cup_{P\in\mathcal{N}} \mathcal{E}_P^0$. Then, using Young's inequality and inserting the definition (2.12) of η_E, we get

$$C\rho_{\mathcal{S}}(e) \le \frac{\theta}{2} \Big(\|\nabla e\|_{0,\Omega}^2 + \mathrm{osc}_1(u_{\mathcal{S}}, \psi)^2 \Big) + \Big(1 + \frac{1}{2\theta} \Big) \Big(\sum_{E\in\mathcal{E}} |\rho_E| \eta_E + \mathrm{osc}_2(u_{\mathcal{S}}, \psi, f)^2 \Big)$$

for arbitrary $\theta > 0$. In view of the first inequality in (3.1), we can chose $\theta \le C$, to finally prove the assertion (3.22). □

Combining Lemma 3.1 and Proposition 3.3 yields our main result.

Theorem 3.4. *Assume that the obstacle* ψ *satisfies condition* (2.2). *Then the hierarchical a posteriori error estimate* $\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ *defined in* (2.10) *provides the upper bound for the energy error*

$$\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) \lesssim -\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}) + \mathrm{osc}(u_{\mathcal{S}}, \psi, f)^2 \tag{3.44}$$

up to the oscillation term defined in (3.11) *and a constant depending only on the shape regularity of* $\mathcal{T}$.

Proof. We estimate

$$\begin{aligned}\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) = -\mathcal{I}(e) \le \rho_{\mathcal{S}}(e) &\lesssim \sum_{E\in\mathcal{E}} |\rho_E|\eta_E + \operatorname{osc}(u_{\mathcal{S}}, \psi, f)^2 \\ &= \rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}}) + \operatorname{osc}(u_{\mathcal{S}}, \psi, f)^2 \le -2\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}) + \operatorname{osc}(u_{\mathcal{S}}, \psi, f)^2\end{aligned}$$

with the help of (2.6), Lemma 3.1, Proposition 3.3, (3.5), and (3.4). □

In light of the discussion in Section 3.2, the oscillation term $\operatorname{osc}(u_{\mathcal{S}}, \psi, f)$ is expected to be of higher order for suitable data so that $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ is asymptotically reliable. Moreover, if no obstacle is present, then the upper bound (3.44) reduces to well-known hierarchical a posterior error estimates and data oscillation for linear elliptic problems [4]. In this case, the above derivation provides a direct proof which does not invoke other a posteriori error estimates.

We conclude this subsection by an a posteriori estimate of the discretization error which is closely related to the estimator (2.12) proposed in [12].

Theorem 3.5. *Assume that the obstacle* ψ *satisfies condition* (2.2). *Then the localized discrete defect problem* (2.9) *provides the upper bound for the discretization error*

$$\|u - u_{\mathcal{S}}\| \lesssim \Big(\sum_{E\in\mathcal{E}} |\rho_E|\eta_E\Big)^{1/2} + \operatorname{osc}(u_{\mathcal{S}}, \psi, f) \tag{3.45}$$

up to the oscillation term defined in (3.11) *and a constant depending only on the shape regularity of* $\mathcal{T}$.

Proof. We estimate

$$\tfrac{1}{2}\|u - u_{\mathcal{S}}\|^2 \le \rho_{\mathcal{S}}(e) \lesssim \sum_{E\in\mathcal{E}} |\rho_E|\eta_E + \operatorname{osc}(u_{\mathcal{S}}, \psi, f)^2$$

by means of Lemma 3.1 and Proposition 3.3. □

Note that the corresponding error estimate $\sum_{E\in\mathcal{E}} |\rho_E|\eta_E$ differs from the hierarchical error estimate (2.12) only for internal edges $E \in \mathcal{E}$ with the property $\rho_E < -d_E < 0$. These edges are contained in the set $\cup_{P\in\mathcal{N}^+\setminus\mathcal{N}^{++}}\omega_P$, which, according to the discussion in Section 3.2, is expected to shrink to the exact free boundary under refinement. Therefore, Theorem 3.5 provides theoretical support for the numerical evidence that the estimator (2.12) is asymptotically reliable.

3.4. **Efficiency.** In this subsection, it is shown that $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ provides a lower bound for $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)$ up to a constant that is explicitely known. To this end, we combine (3.4) with results from [18, 20], which rely on the convexity of $\mathcal{J}$.

Theorem 3.6. *Assume that the obstacle* ψ *satisfies condition* (2.2). *Then the hierarchical a posteriori error estimate* $\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ *defined in* (2.10) *provides the lower bound for the energy error*

$$-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}) \le 6\big(\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)\big).$$

Proof. Combining (3.4) and (3.5), we get

$$-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}) \le \rho_{\mathcal{S}}(\varepsilon_{\mathcal{V}}) = \sum_{E\in\mathcal{E}} \eta_E|\rho_E|.$$

By a variant of [18, Theorem 3.2] (slightly modify the end of the proof to avoid the sums in the claim therein), we have

$$\sum_{E\in\mathcal{E}} \eta_E|\rho_E| \le 6\big(\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)\big).$$

□

4. Verification of a saturation assumption

The purpose of this section is to prove the following variant of the saturation assumption. The quadratic finite element approximation $u_{\mathcal{Q}}$, determined by

$$u_{\mathcal{Q}} \in K_{\mathcal{Q}}: \qquad a(u_{\mathcal{Q}}, v - u_{\mathcal{Q}}) \ge (f, v - u_{\mathcal{Q}}) \quad \forall v \in K_{\mathcal{Q}} \tag{4.1}$$

with

$$K_{\mathcal{Q}} = u_{\mathcal{S}} + \mathcal{A}_{\mathcal{Q}} = \{v \in \mathcal{Q} \mid v(P) \ge \psi(P) \;\; \forall P \in \mathcal{N}_{\mathcal{Q}} \cap \Omega\},$$

satisfies the inequality

$$\mathcal{J}(u_{\mathcal{Q}}) - \mathcal{J}(u) \le \alpha\big(\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)\big) \tag{4.2}$$

with some $\alpha \in (0,1)$, provided that the oscillation term $\operatorname{osc}(u_{\mathcal{S}}, f, \psi)$ is relatively small. Recall that we have $\mathcal{J}(u_{\mathcal{Q}}) - \mathcal{J}(u) = \frac{1}{2}\|u_{\mathcal{Q}} - u\|^2$ and $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) = \frac{1}{2}\|u_{\mathcal{S}} - u\|^2$, if no obstacle is present. Hence, (4.2) can be regarded as a generalization of related results for variational equalities [10]. However, in contrast to the unconstrained case, $\mathcal{J}(u_{\mathcal{Q}}) - \mathcal{J}(u)$ might be negative, because, in general, $K_{\mathcal{Q}}$ is not contained in K. Hence, for obstacle problems (4.2) does not imply that $\mathcal{J}(u_{\mathcal{Q}})$ is more accurate than $\mathcal{J}(u_{\mathcal{S}})$.

The proof of the saturation assumption will be based on the following simple observation.

Lemma 4.1. *Let $\alpha \in (0,1)$. Then the saturation assumption (4.2) holds, if and only if $-\mathcal{I}(e_{\mathcal{Q}})$ is reliable in the sense that*

$$\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) \le \frac{-\mathcal{I}(e_{\mathcal{Q}})}{1-\alpha}. \tag{4.3}$$

Proof. As $e_{\mathcal{Q}}$ and $u_{\mathcal{Q}}$ solve (2.7) and (4.1), respectively, we have $u_{\mathcal{Q}} = u_{\mathcal{S}} + e_{\mathcal{Q}}$. The resulting identity

$$\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) = -\mathcal{I}(e_{\mathcal{Q}}) + \big(\mathcal{J}(u_{Q}) - \mathcal{J}(u)\big).$$

implies the assertion. □

As a consequence of Lemma 4.1, we may prove (4.2) by verifying (4.3). To this end, we note a useful relation between $-\mathcal{I}(e_{\mathcal{Q}})$ and $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$.

Lemma 4.2. *There holds*

$$\mathcal{I}(e_{\mathcal{Q}}) \lesssim \mathcal{I}_{Q}(\varepsilon_{Q}) \le \mathcal{I}_{Q}(\varepsilon_{\mathcal{V}}) \le 0.$$

Proof. The inequalities $\mathcal{I}_{Q}(\varepsilon_{Q}) \le \mathcal{I}_{Q}(\varepsilon_{\mathcal{V}}) \le 0$ follow directly from $0 \in \mathcal{A}_{\mathcal{V}} \subset \mathcal{A}_{\mathcal{Q}}$, respectively. Using (3.2) and (3.3), we get

$$\mathcal{I}(e_{Q}) \le -\tfrac{1}{2}\rho_{\mathcal{S}}(e_{\mathcal{Q}}), \qquad -\rho_{\mathcal{S}}(\varepsilon_{Q}) \le \mathcal{I}_{Q}(\varepsilon_{Q}),$$

so that $\mathcal{I}(e_{Q}) \lesssim \mathcal{I}_{Q}(\varepsilon_{Q})$ follows from $\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}}) \lesssim \rho_{\mathcal{S}}(e_{\mathcal{Q}})$.

In order to show $\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}}) \lesssim \rho_{\mathcal{S}}(e_{\mathcal{Q}})$, we proceed similarly as in the proof of [13, Theorem 4.1]. Choosing $v = \varepsilon_{\mathcal{Q}} \in \mathcal{A}_{\mathcal{Q}}$ in (2.7) and utilizing (3.2), we get

$$\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}) \leq a(e_{\mathcal{Q}}, \varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}) \leq \|e_{\mathcal{Q}}\|\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\| \leq \rho_{\mathcal{S}}(e_{\mathcal{Q}})^{1/2}\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\|. \tag{4.4}$$

To bound $\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\|$ we combine the first inequality in (4.4) and (2.8) with $v = e_{\mathcal{Q}} \in \mathcal{A}_{\mathcal{Q}}$ to obtain

$$-a(e_{\mathcal{Q}}, \varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}) \leq \rho_{\mathcal{S}}(e_Q - \varepsilon_Q) \leq a_{\mathcal{Q}}(\varepsilon_{\mathcal{Q}}, e_{\mathcal{Q}} - \varepsilon_{\mathcal{Q}}).$$

In combination with the equivalence of the energy norms $\|\cdot\|$ and $\|\cdot\|_{\mathcal{Q}}$, see, e.g., [4, 9], we derive

$$\begin{aligned}\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\|^2 &= a(\varepsilon_{\mathcal{Q}}, \varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}) - a(e_{\mathcal{Q}}, \varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}) \leq a(\varepsilon_{\mathcal{Q}}, \varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}) + a_{\mathcal{Q}}(\varepsilon_{\mathcal{Q}}, e_{\mathcal{Q}} - \varepsilon_{\mathcal{Q}}) \\ &\leq \|\varepsilon_{\mathcal{Q}}\|\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\| + \|\varepsilon_{\mathcal{Q}}\|_{\mathcal{Q}}\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\|_{\mathcal{Q}} \lesssim \|\varepsilon_{\mathcal{Q}}\|_{\mathcal{Q}}\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\|.\end{aligned}$$

In light of (3.3), we have shown

$$\|\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}\| \lesssim \|\varepsilon_{\mathcal{Q}}\|_{\mathcal{Q}} \leq \rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}})^{1/2}.$$

Inserting this bound into (4.4), we get

$$\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}}) = \rho_{\mathcal{S}}(e_{\mathcal{Q}}) + \rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}} - e_{\mathcal{Q}}) \leq \rho_{\mathcal{S}}(e_{\mathcal{Q}}) + C\rho_{\mathcal{S}}(e_{\mathcal{Q}})^{1/2}\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}})^{1/2}$$

with a constant $C > 0$ depending only on the shape regularity of $\mathcal{T}$. Invoking Young's inequality, we finally obtain

$$\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}}) \leq \big(1 + \tfrac{C^2}{2}\big)\rho_{\mathcal{S}}(e_{\mathcal{Q}}) + \tfrac{1}{2}\rho_{\mathcal{S}}(\varepsilon_{\mathcal{Q}})$$

which proves the assertion. □

After these preparations we are ready to prove the main result of this section.

Theorem 4.3. *There are constants $C > 0$ and $\alpha \in (0, 1)$ depending only on the shape regularity of $\mathcal{T}$, such that relatively small oscillation*

$$\operatorname{osc}(u_{\mathcal{S}}, f, \psi)^2 \leq C\big(\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)\big) \tag{4.5}$$

implies the saturation assumption

$$\mathcal{J}(u_{\mathcal{Q}}) - \mathcal{J}(u) \leq \alpha\big(\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)\big). \tag{4.6}$$

Proof. Theorem 3.4 and Lemma 4.2 yield

$$\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u) \leq -C_1\mathcal{I}(e_{\mathcal{Q}}) + C_2 \operatorname{osc}(u_{\mathcal{S}}, f, \psi)^2,$$

where C_1 and C_2 are constants depending only on the shape regularity of $\mathcal{T}$, and we may assume $C_1 \geq 1$. Hence, selecting

$$C = \frac{1 - C_1(1-\alpha)}{C_2}, \qquad \alpha \in \left(\frac{C_1 - 1}{C_1}, 1\right),$$

the assertion is a consequence of Lemma 4.1. In particular, one might chose $\alpha = (2C_1 - 1)/(2C_1)$. □

In view of the discussion in Section 3.2, condition (4.5) is expected to hold for sufficiently smooth data and sufficiently fine global refinements of a triangulation with (2.2).

Utilizing Theorem 3.6, the condition (4.5) follows from

$$\operatorname{osc}(u_{\mathcal{S}}, f, \psi)^2 \leq -\frac{C}{6}\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}}), \tag{4.7}$$

with the same constant C. Apart from the constant C, which, in turn, depends on C_1, C_2 from Theorem 3.4 and Lemma 4.2, all quantities in (4.7) are computable. Thus, for given upper bounds for C_1, the condition (4.5) can be verified in practice.

5. Numerical Example

Following [16], we consider the piecewise affine, concave obstacle

$$\psi(x) = \operatorname{dist}(x, \partial\Omega) - \tfrac{1}{5},$$

the domain $\Omega = \{(x_1, x_2) \in \mathbb{R} \mid |x_1| + |x_2| < 1\}$, and the constant load $f = -5$. The triangulations $\mathcal{T}_j$, $j = 1, \ldots, 9$, are obtained by uniform refinement of an initial triangulation $\mathcal{T}_0$, consisting of four congruent triangles. Observe that ψ is piecewise affine over $\mathcal{T}_j$ for all $j = 1, \ldots, 9$. As the exact solution u is not explicitely known, we use the finite element approximation $\tilde{u}$ on level $\mathcal{T}_{11}$ as a substitute. The left picture in Figure 1 shows the 'exact' energy error $\mathcal{J}(u_{\mathcal{S}_j}) - \mathcal{J}(\tilde{u})$ in comparison with our hierarchical a posteriori error estimate $\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}_j})$ and the oscillation term $\operatorname{osc}(u_{\mathcal{S}_j}, \psi, f)$ over the number of unknowns. Both the exact error and the estimator are proportional to h. More precisely, the 'exact' error is asymptotically overestimated by a factor of about 1.5. Similar to the unconstrained case, the oscillation term initially dominates, but vanishes with higher (second) order under refinement. For an explanation, first note that the set $\mathcal{N}^{0+}$ of isolated contact nodes is empty in this example. Hence, $\operatorname{osc}(u_{\mathcal{S}_j}, \psi, f) = \operatorname{osc}_2(u_{\mathcal{S}_j}, \psi, f)$. The set $\bigcup_{(\mathcal{N}_j \setminus (\mathcal{N}_j^{0-} \cup \mathcal{N}_j^{++})} \omega_P$, whose contributions to $\operatorname{osc}_2(u_{\mathcal{S}_j}, \psi, f)$ do not involve local means is depicted in the right picture of Figure 1 for the final level $j = 9$. Obviously, it concentrates at $\partial\Omega$ and at the free boundary, confirming nicely our heuristic reasoning in Section 3.2.

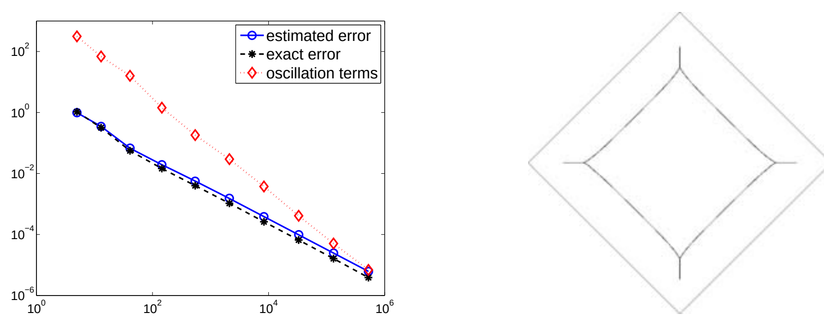

Figure 1. Comparison of the hierarchical error estimator $-\mathcal{I}_{\mathcal{Q}}(\varepsilon_{\mathcal{V}})$ with the exact error $\mathcal{J}(u_{\mathcal{S}}) - \mathcal{J}(u)$ and the oscillation term $\operatorname{osc}(u_{\mathcal{S}_j}, \psi, f)$ (left). The set $\bigcup_{(\mathcal{N}_j \setminus \mathcal{N}_j^{0-}) \setminus \mathcal{N}_j^{++}} \omega_P$, $j = 9$, whose contributions to $\operatorname{osc}_2(u_{\mathcal{S}_j}, \psi, f)$ do not involve local means (right).

References

[1] M. Ainsworth and J.T. Oden. *A posteriori error estimation in finite element analysis.* Wiley, 2000.

[2] R.E. Bank and R.K. Smith. A posteriori error estimates based on hierarchical bases. *SIAM J. Num. Anal.*, 30:921–935, 1993.

[3] S. Bartels and C. Carstensen. Averaging techniques yield reliable a posteriori finite element error control for obstacle problems. *Numer. Math.*, 99(2):225–249, 2004.

[4] F.A. Bornemann, B. Erdmann, and R. Kornhuber. A posteriori error estimates for elliptic problems in two and three space dimensions. *SIAM J. Numer. Anal.*, 33:1188–1204, 1996.

[5] D. Braess. A posteriori error estimators for obstacle problems – another look. *Numer. Math.*, 101:415–421, 2005.

[6] D. Braess, R.H.W. Hoppe, and J. Schöberl. A posteriori estimator for obstacle problems by the hypercircle method. *Submitted to Comp. Visu. Sci.*, 2008.

[7] F. Brezzi and L.A. Caffarelli. Convergence of the discrete free boundaries for finite element approximations. *RAIRO Numer. Anal.*, 17:385–395, 1983.

[8] F. Brezzi, W.W. Hager, and P.A. Raviart. Error estimates for the finite element solution of variational inequalities I. *Numer. Math.*, 28:431–443, 1977.

[9] P. Deuflhard, P. Leinen, and H. Yserentant. Concepts of an adaptive hierarchical finite element code. *IMPACT Comput. Sci. Engrg.*, 1:3–35, 1989.

[10] W. Dörfler and R.H. Nochetto. Small data oscillation implies the saturation assumption. *Numer. Math.*, 91:1–12, 2002.

[11] F. Fierro and A. Veeser. A posteriori error estimators for regularized total variation of characteristic functions. *SIAM J. Numer. Anal.*, 41(6):2032–2055 (electronic), 2003.

[12] R.H.W. Hoppe and R. Kornhuber. Adaptive multilevel–methods for obstacle problems. *SIAM J. Numer. Anal.*, 31(2):301–323, 1994.

[13] R. Kornhuber. A posteriori error estimates for elliptic variational inequalities. *Computers Math. Applic.*, 31:49–60, 1996.

[14] R. Kornhuber. *Adaptive Monotone Multigrid Methods for Nonlinear Variational Problems*. Teubner, Stuttgart, 1997.

[15] R. Kornhuber and Q. Zou. Efficient and reliable hierarchical error estimates for the discretization error of elliptic obstacle problems. Preprint 519, DFG Research Center Matheon, 2008.

[16] R.H. Nochetto, K.G. Siebert, and A. Veeser. Pointwise a posteriori error control for elliptic obstacle problems. *Numer. Math.*, 95:631–658, 2003.

[17] O. Sander. *Multidimensional Coupling in a Human Knee Model*. PhD thesis, FU Berlin, 2008.

[18] K. G. Siebert and A. Veeser. A unilaterally constrained quadratic minimization with adaptive finite elements. *SIAM J. Optim.*, 18:260–289, 2007.

[19] A. Veeser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems. *SIAM J. Numer. Anal.*, 39(1):146–167, 2001.

[20] A. Veeser. Convergent adaptive finite elements for the nonlinear Laplacian. *Numer. Math.*, 92(4):743–770, 2002.

[21] A. Veeser and R. Verfürth. Poincaré constant for non-convex domains. in preparation.

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

[23] O.C. Zienkiewicz, J.P. De S.R. Gago, and D.W. Kelly. The hierarchical concept in finite element analysis. *Computers & Structures*, 16:53–65, 1983.