SIAM J. NUMER. ANAL.
Vol. 49, No. 1, pp. 346–367

ON THE CONVERGENCE OF ADAPTIVE NONCONFORMING FINITE ELEMENT METHODS FOR A CLASS OF CONVEX VARIATIONAL PROBLEMS*

CHRISTOPH ORTNER† AND DIRK PRAETORIUS‡

Abstract. We formulate and analyze an adaptive nonconforming finite element method for the solution of convex variational problems. The class of minimization problems we admit includes highly singular problems for which no Euler–Lagrange equation (or inequality) is available. As a consequence, our arguments only use the structure of the energy functional. We are nevertheless able to prove convergence of an adaptive algorithm, using even refinement indicators that are not reliable error indicators.

Key words. calculus of variations, Crouzeix–Raviart finite element method, convergence of adaptive finite element methods, Lavrentiev phenomenon

AMS subject classifications. 49J45, 65N30, 49M25

DOI. 10.1137/090781073

1. Introduction.

We present a new convergence proof for adaptive nonconforming finite element methods, which is applicable to a wide class of convex variational problems.

For fixed $n \geq 2$ and $m \geq 1$, let Ω be a bounded polygonal domain in $\mathbb{R}^n$ and let $W : \mathbb{R}^{m\times n} \to [0, +\infty]$ be a convex stored energy function, which satisfies a p-growth condition from below, i.e.,

$$W(\xi) \geq C(|\xi|^p - 1) \quad \text{for all } \xi \in \mathbb{R}^{m\times n} \text{ and some } p > 1. \tag{1}$$

For a fixed dead load $f \in \mathrm{L}^{p'}(\Omega)$, where $1/p'+1/p = 1$, we define the energy functional $\mathcal{J} : \mathrm{W}^{1,p}(\Omega)^m \to \mathbb{R} \cup \{+\infty\}$ by

$$\mathcal{J}(v) = \int_\Omega \big(W(\nabla v) - f \cdot v\big)\,\mathrm{d}x. \tag{2}$$

Given some $g \in \mathrm{W}^{1,p}(\Omega)^m$ with $\mathcal{J}(g) < +\infty$ and sets $\Gamma^{(i)} \subset \partial\Omega$, $i = 1, \ldots, m$, with positive surface measure $|\Gamma^{(i)}| > 0$, we define the *admissible set*

$$\mathcal{A} = \big\{v \in \mathrm{W}^{1,p}(\Omega)^m \,:\, v^{(i)} = g^{(i)} \text{ on } \Gamma^{(i)}, i = 1, \ldots, m\big\} \neq \emptyset; \tag{3}$$

here and throughout we use superscripts to denote components of a vector-valued function. We note that admissible functions satisfy the Poincaré-type inequality

$$\|v - g\|_{\mathrm{L}^p(\Omega)} \leq C_p \|\nabla v - \nabla g\|_{\mathrm{L}^p(\Omega)} \quad \text{for all } v \in \mathcal{A}, \tag{4}$$

where C_p is some fixed constant.

*Received by the editors December 23, 2009; accepted for publication (in revised form) November 29, 2010; published electronically February 22, 2011.

http://www.siam.org/journals/sinum/49-1/78107.html

†Mathematical Institute, University of Oxford, 24–29 St. Giles', Oxford OX1 3LB, UK (ortner@maths.ox.ac.uk). This author's research was supported by the EPSRC Critical Mass Programme "New Frontiers in the Mathematics of Solids."

‡Institute for Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstraße 8-10, A-1040 Wien, Austria (Dirk.Praetorius@tuwien.ac.at).

In this paper, we analyze the numerical solution of the minimization problem

$$u \in \operatorname{argmin} \mathcal{J}(\mathcal{A}) \tag{5}$$

by means of an adaptive nonconforming finite element method.

Under the conditions we imposed, the functional $\mathcal{J}$ may be nondifferentiable at a solution, even if W itself is smooth [3], and hence the associated Euler–Lagrange equation is unavailable to us. In this situation, the method of choice for the analysis of (5) (and its discretization) is the *direct method of the calculus of variations* [16], whose application immediately gives the following basic existence result. Since it is helpful to understand our subsequent analysis, we give a brief outline of the proof. We refer to Dacorogna's monograph [16] (in particular Theorem 4.1) for further details.

PROPOSITION 1. *There exists at least one solution of* (5).

Proof. For an admissible function $v \in \mathcal{A}$, the Poincaré inequality (4) implies

$$\|v\|_{\mathrm{L}^p(\Omega)} \lesssim \|\nabla(v-g)\|_{\mathrm{L}^p(\Omega)} + \|g\|_{\mathrm{L}^p(\Omega)} \lesssim \|\nabla v\|_{\mathrm{L}^p(\Omega)} + \|g\|_{\mathrm{W}^{1,p}(\Omega)},$$

where $a \lesssim b$ abbreviates $a \le C\, b$ with a generic constant $C > 0$. From the p-growth condition (1) and Young's inequality, we infer

$$\mathcal{J}(v) \gtrsim \|\nabla v\|^p_{\mathrm{L}^p(\Omega)} - \|f\|_{\mathrm{L}^{p'}(\Omega)}\|v\|_{\mathrm{L}^p(\Omega)} - |\Omega| \gtrsim \|\nabla v\|^p_{\mathrm{L}^p(\Omega)} - C(f,g,\Omega)$$

with an additive constant $C(f,g,\Omega) \ge 0$. Applying the Poincaré-type inequality (4) again, we find that $\mathcal{J}$ is coercive, i.e.,

$$\|v\|^p_{\mathrm{W}^{1,p}(\Omega)} \lesssim \mathcal{J}(v) + 1 \quad \text{for all } v \in \mathcal{A}. \tag{6}$$

Suppose now that $(u_\ell)_{\ell\in\mathbb{N}} \subset \mathcal{A}$ is a minimizing sequence for $\mathcal{J}$, i.e., $\mathcal{J}(u_\ell) \to \inf \mathcal{J}(\mathcal{A})$ as $\ell \to \infty$. By (6) and reflexivity of $\mathrm{W}^{1,p}(\Omega)^m$, we may assume that $(u_\ell)_{\ell\in\mathbb{N}}$ converges weakly to some limit u in $\mathrm{W}^{1,p}(\Omega)^m$. Since $\mathcal{A}$ is convex and closed, it is weakly closed, and hence it follows that $u \in \mathcal{A}$. Finally, convexity and nonnegativity of W imply the weak lower semicontinuity of $\mathcal{J}$, i.e.,

$$\mathcal{J}(u) \le \liminf_{\ell\to\infty} \mathcal{J}(u_\ell);$$

cf. [16, Theorem 3.4]. Thus, we conclude that $\mathcal{J}(u) = \inf \mathcal{J}(\mathcal{A})$, i.e., that $u \in \operatorname{argmin} \mathcal{J}(\mathcal{A})$. ∎

The class of problems included in our analysis is surprisingly general. It includes not only standard variational model problems such as the Dirichlet problem [11, 14, 19, 23], or the p-Laplacian [18, 26], or convex problems with control of stresses [4, 10] (all of these works require uniform p-growth from below *and* above), but is in fact suited for any convex functional of the type (2). Even our p-growth condition from below can be relaxed to some extent [25]. For simplicity we have restricted our presentation to Dirichlet constraints, but this constraint is easily lifted as well (cf. section 5).

One of the most important features of our analysis is that we do not require $\mathcal{J}$ to be continuous in the strong topology of $\mathrm{W}^{1,p}(\Omega)^m$ and hence, using ideas from [25], we are even able to address problems that exhibit a *Lavrentiev gap phenomenon*. If $g \in \mathrm{W}^{1,\infty}(\Omega)^m$, then $\mathcal{A}_\infty := \mathcal{A} \cap \mathrm{W}^{1,\infty}(\Omega)^m$ is nonempty, and we say that (5) exhibits the Lavrentiev gap phenomenon if

$$\inf \mathcal{J}(\mathcal{A}_\infty) > \min \mathcal{J}(\mathcal{A}). \tag{7}$$

Foss, Hrusa, and Mizel [20] have shown that this effect can indeed occur under the conditions we have posed. Note that, since typical conforming finite element functions are Lipschitz continuous, nonconformity of the numerical method is essential for treating problems in that class. Furthermore, since the Lavrentiev effect is closely linked to singularities in the solution of nonlinear variational problems, adaptive solution techniques are particularly important. We refer to [3, 13, 25] for overviews of this fascinating subdiscipline of the calculus of variations.

To the best of our knowledge, three classes of convergence proofs for adaptive finite element methods for linear problems exist to date. The first idea [19, 23] used the so-called inner node property from which a lower bound on the discrete error in terms of the estimator can be derived. This implies, up to oscillation terms, the reduction of the error at each step of the adaptive algorithm. A further step was taken in [24] by circumventing the inner node property at a given step but still requiring that it should be obtained after a fixed number of refinements. The first proof that completely circumvented the use of lower bounds but only requires reliability of the error estimator is given in [14]. Extensions of this convergence analysis to the nonlinear Laplacian, or more general convex variational problems, can be found in [4, 10, 18, 26]. These works require the use of some additional structure of the problems but are still heavily based on the analysis for the Dirichlet problem.

Proofs of the convergence of adaptive nonconforming and mixed finite element methods can be found, for example, in [11, 12]. The analysis contained therein is largely adapted from the conforming case, the crucial modification being a control of nonconformity, which leads to so-called quasi-Galerkin orthogonality relations.

In particular, all previous proofs require the use of Euler–Lagrange equations, and of reliable a posteriori error estimates, both of which are not available for our problem class. We therefore have to use a completely different approach. Our convergence argument is based on the direct method of the calculus of variations, and is strictly tailored to nonconforming finite element methods. It does not apply in an obvious way to the conforming case. In addition to the lack of Euler–Lagrange equations in our problem, it is also interesting to note that we obtain convergence of our adaptive algorithm even though the refinement indicators we use do not provide reliable error bounds.

The first part of our convergence proof (section 3.1) is motivated by [25] and, to the best of our knowledge, uses techniques that have not previously been employed in the context of adaptive finite element methods. The second part (Theorem 8) is closely related to, and in fact inspired by, the proof given in [24].

Finally, we remark that our main result, Theorem 8, does not include the traditional residual-based a posteriori error estimator. We show in section 3.4 that this is not a restriction of our analysis, but that due to the generality of our assumptions on W, the stated result is false in the "missing" case.

The paper is organized as follows. In section 2, we fix the notation and state some auxiliary results. In section 3.1, we provide sufficient conditions for the convergence of the Crouzeix–Raviart finite element method. This analysis motivates the definition of convergence indicators, which we discuss in some detail in section 3.2. Finally, in section 3.3 we formulate an adaptive algorithm and prove its convergence. To conclude, we present several numerical experiments in section 4.

2. Preliminaries.

2.1. Function spaces. Let A be an open subset of $\mathbb{R}^n$. We use $\mathrm{L}^p(A)$ and $\mathrm{W}^{1,p}(A)$ to denote the standard Lebesgue and Sobolev spaces and equip them with their usual norms. The space of distributions is denoted by $\mathcal{D}'(A)$ [2]. The distributional gradient operator is denoted D, while the weak gradient operator is denoted ∇. The space of k-times continuously differentiable functions with compact support in A is denoted $\mathrm{C}_0^k(A)$.

2.2. Triangulation of Ω. For every $\ell \in \mathbb{N}$, we assume that $\mathcal{T}_\ell$ is a regular triangulation of $\bar{\Omega}$ into closed simplices $T \in \mathcal{T}_\ell$. In particular, $\mathcal{T}_\ell$ has no hanging nodes in 2D, no hanging nodes or edges in 3D, and so forth. Let $\mathcal{E}_\ell$ denote the collection of $n-1$ dimensional faces of elements and $\mathcal{E}_\ell^{\mathrm{int}}$ the collection of interior faces. Let $\mathcal{N}_\ell$ denote the set of vertices and $\mathcal{N}_\ell^{\mathrm{nc}}$ denote the set of barycenters of the faces. We assume throughout that, up to surface measure zero, the sets $\Gamma^{(i)}$ are the union of faces $\mathcal{E}_\ell^{d,i} \subset \mathcal{E}_\ell$, $i = 1, \ldots, m$, and we set $\mathcal{E}_\ell^{\mathrm{int},i} = \mathcal{E}_\ell \setminus \mathcal{E}_\ell^{d,i}$.

For each element $T \in \mathcal{T}_\ell$, we set $h_T = \mathrm{diam}(T)$. For each face $E \in \mathcal{E}_\ell$, we set $h_E = \mathrm{diam}(E)$. The mesh-size function $h_\ell : \Omega \to \mathbb{R}_{>0}$ is then almost everywhere (a.e.) defined by $h_\ell(x) = h_T$ for x in the interior of an element $T \in \mathcal{T}_\ell$ and $h_\ell(x) = h_E$ for x in the relative interior of a face $E \in \mathcal{E}_\ell$.

The shape regularity constant $\sigma(\mathcal{T}_\ell)$ is the smallest number $C > 0$ such that

$$C^{-1}\, h_T^n \le |T| \le h_T^n, \quad C^{-1}\, h_E^{n-1} \le |E| \le h_E^{n-1}, \quad \text{and} \quad h_E \le h_T \le C\, h_E$$

for all elements $T \in \mathcal{T}_\ell$ and faces $E \in \mathcal{E}_\ell$ with $E \subset \partial T$. A family $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}}$ of regular triangulations is uniformly shape-regular if $\sup_{\ell\in\mathbb{N}} \sigma(\mathcal{T}_\ell) < \infty$.

2.3. Finite element spaces. We briefly describe the Crouzeix–Raviart finite element space used to discretize (5); see [6, 8, 15] for further details.

The space of all $\mathcal{T}_\ell$-piecewise affine functions is denoted by

$$\mathrm{P}_1(\mathcal{T}_\ell) = \big\{ v \in \mathrm{L}^1(\Omega) \,:\, v|_T \text{ is affine for all } T \in \mathcal{T}_\ell \big\}.$$

$\mathrm{CR}(\mathcal{T}_\ell)$ denotes the Crouzeix–Raviart finite element space,

$$\mathrm{CR}(\mathcal{T}_\ell) = \big\{ v_\ell \in \mathrm{P}_1(\mathcal{T}) \,:\, v_\ell \text{ is continuous at all face barycenters } z \in \mathcal{N}_\ell^{\mathrm{nc}} \big\}.$$

For each barycenter $z \in \mathcal{N}_\ell^{\mathrm{nc}}$, let $E_z \in \mathcal{E}_\ell$ be the unique face containing z. The interpolation operator $\Pi_\ell : \mathrm{W}^{1,1}(\Omega)^m \to \mathrm{CR}(\mathcal{T}_\ell)^m$ is defined via [15]

$$\Pi_\ell v(z) = |E_z|^{-1} \int_{E_z} v \,\mathrm{d}s \quad \text{for all } z \in \mathcal{N}_\ell^{\mathrm{nc}}$$

and thus satisfies $\int_E \Pi_\ell v \,\mathrm{d}s = \int_E v \,\mathrm{d}s$ for all faces $E \in \mathcal{E}_\ell$. We summarize its most important properties for our purpose in the following lemma. It is worth noting that neither the stability of Π_ℓ, nor the interpolation error estimate depend on the shape regularity $\sigma(\mathcal{T}_\ell)$. We remark that particularly the (well-known) mean value identity (9) lies at the heart of our analysis.

LEMMA 2 (properties of the Crouzeix–Raviart interpolant). *Let $v \in \mathrm{W}^{1,p}(\Omega)^m$ and $T \in \mathcal{T}_\ell$, then Π_ℓ has the first-order approximation property*

$$\|v - \Pi_\ell v\|_{\mathrm{L}^p(T)} \le C_{\mathrm{apx}} h_T \, \|\nabla v\|_{\mathrm{L}^p(T)} \tag{8}$$

with $C_{\mathrm{apx}} = 1 + 2/n \le 2$. Furthermore, it satisfies the mean value property

$$|T|^{-1} \int_T \nabla v \,\mathrm{d}x = |T|^{-1} \int_T \nabla(\Pi_\ell v) \,\mathrm{d}x = \nabla(\Pi_\ell v)|_T, \tag{9}$$

which implies the stability estimate

$$\|\nabla(\Pi_\ell v)\|_{\mathrm{L}^p(T)} \le \|\nabla v\|_{\mathrm{L}^p(T)}. \tag{10}$$

Proof. Since the outer unit normal ν to T is constant on each of its faces, integration by parts in T yields

$$\int_T \frac{\partial}{\partial x_j} v \, \mathrm{d}x = \int_{\partial T} v\nu_j \, \mathrm{d}s = \int_{\partial T} (\Pi_\ell v)\nu_j \, \mathrm{d}s = \int_T \frac{\partial}{\partial x_j} (\Pi_\ell v) \, \mathrm{d}x,$$

which proves (9). In particular, we have

$$\|\nabla(\Pi_\ell v)\|_{\mathrm{L}^p(T)} = |T|^{1/p-1} \left| \int_T \nabla v \, \mathrm{d}x \right| \le \|\nabla v\|_{\mathrm{L}^p(T)},$$

which establishes (10).

Next, we recall the well-known trace identity

$$\frac{1}{|E|} \int_E w \, \mathrm{d}s = \frac{1}{|T|} \int_T w \, \mathrm{d}x + \frac{1}{n|T|} \int_T (x-z) \cdot \nabla w \, \mathrm{d}x \quad \text{for all } w \in \mathrm{W}^{1,1}(T),$$

where $z \in T \cap \mathcal{N}$ is the vertex opposite to E, i.e., $T = \mathrm{conv}(E \cup \{z\})$. By definition, $w := v - \Pi_\ell v$ satisfies $\int_E w \, \mathrm{d}s = 0$. Therefore, the integral mean $w_T := |T|^{-1} \int_T w \, \mathrm{d}x$ can be estimated by

$$|w_T| = \frac{1}{|T|} \left| \int_T w \, \mathrm{d}x \right| \le \frac{h_T}{n|T|} \|\nabla w\|_{L^1(T)} \le \frac{h_T |T|^{1/p'-1}}{n} \|\nabla w\|_{\mathrm{L}^p(T)},$$

whence $\|w_T\|_{\mathrm{L}^p(T)} \le (h_T/n) \|\nabla w\|_{\mathrm{L}^p(T)}$. Next, we use the Poincaré inequality $\|w - w_T\|_{\mathrm{L}^p(T)} \le (h_T/2) \|\nabla w\|_{\mathrm{L}^p(T)}$ on the convex domain T. (The case $p = \infty$ is trivial. For the case $p = 1$, see [1]; the result for general p follows from Riesz's interpolation theorem.) This gives

$$\|w\|_{\mathrm{L}^p(T)} \le \|w - w_T\|_{\mathrm{L}^p(T)} + \|w_T\|_{\mathrm{L}^p(T)} \le \left(\frac{1}{2} + \frac{1}{n}\right) h_T \|\nabla w\|_{\mathrm{L}^p(T)}.$$

Hence, we can deduce (8) from (10) and a triangle inequality. $\square$

Since $v_\ell \in \mathrm{CR}(\mathcal{T}_\ell)^m$ may be discontinuous across faces $E \in \mathcal{E}_\ell$, v_ℓ is not weakly differentiable; nevertheless we use ∇v_ℓ to denote its $\mathcal{T}_\ell$-elementwise gradient. We also require a notation for the jumps across interior faces. For $E = T^+ \cap T^- \in \mathcal{E}_\ell^{\mathrm{int}}$, we fix the labeling of the elements $T^\pm$, we let $v_\ell^\pm$ denote the traces from $T^\pm$, and $\nu = \nu_E$ the outer unit normals to T^+. We define the jump across E by

$$[v_\ell] = v_\ell^+ - v_\ell^-.$$

For a boundary face $E \subset \partial\Omega$, $E = \partial\Omega \cap T$, we let $\nu = \nu_E$ be the outer unit normal to Ω, and we define

$$[v_\ell]^{(i)} = \begin{cases} v_\ell^{(i)} - g^{(i)} & \text{if } E \subset \Gamma^{(i)}, \\ 0 & \text{otherwise.} \end{cases}$$

With this notation, the distributional gradient reads

$$\langle Dv_\ell, \varphi\rangle = -\int_\Omega v_\ell \cdot \mathrm{div}\varphi \,\mathrm{d}x = \int_\Omega \nabla v_\ell : \varphi \,\mathrm{d}x - \int_{\cup \mathcal{E}_\ell^{\mathrm{int}}} ([v_\ell] \otimes \nu) : \varphi \,\mathrm{d}s \tag{11}$$

for $v_\ell \in \mathrm{P}_1(\mathcal{T}_\ell)^m$ and $\varphi \in \mathrm{C}_0^1(\Omega)^{m\times n}$. The representation formula (11) can be verified by integration by parts on each element. The symbol $\otimes$ denotes the tensor product $a \otimes b \in \mathbb{R}^{m\times n}$, i.e., $(a\otimes b)_{ij} = a_i b_j$.

3. Adaptive solution. Let $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}}$ be a uniformly shape-regular family of triangulations of Ω, which will subsequently be generated by an adaptive algorithm. We extend the definition of the energy functional $\mathcal{J}$ to the Crouzeix–Raviart finite element space $\mathrm{CR}(\mathcal{T}_\ell)^m$ by setting

$$\mathcal{J}(v_\ell) = \int_\Omega \Big[W(\nabla v_\ell) - f\cdot v_\ell\Big]\,\mathrm{d}x.$$

We stress, however, that ∇v_ℓ now denotes the $\mathcal{T}_\ell$-piecewise gradient.

Since $\mathrm{CR}(\mathcal{T}_\ell)$ is not a subspace of $\mathrm{W}^{1,p}(\Omega)$ we need to take care when defining the set of discrete admissible functions. A natural definition is to impose the Dirichlet condition on the face barycenters,

$$\mathcal{A}_\ell = \big\{v_\ell \in \mathrm{CR}(\mathcal{T}_\ell)^m : v_\ell^{(i)}(z) = \Pi_\ell g^{(i)}(z) \text{ for } z \in \Gamma^{(i)} \cap \mathcal{N}_\ell^{\mathrm{nc}}, i = 1,\dots,m\big\}. \tag{12}$$

We note that $\Pi_\ell v \in \mathcal{A}_\ell$, for each $v \in \mathcal{A}$, and hence, $\mathcal{A}_\ell$ is sufficiently rich to be a "good" approximation of $\mathcal{A}$.

The Crouzeix–Raviart finite element discretization of (5) is to find a minimizer

$$u_\ell \in \operatorname{argmin} \mathcal{J}(\mathcal{A}_\ell). \tag{13}$$

As in the continuous case, we have the following result.

PROPOSITION 3. *There exists at least one solution to* (13).

Proof. We simply adapt the proof of Proposition 1. As before, $\mathcal{A}_\ell$ is convex and closed, and $\mathcal{J}$ is (weakly) lower semicontinuous on the finite dimensional space $\mathrm{CR}(\mathcal{T}_\ell)^m$. It remains only to prove the coercivity of $\mathcal{J}$ in $\mathcal{A}_\ell$: a broken Poincaré inequality [9, Corollary 4.3] (for the case $p = 2$ see also [7]) provides

$$\|v_\ell\|_{\mathrm{L}^p(\Omega)} \le \|v_\ell - \Pi_\ell g\|_{\mathrm{L}^p(\Omega)} + \|\Pi_\ell g\|_{\mathrm{L}^p(\Omega)} \lesssim \|\nabla(v_\ell - \Pi_\ell g)\|_{\mathrm{L}^p(\Omega)} + \|\Pi_\ell g\|_{\mathrm{L}^p(\Omega)},$$

so that stability of Π_ℓ yields

$$\|v_\ell\|_{\mathrm{L}^p(\Omega)} \lesssim \|\nabla v_\ell\|_{\mathrm{L}^p(\Omega)} + \|g\|_{\mathrm{W}^{1,p}(\Omega)}.$$

With the same arguments as in the proof of Proposition 1, we obtain the discrete analogue of (6),

$$\|v_\ell\|_{\mathrm{L}^p(\Omega)}^p + \|\nabla v_\ell\|_{\mathrm{L}^p(\Omega)}^p \lesssim \mathcal{J}(v_\ell) + 1 \quad \text{for all } v_\ell \in \mathcal{A}_\ell. \tag{14}$$

Arguing as above, we conclude the proof. $\square$

3.1. Sufficient conditions for convergence. In this section, we derive conditions under which a sequence $(u_\ell)_{\ell\in\mathbb{N}}$ of discrete solutions converges to a solution of (5). Lemma 4 is the main observation that led to the convergence theorem for uniformly refined meshes [25, equation (31)] and will again play a prominent role here. Lemma 5 is a refinement of [25, Lemma 8], which allows us to adapt the convergence argument to adaptively refined meshes.

LEMMA 4 (upper bound). *For every $v \in \mathrm{W}^{1,p}(\Omega)^m$, it holds that*

$$\mathcal{J}(\Pi_\ell v) \le \mathcal{J}(v) + C_{\mathrm{apx}}\, \|h_\ell f\|_{\mathrm{L}^{p'}(\Omega)}\, \|\nabla v\|_{\mathrm{L}^p(\Omega)}. \tag{15}$$

Proof. Jensen's inequality yields $W\big(|T|^{-1}\int_T \nabla v\,\mathrm{d}x\big) \le |T|^{-1}\int_T W(\nabla v)\,\mathrm{d}x$. From the mean value property (9), we infer

$$\int_T W(\nabla \Pi_\ell v)\,\mathrm{d}x \le \int_T W(\nabla v)\,\mathrm{d}x,$$

and hence,

$$\mathcal{J}(\Pi_\ell v) \le \mathcal{J}(v) + \int_\Omega f\cdot(v - \Pi_\ell v)\,\mathrm{d}x.$$

Elementwise application of the approximation error estimate (8) results in (15). ∎

Remark 1. The upper bound (15) should be expected to be suboptimal. For example, for the standard Dirichlet problem (where $W(F) = |F|^2$), the energy error is formally of order $O(h^2)$ while the estimate in (15) is only of order $O(h)$. One reason for this is that (15) provides an upper bound for *any* v and not only for the energy minimum. However, since we cannot assume any differentiabiliy properties on $\mathcal{J}$ (this is due to the lack of a growth condition on W from above), there is little hope to recover an $O(h^2)$ estimate even if v is a minimizer of $\mathcal{J}$.

LEMMA 5 (compactness of sublevel sets). *Let $v_\ell \in \mathcal{A}_\ell$, $\ell \in \mathbb{N}$, be a sequence satisfying* $\sup_{\ell\in\mathbb{N}} \mathcal{J}(v_\ell) < \infty$, *and assume that*

$$\big\|h_\ell[v_\ell]\big\|_{\mathrm{L}^1(\cup\mathcal{E}_\ell)} \xrightarrow{\ell\to\infty} 0. \tag{16}$$

(We stress that the skeleton $\bigcup\mathcal{E}_\ell$ includes $\partial\Omega$.) Then, there exists a subsequence $(v_{\ell_k})_{k\in\mathbb{N}}$ and a limit $v \in \mathcal{A}$ such that

$$\begin{aligned} v_{\ell_k} &\rightharpoonup v && \textit{weakly in } \mathrm{L}^p(\Omega)^m,\\ \nabla v_{\ell_k} &\rightharpoonup \nabla v && \textit{weakly in } \mathrm{L}^p(\Omega)^{m\times n}. \end{aligned} \tag{17}$$

Proof. As in the proof of Proposition 3 (cf. (14)) boundedness of the energy gives

$$\sup_{\ell\in\mathbb{N}}\big(\|v_\ell\|_{\mathrm{L}^p(\Omega)} + \|\nabla v_\ell\|_{\mathrm{L}^p(\Omega)}\big) < \infty.$$

Since $\mathrm{L}^p(\Omega)$ is reflexive, we may assume without loss of generality that v_ℓ as well as ∇v_ℓ are weakly convergent with limits $v \in \mathrm{L}^p(\Omega)^m$ and $F \in \mathrm{L}^p(\Omega)^{m\times n}$. We now aim to show that v is weakly differentiable with $\nabla v = F$. To this end, we fix $\varphi \in \mathrm{C}_0^\infty(\Omega)^{m\times n}$ and recall the representation formula (11) to obtain

$$\langle Dv_\ell, \varphi\rangle = \int_\Omega \nabla v_\ell : \varphi\,\mathrm{d}x - \int_{\cup\mathcal{E}_\ell} ([v_\ell]\otimes\nu) : \varphi\,\mathrm{d}s.$$

For the first term, weak convergence of ∇v_ℓ to F implies

$$\int_\Omega \nabla v_\ell : \varphi \,\mathrm{d}x \xrightarrow{\ell\to\infty} \int_\Omega F : \varphi \,\mathrm{d}x.$$

For each face $E \in \mathcal{E}_\ell$, let $\varphi_E := |E|^{-1} \int_E \varphi \,\mathrm{d}s$ denote the integral mean of φ over E. Using the fact that $\int_E [v_\ell] \,\mathrm{d}s = 0$ for all interior faces $E \in \mathcal{E}_\ell^{\mathrm{int}}$, we estimate the second term by

$$\begin{aligned} \left| \int_{\cup\mathcal{E}_\ell} ([v_\ell] \otimes \nu) : \varphi \,\mathrm{d}s \right| &= \left| \sum_{E\in\mathcal{E}_\ell} \int_E ([v_\ell] \otimes \nu) : \big(\varphi - \varphi_E\big) \,\mathrm{d}s \right| \\ &\le \sum_{E\in\mathcal{E}_\ell} h_E \int_E \big|[v_\ell]\big| \,\mathrm{d}s \,\|\nabla\varphi\|_{\mathrm{L}^\infty(\Omega)} \\ &= \int_{\cup\mathcal{E}_\ell} h_\ell \big|[v_\ell]\big| \,\mathrm{d}s \,\|\nabla\varphi\|_{\mathrm{L}^\infty(\Omega)}. \end{aligned}$$

By hypothesis (16) and the definition of the distributional gradient, we obtain

$$\int_\Omega F : \varphi \,\mathrm{d}x = \lim_{\ell\to\infty} \langle D v_\ell, \varphi\rangle = -\lim_{\ell\to\infty} \int_\Omega v_\ell \cdot \mathrm{div}\varphi \,\mathrm{d}x = -\int_\Omega v \cdot \mathrm{div}\varphi \,\mathrm{d}x = \langle Dv, \varphi\rangle,$$

which proves $v \in \mathrm{W}^{1,p}(\Omega)^m$ with $\nabla v = Dv = F$.

It remains to show that $v|_{\Gamma^{(i)}} = g|_{\Gamma^{(i)}}$. Here it is crucial that (16) includes the condition that $\|h_\ell(v_\ell - g)^{(i)}\|_{\mathrm{L}^1(\Gamma^{(i)})} \to 0$. The result then follows upon combining the arguments from [25, Lemma 8] with the generalization presented above. ∎

Theorem 6 (convergence of discrete minimizers). *Suppose that a sequence $u_\ell \in \operatorname{argmin} \mathcal{J}(\mathcal{A}_\ell)$ of discrete minimizers satisfies*

$$\|h_\ell f\|_{\mathrm{L}^{p'}(\Omega)} + \|h_\ell [u_\ell]\|_{\mathrm{L}^1(\cup\mathcal{E}_\ell)} \xrightarrow{\ell\to\infty} 0. \tag{18}$$

Then there exists a subsequence $(u_{\ell_k})_{k\in\mathbb{N}}$ and $u \in \operatorname{argmin} \mathcal{J}(\mathcal{A})$ such that

$$\begin{aligned} u_{\ell_k} &\rightharpoonup u \quad \textit{weakly in } \mathrm{L}^p(\Omega)^m, \\ \nabla u_{\ell_k} &\rightharpoonup \nabla u \quad \textit{weakly in } \mathrm{L}^p(\Omega)^{m\times n}, \quad \textit{and} \\ \mathcal{J}(u_\ell) &\to \mathcal{J}(u) = \inf \mathcal{J}(\mathcal{A}). \end{aligned} \tag{19}$$

Moreover, unique solvability (i.e., $\#\mathrm{argmin}\mathcal{J}(\mathcal{A}) = 1$) implies weak convergence $u_\ell \rightharpoonup u$ of the entire sequence. Finally, if W is strictly convex, it even holds that

$$\nabla u_\ell \to \nabla u \qquad \textit{strongly in } \mathrm{L}^p(\Omega)^{m\times n}.$$

Proof. Fix an arbitrary $v \in \mathcal{A}$ with finite energy. From Lemma 4, we infer

$$\mathcal{J}(u_\ell) \le \mathcal{J}(\Pi_\ell v) \le \mathcal{J}(v) + C_{\mathrm{apx}} \|h_\ell f\|_{\mathrm{L}^{p'}(\Omega)} \|\nabla v\|_{\mathrm{L}^p(\Omega)}.$$

Thus, the sequence $(u_\ell)_{\ell\in\mathbb{N}}$ has uniformly bounded energy, and hence, Lemma 5 provides a weakly convergent subsequence $(u_{\ell_k})_{k\in\mathbb{N}}$ with limit $u \in \mathcal{A}$. Since W is convex, $\mathcal{J}$ is lower semicontinuous along the sequence u_{ℓ_k} [16, Theorem 3.4]. This gives

$$\mathcal{J}(u) \le \liminf_{k\to\infty} \mathcal{J}(u_{\ell_k}) \le \limsup_{k\to\infty} \mathcal{J}(u_{\ell_k}) \le \limsup_{\ell\to\infty} \mathcal{J}(u_\ell) \le \mathcal{J}(v). \tag{20}$$

Since $v \in \mathcal{A}$ was arbitrary, we deduce $u \in \operatorname{argmin} \mathcal{J}(\mathcal{A})$. Moreover, the choice $v = u$ yields equality in the latter estimate, and hence, $\mathcal{J}(u) = \lim_k \mathcal{J}(u_{\ell_k}) = \inf \mathcal{J}(\mathcal{A})$.

The convergence $\mathcal{J}(u) = \lim_\ell \mathcal{J}(u_\ell)$ follows from the fact that $\mathcal{J}(u) = \inf \mathcal{J}(\mathcal{A})$, i.e., that the limit is independent of the subsequence. Namely, if $(\widetilde{u}_\ell)_{\ell\in\mathbb{N}}$ is an arbitrary subsequence of $(u_\ell)_{\ell\in\mathbb{N}}$ for which $(\mathcal{J}(\widetilde{u}_\ell))_{\ell\in\mathbb{N}}$ is convergent, then the preceding arguments show that $\lim_\ell \mathcal{J}(\widetilde{u}_\ell) = \inf \mathcal{J}(\mathcal{A})$. In particular, $\liminf_\ell \mathcal{J}(u_\ell) = \limsup_\ell \mathcal{J}(u_\ell) = \inf \mathcal{J}(\mathcal{A})$.

If the minimizer $u \in \operatorname{argmin} \mathcal{J}(\mathcal{A})$ is unique, we can use the same kind of uniqueness argument to show that the entire sequence $(u_\ell)_{\ell\in\mathbb{N}}$ converges weakly to u: More precisely, the preceding argument shows that any subsequence $(\widetilde{u}_\ell)_{\ell\in\mathbb{N}}$ of $(u_\ell)_{\ell\in\mathbb{N}}$ has a weakly convergent subsequence $(\widetilde{u}_{\ell_k})_{k\in\mathbb{N}}$, whose limit is the unique minimizer $u \in \mathcal{A}$. Consequently, the whole sequence $(u_\ell)_{\ell\in\mathbb{N}}$ converges weakly to u.

Finally, if W is strictly convex, a result of Visintin [28] shows that weak convergence together with convergence of the energy implies strong convergence. ∎

3.2. Refinement indicators. The analysis of the previous section has demonstrated that condition (18) is sufficient in order to obtain convergence of the CR-FEM. It is therefore natural to use the quantities featured therein to steer the mesh refinement. Since the origin of the two quantities, $\|h_\ell f\|_{\mathrm{L}^{p'}}$ and $\|h_\ell [u_\ell]\|_{\mathrm{L}^1}$ is somewhat unusual (in particular, they do not arise from an a posteriori error estimate), we make a remark on their origin and interpretation.

Remark 2. The two quantities $\|h_\ell f\|_{\mathrm{L}^{p'}}$ and $\|h_\ell [u_\ell]\|_{\mathrm{L}^1}$ in (18) are closely linked to two conditions known in the calculus of variations as the *limsup condition* and the *liminf condition* of Γ-convergence [17] (or, simply, the *upper bound* and the *lower bound*) and which guarantee convergence of minimizers of a sequence of minimization problems to a minimizer of the correct limit problem.

The main step in the Γ-convergence argument is (20) in the proof of Theorem 6. Note that guaranteeing $\|h_\ell f\|_{\mathrm{L}^{p'}} \to 0$ provides the last inequality in (20) (the *limsup condition*), while guaranteeing $\|h_\ell [u_\ell]\|_{\mathrm{L}^1} \to 0$ establishes weak convergence of the broken gradients, which, together with convexity of W, provides the first inequality in (20) (the *liminf condition*).

Thus, our convergence indicators are not linked to any a posteriori error estimate in the usual sense, but arise from the use of the direct method of the calculus of variations (or Γ-convergence) in the weak convergence argument of Theorem 6.

In what follows, we discuss straightforward modifications of the convergence indicators that are more suitable for steering an adaptive algorithm, but for which our theory still applies. In order to associate the quantity $\|h_\ell f\|_{\mathrm{L}^{p'}(\Omega)}$ to faces $E \in \mathcal{E}_\ell$, it is natural to define a related convergence indicator as

$$\eta_\ell = \sum_{E\in\mathcal{E}_\ell} \eta_\ell(E) = \sum_{E\in\mathcal{E}_\ell} h_E^{p'} \|f\|_{\mathrm{L}^{p'}(\omega_E)}^{p'}, \tag{21}$$

where $\omega_E = \bigcup\{T \in \mathcal{T}_\ell : E \subset T\}$ denotes the patch of elements adjacent to E.

Next, applying Hölder's inequality on each face E shows

$$\begin{aligned} \sum_{E\in\mathcal{E}_\ell} \int_E h_E |[u_\ell]| \,\mathrm{d}s &\le \sum_{E\in\mathcal{E}_\ell} \left(\int_E h_E \,\mathrm{d}s \right)^{1/p'} \left(\int_E h_E |[u_\ell]|^p \,\mathrm{d}s \right)^{1/p} \\ &\lesssim |\Omega|^{1/p'} \left(\sum_{E\in\mathcal{E}_\ell} h_E \int_E \big|[u_\ell]\big|^p \,\mathrm{d}s \right)^{1/p}. \end{aligned} \tag{22}$$

Thus, a straightforward generalization of the indicator $\|h_\ell[u_\ell]\|_{\mathrm{L}^1(\cup\mathcal{E}_\ell)}$ is given by

$$\mu_\ell^{(0)} = \sum_{E\in\mathcal{E}_\ell} \mu_\ell^{(0)}(E) = \sum_{E\in\mathcal{E}_\ell} h_E \|[u_\ell]\|_{\mathrm{L}^p(E)}^p. \tag{23}$$

In many situations, this quantity is a bad candidate for steering the mesh refinement. To see this, let us consider the Dirichlet problem

$$-\Delta u = f \text{ in } \Omega \quad \text{with homogeneous boundary conditions} \quad u = 0 \text{ on } \partial\Omega, \tag{24}$$

where $W(F) = \frac{1}{2}|F|^2$, $p = 2$, $\Gamma^{(i)} = \partial\Omega$, and $g = 0$. For this problem, the "natural" error indicator is given by

$$\varrho_\ell^2 = \sum_{E\in\mathcal{E}_\ell\cap\Omega} h_E \int_E \big|[\nabla u_\ell]\big|^2 \,\mathrm{d}s + \sum_{E\in\mathcal{E}_\ell\cap\partial\Omega} h_E \int_E \big|\partial_\tau u_\ell\big|^2 \,\mathrm{d}s + \sum_{E\in\mathcal{E}_\ell} \eta_\ell(E),$$

where $\eta_\ell(E)$ is defined in (21), and where $\partial_\tau u_\ell$ denotes the tangential part of the gradient. This indicator is reliable and efficient (up to data oscillations) in the sense that

$$C_{\mathrm{rel}}^{-1}\|u - u_\ell\|_{\mathrm{W}^{1,2}(\Omega)} \le \varrho_\ell \le C_{\mathrm{eff}}\big(\|u - u_\ell\|_{\mathrm{W}^{1,2}(\Omega)} + \|h_\ell(f - \Pi_\ell f)\|_{\mathrm{L}^2(\Omega)}\big),$$

and it leads to a convergent adaptive algorithm [11]. Furthermore, using the fact that normal jumps can be estimated above by the $\eta_\ell(E)$ terms (cf. [11, Theorem 3.5]) and that the tangential jump of the gradient can be estimated by the tangential jump of the function, it follows that this indicator is equivalent to

$$\widetilde{\varrho}_\ell^2 = \sum_{E\in\mathcal{E}_\ell} \left\{ h_E^{-1} \int_E \big|[u_\ell]\big|^2 \,\mathrm{d}s + \eta_\ell(E)^2 \right\} = \sum_{E\in\mathcal{E}_\ell} \Big\{ h_E^{-1}\|[u_\ell]\|_{\mathrm{L}^2(E)}^2 + \eta_\ell(E)^2 \Big\}.$$

This argument suggests that the indicator $\mu_\ell^{(0)}$ from (23) is not suitable, since it uses the "wrong" scaling of the mesh size. It therefore appears natural to us to use the generalization

$$\mu_\ell^{(1)} = \sum_{E\in\mathcal{E}_\ell} h_E^{1-p} \int_E \big|[u_\ell]\big|^p \,\mathrm{d}s = \sum_{E\in\mathcal{E}_\ell} h_E^{1-p}\|[u_\ell]\|_{\mathrm{L}^p(E)}^p \tag{25}$$

as a convergence indicator. Simple scaling arguments show why this is, in fact, the correct generalization (cf. [9] and Lemma 7). We will now define a further refinement indicator that can be thought of as an interpolation between $\mu_\ell^{(0)}$ and $\mu_\ell^{(1)}$: for some fixed parameter $\alpha \in [0, 1]$, let

$$\mu_\ell^{(\alpha)} = \sum_{E\in\mathcal{E}_\ell} \mu_\ell^{(\alpha)}(E) = \sum_{E\in\mathcal{E}_\ell} h_E^{1-\alpha p}\|[u_\ell]\|_{\mathrm{L}^p(E)}^p. \tag{26}$$

Although we have initially motivated the definition of $\mu_\ell^{(\alpha)}$ through the error estimator for the Dirichlet problem, we can give an alternative interpretation. On the Dirichlet boundary, $\mu_\ell^{(\alpha)}(E)$ weakly imposes the Dirichlet condition, while in the interior, it can be thought of as a measure of the local nonconformity of the solution u_ℓ. In this sense, it seems a reasonable indicator that is independent of the problem solved.

We conclude this discussion with two simple observations. The first allows us to replace the term $\|h_\ell[u_\ell]\|_{\mathrm{L}^1(\cup\mathcal{E}_\ell)}$ in Theorem 6 by $\mu_\ell^{(\alpha)}$, while the second is intended to simplify the subsequent analysis.

LEMMA 7. *Suppose that* $0 \le \alpha \le 1$, $\ell \in \mathbb{N}$, *and* $u_\ell \in \mathcal{A}_\ell$; *then*

$$\|h_\ell[u_\ell]\|_{\mathrm{L}^1(\cup\mathcal{E}_\ell)}^p \le C_\mu\, \mu_\ell^{(\alpha)}. \tag{27}$$

Furthermore, we have the bounds

$$\mu_\ell^{(\alpha)}(E) \le C'_\mu \|h_\ell^{1-\alpha}\nabla u_\ell\|_{\mathrm{L}^p(\omega_E)}^p \quad \textit{for all interior faces } E \in \mathcal{E}_\ell^{\mathrm{int}}, \tag{28}$$

$$\mu_\ell^{(\alpha)}(E) \le C''_\mu \big(\|h_\ell^{1-\alpha}\nabla u_\ell\|_{\mathrm{L}^p(\omega_E)}^p + \|h_\ell^{1-\alpha}\nabla g\|_{\mathrm{L}^p(\omega_E)}^p\big) \quad \textit{for all } E \in \mathcal{E}_\ell\backslash\mathcal{E}_\ell^{\mathrm{int}}. \tag{29}$$

The constants C_μ, C'_μ, *and* C''_μ *depend on the shape regularity* $\sigma(\mathcal{T}_\ell)$, *and* C_μ *additionally on* $|\Omega|$ *and on* $\mathrm{diam}(\Omega)$.

Proof. The first bound follows from (22). To prove (28), let $T^\pm \in \mathcal{T}_\ell$ denote the unique elements with $E = T^+ \cap T^- \in \mathcal{E}_\ell^{\mathrm{int}}$ and $\omega_E = T^+ \cup T^-$. If z_E denotes the barycenter of E, then $[u_\ell](z_E) = 0$ yields

$$|[u_\ell]| \le h_E|\nabla[u_\ell]| \le h_E(|\nabla u_\ell|_{T^+}| + |\nabla u_\ell|_{T^-}|).$$

Shape regularity of $\mathcal{T}_\ell$ gives $|T^\pm|h_E^{(1-\alpha)p} \approx |E|h_E^p h_E^{1-\alpha p}$ and results in

$$\mu_\ell^{(\alpha)}(E) \lesssim h_E^{1-\alpha p}|E|h_E^p(|\nabla u_\ell|_{T^+}|^p + |\nabla u_\ell|_{T^-}|^p) \approx h_E^{(1-\alpha)p}\|\nabla u_\ell\|_{\mathrm{L}^p(\omega_E)}^p.$$

To prove (29) note that $[u_\ell]^{(i)} = 0$ on E if $E \cap \Gamma^{(i)} = \emptyset$. We may therefore assume, without loss of generality, that $\Gamma^{(i)} = \partial\Omega$ for all $i = 1, \dots, m$, and hence $[u_\ell] = g - u_\ell$ on $\partial\Omega$. The trace inequality reads

$$h_E\|g - u_\ell\|_{\mathrm{L}^p(E)}^p \lesssim \|g - u_\ell\|_{\mathrm{L}^p(T)}^p + h_E^p\|\nabla(g - u_\ell)\|_{\mathrm{L}^p(T)}^p,$$

which follows immediately from the trace theorem for a reference simplex and a scaling argument. See also [27, Lemma 3.2] for a proof in the case $p = 2$ yielding explicit constants, which is immediately generalized to general $p \in [1, \infty]$.

Note that $\int_E (g - u_\ell)\,\mathrm{d}s = 0$ by definition of $\mathcal{A}_\ell \ni u_\ell$, and recall that in the proof of Lemma 2 it was sufficient to have mean zero on one single face to obtain the first-order approximation property. This provides $\|g - u_\ell\|_{\mathrm{L}^p(T)} \lesssim h_T\|\nabla(g - u_\ell)\|_{\mathrm{L}^p(T)}$, which gives

$$\mu_\ell^{(\alpha)}(E) = h_E^{1-\alpha p}\|[u_\ell]\|_{\mathrm{L}^p(E)}^p = h_E^{1-\alpha p}\|g - u_\ell\|_{\mathrm{L}^p(E)}^p \lesssim h_E^{p-\alpha p}\|\nabla(g - u_\ell)\|_{\mathrm{L}^p(T)}^p$$

and immediately implies (29). ∎

3.3. Adaptive strategy. We are now in position to formulate an adaptive mesh refinement strategy for the solution of (13). In what follows, $\alpha \in [0, 1]$ is an arbitrary but fixed parameter of the algorithm.

ALGORITHM 1. **Input:** *Marking parameters* $\theta \in (0, 1]$, $\alpha \in [0, 1]$; *Initial mesh* $\mathcal{T}_0$. *Set* $\ell = 0$.

(a) *Compute a discrete minimizer* $u_\ell \in \mathrm{argmin}\,\mathcal{J}(\mathcal{A}_\ell)$.

(b) *Compute refinement indicators* η_ℓ *and* $\mu_\ell^{(\alpha)}$ *from* (21) *and* (26), *respectively.*

(c) *Generate a set of marked faces* $\mathcal{M}_\ell \subseteq \mathcal{E}_\ell$ *such that*

$$\sum_{E \in \mathcal{M}_\ell} \big(\eta_\ell(E) + \mu_\ell^{(\alpha)}(E)\big) \geq \theta(\eta_\ell + \mu_\ell^{(\alpha)}). \tag{30}$$

(d) *Generate a regular triangulation* $\mathcal{T}_{\ell+1}$*, where at least the marked faces* $E \in \mathcal{M}_\ell$ *are refined.*

(e) *Increase* $\ell \mapsto \ell + 1$ *and go to* (a).

A marking strategy satisfying (30) is often called *Dörfler marking.* It was a crucial ingredient in the first convergence proofs of the adaptive finite element method and has been identified to also play an important role in obtaining optimal convergence rates [14], in which case the cardinality of the set $\mathcal{M}_\ell$ should be minimal. Generically, the value $\theta = 1$ corresponds to uniform refinement, whereas small θ leads to highly adapted meshes.

Our 2D implementation uses newest vertex bisection [22] in step (4) to ensure that the sequence of triangulations $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}}$ generated by Algorithm 1 is uniformly shape-regular. However, besides the uniform shape-regularity, the following convergence result requires only that marked faces are reduced by a uniform factor $\kappa \in (0, 1)$, i.e.,

$$h_{E'} \leq \kappa\, h_E \quad \text{for all } E' \in \mathcal{E}_{\ell+1} \text{ with } E' \subset E \in \mathcal{M}_\ell. \tag{31}$$

Thus, the precise refinement rule in step (d) is fairly arbitrary.

Note that in the following theorem we require that $\alpha \in [0, 1)$, that is, the scaling $\alpha = 1$, which is intuitively optimal, is not included. We see in section 3.4 that this is not a restriction of our analysis, but that the result is false for the case $\alpha = 1$.

Theorem 8 (convergence of the adaptive algorithm). *Suppose that the sequence* $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}}$ *generated by Algorithm* 1 *is uniformly shape-regular, that it satisfies* (31)*, and assume in addition that* $0 \leq \alpha < 1$*. Then the refinement indicators converge to zero, i.e.,*

$$\eta_\ell + \mu_\ell^{(\alpha)} \xrightarrow{\ell\to\infty} 0, \tag{32}$$

and the sequence $(u_\ell)_{\ell\in\mathbb{N}}$ *of discrete minimizers satisfies the conditions of Theorem* 6*.*

Proof. This proof is largely inspired by [24].

As in the proof of Theorem 6, it follows that $\mathcal{J}(u_\ell)$ and hence $\|u_\ell\|_{\mathrm{L}^p(\Omega)} + \|\nabla u_\ell\|_{\mathrm{L}^p(\Omega)}$ are bounded sequences.

To abbreviate the notation, we now drop the superscript (α) in $\mu_\ell^{(\alpha)}$, and we write

$$\eta_\ell(\mathcal{S}_\ell) := \sum_{E\in\mathcal{S}_\ell} \eta_\ell(E) \quad \text{and} \quad \mu_\ell(\mathcal{S}_\ell) := \sum_{E\in\mathcal{S}_\ell} \mu_\ell(E) \quad \text{for all } \mathcal{S}_\ell \subseteq \mathcal{E}_\ell.$$

We consider the set of all faces, respectively, all elements that are eventually not refined, i.e.,

$$\widetilde{\mathcal{E}} := \bigcap_{k\geq 0} \bigcup_{\ell \geq k} \mathcal{E}_\ell \quad \text{and} \quad \widetilde{\mathcal{T}} := \bigcap_{k\geq 0} \bigcup_{\ell \geq k} \mathcal{T}_\ell.$$

It is evident that $T \in \widetilde{\mathcal{T}}$ if and only if all faces of T belong to $\widetilde{\mathcal{E}}$. For the proof of (32), we split the indicators into

$$\eta_\ell + \mu_\ell = \big(\eta_\ell(\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}}) + \mu_\ell(\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}})\big) + \big(\eta_\ell(\mathcal{E}_\ell\cap\widetilde{\mathcal{E}}) + \mu_\ell(\mathcal{E}_\ell\cap\widetilde{\mathcal{E}})\big).$$

Step 1. In the first step, we will prove that

$$\eta_\ell(\mathcal{E}_\ell \setminus \widetilde{\mathcal{E}}) + \mu_\ell(\mathcal{E}_\ell \setminus \widetilde{\mathcal{E}}) \xrightarrow{\ell\to\infty} 0. \tag{33}$$

Recall that $\omega_E := \bigcup\{T \in \mathcal{T}_\ell \,:\, E \subset \partial T\}$. Setting $\widetilde{\Omega}_\ell = \bigcup\{\omega_E \,:\, E \in \mathcal{E}_\ell \setminus \widetilde{\mathcal{E}}\}$, we first claim that $\chi_{\widetilde{\Omega}_\ell} h_\ell \xrightarrow{\ell\to\infty} 0$ a.e. in Ω. To see this, fix $x \in \Omega\backslash\big(\bigcup_\ell \bigcup \mathcal{E}_\ell\big)$ outside of the skeletons of all $\mathcal{T}_\ell$, which form a null-set. For each ℓ, there is a unique element $T_\ell \in \mathcal{T}_\ell$ with $x \in T_\ell$. If $\lim_\ell h_{T_\ell} = 0$, we conclude $\lim_\ell(\chi_{\widetilde{\Omega}_\ell} h_\ell)(x) = 0$. Otherwise, T_ℓ is refined only finitely many times, i.e., there holds $T_\ell = T_{\ell_0}$ for some $\ell_0 \in \mathbb{N}$ and all $\ell \geq \ell_0$, i.e., $T_\ell \in \widetilde{\mathcal{T}}$ and therefore its faces belong to $\widetilde{\mathcal{E}}$. Consequently, $x \notin \widetilde{\Omega}_\ell$ for all $\ell \geq \ell_0$, and hence, $(\chi_{\widetilde{\Omega}_\ell} h_\ell)(x) = 0$ for $\ell \geq \ell_0$. We have, therefore, shown that

$$\lim_{\ell\to\infty} \chi_{\widetilde{\Omega}_\ell} h_\ell = 0 \qquad \text{pointwise a.e. in } \Omega.$$

During mesh refinement, the local mesh-size h_ℓ is pointwise decreasing. Consequently, the monotone convergence theorem yields

$$\chi_{\widetilde{\Omega}_\ell} h_\ell^\beta \psi \xrightarrow{\ell\to\infty} 0 \qquad \text{strongly in } \mathrm{L}^q(\Omega), \tag{34}$$

for all $\beta > 0$, and $\psi \in \mathrm{L}^q(\Omega)$. With $\beta = 1$ and $\psi = f$, we infer

$$\eta_\ell(\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}}) = \sum_{E\in\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}}} \|h_\ell f\|_{\mathrm{L}^{p'}(\omega_E)}^{p'} \lesssim \|h_\ell f\|_{\mathrm{L}^{p'}(\widetilde{\Omega}_\ell)}^{p'} = \|\chi_{\widetilde{\Omega}_\ell} h_\ell f\|_{\mathrm{L}^{p'}(\Omega)}^{p'} \xrightarrow{\ell\to\infty} 0.$$

Before we prove convergence of $\mu_\ell(\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}})$ to zero, it is instructive to consider the refinement indicator $\|h_\ell[u_\ell]\|_{\mathrm{L}^1(\cup\mathcal{E}_\ell)}$ first. Using the facts that $h_E \approx h_\ell$ in ω_E, and that $[u_\ell](z_E) = 0$, we can estimate

$$\int_E h_E|[u_\ell]|\,\mathrm{d}s \lesssim h_E^2 \int_E \big|[\nabla u_\ell]\big|\,\mathrm{d}s \lesssim \int_{\omega_E} h_\ell|\nabla u_\ell|\,\mathrm{d}x.$$

Summing over $E \in \mathcal{E}_\ell \setminus \widetilde{\mathcal{E}}$ and using Hölder's inequality, we obtain

$$\begin{aligned}\|h_\ell[u_\ell]\|_{\mathrm{L}^1(\cup(\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}}))} &\lesssim \sum_{E\in\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}}} \|h_\ell\nabla u_\ell\|_{\mathrm{L}^1(\omega_E)} \\ &\lesssim \|h_\ell\nabla u_\ell\|_{\mathrm{L}^1(\widetilde{\Omega}_\ell)} \leq \|h_\ell\|_{\mathrm{L}^{p'}(\widetilde{\Omega}_\ell)}\|\nabla u_\ell\|_{\mathrm{L}^p(\widetilde{\Omega}_\ell)}.\end{aligned}$$

According to (34), with $\beta = 1$ and $\psi = 1$, the upper bound tends to zero as $\ell \to \infty$.

If we attempt to use the same idea for proving that $\mu_\ell(\mathcal{E}_\ell\backslash\widetilde{\mathcal{E}}) \to 0$, then, using (28) and (29), we first obtain the following bound:

$$\mu_\ell(\mathcal{E}_\ell \setminus \widetilde{\mathcal{E}}) \lesssim \|h_\ell^{(1-\alpha)}\nabla u_\ell\|_{\mathrm{L}^p(\widetilde{\Omega}_\ell)}^p + \|h_\ell^{(1-\alpha)}\nabla g\|_{\mathrm{L}^p(\widetilde{\Omega}_\ell)}^p. \tag{35}$$

Using (34), with $\beta = 1 - \alpha > 0$ and $\psi = \nabla g$, we immediately find that the second integral on the right-hand side of (35) converges to zero. It thus remains only to treat the first term on the right-hand side. Unfortunately, we control ∇u_ℓ only in $\mathrm{L}^p(\Omega)^{m\times n}$. Therefore, we cannot immediately use Hölder's inequality as before to verify that the first term on the right-hand side of (35) tends to zero. Since we do not know whether ∇u_ℓ converges pointwise a.e., we have no hope of using Fatou's lemma either. Instead, we make use of the additional flexibility provided by the condition $\alpha < 1$ to derive an inverse estimate,

$$\begin{aligned}\|h_\ell^{(1-\alpha)}\nabla u_\ell\|^p_{\mathrm{L}^p(\widetilde{\Omega}_\ell)} = \int_{\widetilde{\Omega}_\ell} h_\ell^{(1-\alpha)p}|\nabla u_\ell|^p\,\mathrm{d}x &\le \sum_{\substack{T\in\mathcal{T}_\ell\\ T\subseteq\widetilde{\Omega}_\ell}} h_T^{n+(1-\alpha)p}|\nabla u_\ell|_T|^p \\ &\le \left(\sum_{\substack{T\in\mathcal{T}_\ell\\ T\subseteq\widetilde{\Omega}_\ell}} h_T^{(n+(1-\alpha)p)q/p}|\nabla u_\ell|_T|^q\right)^{p/q} \\ &\lesssim \left(\sum_{\substack{T\in\mathcal{T}_\ell\\ T\subseteq\widetilde{\Omega}_\ell}} h_T^{nq/p+(1-\alpha)q-n}\int_T|\nabla u_\ell|^q\right)^{p/q},\end{aligned}$$

where $1 \le q < p$. In the second estimate above, we used the bound $\|\cdot\|_{\ell^p} \le \|\cdot\|_{\ell^q}$. Setting $\beta = n(q/p-1)+(1-\alpha)q$, we obtain

$$\|h_\ell^{(1-\alpha)}\nabla u_\ell\|^p_{\mathrm{L}^p(\widetilde{\Omega})} \lesssim \left(\int_{\widetilde{\Omega}_\ell} h_\ell^\beta|\nabla u_\ell|^q\,\mathrm{d}x\right)^{p/q} \le \left\|h_\ell^\beta\chi_{\widetilde{\Omega}_\ell}\right\|^{p/q}_{\mathrm{L}^{p/(p-q)}}\|\nabla u_\ell\|^p_{\mathrm{L}^p(\Omega)}$$

by use of Hölder's inequality. For this bound to tend to zero, we require that $\beta > 0$, which can be achieved by choosing q sufficiently close to p and using the fact that $\alpha < 1$. Thus, we have successfully established (33).

Step 2. In the second step, we use the properties of our marking strategy to conclude the proof of (32). Namely, observe that at step ℓ any marked face will be refined during this step, i.e.,

$$\mathcal{M}_\ell \subset \mathcal{E}_\ell \setminus \widetilde{\mathcal{E}}.$$

Therefore, (30) and Step 1 imply that

$$\theta(\eta_\ell+\mu_\ell) \le \eta_\ell(\mathcal{M}_\ell)+\mu_\ell(\mathcal{M}_\ell) \le \eta_\ell(\mathcal{E}_\ell\setminus\widetilde{\mathcal{E}})+\mu_\ell(\mathcal{E}_\ell\setminus\widetilde{\mathcal{E}}) \to 0,$$

as $\ell \to \infty$, which concludes the proof. ∎

Remark 3. We have already remarked in section 3.2 that, for $\alpha < 1$, our refinement indicators are not reliable error indicators, even for a simple Dirichlet problem with homogeneous boundary conditions. Furthermore, our refinement indicators have no information about Neumann boundary conditions, which gives further indication to its "incompleteness." We found it therefore somewhat surprising that we were able to prove convergence of our adaptive strategy.

We also note that our proof does not extend in an obvious way to conforming finite element methods, where the upper bound (15) is false even for quadratic functionals.

The case $\alpha = 1$ will be addressed in the next section.

3.4. The case $\boldsymbol{\alpha = 1}$. In this section, we construct an example demonstrating that the restriction $\alpha < 1$ in Theorem 8 is a true restriction and cannot be removed. To this end, we will construct a convex stored energy function $W : \mathbb{R}^2 \to [0,\infty)$ and a sequence of nonconforming finite element functions u_ℓ so that

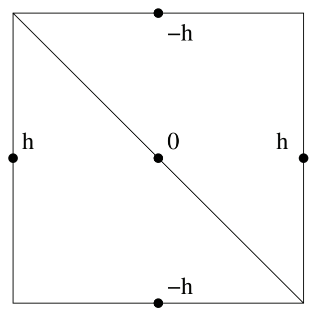

FIG. 1. *Visualization of the construction of an oscillatory test function.*

(i) u_ℓ is a solution of the discrete minimization problem (13) with $f = 0$ and homogeneous Dirichlet data $g = 0$;
(ii) the sequence u_ℓ converges in the sense of (19) to a solution of the exact problem (5);
(iii) the indicator $\mu_\ell^{(\alpha)}$ does not tend to zero if $\alpha = 1$.

Let $\Omega = (0, 1)^2$ and let the mesh $\mathcal{T}_\ell$, $\ell \in \mathbb{N}$, consist of $2^\ell \times 2^\ell$ copies of the elements displayed in Figure 1. We denote $\hat{h}_\ell = 2^{-\ell}$, so that the vertical and horizontal edges have length $\hat{h}_\ell$ while the diagonal edges have length $\sqrt{2}\,\hat{h}_\ell$. We note that we use only that $\hat{h}_\ell \to 0$, but that the rate $\hat{h}_\ell = 2^{-\ell}$ is not important.

We construct an oscillatory function $u_\ell \in \mathrm{CR}(\mathcal{T}_\ell)$ that converges "weakly" to zero. First, we choose $u_\ell(z) = 0$ for all boundary nodes $z \in \mathcal{N}^{\mathrm{nc}} \cap \partial\Omega$, so that u_ℓ satisfies a discrete Dirichlet condition on $\partial\Omega$. Next, on each edge midpoint z we assign nodal values according to Figure 1, that is, we choose $u_\ell(z) = 0$ for midpoints z that lie on a diagonal, $u_\ell(z) = \hat{h}_\ell$ for midpoints z that lie on vertical edges, and $u_\ell(z) = -\hat{h}_\ell$ for midpoints z that lie on horizontal edges. It is easily verified that

$$\|u_\ell\|_{\mathrm{L}^\infty} = 2\hat{h}_\ell \quad \text{and} \quad \|\nabla u_\ell\|_{\mathrm{L}^\infty} = 2,$$

and that

$$\nabla u_\ell(x) \in \big\{(2, -2), (-2, 2)\big\},$$

where the value $(2, -2)$ is attained on elements that lie above a diagonal edge and $(-2, 2)$ on elements that lie below a diagonal edge, excluding of course all elements that lie on the boundary. It is now a simple exercise to show that, for any $p \in [1, \infty)$,

$$\begin{aligned} u_\ell &\to 0 \quad \text{strongly in } \mathrm{L}^p(\Omega) \quad \text{and} \\ \nabla u_\ell &\rightharpoonup 0 \quad \text{weakly in } \mathrm{L}^p(\Omega)^2. \end{aligned}$$

Next, we compute the tangential jumps on diagonal edges, which do not touch the boundary. On such an edge E with upper element T^+ and lower element T^-, a tantengial unit vector is $t = (1/\sqrt{2}, -1/\sqrt{2})$, and we obtain

$$(\nabla u_\ell^+ - \nabla u_\ell^-) \cdot t = (4, -4) \cdot (1/\sqrt{2}, -1/\sqrt{2}) = 8/\sqrt{2}.$$

Translating the tangential jumps of the gradient into jumps of the functions leads to

$$\int_E \big|[u_\ell]\big|^p \,\mathrm{d}s = \, 2\int_{r=0}^{h_E/2} \big|r[\nabla u \cdot t]\big|^p \,\mathrm{d}r = 2(8/\sqrt{2})^p \int_{r=0}^{\hat{h}_\ell/\sqrt{2}} r^p \,\mathrm{d}r = \frac{4^p\sqrt{2}\,\hat{h}_\ell^{p+1}}{p+1}.$$

Upon defining $C_1 = 4^p\sqrt{2}/(p+1)$, we obtain

$$\int_E \big|[u_\ell]\big|^p \,\mathrm{d}s = C_1 \hat{h}_\ell^{p+1}.$$

Moreover, there exist $(1/\hat{h}_\ell - 2)^2$ diagonal edges that do not touch the boundary, and hence

$$\mu_\ell^{(\alpha)} = \sum_{E\in\mathcal{E}_\ell} h_E^{1-\alpha p}\big\|[u_\ell]\big\|_{\mathrm{L}^p(E)}^p \geq C_2' \left(\tfrac{1}{\hat{h}_\ell} - 2\right)^2 \hat{h}_\ell^{2+p-\alpha p} \geq C_2 \hat{h}_\ell^{p(1-\alpha)}$$

for some constant $C_2 > 0$. In particular, we conclude that, if $\alpha = 1$, then

$$\mu_\ell^{(\alpha)} \not\to 0 \quad \text{as} \quad \ell \to \infty.$$

Finally, it remains only to construct a convex variational problem for which the functions u_ℓ are discrete minimizers. To this end, we define a function $w : \mathbb{R} \to [0, +\infty)$ such that

$$w(t) = \begin{cases} 0, & t \in [-2, 2], \\ |t-2|^p, & t > 2, \\ |t+2|^p, & t < -2, \end{cases}$$

and the stored energy function

$$W(F) = w(F_1) + w(F_2).$$

It is clear the W is convex and exhibits p-growth from below (and above), and that

$$\mathcal{J}(u_\ell) = \int_\Omega W(\nabla u_\ell)\,\mathrm{d}x = 0.$$

Thus, the functions u_ℓ are discrete minimizers, and converge in the sense of (19) to a minimizer of the variational problem.

Remark 4. The counterexample produced in this section clearly shows that Theorem 8 is false for $\alpha = 1$. However, we point out that our counterexample uses a stored energy function that is convex but not strictly convex. Naturally, the question arises whether it would be possible to include $\alpha = 1$ if W were strictly convex. The analysis in [24] shows that obtaining strong convergence of the sequence $(\nabla u_\ell)_{\ell\in\mathbb{N}}$ to some $\nabla\widetilde{u}$ a priori (instead of merely weak convergence) is the key. Preliminary investigations suggest that such a result should be true, but that some additional technicalities need to be overcome to obtain it. Hence, we will not pursue this question in the present work.

However, we can remark that in our numerical experiments in the next section, which are exclusively performed for strictly convex stored energy functions, we observe that choosing $\alpha = 1$ always leads to a convergent algorithm and, indeed, to improved convergence rates.

4. Numerical experiments. We have implemented Algorithm 1 for two two-dimensional model problems: the Laplace problem with Dirichlet and with Neumann boundary conditions as well as the example of Foss, Hrusa, and Mizel [20] that exhibits a Lavrentiev gap. Before we present the computational experiments, we briefly outline the details of our implementation.

(a) The solution of the optimization problem is achieved by a damped Newton method if it is nonlinear and a direct solver if it is linear.

(b) We have found that Dörfler's marking strategy with a minimal set $\mathcal{M}_\ell$ yields very slow mesh growth for the highly nonlinear and singular problems that we consider here. Therefore, our strategy is to mark a fixed fraction of edges (with largest indicators) for refinement, i.e.,

$$\#\mathcal{M}_\ell \geq \theta \#\mathcal{E}_\ell \quad \text{with} \quad \min_{E \in \mathcal{M}_\ell} \big(\eta_\ell(E) + \mu_\ell^{(\alpha)}(E)\big) \geq \max_{E \in \mathcal{E}_\ell \setminus \mathcal{M}_\ell} \big(\eta_\ell(E) + \mu_\ell^{(\alpha)}(E)\big).$$

Note that then,

$$\sum_{E \in \mathcal{M}_\ell} \big(\eta_\ell(E) + \mu_\ell^{(\alpha)}(E)\big) \geq \frac{\theta}{1+\theta} \sum_{E \in \mathcal{E}_\ell} \big(\eta_\ell(E) + \mu_\ell^{(\alpha)}(E)\big),$$

so that Dörfler marking (30) still holds with θ replaced by $\theta/(1+\theta)$. We usually chose $\theta = 0.25$, which roughly doubles the number of elements at each iteration.

(c) The mesh refinement is achieved via *newest vertex bisection*, which halves every marked edge and which preserves shape regularity.

(d) We terminate the algorithm when a prescribed number of elements is attained.

For the computations in section 4.1, we estimate the error for the energy by comparing it to a conforming computation. If $f \equiv 0$, then

$$\mathcal{J}(u_\ell) \leq \inf \mathcal{J}(\mathcal{A}) \leq \mathcal{J}(\bar{u}) \quad \text{for all } \bar{u} \in \mathcal{A}.$$

If f is nonzero, then the above estimate depends on an unknown quantity, namely $\|\nabla u\|_{\mathrm{L}^p}$ where $u \in \operatorname{argmin} \mathcal{J}(\mathcal{A})$. However, we have observed that even in that case, $\mathcal{J}(u_\ell)$ is monotonically increasing towards the energy of the exact solution. Therefore, we compute a conforming $\bar{u}$ using a standard adaptive P_1-finite element method [10] and take

$$\inf \mathcal{J}(\mathcal{A}) - \mathcal{J}(u_\ell) \leq \mathcal{J}(\bar{u}) - \mathcal{J}(u_\ell)$$

as a slightly heuristic energy error estimate.

4.1. Linear Laplacian. We begin our experiments with the Laplace equation on the slit domain $\Omega = (-1,1)^2 \setminus [0,1) \times \{0\}$. It is equivalently formulated by setting $W(F) = \frac{1}{2}|F|^2$ with $m = 1$ and $n = 2$. First, we consider the pure Dirichlet problem

$$-\Delta u = 1 \text{ in } \Omega \quad \text{with homogeneous boundary conditions} \quad u = 0 \text{ on } \partial\Omega,$$

where in the energy formulation

$$\Gamma^{(1)} = \partial\Omega, \quad f = 1, \quad g = 0. \tag{36}$$

In order to investigate the effect of a Neumann boundary, we also consider the mixed boundary value problem

$$-\Delta u = 0 \text{ in } \Omega \quad \text{with} \quad \partial u/\partial n = 0 \text{ on } \Gamma_N \quad \text{and} \quad u = g \text{ on } \Gamma_D,$$

where $|\Gamma_D \cap \Gamma_N| = 0$ and $\partial\Omega = \Gamma_D \cup \Gamma_N$. We choose

$$\Gamma^{(1)} = \Gamma_D = \partial\Omega \cap \{x_1 = 1\}, \quad f = 0, \quad g(1, x_2) = \operatorname{sign}(x_2). \tag{37}$$

The exact solutions to both the Dirichlet and the Neumann problems can be decomposed as $u = kS + \bar{u}$, where ∇S has an $r^{-1/2}$ singularity at the re-entrant corner, and where $\bar{u} \in \mathrm{H}^2(\Omega)$. From these facts one can deduce that the best possible convergence rate for the energy, using an adaptive P_1-finite element method, is $O((\#\mathcal{T}_\ell)^{-1})$ [5].

The convergence rates for problems (36) and (37) are shown in Figures 2 and 3, respectively. As expected, we observe that the accuracy improves as α approaches 1.0. In the Dirichlet problem, the convergence rate for $\alpha = 1.0$ and for $\alpha = 0.9$ can barely be distinguished. What is surprising though is that, for the Neumann problem, the value of α does not seem to affect the convergence rate at all. We have no explanation for this effect, but we note that we will also observe it in our second model problem.

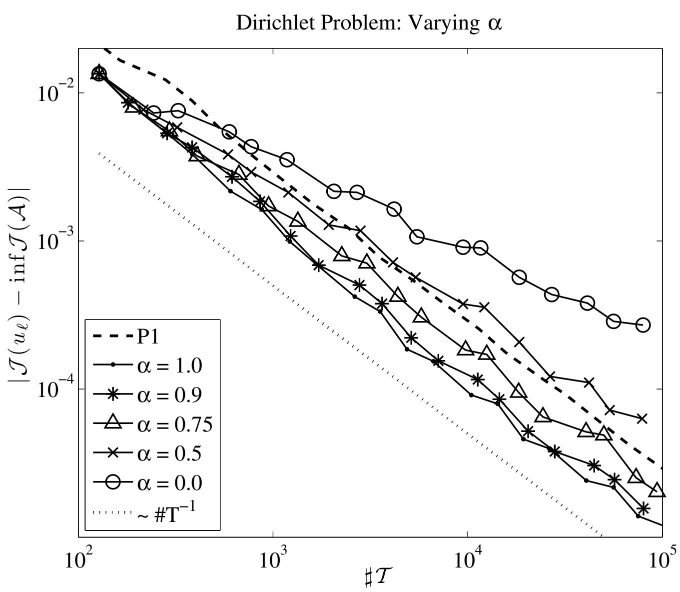

FIG. 2. *Convergence rates for Algorithm* 1 *applied to problem* (36). *As* $\alpha \nearrow 1$, *the convergence rate approaches the optimal rate* $\#\mathcal{T}^{-1}$.

4.2. Lavrentiev phenomenon. For our second numerical experiment, we use an example that exhibits the Lavrentiev gap phenomenon—the focus of our investigation. To this end, we slightly modify the example of Foss, Hrusa, and Mizel [20]. Let $m = n = 2$, let the domain be the half disk $\Omega = \{|x| < 1, x_2 > 0\}$, and let

$$\Gamma^{(1)} = (-1, 0) \times \{0\} \cup \{|x| = 1, x_2 > 0\} \quad \text{and} \quad \Gamma^{(2)} = (0, 1) \times \{0\} \cup \{|x| = 1, x_2 > 0\}.$$

Furthermore, $f \equiv 0$, $g^{(i)} = 0$ on $\{x_2 = 0\}$, and $g = (\cos \frac{\theta}{2}, \sin \frac{\theta}{2})$ on $\{|x| = 1\}$ in polar coordinates (r, θ). Thus, admissible functions are deformations of the half disk Ω into the quarter disk $\{|x| < 1,\ x_1 > 0,\ x_2 > 0\}$. Suppose, for the moment, that the stored energy density is given by

$$W(F) = \big(|F|^2 - 2 \det F\big)^4.$$

Convexity of W follows immediately from the fact that $F \mapsto (|F|^2 - 2 \det F)$ is a nonnegative quadratic form. However, the associated energy functional is not coercive.

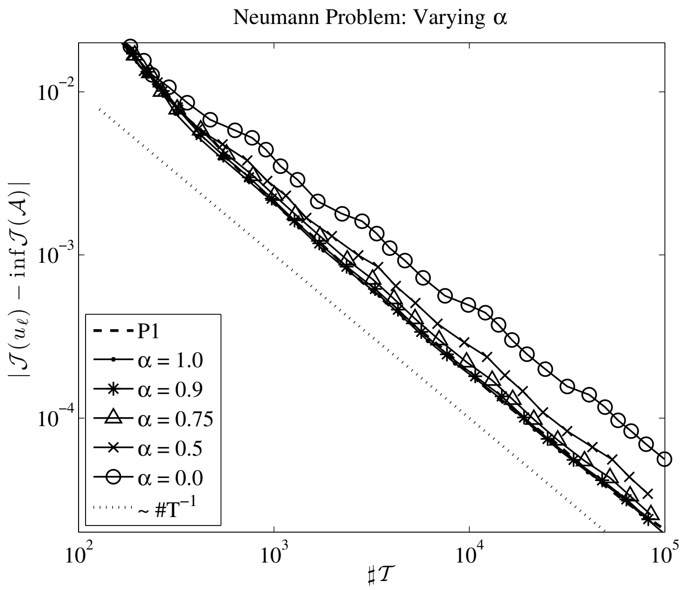

FIG. 3. *Convergence rates for Algorithm* 1 *applied to problem* (37)*. Contrary to intuition, the convergence rate seems to be optimal, independent of the parameter* α*.*

Nevertheless, Foss, Hrusa, and Mizel showed in [20] that it exhibits the Lavrentiev gap phenomenon. The global minimum of $\mathcal{J}$ in $\mathcal{A}$ is the function

$$\bar{u} = r^{1/2}\big(\cos\tfrac{\theta}{2}, \sin\tfrac{\theta}{2}\big),$$

for which $\mathcal{J}(\bar{u}) = 0$. It is furthermore easy to verify that $\bar{u}$ also minimizes the Dirichlet integral for the same boundary conditions. Consequently, for $p = 2$, $\bar{u}$ is also the global minimizer of

$$\mathcal{J}_p(v) = \int_\Omega W_p(\nabla v)\,\mathrm{d}x \tag{38}$$

in $\mathcal{A}$, where

$$W_p(F) = \big(|F|^2 - 2\det F\big)^4 + \tfrac{1}{p}\big(|F_1|^p + |F_2|^p\big).$$

Since we know the solution for $p = 2$ explicitly, we can explicitly compute the energy minimum,

$$\inf \mathcal{J}_2(\mathcal{A}) = \mathcal{J}_2(\bar{u}) = \pi/4.$$

In Figure 4, we plot the convergence rate for the minimization problem, for varying α. We observe the same effect as for the Neumann problem in section 4.1: surprisingly, the convergence rate seems to be independent of the parameter α. While it is encouraging that the convergence rate for the energy appears to be linear (see [18] for a rigorous result for nonlinear convex problems approximated by conforming finite element methods), despite the fact that we are solving a highly nonlinear and singular problem (note that $\mathcal{J}_2$ is not even continuous in the strong topology of $\mathrm{W}^{1,2}(\Omega)^2$), we strongly suspect that this is related to the particularly simple structure of W_2 and the fact that $\nabla\bar{u}$ minimizes the first term of $W_2(\nabla\bar{u})$ pointwise.

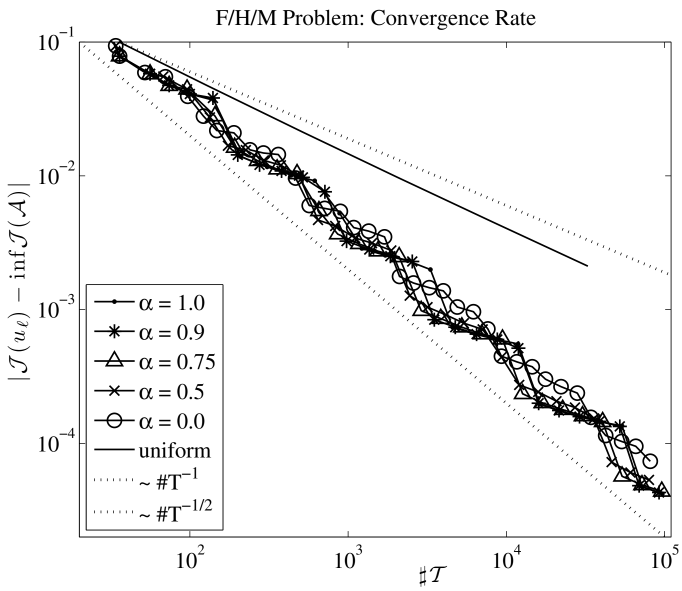

FIG. 4. *Convergence rates for Algorithm* 1 *applied to the minimization problem* $u \in \operatorname{argmin}\mathcal{J}_2(\mathcal{A})$*, with varying marking parameter* α*.*

4.3. Verification of Lavrentiev gaps. In our final experiment, we demonstrate how one could verify whether a given minimization problem exhibits a Lavrentiev gap. We consider the energy functional $\mathcal{J}_p$ from (38). For the parameters $p = 2, 3, 4, 6$, we apply Algorithm 1 with the minimization problem $\operatorname{argmin}\mathcal{J}_p(\mathcal{A})$ and obtain discrete solutions u_ℓ. In addition, we compute an adaptive P_1-solution $\bar{u}_\ell$ for the same problem, though possibly on different meshes, and we plot the difference in energy $\mathcal{J}_p(\bar{u}_\ell) - \mathcal{J}_p(u_\ell)$. (If we were to obtain the P_1-solution on the mesh generated by the adaptive nonconforming algorithm, then the mesh would be adapted to the singularity of the $\mathrm{W}^{1,1}$-minimizer as opposed to that of the $\mathrm{W}^{1,\infty}$-minimizer, and thus lead to suboptimal convergence rates for the conforming solution.)

The theory in [20] would lead us to expect (but except for the case $p = 2$ this is not at all clear) that, for $p = 2, 3$ a Lavrentiev gap occurs, while for $p = 4, 6$ no gap occurs. The computations that we show in Figure 5 agree with this prediction, except possibly in the case $p = 4$, where they suggest that a Lavrentiev gap may, in fact, be present.

5. Conclusion. We have presented an adaptive finite element algorithm for the solution of convex variational problems. Despite the fact that the refinement indicators are *not* reliable error indicators in any classical sense, we have succeeded in proving convergence of our adaptive scheme. The main question we have left open is whether the case $\alpha = 1$ can be included in Theorem 8 if we additionally assume that W is strictly convex. To conclude, we briefly mention some further possible generalizations of our analysis.

In order for the minimization problem (5) to be well-posed (or, rather, for the direct method technique to apply), it is necessary that $\mathcal{J}$ is coercive in $\mathcal{A}$, which, in particular, requires that elements of $\mathcal{A}$ satisfy a Poincaré-type inequality. Similarly, we require a broken Poincaré-type inequality for the discrete admissible set $\mathcal{A}_\ell$ in

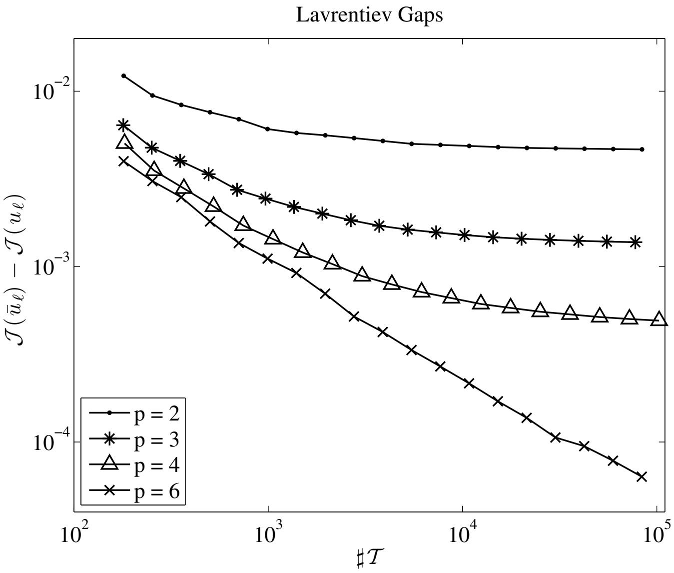

FIG. 5. *Adaptive computation of the Lavrentiev gap* $\inf \mathcal{J}_p(\mathcal{A}_\infty) - \inf \mathcal{J}_p(\mathcal{A})$ *for* $p = 2, 3, 4, 6$. *Contrary to intuition, for* $p = 4$, *the computation suggests that a Lavrentiev gap is present. The scale for the x-axis is the number of elements in the adaptive grid generated for the nonconforming solution. The number of elements used for the* P_1*-FEM solution is typically different.*

order to be able to extract weakly convergent subsequences. The entire analysis applies whenever such a broken Poincaré-type inequality is available for $\mathcal{A}_\ell$. It is therefore straightforward to generalize the results, for example, to problems involving pointwise constraints on the function (e.g., an obstacle problem) or on the gradient (e.g., problems arising in plasticity).

A second important generalization is to allow W to depend on x and on u. It is not at all clear in which generality this can be achieved. Mild dependencies such as piecewise constant dependence on x are easily included in the analysis, however, a strong coupling of (x, u) to ∇u must be avoided. This follows immediately upon considering Manià's functional [21]

$$\mathcal{J}(u) = \int_0^1 u_x^6 (u^3 - x)^2 \,\mathrm{d}x,$$

which is to be minimized subject to $u(0) = 0, u(1) = 1$. Since the Crouzeix–Raviart finite element method reduces to the P_1-finite element method in one dimension, and since Manià's example exhibits a Lavrentiev gap, it follows that the method is not convergent in this case.

Finally, the generalization to polyconvex or even quasiconvex W is even more difficult. Here, both the upper bound (15) and the lower bound (20), i.e., the weak lower-semicontinuity of $\mathcal{J}$ along the sequence u_ℓ, are completely open.

Acknowledgment. Both authors wish to acknowledge the support and hospitality of the Hausdorff Centre for Mathematics (University of Bonn) where this work was initiated.

REFERENCES

[1] G. Acosta and R. G. Durán, *An optimal Poincaré inequality in L^1 for convex domains*, Proc. Amer. Math. Soc., 132 (2004), pp. 195–202.

[2] R. Adams and J. Fournier, *Sobolev Spaces*, 2nd ed., Elsevier Science/Academic Press, Amsterdam, 2003.

[3] J. M. Ball, *Singularities and computation of minimizers for variational problems*, in Foundations of Computational Mathematics (Oxford, 1999), London Math. Soc. Lecture Note Ser. 284, Cambridge University Press, Cambridge, UK, 2001, pp. 1–20.

[4] S. Bartels and C. Carstensen, *A convergent adaptive finite element method for an optimal design problem*, Numer. Math., 108 (2008), pp. 359–385.

[5] P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math., 97 (2004), pp. 219–268.

[6] D. Braess, *Finite Elements, Theory, Fast Solvers, and Applications in Elasticity Theory*, 3rd ed., Cambridge University Press, Cambridge, UK, 2007 (translated from the German by Larry L. Schumaker).

[7] S. C. Brenner, *Poincaré-Friedrichs inequalities for piecewise H^1 functions*, SIAM J. Numer. Anal., 41 (2003), pp. 306–324.

[8] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Ser. Comput. Math. 15, Springer-Verlag, New York, 1991.

[9] A. Buffa and C. Ortner, *Compact embeddings of broken Sobolev Spaces and applications*, IMA J. Numer. Anal., 29 (2009), pp. 827–855.

[10] C. Carstensen, *Convergence of an adaptive FEM for a class of degenerate convex minimization problems*, IMA J. Numer. Anal., 28 (2008), pp. 423–439.

[11] C. Carstensen and R. H. W. Hoppe, *Convergence analysis of an adaptive nonconforming finite element method*, Numer. Math., 103 (2006), pp. 251–266.

[12] C. Carstensen and R. H. W. Hoppe, *Error reduction and convergence for an adaptive mixed finite element method*, Math. Comp., 75 (2006), pp. 1033–1042.

[13] C. Carstensen and C. Ortner, *Analysis of a class of penalty methods for computing singular minimizers*, Comput. Methods Appl. Math., 10 (2010), pp. 137–163.

[14] J. M. Cascon, 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), pp. 2524–2550.

[15] M. Crouzeix and P.-A. Raviart, *Conforming and nonconforming finite element methods for solving the stationary Stokes equations*. I, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge, 7 (1973), pp. 33–75.

[16] B. Dacorogna, *Direct Methods in the Calculus of Variations*, Appl. Math. Sci. 78, Springer-Verlag, Berlin, 1989.

[17] G. Dal Maso, *An Introduction to Γ-convergence*, Progr. Nonlinear Differential Equations Appl. 8, Birkhäuser Boston, Boston, 1993.

[18] L. Diening and C. Kreuzer, *Linear convergence of an adaptive finite element method for the p-Laplacian equation*, SIAM J. Numer. Anal., 46 (2008), pp. 614–638.

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

[20] M. Foss, W. J. Hrusa, and V. J. Mizel, *The Lavrentiev gap phenomenon in nonlinear elasticity*, Arch. Ration. Mech. Anal., 167 (2003), pp. 337–365.

[21] B. Maniá, *Sopra un esempio di Lavrentieff*, Boll. Un. Mat. Ital., 13 (1934), pp. 147–153.

[22] W. F. Mitchell, *A comparison of adaptive refinement techniques for elliptic problems*, ACM Trans. Math. Software, 15 (1989), pp. 326–347.

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

[24] P. Morin, K. G. Siebert, and A. Veeser, *A basic convergence result for conforming adaptive finite elements*, Math. Models Methods Appl. Sci., 18 (2008), pp. 707–737.

[25] C. Ortner, *Nonconforming finite-element discretization of convex variational problems*, IMA J. Numer. Anal., to appear.

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

[27] R. Verfürth, *Error estimates for some quasi-interpolation operators*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 695–713.

[28] A. Visintin, *Strong convergence results related to strict convexity*, Comm. Partial Differential Equations, 9 (1984), pp. 439–466.