Adaptive regularization, discretization, and linearization for nonsmooth problems based on primal-dual gap estimators

François Févotte, Ari Rappaport, Martin Vohralík

Adaptive regularization, discretization, and linearization for nonsmooth problems based on primal-dual gap estimators

François Févotte[‡] Ari Rappaport[*†] Martin Vohralík[*†]

September 19, 2023

Abstract

We consider nonsmooth partial differential equations associated with a minimization of an energy functional. We adaptively regularize the nonsmooth nonlinearity so as to be able to apply the usual Newton linearization, which is not always possible otherwise. We apply the finite element method as a discretization. We focus on the choice of the regularization parameter and adjust it on the basis of an a posteriori error estimate for the difference of energies of the exact and approximate solutions. Importantly, our estimates distinguish the different error components, namely those of regularization, linearization, and discretization. This leads to an algorithm that steers the overall procedure by adaptive stopping criteria with parameters for the regularization, linearization, and discretization levels. We prove guaranteed upper bounds for the energy difference and discuss the robustness of the estimates with respect to the magnitude of the nonlinearity when the stopping criteria are satisfied. Numerical results illustrate the theoretical developments.

Key words: nonlinear elliptic problem, nonsmooth nonlinearity, adaptive regularization, finite elements, primal-dual gap, equilibrated flux reconstruction

Contents

[*]Inria, 2 rue Simone Iff, 75589 Paris, France.
[†]Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée, France.
[‡]Triscale innov, 7 rue de la Croix Martre, 91120 Palaiseau, France.

1 Introduction

Given a Hilbert space V, consider the abstract minimization problem

$$u := \operatorname*{arg\,min}_{v \in V} \mathcal{J}(v)$$

where $\mathcal{J}$ is a convex functional. When $\mathcal{J}$ satisfies certain regularity conditions, one can form the associated Euler–Langrange conditions which are expressed as a nonlinear elliptic partial differential equation (PDE). We are particularly interested in cases where it is difficult to iteratively solve the resulting nonlinear PDE by the standard Newton method, cf. [29, 15], due to non-smoothness. More specifically, applying the standard Newton method can lead to slow convergence or even failure. The difficulty in many cases is the appearance of kink functions, i.e., continuous functions that are non-differentiable on a finite set. In Figure 1, we give three examples in the context of degenerate PDEs. In this work, we, in particular, seek to recover good convergence of Newton's method by adaptively regularizing the nonlinear PDE.

By replacing the original problem by a regularized one, a regularization error appears. For inverse problems [27, 32, 20], regularization is a common strategy and the error due to regularization has been extensively studied. In [32], the authors study the so-called Tikhonov regularization and its associated error. The regularization parameter is chosen adaptively and various criteria are discussed. Regularization is also considered for degenerate PDEs where the operators change type as a function of either space or time [33, 16, 31]. In these cases, a regularized problem is introduced that does not suffer from degeneracy. It is proven in [33] that the regularized solutions converge to the true solution in an approximate sense.

In a similar spirit, for Newton-type methods, regularization (smoothing) Newton methods replace non-differentiable nonlinearities with smooth counterparts, see [46, 36, 35] and the references therein. In this case, the amount of added regularization is proportional to a parameter that is driven to zero as the Newton iterations progress, thereby approaching the original problem.

From a practical point of view, the choice of the regularization parameter should ideally be updated in a dynamic way as the chosen numerical method converges. This leads to the question of how to choose the regularization parameter based on information from a solution iterate. In this work, we adaptively update the regularization parameter based on information from a posteriori error estimators.

A posteriori error estimation for PDEs is a well established subject, see for example the books of Verfürth [44], Ainsworth and Oden [1], Repin [40], and the references therein. A posteriori errors estimators can be utilized to 1) certify the error; 2) drive adaptive refinement strategies; and 3) provide

stopping criteria for iterative solvers. In general, important properties of such estimators are their *reliability* (upper bound for the error) and *efficiency* (lower bound for the error), where the constants in the upper bounds are ideally explicit, independent of the PDE data and finite-dimensional approximation parameters. More specifically in the context of adaptive stopping criteria, it is especially attractive to have a *constant-free* upper bound. As for the error lower bound, the goal is to obtain a generic constant ideally independent of the model parameters. In the case of strongly monotone and Lipschitz continuous operators, robustness with respect to the ratio of the Lipschitz constant to the monotonicity constant is of particular interest. We will present numerical evidence of robustness in such a setting, where theoretical developments are presented in [26]. However, unlike in [26], we consider regularization, explicitly estimate the regularization error, and introduce a solver strategy with adaptive regularization when the nonlinearity does not satisfy the hypotheses for Newton's method to converge. Furthermore, in this work we identify component error estimators and show that these estimators converge to zero in their respective limits. This is turn shows that the total error (measured in the same way as in [26]) converges to zero in a triple limit as detailed later.

In the context of energy minimization, it is advantageous to study a certain class of a posteriori estimators, namely the so-called primal-dual gap estimators [39, 38, 40, 6, 47]. These estimators rely on results from convex analysis to bound the "difference of energies" which we make precise in §4.1. In particular, these estimators do indeed provide a constant-free upper bound on the difference of energies.

In the recent works of [6, 5] Bartels et. al. employ the primal-dual gap estimator to drive a posteriori mesh refinement for singular solutions. It is also applied directly at the level of the energy minimization so that rough problems, e.g., posed in the space of functions of bounded variation (BV), can be treated without appealing to the Euler–Lagrange conditions. The energy minimization is solved directly via the so-called variable-alternating direction method of multipliers [5]. In this method the primal and dual problems are solved in a globally coupled, iterative manner. We note here that duality refers to the dual optimization problem and should not be confused with the notion of duality in adjoint-based a posteriori error analysis.

In the present context, we use a continuous Galerkin finite element discretization for the primal problem and perform a local equilibrated flux reconstruction to obtain a vector field in $\boldsymbol{H}(\mathrm{div}, \Omega)$ with the divergence prescribed by the load. This is achieved by solving linear, local, and mutually independent problems on patches of mesh elements. This resulting object is referred to as the equilibrated flux and is based on principles first established in Prager and Synge [34] and more recently in the works of Ladevèze and Leguillon [30], Destuynder and Métivet [14], Braess and Schöberl [9], and Ern and Vohralík [22]. One major advantage of this strategy is the so-called p-robustness, i.e., the resulting estimator is uniformly efficient for arbitrary polynomial order.

The main contribution of this work is an adaptive algorithm for solving nonsmooth problems by incorporating regularization into the algorithm. This in particular allows us to apply the standard Newton method to nonsmooth nonlinearities. This adaptive solution strategy resembles the one in [22], where the authors distinguish discretization, linearization, algebraic, and quadrature errors through computable component estimators. However, in [22] no regularization is considered when the nonlinearity is not differentiable, and the question of what to do in this case is not addressed. Here, we construct estimators for the errors due to regularization, discretization, and linearization. These estimators then lead to adaptive stopping criteria to steer an adaptive algorithm. We test our algorithm numerically and recover the optimal convergence rate under uniform refinement for a known smooth solution. We also consider a numerical test for an unknown solution on an L-shaped domain and observe the optimal rate of convergence with respect to total degrees of freedom (DOFs) as well as with respect to the cost for our estimator.

The rest of the paper is organized as follows. In §2 we introduce the relevant mathematical details of the problem, as well as our regularization strategy. In §3 we define the discrete spaces as well as the particular form of the Newton algorithm. In §4 we discuss some common notions of error and their relations to the difference of energies. Next, in §5 we introduce the necessary ideas from duality theory to describe the primal-dual gap estimators. In §6 we give the details of the flux reconstruction in the current setting. We introduce our decomposition of the upper bound provided by the primal-dual estimator in §7. We discuss the efficiency of the estimators in §8. We subsequently introduce the adaptive algorithm in §9 and we present numerical results in §10. Finally, we conclude in §11 and discuss future work.

2 Continuous problem statement and regularization

In this section we will fix continuous-level notation and then introduce in detail the model problem which we study throughout the rest of the paper.

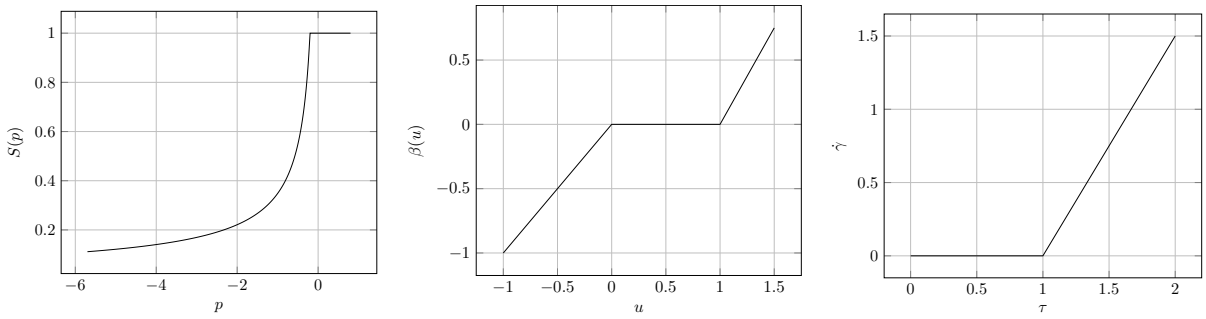

Figure 1 – Examples of kink-type nonlinear algebraic closures arising in the study of degenerate PDEs. From left to right, we have the saturation function for the Richards equation, namely for the Brooks–Corey model [11], the enthalphy function for the Stefan problem [23], and the shear stress/shear rate relation for Bingham plastics [28].

2.1 Notation

For $d = 2, 3$, let $\Omega \subset \mathbb{R}^d$ be a polygon or polyhedron with Lipschitz boundary, $\partial\Omega$. We define the Euclidean norm on $\mathbb{R}^d$ by $|\cdot|$. We introduce the space of Lebesgue square-integrable functions $L^2(\omega)$ with scalar product $(\cdot,\cdot)_\omega$ and norm $\|\cdot\|_\omega$ on $\omega \subseteq \Omega$. We drop the subscript when $\omega = \Omega$. We use the same notation for vector-valued functions. Next, we define, for scalar-valued functions, the standard Sobolev space $H^1(\Omega) = \{v \in L^2 : \partial_{x_i} v \in L^2(\Omega), \forall\, 1 \le i \le d\}$ with $H^1_0(\Omega)$ being the subspace of $H^1(\Omega)$ of functions with vanishing trace on $\partial\Omega$. For vector-valued functions we consider the space $\boldsymbol{H}(\mathrm{div}, \Omega) := \{\boldsymbol{v} \in [L^2(\Omega)]^d; \nabla\cdot\boldsymbol{v} \in L^2(\Omega)\}$.

2.2 Energy minimization and equivalent formulations

Let $\phi : \mathbb{R} \to \mathbb{R}$ be a given function and $f \in L^2(\Omega)$. Consider the functional $\mathcal{J} : H^1_0(\Omega) \to \mathbb{R}$ given by

$$\mathcal{J}(v) := \int_\Omega \phi(|\nabla v|)\,\mathrm{d}\boldsymbol{x} - \int_\Omega f v\,\mathrm{d}\boldsymbol{x}. \tag{2.1}$$

We will make the following assumptions on the function ϕ.

Assumption 2.1 (Assumptions on the energy function)**.** *We assume that the function ϕ is convex and of class $C^1(\mathbb{R})$ with*

$$\phi(0) = \phi'(0) = 0. \tag{2.2}$$

We further assume ϕ satisfies, for real constants $0 < \alpha \le L$,

$$|\phi'(r) - \phi'(s)| \le L|r - s| \quad \forall r, s \in \mathbb{R}, \tag{2.3a}$$

$$(\phi'(r) - \phi'(s))(r - s) \ge \alpha(r - s)^2 \quad \forall r, s \in \mathbb{R}. \tag{2.3b}$$

Note that (2.3b) in particular implies $\phi'(r) \ge \alpha r$, $\forall r \ge 0$, so that $\phi' : \mathbb{R}^+ \to \mathbb{R}^+$.

We will be interested in the solution to the minimization problem

$$u := \operatorname*{arg\,min}_{v \in H^1_0(\Omega)} \mathcal{J}(v). \tag{2.4}$$

Due to the convexity of the functional $\mathcal{J}$ following from Assumption 2.1 and the fact that $H^1_0(\Omega)$ is complete, the solution exists and is unique, see e.g. [45]. Another way to characterize the solution to problem (2.4) is through its Euler–Langrange equations. To this end, we introduce the nonlinear functions $a : \mathbb{R} \to \mathbb{R}$, $\boldsymbol{A} : \mathbb{R}^d \to \mathbb{R}^d$,

$$a(s) := \frac{\phi'(s)}{s}, \quad \boldsymbol{A}(\boldsymbol{q}) := a(|\boldsymbol{q}|)\boldsymbol{q}. \tag{2.5}$$

A consequence of this definition is the following.

Lemma 2.2 (Strongly monotone and Lipschitz continuous operator)**.** *For L and α from Assumption 2.1, the operator $\boldsymbol{A}$ given by (2.5) is strongly monotone*

$$\alpha\|\nabla(v - w)\|^2 \le (\boldsymbol{A}(\nabla v) - \boldsymbol{A}(\nabla w), \nabla(v - w)) \quad \forall v, w \in H^1_0(\Omega). \tag{2.6}$$

It is also Lipschitz continuous

$$\|\boldsymbol{A}(\nabla v) - \boldsymbol{A}(\nabla w)\| \leq L\|\nabla(v-w)\| \quad \forall v, w \in H_0^1(\Omega). \tag{2.7}$$

The proof is standard and is detailed in, e.g., [26, Proposition A.1]. Then the solution to (2.4) also solves the following weak formulation (the Euler–Langrange equations of (2.4)): find $u \in H_0^1(\Omega)$ such that

$$(\boldsymbol{A}(\nabla u), \nabla v) = (f, v) \quad \forall v \in H_0^1(\Omega). \tag{2.8}$$

Consequently, the strong formulation of (2.8) and (2.4) is given by the boundary-value problem

$$-\nabla\cdot\boldsymbol{A}(\nabla u) = f \quad \text{in } \Omega, \tag{2.9a}$$
$$u = 0 \quad \text{on } \partial\Omega. \tag{2.9b}$$

2.3 Regularization

According to Assumption 2.1 it is possible that the function ϕ belongs to the class $C^1(\mathbb{R})$ but not $C^2(\mathbb{R})$. In particular, this means that the nonlinear functions a and subsequently $\boldsymbol{A}$ defined by (2.5) are not necessarily (Fréchet) differentiable. We give an example in §2.4 below. Thus the classical Newton method to iteratively linearize (2.8) can struggle to converge as we demonstrate later with numerical examples. To overcome this issue, our approach will be to introduce an auxiliary regularized problem that we use to create a sequence of solutions that approach the solution to the non-regularized problem. We will more precisely introduce a regularized function defined from ϕ', parameterized by $\epsilon > 0$, which we call ϕ'_ϵ. We will make the following assumption.

Assumption 2.3 (Regularization of ϕ)**.** *For every $\epsilon > 0$, the regularized function satisfies*

$$\phi'_\epsilon(0) = 0, \tag{2.10}$$
$$\phi'_\epsilon \in C^2(\mathbb{R}). \tag{2.11}$$

Next, the regularized function satisfies inequalities similar to (2.3):

$$|\phi'_\epsilon(r) - \phi'_\epsilon(s)| \leq \overline{L}|r-s| \quad \forall r, s \in \mathbb{R}, \tag{2.12a}$$
$$(\phi'_\epsilon(r) - \phi'_\epsilon(s))(r-s) \geq \underline{\alpha}(r-s)^2 \quad \forall r, s \in \mathbb{R}, \tag{2.12b}$$

where $0 < \underline{\alpha} \leq \overline{L}$ are real constants independent of ϵ. Moreover, in all points $s \in \mathbb{R}$, there holds

$$(\phi'_\epsilon - \phi')(s) \overset{\epsilon\to 0}{\to} 0. \tag{2.13}$$

The regularized function implicitly defines regularized versions a_ϵ and $\boldsymbol{A}_\epsilon$ as in (2.5) through

$$a_\epsilon(s) := \frac{\phi'_\epsilon(s)}{s}, \boldsymbol{A}_\epsilon(\boldsymbol{q}) := a_\epsilon(|\boldsymbol{q}|)\boldsymbol{q}, \tag{2.14}$$

which, by the same reasoning as in Lemma 2.2, satisfy

$$\underline{\alpha}\|\nabla(v-w)\|^2 \leq (\boldsymbol{A}_\epsilon(\nabla v) - \boldsymbol{A}_\epsilon(\nabla w), \nabla(v-w)) \quad \forall v, w \in H_0^1(\Omega), \tag{2.15a}$$
$$\|\boldsymbol{A}_\epsilon(\nabla v) - \boldsymbol{A}_\epsilon(\nabla w)\| \leq \overline{L}\|\nabla(v-w)\| \quad \forall v, w \in H_0^1(\Omega). \tag{2.15b}$$

We now prove a consequence of Assumption 2.3 which will be useful later.

Lemma 2.4 (L^2 convergence of the regularization)**.** *For any vector field $\boldsymbol{v} \in L^2(\Omega)$ we have that*

$$\lim_{\epsilon\to 0} \|\phi'_\epsilon(|\boldsymbol{v}|) - \phi'(|\boldsymbol{v}|)\| \to 0. \tag{2.16}$$

Proof. We will make use of the Lebesgue dominated convergence theorem in $L^2(\Omega)$, see e.g. [42]. Indeed, consider an arbitrary real sequence $\{\epsilon^n\}_{n\in\mathbb{N}}$ tending to zero and consider the sequence of functions

$$g_n(\boldsymbol{x}) := \phi'_{\epsilon^n}(|\boldsymbol{v}(\boldsymbol{x})|) - \phi'(|\boldsymbol{v}(\boldsymbol{x})|).$$

Then by (2.13) from Assumption 2.3, we have that $g_n \to 0$ pointwise for almost all $\boldsymbol{x} \in \Omega$. Next, to establish a dominating function, observe that since $\phi'(0) = \phi'_\epsilon(0) = 0$ by (2.2) and (2.10), we have

$$|g_n(x)| \le |\phi'_{\epsilon^n}(|\boldsymbol{v}(\boldsymbol{x})|)| + |\phi'(|\boldsymbol{v}(\boldsymbol{x})|)| \overset{(2.3a),(2.12a)}{\le} (\overline{L} + L)|\boldsymbol{v}(\boldsymbol{x})| =: g(\boldsymbol{x}) \in L^2(\Omega),$$

so we can choose g as the dominating function. Finally, since ϵ^n was arbitrary, the sequential criterion for a limit ensures (2.16). □

For algorithmic reasons, we will consider a monotonically decreasing sequence, $\{\epsilon^j\}_{j\ge 0}$ of positive real values which is in particular determined by two values $\epsilon^0 > 0$ and $0 < C_\epsilon < 1$, where, for $j \ge 1$,

$$\epsilon^j := C_\epsilon \epsilon^{j-1}. \tag{2.17}$$

All these considerations lead us to a regularized version of the problem (2.8): for a fixed $j \ge 0$, find $u^j \in H^1_0(\Omega)$ such that

$$(\boldsymbol{A}_{\epsilon^j}(\nabla u^j), \nabla v) = (f, v) \quad \forall v \in H^1_0(\Omega). \tag{2.18}$$

2.4 An example nonsmooth nonlinearity with a kink

To make our notions more concrete, we introduce a simple but instructive example for our study. Consider $\phi \in C^1(\mathbb{R}) \setminus C^2(\mathbb{R})$ given by

$$\phi(s) := \begin{cases} \frac{1}{2}(s - s_0)^2 + s_0 s - \frac{1}{2}s_0^2 & s \le s_0, \\ \frac{m}{2}(s - s_0)^2 + s_0 s - \frac{1}{2}s_0^2, & s > s_0 \end{cases} \tag{2.19}$$

with continuous derivative

$$\phi'(s) = \begin{cases} s, & s \le s_0, \\ m(s - s_0) + s_0, & s > s_0, \end{cases} \tag{2.20}$$

where $s_0 > 0$ determines the location of the discontinuity in the second derivative and $m \ge 1$ determines the slope to the right of s_0. An illustration is given in Figure 2. We call the function (2.20) a kink function due to the fact that $\phi'(s)$ is not strongly differentiable at the point s_0. This function satisfies Assumption 2.1 with $L = m$ and $\alpha = 1$ since the weak derivative of ϕ' is given by

$$\phi''(s) = \begin{cases} 1, & s < s_0, \\ m, & s > s_0. \end{cases} \tag{2.21}$$

For this particular choice of function ϕ, applying the standard Newton method leads to failure of convergence as illustrated in the example of §10.1.2.

We now introduce a regularized version of the function ϕ' of (2.20) that in turn defines regularized versions of the nonlinear functions a and $\boldsymbol{A}$ in (2.14). We first notice that (2.19) can be equivalently rewritten as

$$\phi'(s) = \frac{m-1}{2}|s - s_0| + \frac{m+1}{2}(s - s_0) + s_0. \tag{2.22}$$

We then consider, for a fixed value of $\epsilon > 0$, the smooth approximation of the absolute value function

$$|s|_\epsilon := \sqrt{s^2 + \epsilon^2}. \tag{2.23}$$

We then replace the absolute value in (2.22) by the smooth version

$$\hat{\phi}'_\epsilon(s) := \frac{m-1}{2}|s - s_0|_\epsilon + \frac{m+1}{2}(s - s_0) + s_0. \tag{2.24}$$

We then set

$$\phi'_\epsilon(s) := \hat{\phi}'_\epsilon(s) - \hat{\phi}'_\epsilon(0) \tag{2.25}$$

to achieve $\phi'_\epsilon(0) = 0$. An illustration is given in Figure 3. We now show the following

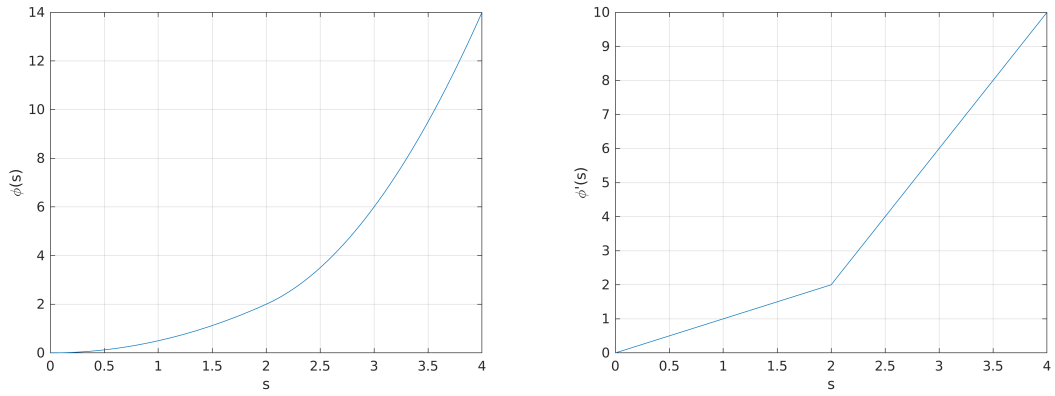

Figure 2 – [Kink function (2.19) with $m = 4$, $s_0 = 2$] The kink function ϕ (2.19) and its derivative ϕ' (2.20) Notice that ϕ' is not strongly differentiable at s_0.

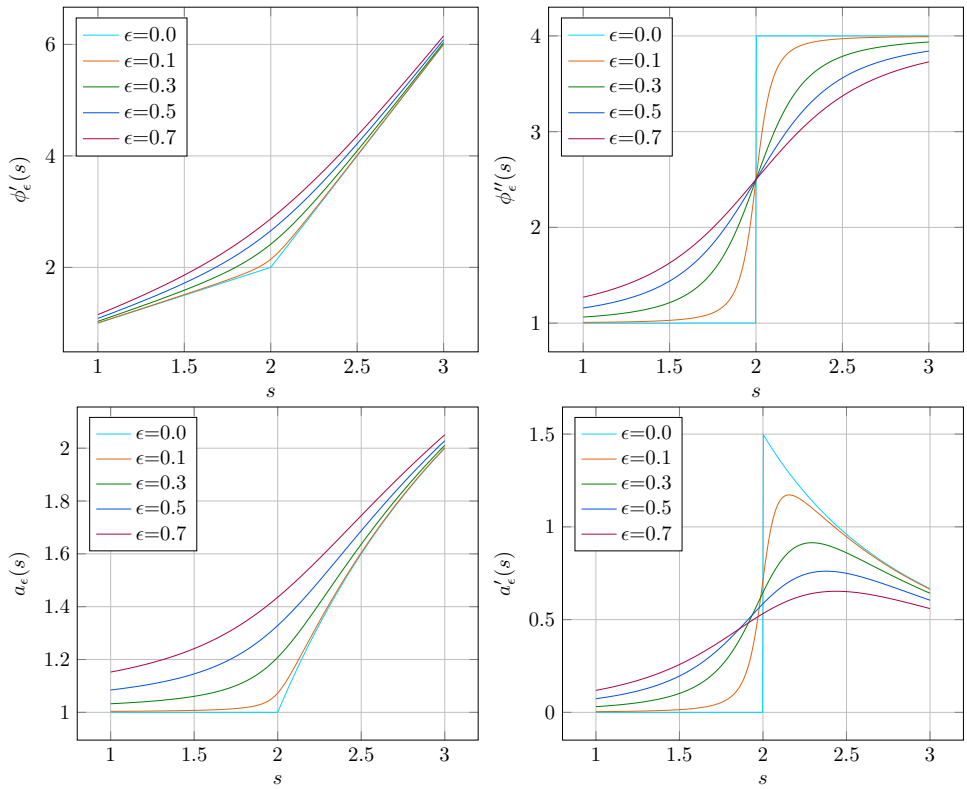

Figure 3 – [Kink function (2.19) with $m = 4$, $s_0 = 2$] Regularization of the kink function (2.19) by replacing the absolute value function with its differentiable counterpart, see (2.25).

Lemma 2.5 (Example regularization (2.25))**.** *The definition* (2.25) *satisfies Assumption 2.3.*

Proof. The requirement (2.10) is obvious from the definition. Next, the function is actually $C^\infty(\mathbb{R})$—in particular $C^2(\mathbb{R})$—so (2.11) is satisfied. Next, for (2.12), observe that

$$\phi''_\epsilon(s) = \frac{m-1}{2}\frac{s}{\sqrt{s^2+\epsilon^2}} + \frac{m+1}{2}$$

is strictly increasing and $\lim_{s\to-\infty}\phi''_\epsilon(s) = 1$, as well as $\lim_{s\to\infty}\phi''_\epsilon(s) = m$. Thus, $1 \le \phi''_\epsilon \le m$ and appealing to, e.g., [26, Proposition A.2], we conclude that (2.12) is satisfied with $\overline{L} = m$ and $1 = \underline{\alpha}$,

Finally, we show that for all $s \in \mathbb{R}$,

$$|s|_\epsilon - |s| = \sqrt{s^2+\epsilon^2} - \sqrt{s^2} = \frac{\epsilon^2}{\sqrt{s^2+\epsilon^2}+\sqrt{s^2}} \to 0$$

when $\epsilon \to 0$, which confirms (2.13). □

3 Discrete problem and linearization

We now give details for the discretization of the regularized problem (2.18) via the continuous Galerkin finite element method [21] and subsequent linearization.

3.1 Finite element discretization

Let $\mathcal{T}_0$ be a simplicial mesh of the physical domain Ω with no "hanging nodes" i.e., $\mathcal{T}_0 = \cup_K \{K\}$, where the intersection of (the closure of) two arbitrary simplices $K, K' \in \mathcal{T}_0$ are either empty or an l-dimensional simplex for $0 \le l \le d-1$. From the initial mesh $\mathcal{T}_0$, we generate a hierarchy $\{\mathcal{T}_\ell\}_{\ell=1}^L$ of nested meshes, i.e., $\mathcal{T}_\ell \subset \mathcal{T}_{\ell+1}$ for all $\ell \ge 0$. We assume that each mesh in the hierarchy is also free of hanging nodes in the same sense as for $\mathcal{T}_0$. We also assume that the hierarchy of meshes is shape regular, i.e., that there exists a constant $\kappa_\mathcal{T}$ such that, for all ℓ,

$$\max_{K\in\mathcal{T}_\ell} \kappa_K \le \kappa_\mathcal{T}, \tag{3.1}$$

where $\kappa_K := \frac{h_K}{\rho_K}$, h_K is the diameter of K, and ρ_K is the radius of the largest inscribed ball of K. For an arbitrary collection of simplices $\mathcal{T}$ of some mesh $\mathcal{T}_\ell$ and its corresponding subdomain $\omega \subset \Omega$, we define the broken polynomial space of order $p \ge 0$ by

$$\mathcal{P}_p(\mathcal{T}) := \{v \in L^2(\omega) : v|_K \in \mathcal{P}_p(K), \forall K \in \mathcal{T}\}. \tag{3.2}$$

Finally, we introduce, for a fixed polynomial degree $p \ge 1$ and for each ℓ an $H_0^1(\Omega)$-conforming finite-dimensional space,

$$V_\ell^p := H_0^1(\Omega) \cap \mathcal{P}_p(\mathcal{T}_\ell). \tag{3.3}$$

We now consider a discrete equivalent of the regularized continuous problem (2.18) where we seek, for $j, \ell \ge 0$ the solution $u_\ell^j \in V_\ell^p$ such that

$$(\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j), \nabla v_\ell) = (f, v_\ell) \quad \forall v_\ell \in V_\ell^p. \tag{3.4}$$

Note that while this problem is finite-dimensional, it is still nonlinear.

3.2 Linearization

We now define a linearization scheme for the regularized finite-dimensional problem (3.4). For fixed $\ell, j \ge 0$, let $u_\ell^{j,0} \in V_\ell^p$ be the initial guess. Denoting by $k \ge 1$ the linearization iterate, a step of the linearization procedure to approximately solve (3.4) takes the form: find $u_\ell^{j,k} \in V_\ell^p$ such that

$$(\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}), \nabla v) = (f, v) \quad \forall v \in V_\ell^p, \tag{3.5}$$

where the operator $\boldsymbol{A}_{\epsilon^j}^{k-1}$ is affine and takes the form

$$\boldsymbol{A}_{\epsilon^j}^{k-1}(\boldsymbol{q}) := \mathbb{A}_{\epsilon^j}^{k-1}\boldsymbol{q} - \boldsymbol{b}_{\epsilon^j}^{k-1} \tag{3.6}$$

for a matrix-valued function $\mathbb{A}_{\epsilon^j}^{k-1} : \Omega \to \mathbb{R}^{d\times d}$ and a vector-valued function $\boldsymbol{b}_{\epsilon^j}^{k-1} : \Omega \to \mathbb{R}^d$. Once a basis of V_ℓ^p is chosen, the problem (3.5) is equivalent to solving a linear system of algebraic equations.

We make the following assumptions on the linearization.

Assumption 3.1 (Assumptions on the linearization)**.** *Let the regularization step $j \ge 0$ and mesh level $\ell \ge 0$ be fixed. We assume that the linearized operator converges in the sense that*

$$\lim_{k\to\infty} \|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j)\| = 0. \tag{3.7}$$

Finally, we assume that $\mathbb{A}_{\epsilon^j}^{k-1}$ is uniformly bounded and symmetric positive definite, that is, the following conditions hold for all $\boldsymbol{x} \in \Omega$ and all $\boldsymbol{v} \in \mathbb{R}^d$,

$$|\mathbb{A}_{\epsilon^j}^{k-1}(\boldsymbol{x})\boldsymbol{v}| \le \overline{\lambda}|\boldsymbol{v}| \qquad \textit{(boundedness)}, \tag{3.8a}$$

$$\underline{\lambda}|\boldsymbol{v}|^2 \le (\mathbb{A}_{\epsilon^j}^{k-1}(\boldsymbol{x})\boldsymbol{v}) \cdot \boldsymbol{v} \qquad \textit{(positive definiteness)}, \tag{3.8b}$$

where $0 < \underline{\lambda} \le \overline{\lambda}$ are real constants.

A prototypical example is the Picard, or fixed point, linearization where

$$\boldsymbol{A}_{\epsilon^j}^{k-1}(\boldsymbol{q}) := a_{\epsilon^j}(|\nabla u_\ell^{j,k-1}|)\boldsymbol{q}, \tag{3.9}$$

whereas for a Newton step the linearized function is given by

$$\boldsymbol{A}_{\epsilon^j}^{k-1}(\boldsymbol{q}) := a_{\epsilon^j}(|\nabla u_\ell^{j,k-1}|)\boldsymbol{q} + \big(\partial_{\boldsymbol{q}} a_{\epsilon^j}(|\nabla u_\ell^{j,k-1}|) \otimes \nabla u_\ell^{j,k-1}\big)(\boldsymbol{q} - \nabla u_\ell^{j,k-1}), \tag{3.10}$$

and, in turn,

$$\partial_{\boldsymbol{q}} a_{\epsilon^j}(|\boldsymbol{q}|) = \frac{a'_{\epsilon^j}(|\boldsymbol{q}|)}{|\boldsymbol{q}|}\boldsymbol{q} = \big(\phi''_\epsilon(|\boldsymbol{q}|) - \phi'_\epsilon(|\boldsymbol{q}|)|\boldsymbol{q}|^{-1}\big)|\boldsymbol{q}|^{-2}\boldsymbol{q}. \tag{3.11}$$

In [26, Section 2.3.2] it is demonstrated that these two linearization schemes satisfy (3.8) for $\underline{\lambda} = \underline{\alpha}$ and $\overline{\lambda} = \overline{L}$.

Remark 3.2 (Newton linearization)**.** *Note that the derivative $\phi''_\epsilon(s)$ appears in (3.11). For nonsmooth nonlinearities, where merely the function $\phi \in C^1(\mathbb{R}) \setminus C^2(\mathbb{R})$, the second derivative ϕ'' does not exist in a strong sense. This is the motivation for introducing the regularization.*

4 Three ways of measuring the error and their mutual relations

We have now established and characterized different solutions, namely the true solution u of (2.8) and the approximate solution to the regularized, discretized, and linearized problem of (3.5) $u_\ell^{j,k}$. The next step is to define a concrete notion of error between these two objects, and then, ideally, to have a computable means of estimating it. We postpone the discussion of error estimation to §7. In this section we will recall, following [40, 47, 3], three notions of error between u and $u_\ell^{j,k}$, namely the difference of energies (2.1), a notion related to a (weighted by $\alpha^{1/2}$) $H_0^1(\Omega)$ norm, and the dual (weighted by $\alpha^{-1/2}$) norm of the residual. One salient feature when comparing these error measures is that they all coincide in the linear case where $\phi(s) = s^2/s$ and $\boldsymbol{A}(\boldsymbol{q}) = \boldsymbol{q}$, as will be made precise in §4.4. Recall that we assume that the constants L and α are given by Assumption 2.1.

4.1 Energy difference

A physically-motivated notion of error is the difference of energies. For $v \in H_0^1(\Omega)$, this is given as

$$0 \leq \mathcal{J}(v) - \mathcal{J}(u), \tag{4.1}$$

where the energy functional $\mathcal{J}$ is defined in (2.1). Since the true solution u is the unique minimum of $\mathcal{J}$ in $H_0^1(\Omega)$ as per (2.4), this quantity is guaranteed to be nonnegative and only 0 when $v = u$.

4.2 Energy norm

A second type of error measure we will consider will be that of the energy norm. For $v \in H_0^1(\Omega)$ we namely consider

$$|||v||| := \alpha^{1/2}\|\nabla v\|, \tag{4.2}$$

where $\alpha > 0$ is the monotonicity constant from (2.3b). Thus, the error between $v \in H_0^1(\Omega)$ and the solution u of (2.8) is here expressed as

$$0 \leq \tfrac{1}{2}|||u - v|||^2. \tag{4.3}$$

The reason for the choice of squaring and dividing by two will become clear in §4.4.

4.3 Dual norm of the residual

Finally we consider the abstract error quantity obtained through the dual norm of the residual. First, we define, for a fixed $v \in H_0^1(\Omega)$, the residual functional $\mathcal{R}(v) \in H^{-1}(\Omega)$ by

$$\langle \mathcal{R}(v), w\rangle := (f, w) - (\boldsymbol{A}(\nabla v), \nabla w), \quad w \in H_0^1(\Omega), \tag{4.4}$$

where the duality pairing $\langle\cdot,\cdot\rangle$ is between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$. Next we introduce the dual norm for $\mathcal{R} \in H^{-1}(\Omega)$

$$|||\mathcal{R}(v)|||_{-1} := \sup_{\varphi \in H_0^1(\Omega),\, |||\varphi||| = 1} \langle \mathcal{R}(v), \varphi\rangle. \tag{4.5}$$

Here the error in the energy dual norm of the residual is given, for $v \in H^1_0(\Omega)$,

$$0 \leq \tfrac{1}{2}|||\mathcal{R}(v)|||^2_{-1}. \tag{4.6}$$

As in the previous section, note that $\mathcal{R}(v) = 0$ if and only if $v = u$ solves the continuous problem (2.8). The choice of squaring and dividing by two will again be made clear in §4.4. Finally, if we consider the standard definition of the dual norm,

$$\|\mathcal{R}(v)\|_{H^{-1}(\Omega)} := \sup_{\varphi \in H^1_0(\Omega),\, \|\nabla\varphi\|=1} \langle \mathcal{R}(v), \varphi \rangle, \tag{4.7}$$

our definition satisfies the scaling

$$|||\mathcal{R}(v)|||_{-1} = \alpha^{-1/2}\|\mathcal{R}(v)\|_{H^{-1}(\Omega)}. \tag{4.8}$$

4.4 Equivalence in the linear case

In this section, we recall the special relationship between the three different error measures of the previous sections in the case where $\phi(s) = s^2/2$, i.e., $\alpha = L = 1$, which writes as

$$\mathcal{J}(v) - \mathcal{J}(u) = \tfrac{1}{2}|||v-u|||^2 = \tfrac{1}{2}|||\mathcal{R}(v)|||^2_{-1}. \tag{4.9}$$

For the sake of completeness, we recall the proof of (4.9). As for the first equality,

$$\begin{aligned}
\mathcal{J}(v) - \mathcal{J}(u) &= \int_\Omega \tfrac{1}{2}|\nabla v|^2 - fv \,\mathrm{d}\boldsymbol{x} - \left(\int_\Omega \tfrac{1}{2}|\nabla u|^2 - fu \,\mathrm{d}\boldsymbol{x} \right) \\
&\overset{(2.8)}{=} \int_\Omega \tfrac{1}{2}|\nabla v|^2 - \nabla u \cdot \nabla v \,\mathrm{d}\boldsymbol{x} - \left(\int_\Omega \tfrac{1}{2}|\nabla u|^2 - \nabla u \cdot \nabla u \,\mathrm{d}\boldsymbol{x} \right) \\
&= \int_\Omega \tfrac{1}{2}|\nabla v|^2 - \nabla u \cdot \nabla v + \tfrac{1}{2}|\nabla u|^2 \,\mathrm{d}\boldsymbol{x} = \tfrac{1}{2}|||u-v|||^2.
\end{aligned}$$

The latter one is then simply

$$|||u-v||| = \sup_{\varphi \in H^1_0(\Omega),\, \|\nabla\varphi\|=1} (\nabla(u-v), \nabla\varphi) = \sup_{\varphi \in H^1_0(\Omega),\, \|\nabla\varphi\|=1} \{(f,\varphi) - (\nabla v, \nabla\varphi)\} = |||\mathcal{R}(v)|||_{-1}.$$

4.5 Relations in the nonlinear case

In the nonlinear case, the measures are only equivalent up to factors of L/α. Indeed, between the dual norm of the residual and the energy norm, a factor of exactly L/α relates the two:

Proposition 4.1 (Relation energy norm-dual residual norm)**.** *Let u solve (2.8) and let $v \in H^1_0(\Omega)$ be arbitrary. Then there holds*

$$\tfrac{1}{2}|||v-u|||^2 \leq \tfrac{1}{2}|||\mathcal{R}(v)|||^2_{-1} \leq \frac{L^2}{2\alpha^2}|||v-u|||^2. \tag{4.10}$$

Proof. By the definition of the energy norm (4.2) and the dual residual norm (4.5), together with the monotonicity assumption (2.3b) implying (2.6) and definition (2.8),

$$\begin{aligned}
|||v-u||| = \alpha^{1/2}\|\nabla(v-u)\| &\overset{(2.6)}{\leq} \frac{(\boldsymbol{A}(\nabla v) - \boldsymbol{A}(\nabla u), \nabla(v-u))}{\alpha^{1/2}\|\nabla(v-u)\|} \\
&\leq \sup_{\varphi \in H^1_0(\Omega),\, |||\varphi|||=1} (\boldsymbol{A}(\nabla v) - \boldsymbol{A}(\nabla u), \nabla\varphi) \\
&\overset{(2.8)}{=} \sup_{\varphi \in H^1_0(\Omega),\, |||\varphi|||=1} \{(\boldsymbol{A}(\nabla v), \nabla\varphi) - (f,\varphi)\} \\
&= |||\mathcal{R}(v)|||_{-1}.
\end{aligned} \tag{4.11}$$

For the second inequality, observe that, using the Lipschitz continuity assumption (2.3a) implying (2.7),

$$\begin{aligned}
|||\mathcal{R}(v)|||_{-1} &\overset{(4.8)}{=} \alpha^{-1/2}\|\mathcal{R}(v)\|_{H^{-1}(\Omega)} \\
&\overset{(2.8)}{=} \alpha^{-1/2} \sup_{\varphi\in H_0^1(\Omega),\|\nabla\varphi\|=1} (\boldsymbol{A}(\nabla v)-\boldsymbol{A}(\nabla u),\nabla\varphi) \\
&\leq \alpha^{-1/2} \sup_{\varphi\in H_0^1(\Omega),\|\nabla\varphi\|=1} \|\boldsymbol{A}(\nabla v)-\boldsymbol{A}(\nabla u)\|\,\|\nabla\varphi\| \\
&= \alpha^{-1/2}\|\boldsymbol{A}(\nabla v)-\boldsymbol{A}(\nabla u)\| \\
&\overset{(2.7)}{\leq} L\alpha^{-1/2}\|\nabla(v-u)\| \\
&\overset{(4.2)}{=} \frac{L}{\alpha}|||v-u|||.
\end{aligned} \tag{4.12}$$

□

When comparing the energy difference with the energy norm, a factor of $\sqrt{L/\alpha}$ is instead introduced.

Proposition 4.2. *Let u solve (2.8) and let $v \in H_0^1(\Omega)$ be arbitrary. Then there holds*

$$\tfrac{1}{2}|||v-u|||^2 \leq \mathcal{J}(v)-\mathcal{J}(u) \leq \frac{L}{2\alpha}|||v-u|||^2. \tag{4.13}$$

Proof. See [24, Lemma 5.1]. □

5 Duality theory

In order to bound the energy difference as introduced in §4.1, we will proceed using duality for convex functions, following [7, 40, 47].

5.1 Fenchel conjugate and its properties

Let us introduce the Fenchel conjugate (also known as the Legendre transform). For a convex function ϕ, it is given by

$$\phi^*(r) := \int_0^r (\phi')^{-1}(s)\,\mathrm{d}s. \tag{5.1}$$

We also define the conjugate operator

$$\boldsymbol{A}^*(\boldsymbol{q}) := \frac{(\phi^*)'(|\boldsymbol{q}|)}{|\boldsymbol{q}|}\boldsymbol{q}. \tag{5.2}$$

To illustrate, let us compute explicitly the Fenchel conjugate for the kink example (2.19) of §2.4. We first compute the inverse of the derivative

$$(\phi')^{-1}(s) = \begin{cases} s, & s \leq s_0 \\ \frac{s+(m-1)s_0}{m}, & s > s_0 \end{cases}. \tag{5.3}$$

Then the Fenchel conjugate (5.1) takes the form

$$\phi^*(r) = \int_0^r (\phi')^{-1}(s)\,\mathrm{d}s = \begin{cases} \frac{1}{2}r^2, & r \leq s_0 \\ \frac{1}{2m}[(r+(m-1)s_0)^2 - s_0^2 m(m-1)], & r > s_0 \end{cases}. \tag{5.4}$$

Indeed, for the case where $r \geq s_0$,

$$\begin{aligned}
\int_0^r (\phi')^{-1}(s)\,\mathrm{d}s &= \int_0^{s_0} s\,\mathrm{d}s + \int_{s_0}^r \frac{s+(m-1)s_0}{m}\,\mathrm{d}s \\
&= \frac{1}{2m}\left(r^2 + 2(m-1)s_0 r - s_0^2(m-1)\right) \\
&= \frac{1}{2m}[(r+(m-1)s_0)^2 - s_0^2 m(m-1)].
\end{aligned} \tag{5.5}$$

Definition (5.1) yields the following properties, see [40, 7] or Appendix A:

Proposition 5.1 (Properties of the Fenchel conjugate)**.** *Let ϕ be a convex function with $\phi(0) = \phi'(0) = 0$ and let $\phi^* : \mathbb{R} \to \mathbb{R}$ be its Fenchel conjugate given by (5.1). Then the following properties hold.*

$$\phi^*(r) = r(\phi')^{-1}(r) - \phi((\phi')^{-1}(r)) = \max_s\{sr - \phi(s)\}, \tag{5.6a}$$

$$\phi^* \text{ is convex}, \tag{5.6b}$$

$$\phi^* \in C^1(\mathbb{R}) \text{ and } (\phi^*)' = (\phi')^{-1}, \tag{5.6c}$$

$$\phi^*(0) = (\phi^*)'(0) = 0. \tag{5.6d}$$

This proposition can be used to derive the following well-known result [40, 47, 7, 37] or Appendix A:

Corollary 5.2 (Young's inequality for convex functions)**.** *Let $\phi \in C^1(\mathbb{R})$ be convex and let ϕ^* be given by (5.1). Then*

$$sr \le \phi(s) + \phi^*(r) \quad \text{for all } s, r \ge 0, \tag{5.7}$$

where the equality holds for $r = \phi'(s)$ or equivalently $s = (\phi^*)'(r)$.

Next, we consider the relationship between the vector-valued counterparts.

Corollary 5.3 ($\boldsymbol{A}$ and $\boldsymbol{A}^*$)**.** *Let $\boldsymbol{A}$ be given by (2.5) and $\boldsymbol{A}^*$ be given by (5.2). Then the following holds for all $\boldsymbol{q} \in \mathbb{R}^d$*

$$\boldsymbol{A}(\boldsymbol{A}^*(\boldsymbol{q})) = \boldsymbol{q} \tag{5.8a}$$

$$\boldsymbol{A}(\boldsymbol{q}) \cdot \boldsymbol{q} = \phi(|\boldsymbol{q}|) + \phi^*(|\boldsymbol{A}(\boldsymbol{q})|). \tag{5.8b}$$

Finally, there holds:

Lemma 5.4 (Lipschitz continuity of $(\phi^*)'$)**.** *Let ϕ satisfy the Assumption 2.1. Then $(\phi^*)'$ is Lipschitz continuous with Lipschitz constant equal to α^{-1}, i.e.,*

$$|(\phi^*)'(r) - (\phi^*)'(s)| \le \alpha^{-1}|r - s|. \tag{5.9}$$

The proof of these results is again given in Appendix A.

5.2 The two energies principle

We are led to investigate the dual optimization problem to (2.4). The dual problem to (2.4) can be stated as

$$\boldsymbol{\sigma} := \operatorname*{arg\,max}_{\substack{\boldsymbol{\varsigma} \in \boldsymbol{H}(\mathrm{div}, \Omega) \\ \nabla \cdot \boldsymbol{\varsigma} = f}} \mathcal{J}^*(\boldsymbol{\varsigma}), \tag{5.10}$$

where the dual functional $\mathcal{J}^* : \boldsymbol{H}(\mathrm{div}, \Omega) \to \mathbb{R}$ is given by

$$\mathcal{J}^*(\boldsymbol{\varsigma}) := -\int_\Omega \phi^*(|\boldsymbol{\varsigma}|)\, \mathrm{d}\boldsymbol{x}. \tag{5.11}$$

It turns out that the flux $\boldsymbol{\sigma} = -\boldsymbol{A}(\nabla u)$, where u is given by (2.4), or, equivalently, by (2.8). We see this in the following way. First of all, the Euler–Langrange conditions for (5.10) are: find $\boldsymbol{\sigma} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{\sigma} = f$ such that

$$(\boldsymbol{A}^*(\boldsymbol{\sigma}), \boldsymbol{v}) = 0 \quad \forall \boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}, \Omega) \text{ with } \nabla \cdot \boldsymbol{v} = 0. \tag{5.12}$$

We consider the equivalent mixed formulation of (5.12): find $(\boldsymbol{\sigma}, \tilde{u}) \in \boldsymbol{H}(\mathrm{div}, \Omega) \times L^2(\Omega)$ such that

$$(\boldsymbol{A}^*(\boldsymbol{\sigma}), \boldsymbol{v}) - (\tilde{u}, \nabla \cdot \boldsymbol{v}) = 0 \qquad \forall \boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}, \Omega), \tag{5.13a}$$

$$(\nabla \cdot \boldsymbol{\sigma}, q) = (f, q) \qquad \forall q \in L^2(\Omega). \tag{5.13b}$$

By the definition of the weak gradient, (5.13a) implies

$$\nabla \tilde{u} = -\boldsymbol{A}^*(\boldsymbol{\sigma}) \overset{(5.8a)}{\Longrightarrow} -\boldsymbol{A}(\nabla \tilde{u}) = \boldsymbol{\sigma}. \tag{5.14}$$

Finally taking $q \in H_0^1(\Omega) \subset L^2(\Omega)$ as the test function in (5.13b) shows that $\tilde{u} = u$ is the solution to problem (2.8).

To make the connection between the primal and dual problems more precise, we proceed to introduce the saddle point functional by

$$\mathcal{L}(v, \boldsymbol{\varsigma}) := \mathcal{J}^*(\boldsymbol{\varsigma}) - (\nabla v, \boldsymbol{\varsigma}) - (f, v), \quad \boldsymbol{\varsigma} \in \boldsymbol{H}(\mathrm{div}, \Omega),\ v \in H_0^1(\Omega). \tag{5.15}$$

Then we have the following

Lemma 5.5 (Two energies principle)**.** *Let u be the solution to the minimization problem (2.4) and $\boldsymbol{\sigma}$ be the solution to (5.10). Let $\mathcal{L}$ be as in (5.15). Then*

$$\max_{\substack{\boldsymbol{\varsigma}\in\boldsymbol{H}(\mathrm{div},\Omega)\\ \nabla\cdot\boldsymbol{\varsigma}=f}} \mathcal{J}^*(\boldsymbol{\varsigma}) = \mathcal{J}^*(\boldsymbol{\sigma}) = \mathcal{L}(u,\boldsymbol{\sigma}) = \mathcal{J}(u) = \min_{v\in H^1_0(\Omega)} \mathcal{J}(v). \tag{5.16}$$

There also holds

$$\mathcal{L}(u,\boldsymbol{\sigma}) = \max_{\substack{\boldsymbol{\varsigma}\in\boldsymbol{H}(\mathrm{div},\Omega)\\ \nabla\cdot\boldsymbol{\varsigma}=f}} \min_{v\in H^1_0(\Omega)} \mathcal{L}(v,\boldsymbol{\varsigma}). \tag{5.17}$$

Proof. The first and last equalities of (5.16) follow by definition. For the second equality of (5.16), note that from $\boldsymbol{\sigma}\in\boldsymbol{H}(\mathrm{div},\Omega)$ and $\nabla\cdot\boldsymbol{\sigma}=f$, for any $v\in H^1_0(\Omega)$, we have $-(\nabla v,\boldsymbol{\sigma})-(f,v)=0$. For the third equality of (5.16),

$$\begin{aligned}\mathcal{L}(u,\boldsymbol{\sigma}) &= \int_\Omega -\phi^*(|\boldsymbol{\sigma}|) - \nabla u\cdot\boldsymbol{\sigma} - fu\,\mathrm{d}\boldsymbol{x}\\ &\overset{(5.14)}{=} \int_\Omega -\phi^*(|\boldsymbol{A}(\nabla u)|) + \nabla u\cdot\boldsymbol{A}(\nabla u) - fu\,\mathrm{d}\boldsymbol{x} \overset{(5.8b)}{=} \int_\Omega \phi(|\nabla u|) - fu\,\mathrm{d}\boldsymbol{x} \overset{(2.1)}{=} \mathcal{J}(u).\end{aligned}$$

Finally, (5.17) follows since, as above,

$$\max_{\substack{\boldsymbol{\varsigma}\in\boldsymbol{H}(\mathrm{div},\Omega)\\ \nabla\cdot\boldsymbol{\varsigma}=f}} \min_{v\in H^1_0(\Omega)} \mathcal{L}(v,\boldsymbol{\varsigma}) \overset{(5.15)}{=} \max_{\substack{\boldsymbol{\varsigma}\in\boldsymbol{H}(\mathrm{div},\Omega)\\ \nabla\cdot\boldsymbol{\varsigma}=f}} \mathcal{J}^*(\boldsymbol{\varsigma}) = \mathcal{J}^*(\boldsymbol{\sigma}) = \mathcal{L}(u,\boldsymbol{\sigma}).$$

□

The above developments directly lead to an upper bound on the energy difference [39, 47, 4, 7]:

Corollary 5.6 (Two energies principle)**.** *Under the assumptions of Lemma 5.5, the following holds for any $\boldsymbol{\varsigma}\in\boldsymbol{H}(\mathrm{div},\Omega)$ with $\nabla\cdot\boldsymbol{\varsigma}=f$ and any $v\in H^1_0(\Omega)$,*

$$0 \le \mathcal{J}(v) - \mathcal{J}(u) \le \mathcal{J}(v) - \mathcal{J}^*(\boldsymbol{\varsigma}). \tag{5.18}$$

Proof. We apply the properties of the objects involved,

$$0 \overset{(2.4)}{\le} \mathcal{J}(v) - \mathcal{J}(u) \overset{(5.16)}{=} \mathcal{J}(v) - \mathcal{J}^*(\boldsymbol{\sigma}) \overset{(5.10)}{\le} \mathcal{J}(v) - \mathcal{J}^*(\boldsymbol{\varsigma}). \tag{5.19}$$

□

6 Equilibrated flux and its components

In this section, we detail an algorithm to construct a dual object $\boldsymbol{t}_\ell^{j,k}\in\boldsymbol{H}(\mathrm{div},\Omega)$ with $\nabla\cdot\boldsymbol{t}_\ell^{j,k}=f$ as required by the duality theory of §5. We consider patch-wise minimizations corresponding to local Neumann mixed finite element problems. This strategy has already been employed in many contexts cf. [9, 22, 14]. First we make the following assumption to simplify the analysis. The treatment of general f has been studied carefully in e.g. [22, 26].

Assumption 6.1 (No data oscillation)**.** *We suppose for simplicity that the source term is a piecewise polynomial, $f\in\mathcal{P}_p(\mathcal{T}_\ell)$.*

6.1 Equilibrated flux

We begin with some additional geometric information. For a given mesh level $\ell\ge 0$, let $\mathcal{V}_\ell$ be the set of mesh vertices partitioned to $\mathcal{V}_\ell=\mathcal{V}_\ell^{\mathrm{int}}\cup\mathcal{V}_\ell^{\mathrm{ext}}$ by interior and boundary vertices. Next let $\omega_{\boldsymbol{a}}$ be the subdomain corresponding to the set of elements of $\mathcal{T}_\ell$ for which $\boldsymbol{a}$ is a vertex, denoted by $\mathcal{T}_{\boldsymbol{a}}$. We also make use of the hat functions $\psi_{\boldsymbol{a}}\in\mathcal{P}_1(\mathcal{T}_\ell)\cap C^0(\overline{\Omega})$ associated with the vertex $\boldsymbol{a}\in\mathcal{V}_\ell$.

For a collection of simplices $\mathcal{T}$ and their corresponding domain ω, we introduce the broken Raviart–Thomas–Nédélec finite element space [10] of order $p\ge 0$,

$$\boldsymbol{RT}_p(\mathcal{T}) := \{\boldsymbol{v}_\ell\in[L^2(\omega)]^d : \boldsymbol{v}_\ell|_K\in[\mathcal{P}_p(K)]^d + \boldsymbol{x}\mathcal{P}_p(K),\ \forall K\in\mathcal{T}\}. \tag{6.1}$$

Next, to account for normal face continuity, we define the vertex patch space

$$\boldsymbol{V}_\ell^p(\omega_{\boldsymbol{a}}) := \boldsymbol{RT}_p(\mathcal{T}_{\boldsymbol{a}}) \cap \boldsymbol{H}_0(\mathrm{div}, \omega_{\boldsymbol{a}}), \tag{6.2}$$

where $\boldsymbol{H}_0(\mathrm{div}, \omega_{\boldsymbol{a}})$ is the subspace of $\boldsymbol{H}(\mathrm{div}, \omega_{\boldsymbol{a}})$ of functions with vanishing normal trace on $\partial\omega_{\boldsymbol{a}}$ when $\boldsymbol{a} \in \mathcal{V}_\ell^{\mathrm{int}}$ and on $\partial\omega_{\boldsymbol{a}} \setminus \{\psi_{\boldsymbol{a}} > 0\}$ when $\boldsymbol{a} \in \mathcal{V}_\ell^{\mathrm{ext}}$. For $v \in L^2(\Omega)$, define the L^2-projection $\Pi_{\ell,p} v \in \mathcal{P}_p(\mathcal{T}_\ell)$ by $(v - \Pi_{\ell,p} v, v_\ell) = 0$ for all $v_\ell \in \mathcal{P}_p(\mathcal{T}_\ell)$. Note that it acts elementwise. Finally, for $\mathcal{T} \subset \mathcal{T}_\ell$ and the corresponding domain $\omega \subseteq \Omega$, define the mean-free space by $\mathcal{P}_p^*(\mathcal{T}) := \{v \in \mathcal{P}_p(\mathcal{T}) : \int_\omega v \,\mathrm{d}\boldsymbol{x} = 0\}$.

Definition 6.2 (Total flux $\boldsymbol{t}_\ell^{j,k}$)**.** *Let $u_\ell^{j,k}$ be the solution to (3.5). For all vertices $\boldsymbol{a} \in \mathcal{V}_\ell$, define $\boldsymbol{t}_{\boldsymbol{a}}^{j,k} \in \boldsymbol{V}_\ell^p(\omega_{\boldsymbol{a}})$ and $q_{\boldsymbol{a}} \in \mathcal{P}_p^*(\mathcal{T}_{\boldsymbol{a}})$ as the solution to the patch-local mixed finite element problem*

$$(\boldsymbol{t}_{\boldsymbol{a}}^{j,k}, \boldsymbol{v}_\ell)_{\omega_{\boldsymbol{a}}} - (q_{\boldsymbol{a}}, \nabla\cdot\boldsymbol{v}_\ell)_{\omega_{\boldsymbol{a}}} = -(\psi_{\boldsymbol{a}} \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}), \boldsymbol{v}_\ell)_{\omega_{\boldsymbol{a}}}, \tag{6.3a}$$

$$(\nabla\cdot\boldsymbol{t}_{\boldsymbol{a}}^{j,k}, r_\ell)_{\omega_{\boldsymbol{a}}} = (f\psi_{\boldsymbol{a}} - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) \cdot \nabla\psi_{\boldsymbol{a}}, r_\ell)_{\omega_{\boldsymbol{a}}} \tag{6.3b}$$

for all pairs $(\boldsymbol{v}_\ell, r_\ell) \in \boldsymbol{V}_\ell^p(\omega_{\boldsymbol{a}}) \times \mathcal{P}_p^*(\mathcal{T}_{\boldsymbol{a}})$. After solving this local problem on each patch and extending $\boldsymbol{t}_{\boldsymbol{a}}^{j,k}$ by $\boldsymbol{0}$ outside of $\omega_{\boldsymbol{a}}$, assemble the global flux by

$$\boldsymbol{t}_\ell^{j,k} = \sum_{\boldsymbol{a}\in\mathcal{V}_\ell} \boldsymbol{t}_{\boldsymbol{a}}^{j,k}. \tag{6.3c}$$

The patch problem (6.3) is equivalent to solving the local minimization problem,

$$\boldsymbol{t}_{\boldsymbol{a}}^{j,k} := \operatorname*{arg\,min}_{\substack{\boldsymbol{v}_\ell \in \boldsymbol{V}_\ell^p(\omega_{\boldsymbol{a}}) \\ \nabla\cdot\boldsymbol{v}_\ell = \Pi_{\ell,p}(\psi_{\boldsymbol{a}} f - \nabla\psi_{\boldsymbol{a}} \cdot \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}))}} \|\psi_{\boldsymbol{a}} \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) + \boldsymbol{v}_\ell\|_{\omega_{\boldsymbol{a}}}. \tag{6.4}$$

We study the wall time cost of constructing the total flux (6.3) in Appendix B. The flux satisfies the divergence constraint.

Lemma 6.3 (Divergence of the equilibrated flux)**.** *Given Assumption 6.1, the flux $\boldsymbol{t}_\ell^{j,k}$ given by Definition 6.2 satisfies*

$$\nabla\cdot\boldsymbol{t}_\ell^{j,k} = f. \tag{6.5}$$

Proof. By construction,

$$\nabla\cdot\boldsymbol{t}_\ell^{j,k} = \sum_{\boldsymbol{a}\in\mathcal{V}_\ell} \nabla\cdot\boldsymbol{t}_{\boldsymbol{a}}^{j,k} = \Pi_{\ell,p}\left(\sum_{\boldsymbol{a}\in\mathcal{V}_\ell} (\psi_{\boldsymbol{a}} f - \nabla\psi_{\boldsymbol{a}} \cdot \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}))\right) = \Pi_{\ell,p} f = f,$$

see e.g. [9, 22]. □

We have the following stability result obtained by proceeding as in [41].

Lemma 6.4 (Stability of the equilibrated flux)**.** *For a fixed vertex $\boldsymbol{a} \in \mathcal{V}_\ell$, the solution to the patch problem (6.3) satisfies*

$$\|\boldsymbol{t}_{\boldsymbol{a}}^{j,k}\|_{\omega_{\boldsymbol{a}}} \lesssim \|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k})\|_{\omega_{\boldsymbol{a}}} + h_{\omega_{\boldsymbol{a}}} \|\psi_{\boldsymbol{a}} f\|_{\omega_{\boldsymbol{a}}}. \tag{6.6}$$

The hidden constant only depends on the space dimension d and the mesh shape regularity constant $\kappa_{\mathcal{T}}$ of (3.1).

Proof. We first recall the result used in [41, Lemma 4.1], i.e., for any $\boldsymbol{\tau}_{\boldsymbol{a}} \in \boldsymbol{RT}_p(\mathcal{T}_{\boldsymbol{a}})$ and $g_{\boldsymbol{a}} \in \mathcal{P}_p(\mathcal{T}_{\boldsymbol{a}})$,

$$\min_{\substack{\boldsymbol{v}_\ell \in \boldsymbol{V}_\ell^p(\omega_{\boldsymbol{a}}) \\ \nabla\cdot\boldsymbol{v}_\ell = g_{\boldsymbol{a}}}} \|\boldsymbol{\tau}_{\boldsymbol{a}} + \boldsymbol{v}_\ell\|_{\omega_{\boldsymbol{a}}} \le \sup_{\substack{v \in H_*^1(\omega_{\boldsymbol{a}}) \\ \|\nabla v\|_{\omega_{\boldsymbol{a}}} = 1}} \{(g_{\boldsymbol{a}}, v)_{\omega_{\boldsymbol{a}}} - (\boldsymbol{\tau}_{\boldsymbol{a}}, \nabla v)_{\omega_{\boldsymbol{a}}}\}, \tag{6.7}$$

where $H_*^1(\omega_{\boldsymbol{a}})$ is the subspace of functions in $H^1(\omega_{\boldsymbol{a}})$ that have mean value zero on the patch subdomain $\omega_{\boldsymbol{a}}$ if $\boldsymbol{a} \in \mathcal{V}_\ell^{\mathrm{int}}$ is an interior vertex, or that vanish on $\partial\omega_{\boldsymbol{a}} \cap \{\psi_{\boldsymbol{a}} > 0\}$ when $\boldsymbol{a} \in \mathcal{V}_\ell^{\mathrm{ext}}$ is a boundary vertex.

Next, set

$$g_{\boldsymbol{a}}^{j,k} := \Pi_{\ell,p}(\psi_{\boldsymbol{a}} f - \nabla\psi_{\boldsymbol{a}} \cdot \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k})), \quad \boldsymbol{\tau}_{\boldsymbol{a}}^{j,k} := \psi_{\boldsymbol{a}} \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}). \tag{6.8}$$

For technical reasons, we will need to introduce another auxiliary problem. First, we introduce $\boldsymbol{\Pi}^{\boldsymbol{RT}}_{\ell,p-1}$, the $[L^2]^d$-orthogonal projection from $[L^2(\Omega)]^d$ to $\boldsymbol{RT}_{p-1}(\mathcal{T}_\ell)$. Note that it acts elementwise. We consider the vector-valued data

$$\tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}} := \psi_{\boldsymbol{a}} \boldsymbol{\Pi}^{\boldsymbol{RT}}_{\ell,p-1}(\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)) \tag{6.9}$$

and the associated minimization problem

$$\tilde{\boldsymbol{t}}^{j,k}_{\boldsymbol{a}} := \min_{\substack{\boldsymbol{v}_\ell \in \boldsymbol{V}^p_\ell(\omega_{\boldsymbol{a}}) \\ \nabla\cdot \boldsymbol{v}_\ell = g^{j,k}_{\boldsymbol{a}}}} \|\tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}} + \boldsymbol{v}_\ell\|_{\omega_{\boldsymbol{a}}}. \tag{6.10}$$

We are now prepared to derive the bound (6.6). We start with

$$\begin{aligned}
\|\boldsymbol{t}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} &\le \|\boldsymbol{t}^{j,k}_{\boldsymbol{a}} + \boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \\
&\overset{(6.4)}{\le} \|\tilde{\boldsymbol{t}}^{j,k}_{\boldsymbol{a}} + \boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \\
&\le \|\tilde{\boldsymbol{t}}^{j,k}_{\boldsymbol{a}} + \tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}} - \tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \\
&\overset{(6.10)}{=} \min_{\substack{\boldsymbol{v}_\ell \in \boldsymbol{V}^p_\ell(\omega_{\boldsymbol{a}}) \\ \nabla\cdot \boldsymbol{v}_\ell = g^{j,k}_{\boldsymbol{a}}}} \|\tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}} + \boldsymbol{v}_\ell\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}} - \tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \\
&\overset{(6.7)}{\le} \underbrace{\sup_{\substack{v\in H^1_*(\omega_{\boldsymbol{a}}) \\ \|\nabla v\|_{\omega_{\boldsymbol{a}}}=1}} \{\underbrace{(g^{j,k}_{\boldsymbol{a}}, v)_{\omega_{\boldsymbol{a}}}}_{a} - \underbrace{(\tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}, \nabla v)_{\omega_{\boldsymbol{a}}}}_{b}\}}_{\mathfrak{T}_1} + \underbrace{\|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}} - \tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}}}_{\mathfrak{T}_2} + \underbrace{\|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}}}_{\mathfrak{T}_3}.
\end{aligned}$$

We now bound these three terms individually. For the third term,

$$\mathfrak{T}_3 = \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \overset{(6.8)}{=} \|\psi_{\boldsymbol{a}} \boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}} \le \|\psi_{\boldsymbol{a}}\|_{L^\infty(\omega_{\boldsymbol{a}})} \|\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}} = \|\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}}. \tag{6.11}$$

For the second term,

$$\begin{aligned}
\mathfrak{T}_2 &\le \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\| + \|\tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \\
&\overset{(6.9)}{=} \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\| + \|\psi_{\boldsymbol{a}} \boldsymbol{\Pi}^{\boldsymbol{RT}}_{\ell,p-1}(\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)))\|_{\omega_{\boldsymbol{a}}} \\
&\le \|\boldsymbol{\tau}^{j,k}_{\boldsymbol{a}}\| + \|\boldsymbol{\Pi}^{\boldsymbol{RT}}_{\ell,p-1}(\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell))\|_{\omega_{\boldsymbol{a}}} \\
&\overset{(6.11)}{\le} \|\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}}.
\end{aligned}$$

Finally, for the first term, fix $v \in H^1_*(\omega_{\boldsymbol{a}})$ with $\|\nabla v\|_{H^1_*(\omega_{\boldsymbol{a}})} = 1$. For term b,

$$(\tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}, \nabla v)_{\omega_{\boldsymbol{a}}} \overset{\text{C.S.}}{\le} \|\tilde{\boldsymbol{\tau}}^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}},$$

whence the preceding arguments can be applied again. For term a,

$$\begin{aligned}
(g^{j,k}_{\boldsymbol{a}}, v)_{\omega_{\boldsymbol{a}}} &\overset{\text{C.S.}}{\le} \|g^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \|v\|_{\omega_{\boldsymbol{a}}} \\
&\overset{\text{Poincaré}}{\lesssim} h_{\omega_{\boldsymbol{a}}} \|g^{j,k}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \\
&\overset{(6.8)}{=} h_{\omega_{\boldsymbol{a}}} \|\Pi_{\ell,p}(\psi_{\boldsymbol{a}} f - \nabla\psi_{\boldsymbol{a}} \cdot \boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell))\|_{\omega_{\boldsymbol{a}}} \\
&\le h_{\omega_{\boldsymbol{a}}} \|\psi_{\boldsymbol{a}} f - \nabla\psi_{\boldsymbol{a}} \cdot \boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}} \\
&\le h_{\omega_{\boldsymbol{a}}} \left(\|\psi_{\boldsymbol{a}} f\|_{\omega_{\boldsymbol{a}}} + \|\nabla\psi_{\boldsymbol{a}} \cdot \boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}}\right) \\
&\le h_{\omega_{\boldsymbol{a}}} \left(\|\psi_{\boldsymbol{a}} f\|_{\omega_{\boldsymbol{a}}} + \|\nabla\psi_{\boldsymbol{a}}\|_{L^\infty(\omega_{\boldsymbol{a}})} \|\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}}\right) \\
&\lesssim h_{\omega_{\boldsymbol{a}}} \left(\|\psi_{\boldsymbol{a}} f\|_{\omega_{\boldsymbol{a}}} + h^{-1}_{\omega_{\boldsymbol{a}}} \|\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}}\right) \\
&= h_{\omega_{\boldsymbol{a}}} \|\psi_{\boldsymbol{a}} f\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{A}^{k-1}_{\epsilon^j}(\nabla u^{j,k}_\ell)\|_{\omega_{\boldsymbol{a}}}.
\end{aligned}$$

Combining these terms concludes the proof. □

6.2 Component fluxes

In addition to the flux in Definition 6.2, we introduce three more fluxes. The idea, as in [22], is that $\boldsymbol{t}_\ell^{j,k}$ will contain information about the total error, and the additional fluxes will isolate components of the error. This definition is more precisely intended to distinguish the errors coming from regularization, linearization, and discretization.

Definition 6.5 (Decomposition of $\boldsymbol{t}_\ell^{j,k}$ into components)**.** *Let $u_\ell^{j,k}$ be the solution to (3.5), for $\ell \geq 0, j \geq 0$, and $k \geq 1$. Let $\boldsymbol{A}, \boldsymbol{A}_{\epsilon^j}$, and $\boldsymbol{A}_{\epsilon^j}^{k-1}$ be given by (2.5), (2.14), and (3.6), respectively. Then define*

$$\boldsymbol{r}_\ell^{j,k} := \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k}) - \boldsymbol{A}(\nabla u_\ell^{j,k}) \qquad \text{[regularization error flux]}, \tag{6.12a}$$

$$\boldsymbol{l}_\ell^{j,k} := \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k}) \qquad \text{[linearization error flux]}, \tag{6.12b}$$

$$\boldsymbol{d}_\ell^{j,k} := \boldsymbol{t}_\ell^{j,k} - \boldsymbol{r}_\ell^{j,k} - \boldsymbol{l}_\ell^{j,k} = \boldsymbol{t}_\ell^{j,k} + \boldsymbol{A}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) \qquad \text{[discretization flux]}. \tag{6.12c}$$

Based on our Assumption 2.3 on ϕ'_ϵ, we have the following result.

Lemma 6.6 (H_0^1-convergence of the regularized approximation)**.** *Consider a discrete version of (2.8), i.e., find $u_\ell \in V_\ell^p$ such that*

$$(\boldsymbol{A}(\nabla u_\ell), \nabla v_\ell) = (f, v_\ell) \quad \forall v \in V_\ell^p. \tag{6.13}$$

Then the solution to the regularized discrete problem (3.4) satisfies

$$\lim_{j\to\infty} \|\nabla(u_\ell - u_\ell^j)\| = 0. \tag{6.14}$$

Proof. Using the strong monotonicity of $\boldsymbol{A}_{\epsilon^j}$,

$$\begin{aligned}
\underline{\alpha}\|\nabla(u_\ell - u_\ell^j)\|^2 &\overset{(2.15a)}{\leq} \big(\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j), \nabla(u_\ell - u_\ell^j)\big) \\
&= \big(\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell) + \boldsymbol{A}(\nabla u_\ell) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j), \nabla(u_\ell - u_\ell^j)\big) \\
&\overset{(3.4),(6.13)}{=} \big(\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell), \nabla(u_\ell - u_\ell^j)\big) \\
&\overset{\text{C.S.}}{\leq} \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell)\|\|\nabla(u_\ell - u_\ell^j)\| \\
&\overset{(2.5),(2.14)}{=} \|\phi'_{\epsilon^j}(|\nabla u_\ell|) - \phi'(|\nabla u_\ell|)\|\|\nabla(u_\ell - u_\ell^j)\|.
\end{aligned}$$

Thus, $\|\nabla(u_\ell - u_\ell^j)\| \leq \underline{\alpha}^{-1}\|\phi'_{\epsilon^j}(|\nabla u_\ell|) - \phi'(|\nabla u_\ell|)\| \overset{j\to\infty}{\longrightarrow} 0$ by (2.16). □

We have a similar result for the linearized problem.

Lemma 6.7 (H_0^1-convergence of the linearized approximation)**.** *Let the regularization step $j \geq 0$ and mesh level $\ell \geq 0$ be fixed. Then*

$$\lim_{k\to\infty} \|\nabla(u_\ell^j - u_\ell^{j,k})\| = 0. \tag{6.15}$$

Proof. We use the coercivity of the linearization matrix of Assumption 3.1,

$$\begin{aligned}
\underline{\lambda}\|\nabla(u_\ell^{j,k} - u_\ell^j)\|^2 &\overset{(3.8b)}{\leq} (\mathbb{A}_{\epsilon^j}^{k-1}\nabla(u_\ell^{j,k} - u_\ell^j), \nabla(u_\ell^{j,k} - u_\ell^j)) \\
&\overset{(3.6)}{=} (\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j), \nabla(u_\ell^{j,k} - u_\ell^j)) \\
&= (\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) + \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j), \nabla(u_\ell^{j,k} - u_\ell^j)) \\
&\overset{(3.4),(3.5)}{=} (\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j), \nabla(u_\ell^{j,k} - u_\ell^j)) \\
&\overset{\text{C.S.}}{\leq} \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j)\|\|\nabla(u_\ell^{j,k} - u_\ell^j)\|.
\end{aligned}$$

Thus, $\|\nabla(u_\ell^{j,k} - u_\ell^j)\| \leq \underline{\lambda}^{-1}\|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j)\| \overset{k\to\infty}{\longrightarrow} 0$ by (3.7). □

We are now prepared to prove the following.

Lemma 6.8 (Convergence of the regularization error flux)**.** *For a fixed mesh index $\ell \geq 0$, the regularization error flux $\boldsymbol{r}_\ell^{j,k}$ given in (6.12a) satisfies*

$$\lim_{j,k\to\infty} \|\boldsymbol{r}_\ell^{j,k}\| = 0. \tag{6.16}$$

Proof. From the definition of the regularization component flux,

$$\begin{aligned}
\lim_{j,k\to\infty} \|\boldsymbol{r}_\ell^{j,k}\| &\overset{(6.12a)}{=} \lim_{j,k\to\infty} \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k}) - \boldsymbol{A}(\nabla u_\ell^{j,k})\| \\
&\leq \lim_{j,k\to\infty} \Big(\|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell)\| + \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell)\| + \|\boldsymbol{A}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell^{j,k})\| \Big) \\
&\overset{(2.15b),(2.7)}{\leq} \lim_{j,k\to\infty} \Big(\|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell)\| + (L+\overline{L})\|\nabla(u_\ell - u_\ell^{j,k})\| \Big) \\
&\leq \lim_{j,k\to\infty} \Big(\|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell)\| + (L+\overline{L})(\|\nabla(u_\ell - u_\ell^j)\| + \|\nabla(u_\ell^j - u_\ell^{j,k})\|) \Big) \\
&\overset{(6.15),(6.14)}{=} \lim_{j\to\infty} \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell)\| \\
&\overset{(2.5),(2.14)}{=} \lim_{j\to\infty} \|\phi'_{\epsilon^j}(|\nabla u_\ell|) - \phi'(|\nabla u_\ell|)\| \overset{(2.16)}{=} 0.
\end{aligned}$$

□

We now turn our attention to the convergence of the linearization error flux component.

Lemma 6.9 (Convergence of the linearization error flux)**.** *Let the regularization step $j \geq 0$ and mesh level $\ell \geq 0$ be fixed. The linearization error flux $\boldsymbol{l}_\ell^{j,k}$ given by (6.12b) satisfies*

$$\lim_{k\to\infty} \|\boldsymbol{l}_\ell^{j,k}\| = 0. \tag{6.17}$$

Proof. From the definition of the linearization error flux,

$$\begin{aligned}
\lim_{k\to\infty} \|\boldsymbol{l}_\ell^{j,k}\| &\overset{(6.12b)}{=} \lim_{k\to\infty} \|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k})\| \\
&\leq \lim_{k\to\infty} \Big(\|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j)\| + \|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j)\| \\
&\quad + \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k})\| \Big) \\
&\overset{(3.7)}{=} \lim_{k\to\infty} \Big(\|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j)\| + \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k})\| \Big) \\
&\overset{(2.15b)}{\leq} \lim_{k\to\infty} \Big(\|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j)\| + \overline{L}\|\nabla(u_\ell^j - u_\ell^{j,k})\| \Big) \\
&\overset{(3.6)}{=} \lim_{k\to\infty} \Big(\|\mathbb{A}_{\epsilon^j}^{k-1}\nabla(u_\ell^{j,k} - u_\ell^j)\| + \overline{L}\|\nabla(u_\ell^j - u_\ell^{j,k})\| \Big) \\
&\overset{(3.8a)}{\leq} \lim_{k\to\infty} (\overline{\lambda} + \overline{L})\|\nabla(u_\ell^j - u_\ell^{j,k})\| \overset{(6.15)}{=} 0.
\end{aligned}$$

□

Combining these results with the stability of the equilibrated flux in Lemma 6.4 results in the following.

Lemma 6.10 (Boundedness of the total equilibrated flux)**.** *The equilibrated flux $\boldsymbol{t}_\ell^{j,k}$ of (6.3c) is bounded in the indices j and k, i.e.,*

$$\lim_{j,k\to\infty} \|\boldsymbol{t}_\ell^{j,k}\| = C_{f,\ell} < \infty. \tag{6.18}$$

Proof. We first observe that

$$\|\boldsymbol{t}_\ell^{j,k}\|^2 \leq (d+1) \sum_{\boldsymbol{a}\in\mathcal{V}_\ell} \|\boldsymbol{t}_{\boldsymbol{a}}^{j,k}\|_{\omega_{\boldsymbol{a}}}^2. \tag{6.19}$$

Now, for a fixed $\boldsymbol{a}, j, k$, letting $C_{f,\boldsymbol{a}} := h_{\omega_{\boldsymbol{a}}}\|\psi_{\boldsymbol{a}} f\|_{\omega_{\boldsymbol{a}}}$,

$$\begin{aligned}
\|\boldsymbol{t}_{\boldsymbol{a}}^{j,k}\|_{\omega_{\boldsymbol{a}}} &\overset{(6.6)}{\lesssim} \|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k})\|_{\omega_{\boldsymbol{a}}} + C_{f,\boldsymbol{a}} \\
&\leq \|\boldsymbol{A}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j)\|_{\omega_{\boldsymbol{a}}} + C_{f,\boldsymbol{a}} \\
&\leq \|\boldsymbol{A}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}} + \underbrace{\|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}}}_{\mathfrak{T}_1} + \underbrace{\|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j)\|_{\omega_{\boldsymbol{a}}}}_{\mathfrak{T}_2} \\
&+ \underbrace{\|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^j)\|_{\omega_{\boldsymbol{a}}}}_{\mathfrak{T}_3} + C_{f,\boldsymbol{a}}
\end{aligned}$$

First, we have

$$\begin{aligned}
\lim_{j\to\infty} \mathfrak{T}_1 &\leq \lim_{j\to\infty} \left(\|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^j) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}}\right) \\
&\overset{(2.15b)}{\leq} \lim_{j\to\infty} \left(\overline{L}\|\nabla(u_\ell^j - u_\ell)\|_{\omega_{\boldsymbol{a}}} + \|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell) - \boldsymbol{A}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}}\right) \\
&\overset{(2.5)}{=} \lim_{j\to\infty} \left(\overline{L}\|\nabla(u_\ell^j - u_\ell)\|_{\omega_{\boldsymbol{a}}} + \|\phi'_{\epsilon^j}(|\nabla u_\ell|) - \phi'(|\nabla u_\ell|)\|_{\omega_{\boldsymbol{a}}}\right) \overset{(2.16),(6.14)}{=} 0.
\end{aligned}$$

Next,

$$\lim_{j,k\to\infty} \mathfrak{T}_2 \overset{(3.7)}{=} 0.$$

Finally, we have

$$\lim_{j,k\to\infty} \mathfrak{T}_3 \overset{(3.6)}{=} \lim_{j,k\to\infty} \|\mathbb{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k} - u_\ell^j)\|_{\omega_{\boldsymbol{a}}} \overset{(3.8a)}{\leq} \lim_{j,k\to\infty} \overline{\lambda}\|\nabla(u_\ell^{j,k} - u_\ell^j)\|_{\omega_{\boldsymbol{a}}} \overset{(6.15)}{=} 0.$$

Combining these results with (6.19), we conclude

$$\lim_{j,k\to\infty} \|\boldsymbol{t}_\ell^{j,k}\|^2 \leq (d+1) \sum_{\boldsymbol{a}\in\mathcal{V}_\ell} \left(\|\boldsymbol{A}(\nabla u_\ell)\|_{\omega_{\boldsymbol{a}}} + C_{f,\boldsymbol{a}}\right)^2 =: (C_{f,\ell})^2.$$

□

Lemma 6.11 (Convergence of the discretization flux)**.** *For a fixed mesh index $\ell \geq 0$, the discretization flux $\boldsymbol{d}_\ell^{j,k}$ given in (6.12c) satisfies*

$$\lim_{j,k\to\infty} \|\boldsymbol{d}_\ell^{j,k} - \boldsymbol{t}_\ell^{j,k}\| = 0. \tag{6.20}$$

Consequently, we have that the discretization flux satisfies

$$\lim_{j,k\to\infty} \|\boldsymbol{d}_\ell^{j,k}\| = C_{f,\ell} < \infty, \tag{6.21}$$

where $C_{f,\ell}$ is given in (6.18).

Proof. Equation (6.20) is a direct consequence of Lemmas 6.8 and 6.9 taken with the definition (6.12c). For (6.21),

$$\lim_{j,k\to\infty} \|\boldsymbol{d}_\ell^{j,k}\| \leq \lim_{j,k\to\infty} \left(\|\boldsymbol{d}_\ell^{j,k} - \boldsymbol{t}_\ell^{j,k}\| + \|\boldsymbol{t}_\ell^{j,k}\|\right) \overset{(6.20),(6.18)}{\leq} C_{f,\ell}. \tag{6.22}$$

□

We use this fact to justify calling $\boldsymbol{d}_\ell^{j,k}$ the discretization flux. Indeed, the total flux $\boldsymbol{t}_\ell^{j,k}$ is captured by the discretization component $\boldsymbol{d}_\ell^{j,k}$ upon convergence of both the regularization and linearization indices j and k respectively. We now present the results separately for the three ways of measuring the error introduced in §4.

7 A posteriori error estimates distinguishing the error components

In this section we present error estimates that provide an upper bound and decompose the total error in a given numerical solution. The components estimate the error due to regularization, discretization, and linearization.

7.1 Energy difference

The results of §5 lead us directly to the following upper bound on the error in the difference of energies.

Proposition 7.1 (Upper bound on the energy difference)**.** *Let $u \in H^1_0(\Omega)$ be the solution to (2.4), and let $\mathcal{J}$ and $\mathcal{J}^*$ be given by (2.1) and (5.11), respectively. For a mesh index $\ell \geq 0$, a regularization step $j \geq 0$, and linearization step $k \geq 1$, let $u_\ell^{j,k} \in V_\ell^p$ be the solution to (3.5) and $\boldsymbol{t}_\ell^{j,k}$ be given by Definition 6.2. Then there holds*

$$0 \leq \underbrace{\mathcal{J}(u_\ell^{j,k}) - \mathcal{J}(u)}_{\text{total error } (e_{\mathrm{tot}}^{\ell,j,k})^2} \leq \underbrace{\mathcal{J}(u_\ell^{j,k}) - \mathcal{J}^*(\boldsymbol{t}_\ell^{j,k})}_{\text{total est. } (\eta_{\mathrm{tot}}^{\ell,j,k})^2}. \tag{7.1}$$

Proof. We apply Corollary 5.6 with $v = u_\ell^{j,k}$ and $\boldsymbol{\varsigma} = \boldsymbol{t}_\ell^{j,k}$. □

Remark 7.2 (Equivalent definition of the total estimator)**.** *The definition of the total estimator (7.1) has an equivalent form that is more amenable to local adaptive mesh refinement. This has been already discussed in, e.g., [6, Proposition 4.9]. Indeed, we have that*

$$\begin{aligned}
(\eta_{\mathrm{tot}}^{\ell,j,k})^2 &\overset{(7.1)}{=} \mathcal{J}(u_\ell^{j,k}) - \mathcal{J}^*(\boldsymbol{t}_\ell^{j,k}) \\
&\overset{(2.1),(5.11)}{=} \int_\Omega \phi(|\nabla u_\ell^{j,k}|) + \phi^*(|\boldsymbol{t}_\ell^{j,k}|) - f u_\ell^{j,k} \,\mathrm{d}\boldsymbol{x} \\
&\overset{(6.5)}{=} \int_\Omega \phi(|\nabla u_\ell^{j,k}|) + \phi^*(|\boldsymbol{t}_\ell^{j,k}|) - (\nabla\cdot\boldsymbol{t}_\ell^{j,k}) u_\ell^{j,k} \,\mathrm{d}\boldsymbol{x} \\
&\overset{\text{I.B.P.}}{=} \int_\Omega \phi(|\nabla u_\ell^{j,k}|) + \phi^*(|\boldsymbol{t}_\ell^{j,k}|) + \nabla u_\ell^{j,k} \cdot \boldsymbol{t}_\ell^{j,k} \,\mathrm{d}\boldsymbol{x} \\
&= \sum_{K\in\mathcal{T}_\ell} \int_K \underbrace{\phi(|\nabla u_\ell^{j,k}|) + \phi^*(|\boldsymbol{t}_\ell^{j,k}|) + \nabla u_\ell^{j,k} \cdot \boldsymbol{t}_\ell^{j,k}}_{\eta_{\mathrm{tot},K}^{\ell,j,k} \geq 0 \text{ by } (5.7)} \,\mathrm{d}\boldsymbol{x}.
\end{aligned} \tag{7.2}$$

The advantage of this definition is that the last integrand is non-negative by the generalized Young's inequality for convex functions (5.7). Indeed, the reasoning goes as follows:

$$\phi(|\nabla u_\ell^{j,k}|) + \phi^*(|\boldsymbol{t}_\ell^{j,k}|) \overset{(5.7)}{\geq} |\nabla u_\ell^{j,k}||\boldsymbol{t}_\ell^{j,k}| \overset{C.S.}{\geq} -\nabla u_\ell^{j,k} \cdot \boldsymbol{t}_\ell^{j,k}. \tag{7.3}$$

.

We now present a decomposition of total estimator, employing Definition 6.2 and (7.2).

Theorem 7.3 (Decomposition of the energy difference upper bound)**.** *Let the assumptions of Proposition 7.1 hold. Let in addition $\boldsymbol{d}_\ell^{j,k}, \boldsymbol{r}_\ell^{j,k}$, and $\boldsymbol{l}_\ell^{j,k}$ be given by Definition 6.5. Then the total estimator in (7.1) can be further bounded from above as*

$$\begin{aligned}
(\eta_{\mathrm{tot}}^{\ell,j,k})^2 \leq & \underbrace{\left| \int_\Omega \phi(|\nabla u_\ell^{j,k}|) + \phi^*(|\boldsymbol{d}_\ell^{j,k}|) + \nabla u_\ell^{j,k} \cdot \boldsymbol{t}_\ell^{j,k} \,\mathrm{d}\boldsymbol{x} \right|}_{\text{discretization est. } (\eta_{\mathrm{dis}}^{\ell,j,k})^2} \\
& + \underbrace{\left| \int_\Omega \phi^*(|\boldsymbol{d}_\ell^{j,k} + \boldsymbol{r}_\ell^{j,k}|) - \phi^*(|\boldsymbol{d}_\ell^{j,k}|) \,\mathrm{d}\boldsymbol{x} \right|}_{\text{regularization est. } (\eta_{\mathrm{reg}}^{\ell,j,k})^2} \\
& + \underbrace{\left| \int_\Omega \phi^*(|\boldsymbol{t}_\ell^{j,k}|) - \phi^*(|\boldsymbol{d}_\ell^{j,k} + \boldsymbol{r}_\ell^{j,k}|) \,\mathrm{d}\boldsymbol{x} \right|}_{\text{linearization est. } (\eta_{\mathrm{lin}}^{\ell,j,k})^2}.
\end{aligned} \tag{7.4}$$

Proof. The proof follows by adding and subtracting $\phi^*(|\boldsymbol{d}_\ell^{j,k} + \boldsymbol{r}_\ell^{j,k}|)$ and $\phi^*(|\boldsymbol{d}_\ell^{j,k}|)$ to the integrand of (7.2) and using the triangle inequality. □

We now show our definition of the regularization component behaves in the way that we would expect.

Lemma 7.4 (Convergence of the regularization error estimator)**.** *The regularization component estimator* $\eta_{\mathrm{reg}}^{\ell,j,k}$ *of* (7.4) *tends to 0 as* $j,k \to \infty$.

Proof. From the definition of the regularization component estimator,

$$\begin{aligned}
\lim_{j,k\to\infty} (\eta_{\mathrm{reg}}^{\ell,j,k})^2 &\overset{(7.4)}{=} \lim_{j,k\to\infty} \left| \int_\Omega \phi^*(|\boldsymbol{d}_\ell^{j,k}|) - \phi^*(|\boldsymbol{d}_\ell^{j,k} + \boldsymbol{r}_\ell^{j,k}|) \,\mathrm{d}\boldsymbol{x} \right| \\
&= \lim_{j,k\to\infty} \left| \int_\Omega \int_{|\boldsymbol{d}_\ell^{j,k}+\boldsymbol{r}_\ell^{j,k}|}^{|\boldsymbol{d}_\ell^{j,k}|} (\phi^*)'(s) \,\mathrm{d}s \,\mathrm{d}\boldsymbol{x} \right| \\
&\overset{(5.6d),(5.9)}{\leq} \alpha^{-1} \lim_{j,k\to\infty} \left| \int_\Omega \int_{|\boldsymbol{d}_\ell^{j,k}+\boldsymbol{r}_\ell^{j,k}|}^{|\boldsymbol{d}_\ell^{j,k}|} s \,\mathrm{d}s \,\mathrm{d}\boldsymbol{x} \right| \\
&= (2\alpha)^{-1} \lim_{j,k\to\infty} \left| \int_\Omega |\boldsymbol{d}_\ell^{j,k}|^2 - |\boldsymbol{d}_\ell^{j,k} + \boldsymbol{r}_\ell^{j,k}|^2 \,\mathrm{d}\boldsymbol{x} \right| \\
&= (2\alpha)^{-1} \lim_{j,k\to\infty} \left| \|\boldsymbol{d}_\ell^{j,k}\|^2 - \|\boldsymbol{d}_\ell^{j,k} + \boldsymbol{r}_\ell^{j,k}\|^2 \right| \overset{(6.16)}{=} 0
\end{aligned}$$

where we have also used that $\boldsymbol{d}_\ell^{j,k}$ is uniformly bounded in j,k by Lemma 6.11 to interchange the limit and the integral in the last equality. □

The same argument holds for the linearization error component estimator.

Lemma 7.5 (Convergence of the linearization error estimator)**.** *The linearization component estimator* $\eta_{\mathrm{lin}}^{\ell,j,k}$ *of* (7.4) *tends to 0 as* $k \to \infty$.

Finally, we have the following result pertaining to the discretization error component estimator.

Lemma 7.6 (Convergence of the discretization error estimator)**.** *The discretization component estimator satisfies*

$$\lim_{j,k\to\infty} (\eta_{\mathrm{dis}}^{\ell,j,k})^2 = \lim_{j,k\to\infty} (\eta_{\mathrm{tot}}^{\ell,j,k})^2. \tag{7.5}$$

Proof. In addition to (7.4), there holds

$$(\eta_{\mathrm{dis}}^{\ell,j,k})^2 \leq (\eta_{\mathrm{tot}}^{\ell,j,k})^2 + (\eta_{\mathrm{reg}}^{\ell,j,k})^2 + (\eta_{\mathrm{lin}}^{\ell,j,k})^2.$$

Thus, Lemma 7.4 and Lemma 7.5 finish the proof. □

7.2 Dual norm of the residual

Considering the dual norm of the residual as an error estimator leads to a different set of estimators. These have been studied previously in a variety of contexts [22, 16, 19]. First, we consider the total estimator, which provides an upper bound on the quantity (4.6).

Lemma 7.7 (Upper bound on the dual norm of the residual)**.** *Let the Assumption 6.1 hold. Let* $\mathcal{R}$ *be defined by* (4.4) *and let* $u_\ell^{j,k} \in V_\ell^p$ *be the solution of* (3.5)*. Then there holds*

$$0 \leq \tfrac{1}{2} \underbrace{|||\mathcal{R}(u_\ell^{j,k})|||_{-1}^2}_{\text{total error } (\tilde{e}_{\mathrm{tot}}^{\ell,j,k})^2} \leq \tfrac{1}{2} \underbrace{\alpha^{-1} \|\boldsymbol{A}(\nabla u_\ell^{j,k}) + \boldsymbol{t}_\ell^{j,k}\|^2}_{\text{total est. } (\tilde{\eta}_{\mathrm{tot}}^{\ell,j,k})^2}. \tag{7.6}$$

Proof. Let us first fix a function $\varphi \in H_0^1(\Omega)$ with $|||\varphi||| = \|\alpha^{1/2}\nabla\varphi\| = 1$. Then,

$$\begin{aligned}
\langle \mathcal{R}(u_\ell^{j,k}), \nabla\varphi \rangle &\overset{(4.4)}{=} (f,\varphi) - (\boldsymbol{A}(\nabla u_\ell^{j,k}), \nabla\varphi) \overset{(6.5)}{=} -(\boldsymbol{A}(\nabla u_\ell^{j,k}) + \boldsymbol{t}_\ell^{j,k}, \nabla\varphi) \\
&= -\alpha^{-1/2}(\boldsymbol{A}(\nabla u_\ell^{j,k}) + \boldsymbol{t}_\ell^{j,k}, \alpha^{1/2}\nabla\varphi) \leq \alpha^{-1/2} \|\boldsymbol{A}(\nabla u_\ell^{j,k}) + \boldsymbol{t}_\ell^{j,k}\| \|\alpha^{1/2}\nabla\varphi\| \\
&= \tilde{\eta}_{\mathrm{tot}}^{\ell,j,k}.
\end{aligned} \tag{7.7}$$

Since φ was arbitrary, (7.6) follows the definition (4.5). □

Corollary 7.8 (Decomposition of the upper bound)**.** *Let assumptions of Lemma 7.7 hold. Then*

$$\begin{aligned}\tilde{\eta}_{\mathrm{tot}}^{\ell,j,k} \leq & \underbrace{\alpha^{-1/2}\|\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) + \boldsymbol{t}_\ell^{j,k}\|}_{\text{discretization est. } \tilde{\eta}_{\mathrm{dis}}^{\ell,j,k}} + \underbrace{\alpha^{-1/2}\|\boldsymbol{A}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k})\|}_{\text{regularization est. } \tilde{\eta}_{\mathrm{reg}}^{\ell,j,k}} \\ & + \underbrace{\alpha^{-1/2}\|\boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k})\|}_{\text{linearization est. } \tilde{\eta}_{\mathrm{lin}}^{\ell,j,k}}.\end{aligned} \tag{7.8}$$

Proof. From the definition (7.6) of $\tilde{\eta}_{\mathrm{tot}}^{\ell,j,k}$,

$$\begin{aligned}\tilde{\eta}_{\mathrm{tot}}^{\ell,j,k} &= \alpha^{-1/2}\|\boldsymbol{t}_\ell^{j,k} + \boldsymbol{A}(\nabla u_\ell^{j,k})\| \\ &= \alpha^{-1/2}\|\boldsymbol{t}_\ell^{j,k} + \big(\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k}) + \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k}) - \boldsymbol{A}_{\epsilon^j}(\nabla u_\ell^{j,k})\big) + \boldsymbol{A}(\nabla u_\ell^{j,k})\| \\ &\leq \tilde{\eta}_{\mathrm{dis}}^{\ell,j,k} + \tilde{\eta}_{\mathrm{reg}}^{\ell,j,k} + \tilde{\eta}_{\mathrm{lin}}^{\ell,j,k}.\end{aligned}$$

□

Remark 7.9 (Decomposition of the residual)**.** *We can also split the residual $\mathcal{R}(v)$ introduced in (4.4) as*

$$\mathcal{R}(v) = \mathcal{R}_{\mathrm{dis}}^{\ell,j,k}(v) + \mathcal{R}_{\mathrm{reg}}^{\ell,j,k}(v) + \mathcal{R}_{\mathrm{lin}}^{\ell,j,k}(v), \tag{7.9}$$

where

$$\langle \mathcal{R}_{\mathrm{dis}}^{\ell,j,k}(v), w\rangle := (f, w) - (\boldsymbol{A}_{\epsilon^j}^{k-1}\nabla v, \nabla w), \tag{7.10a}$$

$$\langle \mathcal{R}_{\mathrm{reg}}^{\ell,j,k}(v), w\rangle := (\boldsymbol{A}_{\epsilon^j}(\nabla v) - \boldsymbol{A}(\nabla v), \nabla w), \tag{7.10b}$$

$$\langle \mathcal{R}_{\mathrm{lin}}^{\ell,j,k}(v), w\rangle := (\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla v) - \boldsymbol{A}_{\epsilon^j}(\nabla v), \nabla w). \tag{7.10c}$$

Then there holds

$$|||\mathcal{R}(u_\ell^{j,k})|||_{-1} \leq |||\mathcal{R}_{\mathrm{dis}}^{\ell,j,k}(u_\ell^{j,k})|||_{-1} + |||\mathcal{R}_{\mathrm{reg}}^{\ell,j,k}(u_\ell^{j,k})|||_{-1} + |||\mathcal{R}_{\mathrm{lin}}^{\ell,j,k}(u_\ell^{j,k})|||_{-1}. \tag{7.11}$$

7.3 Energy norm

In light of the relation between the dual norm of the residual and the energy norm presented in Proposition 4.1, we observe that the results of the previous section imply,

Lemma 7.10 (Upper bound on the energy error)**.** *Let the assumptions of Lemma 7.7 hold. Then,*

$$|||u_\ell^{j,k} - u||| \leq |||\mathcal{R}(u_\ell^{j,k})|||_{-1} \leq \tilde{\eta}_{\mathrm{tot}}^{\ell,j,k} \leq \tilde{\eta}_{\mathrm{dis}}^{\ell,j,k} + \tilde{\eta}_{\mathrm{reg}}^{\ell,j,k} + \tilde{\eta}_{\mathrm{lin}}^{\ell,j,k}. \tag{7.12}$$

8 Efficiency of the estimators

Up to this point, we have only considered the property that the estimators we construct are global upper bounds of the errors that we have defined and identified the error components. However, it has already been demonstrated that these estimators provide lower bounds to the error as well, even local lower bounds.

8.1 Dual norm of the residual

The setting of [22, 8] covers the present one. In particular, the flux of Definition 6.2, which in turn defines $\tilde{\eta}_{\mathrm{tot}}^{\ell,j,k}$ as in (7.6), provides a local lower bound under certain conditions on stopping criteria of the form (9.1a) and (9.1b) below. Roughly speaking, the efficiency is achieved when discretization component estimator is larger than the other components. This efficiency is robust with respect to the Lipschitz/montonicity parameters L/α, i.e.,

$$\tilde{\eta}_{\mathrm{tot}}^{\ell,\bar{j},\bar{k}} \lesssim |||\mathcal{R}(u_\ell^{\bar{j},\bar{k}})|||_{-1} + \text{oscillation and quadrature terms},$$

where the hidden constant has no dependence on L and α and the oscillation and quadrature terms are discussed in §8.3 below.

8.2 Energy norm

In the case of the energy norm, we can also obtain a (local) lower bound for the estimator $\tilde{\eta}_{\text{tot}}^{\ell,j,k}$ of (7.12) by using the lower bound of (4.10):

$$|||\mathcal{R}(u_\ell^{\bar{j},\bar{k}})|||_{-1} \leq \frac{L}{\alpha}|||u_\ell^{\bar{j},\bar{k}} - u|||. \tag{8.1}$$

Unfortunately, this makes appear the "strength of the nonlinearity" factor L/α. We will show numerically in §10.1.1, that this bound actually appears to be sharp. Thus, if L/α is large, the a posteriori estimate is pessimistic for the error measured in the energy norm.

8.3 Energy difference

In [26, Theorems 3.4 and 4.1], we study the efficiency of $\eta_{\text{tot}}^{\ell,j,k}$ as well as of a related estimator that incorporates the error in the difference of energies for the linear minimization problem corresponding to the linearization step (3.5). In particular, results of the form

$$\mathcal{J}(u_\ell^{j,k}) - \mathcal{J}(u) \leq (\eta_{\text{tot}}^{\ell,j,k})^2 \lesssim (C_\ell^k)^2(\mathcal{J}(u_\ell^{j,k}) - \mathcal{J}(u)) + (\text{oscillation and quadrature terms})^2 \tag{8.2}$$

are obtained where a typical data oscillation term is of the form $\left(\sum_{K\in\mathcal{T}_\ell}\left(\frac{h_K}{\pi}\|(I-\Pi_{\ell,p})f\|_K^2\right)\right)^{1/2}$ and a typical quadrature term is of the form $\|(\boldsymbol{I}-\boldsymbol{\Pi}_\ell^{\boldsymbol{RT}})\boldsymbol{A}_{\epsilon^j}^{k-1}(\nabla u_\ell^{j,k})\|$, where the projection operators $\Pi_{\ell,p}$ and $\boldsymbol{\Pi}_\ell^{\boldsymbol{RT}}$ are defined in §6. Importantly, the constant C_ℓ^k only depends locally on the ratio of the biggest and smallest eigenvalues of $\mathbb{A}_{\epsilon^j}^{k-1}$ and the hidden constant has no dependence on α and L at all.

9 Adaptive algorithm

In this section we will use the estimators based on the energy difference as in §7.1 to devise an adaptive algorithm. In particular, the algorithm will construct a sequence of solutions $u_\ell^{j,k}$ over mesh levels ℓ, regularization iterations j, and Newton iterations k. The main ideas will be to 1) spend the maximum amount of computing time on coarser meshes where computations are cheap and 2) decrease the regularization parameter adaptively so as to make Newton converge but avoid polluting the solution to the approximate problem. The algorithm accepts user-defined parameters $\gamma_{\text{lin}} > 0$ and $\gamma_{\text{reg}} > 0$ that express the requested relative sizes of the corresponding error components. Additionally, the algorithm takes as parameters $C_\epsilon \in (0,1)$ and $\epsilon^0 > 0$ that determine the sequence of regularization parameters according to (2.17) and a user-specified tolerance `tol`, the requested maximal overall error. We consider three stopping criteria with bars denoting the stopping indices as

$$\eta_{\text{lin}}^{\ell,j,\bar{k}} < \gamma_{\text{lin}}\eta_{\text{reg}}^{\ell,j,\bar{k}}, \tag{9.1a}$$

$$\eta_{\text{reg}}^{\ell,\bar{j},\bar{k}} < \gamma_{\text{reg}}\eta_{\text{dis}}^{\ell,\bar{j},\bar{k}}, \tag{9.1b}$$

$$\eta_{\text{tot}}^{\bar{\ell},\bar{j},\bar{k}} < \texttt{tol}. \tag{9.1c}$$

The first criterion (9.1a) indicates that the Newton solver should not continue on a given regularized problem if it has sufficiently converged. The problem should be changed (increasing the difficulty) by lowering the regularization parameter. The second criterion (9.1b) says that once the regularization parameter is sufficiently small on a given mesh, we can then pass to a finer mesh through the refinement procedure `REFINE` (newest vertex bisection or uniform refinement). Finally, the last criterion (9.1c) checks whether the estimator for the total error is below the user-specified threshold. Details are given in Algorithm 1.

10 Numerical experiments

We now present numerical experiments to substantiate the theory developed in the preceding sections. In particular, we compare and contrast the three error measures discussed in §4 for a polynomial manufactured solution defined on the unit square. Next, we explore several solver strategies for this same polynomial solution, including the adaptive Algorithm 1. In this case we will use uniform mesh refinement. Finally, we consider an unknown solution on an L-shaped domain and test the adaptive Algorithm 1

Algorithm 1: Adaptive regularized Newton algorithm

Initialization: Choose an initial guess $u_0^{0,0} \in V_0^p$ and initialize $\ell = j := 0$

Parameters : $\gamma_{\text{reg}}, \gamma_{\text{lin}}, \texttt{tol}, \epsilon^0, C_\epsilon$

Loop for discretization
  Loop for regularization
    Initialize $k := 0$
    Loop for linearization
      Increment $k := k + 1$
      From $u_\ell^{j,k-1}$ compute the linearized operator $\boldsymbol{A}_{\epsilon^j}^{k-1}$ by (3.10)
      Solve for $u_\ell^{j,k}$ in (3.5)
      Compute $\boldsymbol{t}_\ell^{j,k}$ following Definition 6.2 and $\eta_{\text{tot}}^{\ell,j,k}$ following (7.1)
      Compute estimators $\eta_{\text{dis}}^{\ell,j,k}, \eta_{\text{reg}}^{\ell,j,k}, \eta_{\text{lin}}^{\ell,j,k}$ following (7.4)
    until $\eta_{\text{lin}}^{\ell,j,k} < \gamma_{\text{lin}} \eta_{\text{reg}}^{\ell,j,k}$
    Update $\bar{k} := k$
    if $\eta_{\text{reg}}^{\ell,j,\bar{k}} \geq \gamma_{\text{reg}} \eta_{\text{dis}}^{\ell,j,\bar{k}}$ **then**
      Increment $j := j + 1$
      Update $\epsilon^j := C_\epsilon \epsilon^{j-1}$
    end
  until $\eta_{\text{reg}}^{\ell,j,\bar{k}} < \gamma_{\text{reg}} \eta_{\text{dis}}^{\ell,j,\bar{k}}$
  Update $\bar{j} := j$
  Increment $\ell := \ell + 1$
  $V_\ell^p := \texttt{REFINE}(V_{\ell-1}^p, \eta_{\text{tot},K}^{\ell,\bar{j},\bar{k}})$
  $u_{\ell j,0} := u_{\ell-1}^{\bar{j},\bar{k}}$
until $\eta_{\text{tot}}^{\ell,\bar{j},\bar{k}} < \texttt{tol}$
Update $\bar{\ell} := \ell$
return $u_{\bar{\ell}}^{\bar{j},\bar{k}}$

comparing both adaptive and uniform mesh refinement, in addition to adaptivity in regularization and linearization.

We will consider the effectivity index defined as the ratio of the estimator to the error, and in particular we have, using the notation of Proposition 7.1, Lemma 7.7, and Lemma 7.10,

$$I_{\text{tot}}^{\ell,j,k} := \frac{\eta_{\text{tot}}^{\ell,j,k}}{e_{\text{tot}}^{\ell,j,k}}, \quad \tilde{I}_{\text{tot}}^{\ell,j,k} := \frac{\tilde{\eta}_{\text{tot}}^{\ell,j,k}}{\tilde{e}_{\text{tot}}^{\ell,j,k}}, \quad \hat{I}_{\text{tot}}^{\ell,j,k} := \frac{\tilde{\eta}_{\text{tot}}^{\ell,j,k}}{|||u_\ell^{j,k} - u|||}. \tag{10.1}$$

We also consider a relative version of the various quantities (both errors and estimators) by dividing by the energy of the approximate solution, e.g.,

$$(\eta_{\text{lin,rel}}^{\ell,j,k})^2 := \frac{(\eta_{\text{lin}}^{\ell,j,k})^2}{\mathcal{J}(u_\ell^{j,k})}. \tag{10.2}$$

We will start with piecewise linear continuous finite elements i.e., we first set the polynomial order $p = 1$ in (3.3), but later we test adaptivity for $2 \leq p \leq 5$. All numerical experiments are conducted with the help of the `Gridap.jl` library [2, 43] in the Julia programming language.

10.1 Polynomial solution on a square

In this case, we consider a square domain $\Omega = (0, 1)^2 \subset \mathbb{R}^2$ and we take a manufactured solution,

$$u(x) = 10\, x(x-1)y(y-1) \tag{10.3}$$

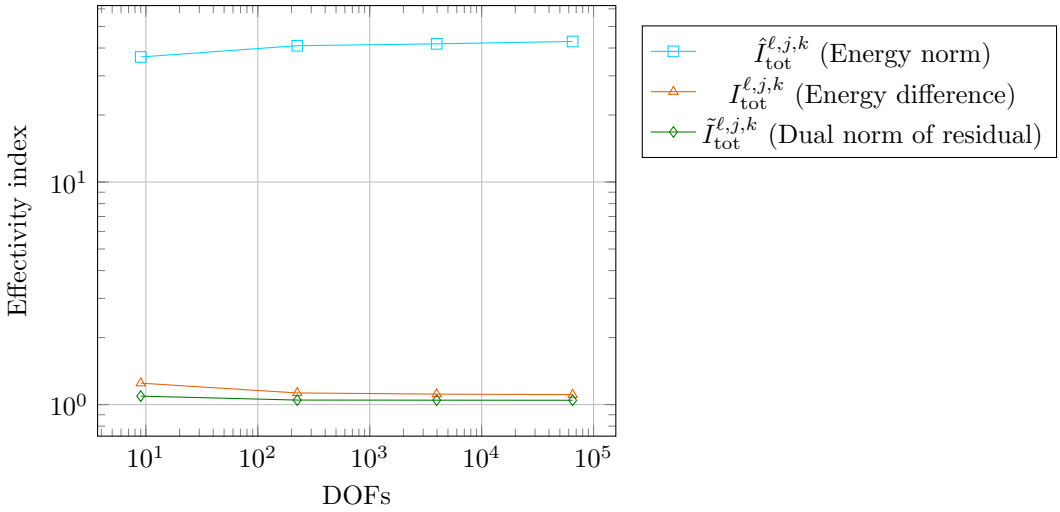

Figure 4 – [Polynomial solution (10.3), kink nonlinearity (2.19) with $s_0 = 1$ and m =64, polynomial degree $p = 1$, #DOFs varying, no regularization $\epsilon = 0$] Robustness with respect to the number of DOFs for the three error measures.

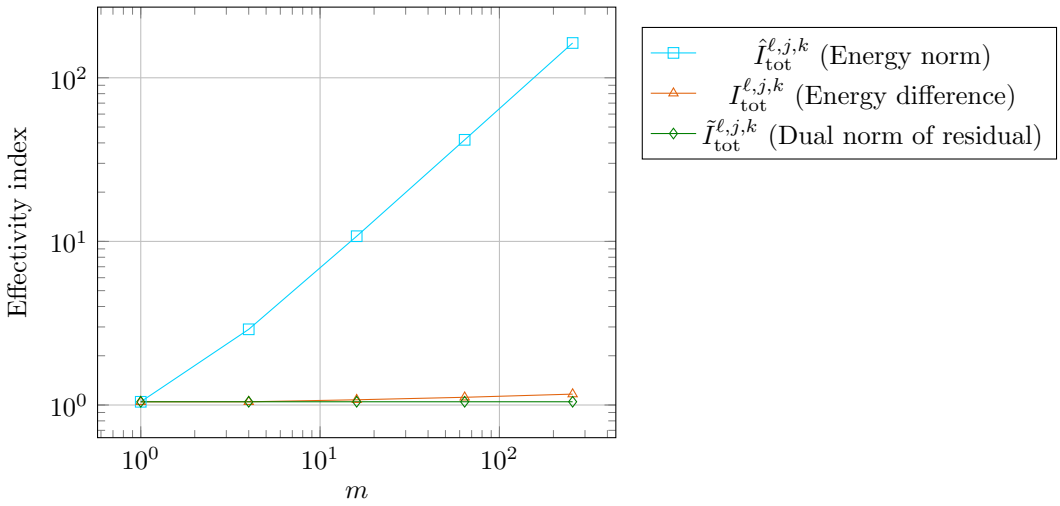

Figure 5 – [Polynomial solution (10.3), kink nonlinearity with $s_0 = 1$ and m varying, polynomial degree $p = 1$, #DOFs=3969, no regularization $\epsilon = 0$] The effectivity associated with the energy norm scales linearly in m whereas the other effectivities remain constant.

to generate the source term f. We neglect that $f \notin \mathcal{P}_0(\mathcal{T}_\ell)$. We will take ϕ in the energy (2.1) as in the definition of the kink function as (2.19). Thus, according to (2.21), we have that the monotonicity constant $\alpha = 1$ and the Lipschitz continuity constant $L = m$.

10.1.1 Comparison of the three error measures of §4

In this section, we will numerically investigate the relationships between the error measures discussed in section §4. The results are given in Figures 4 and 5. For this example we set the regularization parameter to zero, i.e., $\epsilon^0 := 0$. We first consider the scaling of the effectivity indices (10.1) with respect to the number of DOFs. We remark that all three error measures appear to be stable under uniform mesh refinement. This is consistent with the theory since the constants in the reliability and efficiency bounds are independent of the mesh size/number of DOFs. We do, however, note that the effectivity for the energy difference is much larger for each value of the mesh than for the other two error measures.

Now we consider the scaling with respect to the "size of the nonlinearity" i.e., $m = L/\alpha$ in Figure 5. We begin with the dual norm of the residual. We observe that the estimator (7.6) is not only a constant-free upper bound on the dual norm of the residual, but it is also a (local) lower bound [22, 8] robust with respect to m.

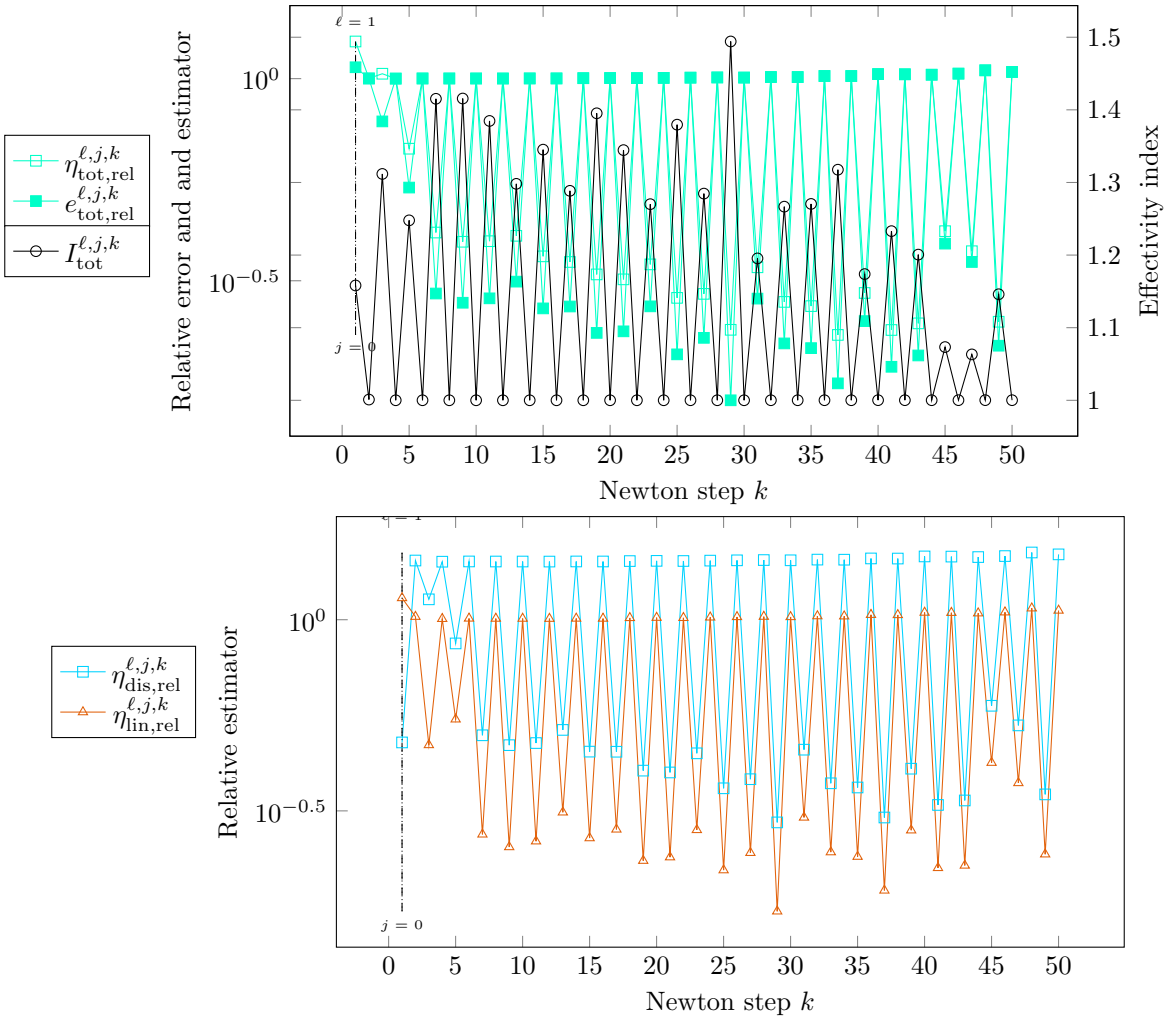

Figure 6 – [Polynomial solution (10.3), kink nonlinearity (2.19) with $s_0 = 1$ and $m = 10000$, polynomial degree $p = 1$, #DOFs = 261121, no regularization $\epsilon^0 = 0$, refinement index ℓ, regularization index j, linearization index k,] The classical Newton method fails to converge for the unregularized problem corresponding to $\epsilon = 0$.

Next we observe that the energy norm effectivity scales like m. This is consistent with the theoretical bounds, (4.10) and (8.1) and confirms non-robustness.

Finally, we consider the effectivity based on the difference of energies. This estimator too appears to be robust with respect to the scaling for this range of parameters. This behavior has been studied in [26]. In particular, a robustness result was demonstrated for a modified estimator, see (8.2).

10.1.2 Need for regularization for large ratios L/α

For the rest of this section, we will only consider the error and estimator based on the difference of energies of Proposition 7.1. We now study a much larger value of the "size of the nonlinearity" $m = L/\alpha$ and test the standard Newton algorithm, i.e., performing Algorithm 1 without regularization (with $\epsilon^0 = 0$). In particular, we set $m = 10000$ and we consider the same manufactured solution (10.3) on a fixed uniform mesh with 261121 DOFs. In Figure 6, we plot both the relative total error and estimator along with the corresponding effectivity index in function of the Newton iterations. In the lower figure, the components as in (7.4) are shown. Based on the behavior of the linearization estimator $\eta^{\ell,j,k}_{\mathrm{lin}}$, we conclude that the Newton solver fails to meet the specified convergence criteria after 50 iterations and we artificially terminate the algorithm. This manifests possible non-convergence of the Newton linearization for nonsmooth nonlinearities.

In Figure 7 we consider the same problem but now we fix a relatively large value for the initial regularization of $\epsilon^0 := 0.125$. We set the parameters $\gamma_{\mathrm{reg}} := 1.0\mathrm{e}16$, $\gamma_{\mathrm{lin}} := 1.0\mathrm{e}\text{-}5$ to ensure the algo-

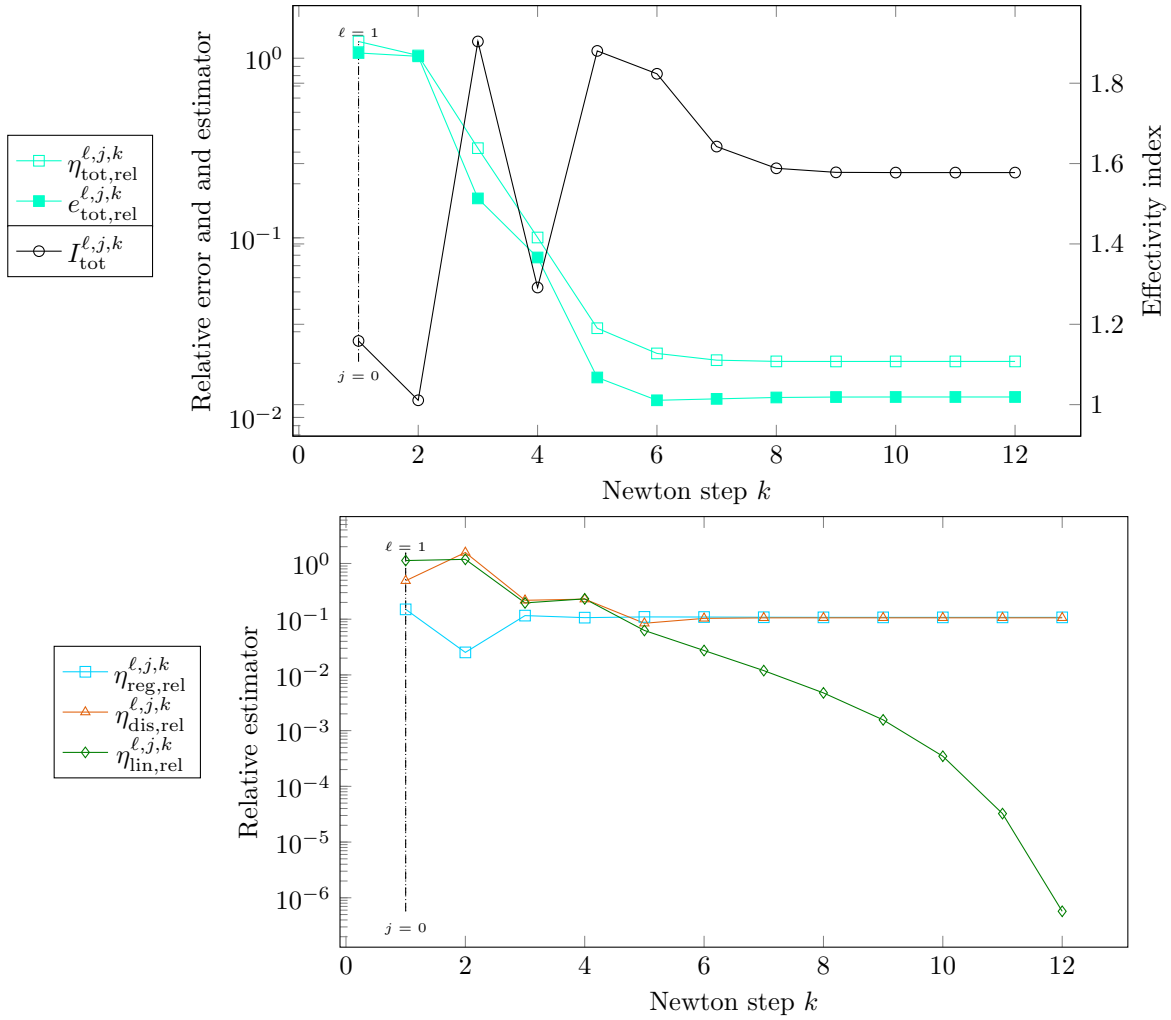

Figure 7 – [Polynomial solution (10.3), kink nonlinearity (2.19) with $s_0 = 1$ and $m = 10000$, #DOFs = 261121, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 1.0\mathrm{e}16$, $\gamma_{\mathrm{lin}} = 1.0\mathrm{e}{-5}$, $C_\epsilon = 1$, polynomial degree $p = 1$, mesh refinement index ℓ, regularization index j, linearization index k] The classical Newton method converges for the regularized problem with $\epsilon = 0.125$ but the errors and estimates other than the linearization stagnate due to the fixed regularization and discretization.

rithm will converge fully in linearization but will not perform any adaptivity in either regularization or discretization.

We first remark that the effectivity index of our a posteriori error estimator oscillates much less, compared to Figure 6, signifying a more stable approximation of the error. It appears further to converge after several iterations to a value near 1.6. We next remark that the Newton linearization exhibits the optimal quadratic convergence according to the values of the linearization estimator $\eta^{\ell,j,k}_{\mathrm{lin}}$. However, the other two estimator components $\eta^{\ell,j,k}_{\mathrm{reg}}$ and $\eta^{\ell,j,k}_{\mathrm{dis}}$ stagnate at similar values: 1.06e-1 and 1.07e-1 respectively. We remark that the reason that these are larger than the total estimator $\eta^{\ell,j,k}_{\mathrm{tot}} = 2.05\mathrm{e}{-2}$ is due to the insertion of absolute values in the definition (7.4). In any case, the Newton linearization now converges, but the regularization component is much too large to be satisfactory. This motivates the adaptive Algorithm 1 where the regularization estimator is decreased adaptively along the iterations.

10.1.3 Adaptive regularization and linearization

We now engage the adaptive regularization and linearization of Algorithm 1 by setting the parameters $\gamma_{\mathrm{reg}} = 0.6$, $\gamma_{\mathrm{lin}} = 0.4$ and $C_\epsilon = 0.5$. We again set $m = 10000$, $\epsilon^0 = 0.125$, and we now start from a uniform coarse mesh with 3969 DOFs for $\ell = 0$ ($64 \times 64 \times 2$ triangles). The results of the adaptive algorithm are presented in Figure 8. We first remark that the effectivity index (10.1) stays bounded below 2, and as the Newton solver converges for a fixed (j, ℓ), the effectivity approaches a value near 1.4. Next, we

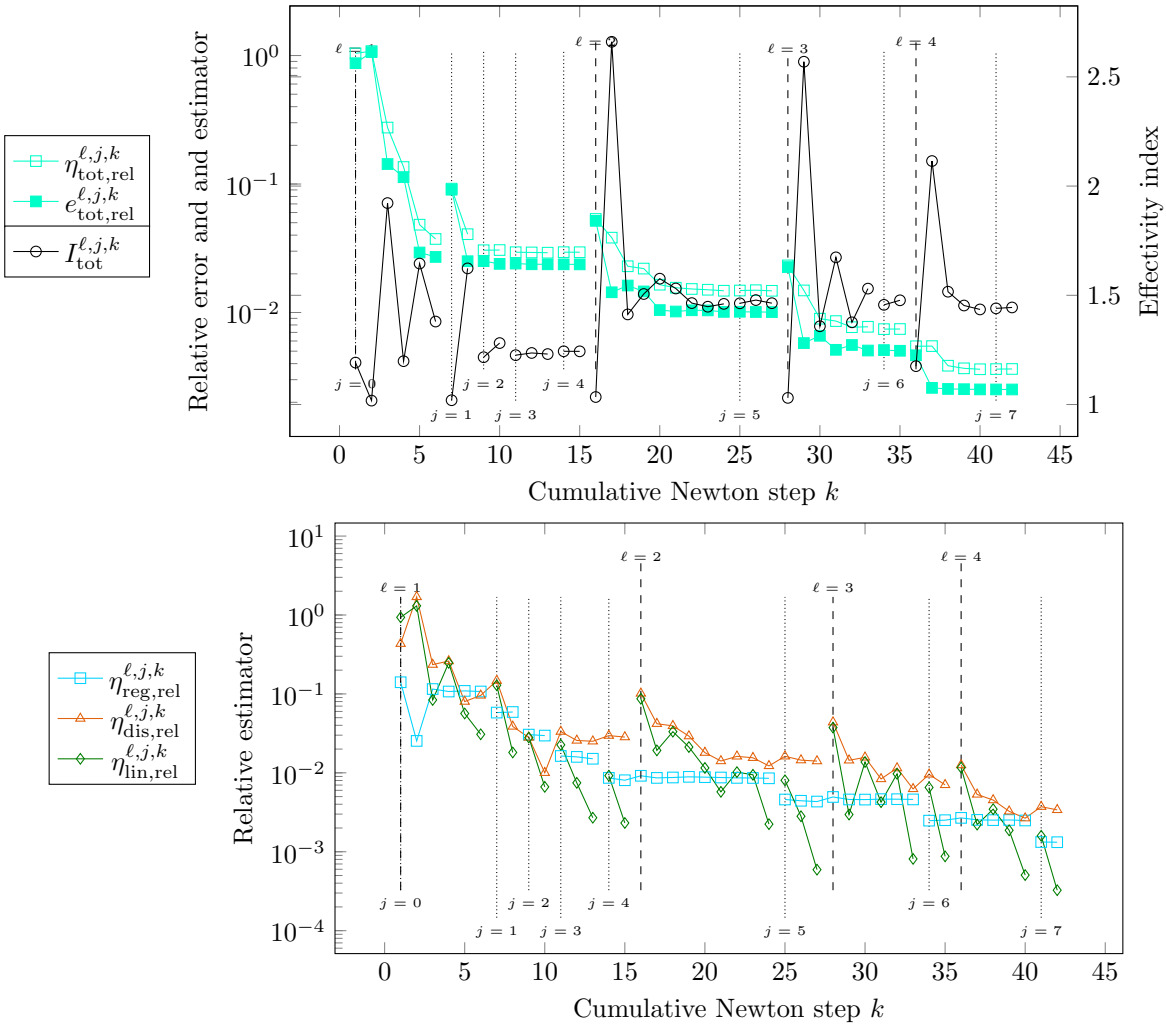

Figure 8 – [Polynomial solution (10.3), kink nonlinearity (2.19) with $s_0 = 1$ and $m = 10000$, final #DOFs = 261121, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.6$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$, polynomial degree $p = 1$, mesh refinement index ℓ, regularization index j, linearization index k] The adaptive Algorithm 1 applied to a polynomial solution. The final value of the regularization parameter is $\epsilon^6 = 1.95\text{e-}3$.

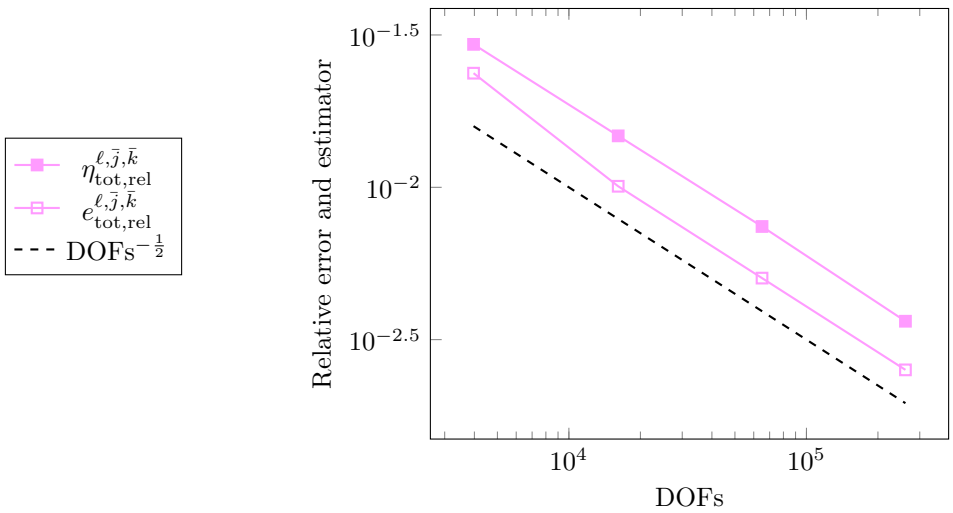

Figure 9 – [Polynomial solution (10.3), kink nonlinearity (2.19) with $s_0 = 1$ and $m = 10000$, polynomial degree $p = 1$, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.6$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$, #DOFs varies] Achieving the optimal rate of convergence for a polynomial solution with uniform mesh refinement.

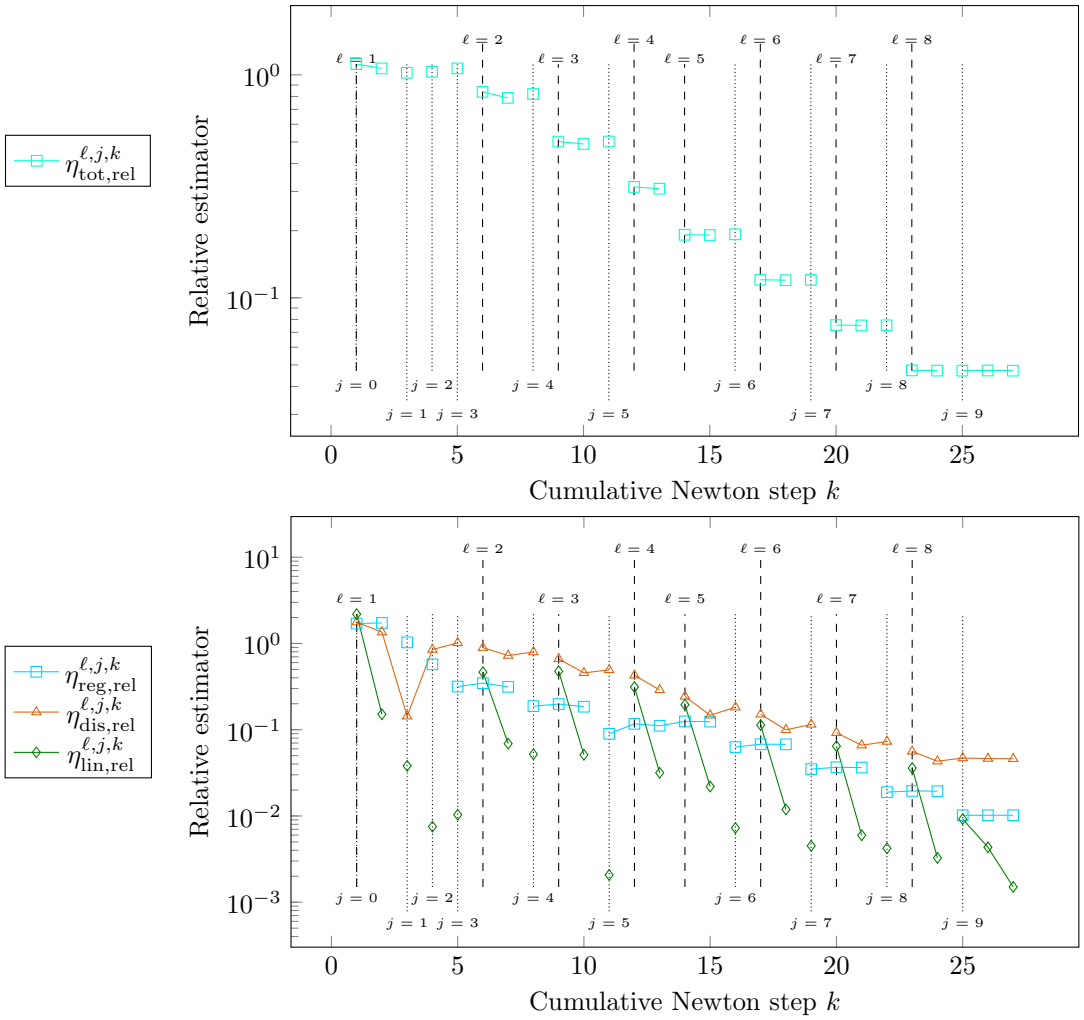

Figure 10 – [Unknown singular solution with data (10.4), kink nonlinearity (2.19) with $s_0 = 0.75$ and $m = 10000$, final #DOFs = 97793, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.4$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$, polynomial degree $p = 1$, mesh refinement index ℓ, regularization index j, linearization index k] Applying Algorithm 1 using uniform refinement.

observe that in accordance with the criteria (9.1b), the regularization component estimator is always γ_{reg}-times below the discretization component estimator. Thus, we can guarantee that in this sense the regularization does not pollute the overall error, in constrast to the previous section. Indeed, this is substantiated in Figure 9, where the optimal rate of convergence of both the error and the estimator with respect to DOFs is observed, for the stopping indices $\bar{k}$ and $\bar{j}$ satisfying respectively (9.1a) and (9.1b). Finally, we see that the majority of the iterations are spent on the meshes $\ell = 0, 1, 2$. This is another advantage of the adaptive algorithm, where the coarser meshes serve as a way to create a good initial guess for the next mesh. For a smooth problem it is not strictly necessary to begin on a coarse mesh, since the refinement procedure is known a priori, but as we will see in the following section, sometimes starting on a coarse mesh in not just useful to create a sequence of initial guesses, but also to efficiently obtain an optimal mesh family using adaptive mesh refinement. Finally, we remark that the final total error is 2.51e-3 as opposed to 1.03e-2 in the previous case, where no adaptivity was used. This confirms that fixing once and for all the regularization parameter can deteriorate the quality of the final solution, which does not happen for Algorithm 1.

10.2 Unknown solution on an L-shaped domain

We now consider an L-shaped domain $\Omega = (-1, 1)^2 \setminus ([0, 1] \times [-1, 0])$ where we impose the boundary condition and right-hand side as

$$u = u_{\mathrm{D}}(r, \theta) := r^\alpha \sin(\alpha\theta) \text{ on } \partial\Omega \text{ and } f = 0 \text{ in } \Omega, \tag{10.4}$$

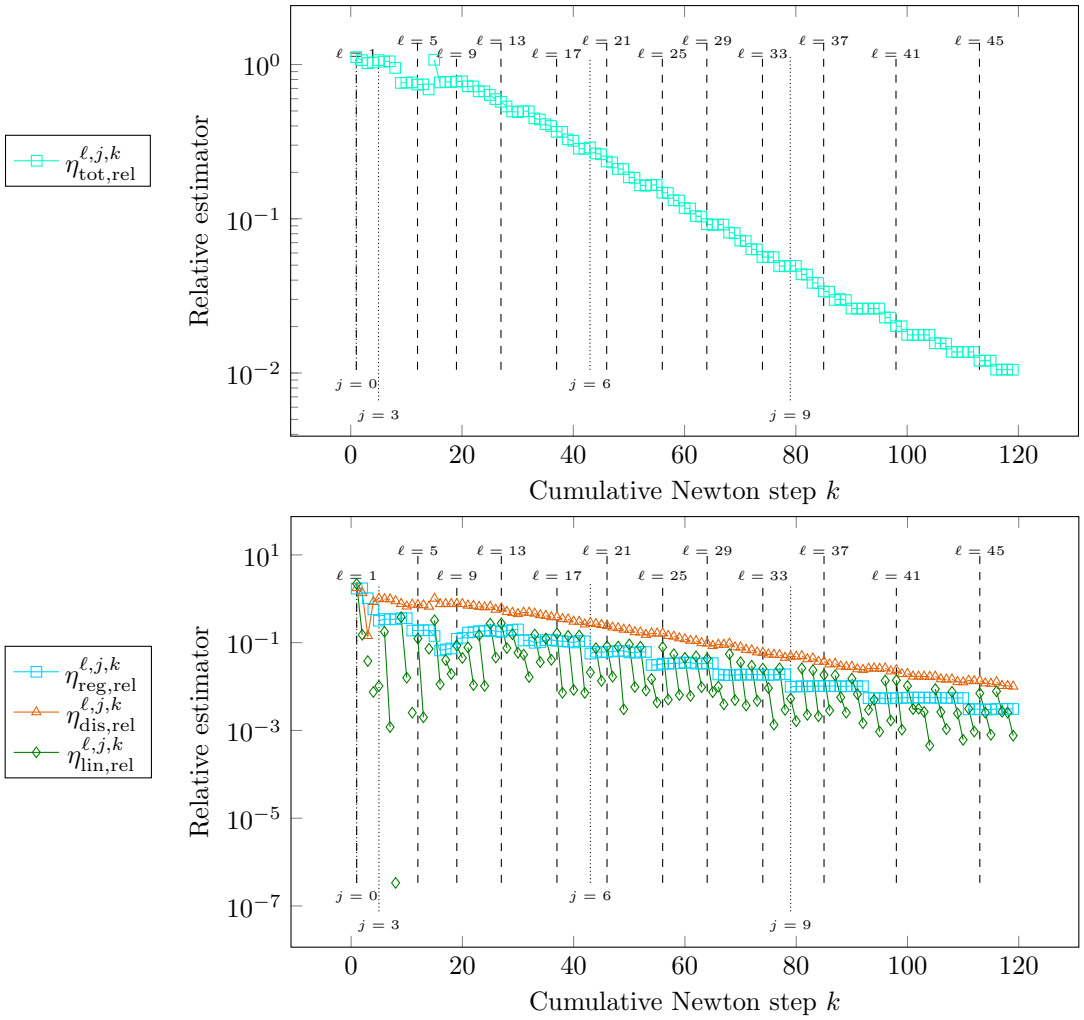

Figure 11 – [Unknown singular solution with data (10.4), kink nonlinearity (2.19) with $s_0 = 0.75$ and $m = 10000$, final #DOFs = 86973, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.4$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$, polynomial degree $p = 1$, mesh refinement index ℓ, regularization index j, linearization index k] Applying Algorithm 1 using adaptive refinement.

where $\alpha = \frac{2}{3}$. We note that the trace of this function is not a piecewise polynomial at the boundary, but we ignore the error due to interpolation of the boundary condition in this case. This has been rigorously studied in, e.g., [17]. We will consider the parameters $\gamma_{\mathrm{reg}} = 0.4$, $\gamma_{\mathrm{lin}} = 0.4$ and $C_\epsilon = 0.5$. The main difference compared to the polynomial solution of §10.1 is that we anticipate that uniform mesh refinement will not achieve the a priori optimal convergence rate due to the re-entrant corner and the nonlinearity that will activate around the curve $|\nabla u| = s_0$. We will first consider uniform refinement of the mesh as in the previous problem, and then adaptive refinement using the estimator $\eta^{\ell,j,k}_{\mathrm{tot},K}$ in (7.2) will be employed on line 19 of Algorithm 1. Note that here, since we do not know the true solution, we do not plot the error and the effectivity index.

In Figure 10, we consider the uniform mesh refinement strategy. We see the estimator along with the component estimators. The adaptive algorithm works as before for the smooth case, with the regularization estimator always below the discretization estimator, and for fixed (j, ℓ) the Newton solver converges very quickly. However, if we now consider convergence with respect to DOFs in Figure 12, we see that we obtain the suboptimal convergence rate of DOFs$^{-1/3}$. This is evidence that the true solution is not H^2 regular, and therefore the optimal rate of DOFs$^{-1/2}$ will not be achieved for uniform mesh refinement. Next, we consider in Figure 11 applying Algorithm1 but now with adaptive mesh refinement using Dörfler marking [18]. The elementwise indicators are given by $\eta^{\ell,j,k}_{\mathrm{tot},K}$ from (7.2) and we use the newest vertex bisection algorithm to enforce mesh conformity, i.e., to ensure no hanging nodes are generated. We see first of all that many more iterations are needed to obtain a similar number of final DOFs. However, the total estimator at the end of the iterations is much lower compared to that of the uniform case.

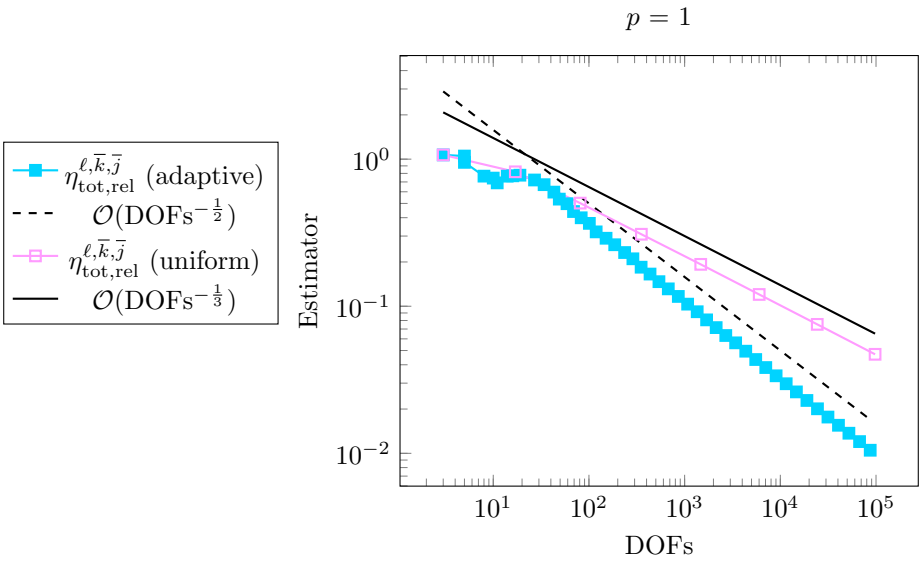

Figure 12 – [Unknown singular solution with data (10.4), kink nonlinearity (2.19) with $s_0 = 0.75$ and $m = 10000$, polynomial degree $p = 1$, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.4$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$, #DOFs varies] Comparison of the suboptimal convergence for uniform refinement and optimal convergence for adaptive mesh refinement based on the estimator $\eta^{\ell,j,k}_{\mathrm{tot}}$ for the lowest order.

This is even more explicit when we plot the number of DOFs versus the total estimator in Figure 12. In particular, we see that upon running the same algorithm with adaptive mesh refinement, we recover the optimal rate of $\mathrm{DOFs}^{-1/2}$ for the estimator with respect to DOFs.

One advantage of adaptive mesh refinement is that it allows in general to recover the optimal rate of convergence for arbitrary polynomial degree with respect to DOFs, i.e., $\mathrm{DOFs}^{-p/d}$, even when the solution does not have sufficient regularity for the a priori theory, see e.g. [13]. We now test the convergence rate and the behavior of the adaptive algorithm for higher polynomial degrees. We first show the convergence plots for $2 \leq p \leq 5$ in Figure 13. We observe that for $p = 2, 3$ the optimal rate of convergence $\mathrm{DOFs}^{-p/d}$ is again achieved for the adaptive mesh refinement. However, for $p \geq 4$, the convergence rate is suboptimal, it appears to be similar to that of $p = 3$, i.e., $\mathrm{DOFs}^{-3/2}$. We can potentially explain this deterioration by the appearance of one-dimensional curve singularities, see e.g., [12]. Indeed, in Figure 14, we notice there is non-trivial refinement along the curves where the norm of the gradient equals s_0. Since the right hand side $f = 0$ and we have chosen a nonsmooth nonlinearity, heuristically the solution must also be nonsmooth along $|\nabla u| = s_0$ to compensate. Finally, we consider comparing the estimator against a notion of cost across all iterations following [24, 25] We define this cost at each step of the algorithm as

$$\mathrm{Cost} = \sum_{\ell=0}^{\overline{\ell}} \sum_{j=0}^{\overline{j}(\ell)} \sum_{k=1}^{\overline{k}(\ell,j)} (\mathrm{DOFs})_\ell. \tag{10.5}$$

We observe in Figure 15 that the rates in this metric are very similar to those observed for the convergence with respect to DOFs of Figure 12; in particular, we shall obtain the optimal -1/2 rate in cost for adaptive mesh refinement.

11 Conclusion and future work

In this paper, we have considered an adaptive algorithm to iteratively solve energy minimization problems with nonsmooth nonlinearities. Our adaptive algorithm is guided by the so-called primal-dual gap error estimator which provides an upper bound for the difference of energies. We construct the necessary dual object required by the estimator by solving mutually independent, patch-local, minimization problems. We introduce a regularization to allow the use of a standard Newton's method as a nonlinear solver for the nonsmooth system of equations associated to the minimization problem. The algorithm adaptively controls the regularization parameter to reduce the model error incurred by regularizing the problem. We perform a decomposition of the total estimator into component estimators related to regularization,

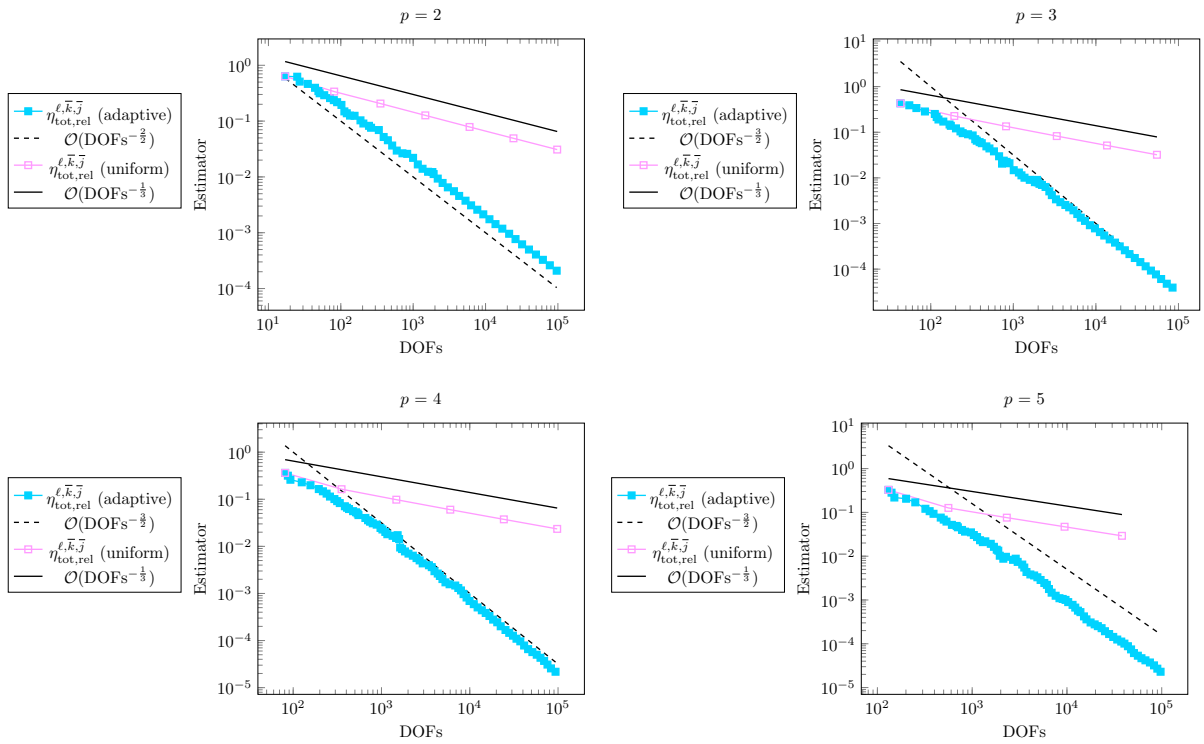

Figure 13 – [Unknown singular solution with data (10.4), kink nonlinearity (2.19) with $s_0 = 0.75$ and $m = 1000$, different polynomial degrees p, #DOFs varying, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.4$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$] Different polynomial degrees p using adaptive mesh refinement based on $\eta^{\ell,j,k}_{\mathrm{tot}}$. The optimal rate of $\mathrm{DOFs}^{-p/d}$ is obtained for adaptive refinement up to $p = 3$, then only suboptimal convergence of $\mathrm{DOFs}^{-3/d}$ is achieved. We explain this by the appearance of one-dimensional (curve) singularities. In the case of uniform mesh refinement, increasing the polynomial degree does not change the suboptimal rate of $\mathrm{DOFs}^{-1/3}$.

discretization, and linearization. In particular, we prove that these component estimators converge to zero in the limit as the number of associated iterations tends to infinity. These component estimators are used to construct stopping criteria for the various components of the algorithm.

We test our algorithm numerically on two examples. In the first example, we show that the regularization restores the (quadratic) convergence of Newton's method, which without regularization failed to converge. Moreover, the adaptivity in the regularization does not influence the optimal rate of convergence of the error in the energy difference with respect to DOFs. In the second example, we consider an unknown solution on an L-shaped domain. We use adaptive mesh refinement to overcome the geometric singularity generated by the re-entrant corner; there also appears a singularity along a curve arising from the nonsmooth nonlinearity. With the help of adaptive mesh refinement, we again obtain the optimal rate of convergence with respect to DOFs for low order cases. However, for higher orders, suboptimal convergence rates are obtained. We attribute this to the appearance of the above-discussed singularity, which is a well known difficulty for isotropic mesh refinement.

In terms of future work, one possible approach to address the singularity problem in the L-shaped domain case would be to employ an anisotropic refinement strategy. In our work we, however, would be missing a number of theoretical tools. It would also be instructive to prove convergence of the adaptive algorithm, i.e., to show rigorously that by decreasing the regularization, we can obtain the optimal rate of convergence with respect to DOFs and cost, which is what we observe numerically. It may also be possible to extend certain aspects of this algorithm to other energy minimization settings posed in different spaces like the p-Laplace problem or the obstacle problem.

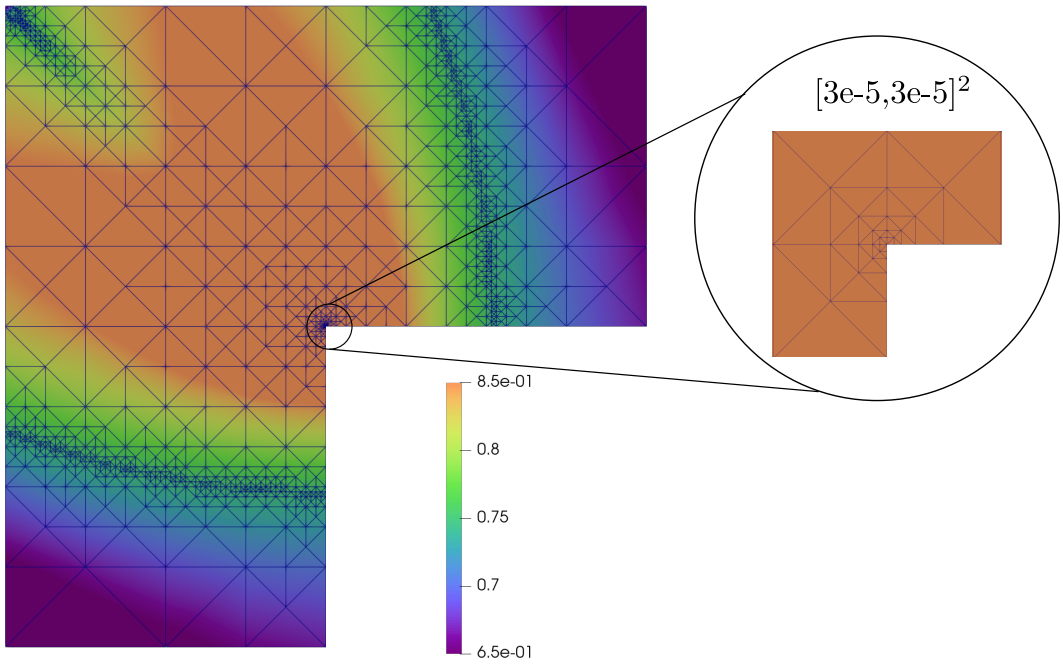

Figure 14 – [Unknown singular solution with data (10.4), kink nonlinearity (2.19) with $s_0 = 0.75$ and $m = 1000$, polynomial degree $p = 4$, #DOFs=93681, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.4$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$,] Coloring corresponding to the norm of the gradient of the approximate solution at the final iteration, i.e., $|\nabla u_\ell^{\bar{j},\bar{k}}|$. We note the aggressive refinement at the re-entrant corner, and the weaker, but still substantial, refinement along the curves corresponding to $s_0 = |\nabla u_\ell^{\bar{j},\bar{k}}|$, i.e., at the kink.

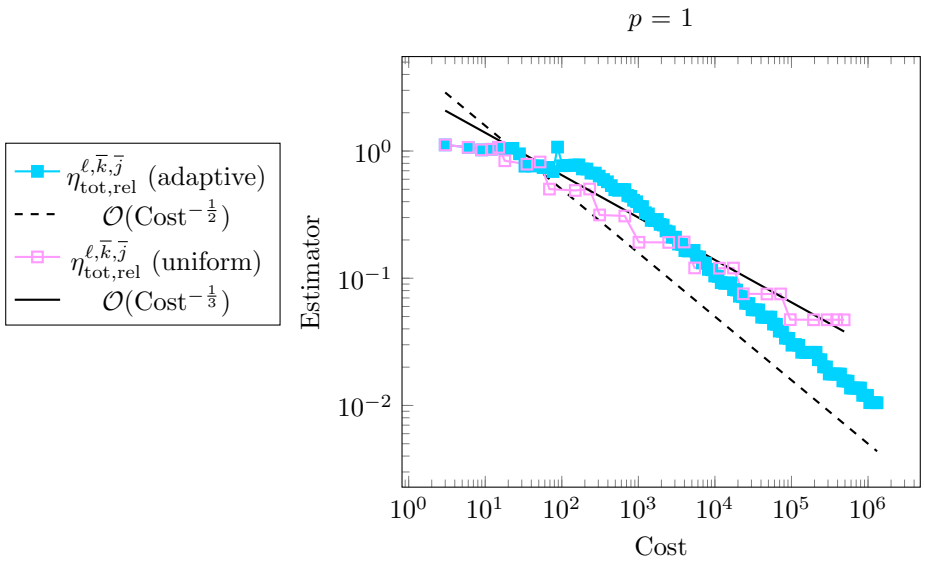

Figure 15 – [Unknown singular solution with data (10.4), kink nonlinearity (2.19) with $s_0 = 0.75$ and $m = 10000$, polynomial degree $p = 1$, $\epsilon^0 = 0.125$, $\gamma_{\mathrm{reg}} = 0.4$, $\gamma_{\mathrm{lin}} = 0.4$, $C_\epsilon = 0.5$, #DOFs varies] Convergence in costs given by the triple sum (10.5). We observe the same rates in function of the cost as we do for the DOFs which is expected theoretically.

Appendix A Proofs from §5

Proof of Proposition 5.1. We begin by proving (5.6a). To simplify notation, define $\xi := \phi'$. We consider the integral in (5.1), with the change of variables $s = \phi'(t) = \xi(t)$

$$\begin{aligned}\int_0^r \xi^{-1}(s)\,\mathrm{d}s &= \int_{\xi^{-1}(0)}^{\xi^{-1}(r)} \xi^{-1}(\xi(t))\xi'(t)\,\mathrm{d}t = \int_0^{\xi^{-1}(r)} t\xi'(t)\,\mathrm{d}t \\ &= t\xi(t)\Big|_0^{\xi^{-1}(r)} - \int_0^{\xi^{-1}(r)} \xi(t)\,\mathrm{d}t = r\xi^{-1}(r) - \phi(\xi^{-1}(r)) + \phi(0),\end{aligned}$$

where we have used our assumptions that $\phi'(0) = 0$ and hence also $(\phi')^{-1}(0) = 0 = \xi^{-1}(0)$. The second equality in (5.6a) follows from the basic fact that for a convex differentiable function, the max is obtained by setting the derivative w.r.t. s in the curly braces equal to zero, and hence $r = \phi'(s)$.

To prove (5.6b), we consider the criterion for convexity. For $r_1, r_2 \in \mathrm{Dom}(\phi^*)$ and $\alpha \in [0,1]$,

$$\begin{aligned}\phi^*(\alpha r_1 + (1-\alpha)r_2) &= \max_s\{s[\alpha r_1 + (1-\alpha)r_2] - \phi(s)\} \\ &= \max_s\{\alpha[sr_1 - \phi(s)] + (1-\alpha)[sr_2 - \phi(s)]\} \\ &\le \alpha \max_s\{sr_1 - \phi(s)\} + (1-\alpha)\max_s\{sr_2 - \phi(s)\} \\ &= \alpha\phi^*(r_1) + (1-\alpha)\phi^*(r_2).\end{aligned}$$

Finally, to prove (5.6c), let now $\zeta(r) := (\phi')^{-1}(r)$, so that

$$\frac{d}{dr}\phi^*(r) = \frac{d}{dr}\big(r\zeta(r) - \phi(\zeta(r))\big) = \zeta(r) + r\zeta'(r) - \phi'(\zeta(r))\zeta'(r) = \zeta(r) = (\phi')^{-1}(r),$$

because $\phi'(\zeta(r)) = r$ by definition, whereas (5.6d) is obvious. □

Proof of Corollary 5.2. The inequality (5.7), follows immediately from the max definition of the transform, i.e., the second equality of (5.6a). The maximum in (5.6a) for $r = \phi'(s)$, as discussed above, which leads to the equality. □

Proof of Corollary 5.3. Since we know $\phi' : \mathbb{R}^+ \to \mathbb{R}^+$, we have

$$\boldsymbol{A}(\boldsymbol{q}) \cdot \boldsymbol{q} = \frac{\phi'(|\boldsymbol{q}|)}{|\boldsymbol{q}|}\boldsymbol{q} \cdot \boldsymbol{q} = \underbrace{|\boldsymbol{A}(\boldsymbol{q})|}_{r}\underbrace{|\boldsymbol{q}|}_{s}. \tag{A.1}$$

Now take r in s as in the Young inequality (5.7), and note that $r = \phi'(s)$ so equality (5.8b) holds. Unpacking the definitions, we find (5.8a) as

$$\boldsymbol{A}(\boldsymbol{A}^*(\boldsymbol{q})) \overset{(5.2)}{=} \frac{\phi'\big(|\boldsymbol{A}^*(\boldsymbol{q})|\big)}{|\boldsymbol{A}^*(\boldsymbol{q})|}\boldsymbol{A}^*(\boldsymbol{q}) = \phi'\big((\phi^*)'(|\boldsymbol{q}|)\big)\frac{\boldsymbol{A}^*(\boldsymbol{q})}{(\phi^*)'(|\boldsymbol{q}|)} \overset{(5.6c)}{=} \frac{|\boldsymbol{q}|\boldsymbol{A}^*(\boldsymbol{q})}{(\phi^*)'(|\boldsymbol{q}|)} \overset{(5.2)}{=} \boldsymbol{q}.$$

□

Proof of Lemma 5.4. Note that for any $x, y \ge 0$,

$$\begin{aligned}&0 \le \alpha(x-y)^2 \overset{(2.3b)}{\le} (x-y)(\phi'(x) - \phi'(y)) \\ \implies &\alpha|x-y|^2 \le |x-y||\phi'(x) - \phi'(y)| \\ \implies &\alpha|x-y| \le |\phi'(x) - \phi'(y)|.\end{aligned}$$

Thus, since we assume ϕ' is bijective on $\mathbb{R}^+$, we may take $x = (\phi')^{-1}(r), y = (\phi')^{-1}(s)$, yielding

$$\alpha|(\phi')^{-1}(r) - (\phi')^{-1}(s)| \le |r-s|.$$

The relationship given by (5.6c) finishes the proof. □

References

[1] Ainsworth, M., and Oden, J. T. *A Posteriori Error Estimation in Finite Element Analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2000.

[2] Badia, S., and Verdugo, F. Gridap: An extensible Finite Element toolbox in Julia. *Journal of Open Source Software 5*, 52 (2020), 2520.

[3] Bartels, S. Error control and adaptivity for a variational model problem defined on functions of bounded variation. *Mathematics of Computation 84*, 293 (2015), 1217–1240.

[4] Bartels, S. *Numerical Methods for Nonlinear Partial Differential Equations*, vol. 47 of *Springer Series in Computational Mathematics*. Springer, Cham, 2015.

[5] Bartels, S., and Milicevic, M. Efficient iterative solution of finite element discretized nonsmooth minimization problems. *Computers & Mathematics with Applications 80*, 5 (2020), 588–603.

[6] Bartels, S., and Milicevic, M. Primal-dual gap estimators for a posteriori error analysis of nonsmooth minimization problems. *ESAIM: Mathematical Modelling and Numerical Analysis 54*, 5 (2020), 1635–1660.

[7] Bauschke, H. H., and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2 ed. CMS Books in Mathematics/Ouvrages de Mathématiques de La SMC. Springer, Cham, 2017.

[8] Blechta, J., Málek, J., and Vohralík, M. Localization of the $W^{-1,q}$ norm for local a posteriori efficiency. *IMA Journal of Numerical Analysis 40*, 2 (2020), 914–950.

[9] Braess, D., and Schöberl, J. Equilibrated residual error estimator for edge elements. *Mathematics of Computation 77*, 262 (2008), 651–673.

[10] Brezzi, F., and Fortin, M., Eds. *Mixed and Hybrid Finite Element Methods*, vol. 15 of *Springer Series in Computational Mathematics*. Springer, New York, NY, 1991.

[11] Brooks, R. H., and Corey, A. T. Properties of Porous Media Affecting Fluid Flow. *Journal of the Irrigation and Drainage Division 92*, 2 (1966), 61–88.

[12] Carstensen, C., Maischak, M., Praetorius, D., and Stephan, E. P. Residual-based a posteriori error estimate for hypersingular equation on surfaces. *Numerische Mathematik 97*, 3 (2004), 397–425.

[13] Cascon, J. M., Kreuzer, C., Nochetto, R. H., and Siebert, K. G. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM Journal on Numerical Analysis 46*, 5 (2008), 2524–2550.

[14] Destuynder, P., and Métivet, B. Explicit error bounds in a conforming finite element method. *Mathematics of Computation 68*, 228 (1999), 1379–1396.

[15] Deuflhard, P. *Newton Methods for Nonlinear Problems*, vol. 35 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2004.

[16] Di Pietro, D., Vohralík, M., and Yousef, S. Adaptive regularization, linearization, and discretization and a posteriori error control for the two-phase Stefan problem. *Mathematics of Computation 84*, 291 (2015), 153–186.

[17] Dolejší, V., Ern, A., and Vohralík, M. *hp*-adaptation driven by polynomial-degree-robust a posteriori error estimates for elliptic problems. *SIAM Journal on Scientific Computing 38*, 5 (2016), A3220–A3246.

[18] Dörfler, W. A convergent adaptive algorithm for Poisson's equation. *SIAM Journal on Numerical Analysis 33*, 3 (1996), 1106–1124.

[19] El Alaoui, L., Ern, A., and Vohralík, M. Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems. *Computer Methods in Applied Mechanics and Engineering 200*, 37 (2011), 2782–2795.

[20] Engl, H. W., Hanke, M., and Neubauer, A. *Regularization of Inverse Problems*, vol. 375 of *Mathematics and Its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[21] Ern, A., and Guermond, J.-L. *Theory and Practice of Finite Elements*, vol. 159 of *Applied Mathematical Sciences*. Springer New York, New York, NY, 2004.

[22] Ern, A., and Vohralík, M. Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM Journal on Scientific Computing 35*, 4 (2013), A1761–A1791.

[23] Friedman, A. The Stefan problem in several space variables. *Transactions of the American Mathematical Society 133*, 1 (1968), 51–87.

[24] Gantner, G., Haberl, A., Praetorius, D., and Stiftner, B. Rate optimal adaptive FEM with inexact solver for nonlinear operators. *IMA Journal of Numerical Analysis 38*, 4 (2018), 1797–1831.

[25] Haberl, A., Praetorius, D., Schimanko, S., and Vohralík, M. Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. *Numerische Mathematik 147*, 3 (2021), 679–725.

[26] Harnist, A., Mitra, K., Rappaport, A., and Vohralík, M. Robust energy a posteriori estimates for nonlinear elliptic problems. preprint hal-04033438, May 2023.

[27] Hofmann, B., Kaltenbacher, B., Pöschl, C., and Scherzer, O. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data 23*, 3 (2007), 987–1010.

[28] Irgens, F. *Rheology and Non-Newtonian Fluids*. Springer International Publishing, Cham, 2014.

[29] Kelley, C. T. *Iterative Methods for Linear and Nonlinear Equations*, vol. 16 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995.

[30] Ladeveze, P., and Leguillon, D. Error estimate procedure in the finite element method and applications. *SIAM Journal on Numerical Analysis 20*, 3 (1983), 485–509.

[31] Nochetto, R. H. Error estimates for multidimensional singular parabolic problems. *Japan Journal of Applied Mathematics 4*, 1 (1987), 111–138.

[32] Pereverzev, S., and Schock, E. On the adaptive selection of the parameter in regularization of ill-posed problems. *SIAM Journal on Numerical Analysis 43*, 5 (2005), 2060–2076.

[33] Pop, I. S., and Schweizer, B. Regularization schemes for degenerate Richards equations and outflow conditions. *Mathematical Models and Methods in Applied Sciences 21*, 08 (2011), 1685–1712.

[34] Prager, W., and Synge, J. L. Approximations in elasticity based on the concept of function space. *Quarterly of Applied Mathematics 5*, 3 (1947), 241–269.

[35] Qi, H.-D., and Liao, L.-Z. A Smoothing Newton Method for General Nonlinear Complementarity Problems. *Computational Optimization and Applications 17*, 2 (2000), 231–253.

[36] Qi, L., and Sun, D. Smoothing Functions and Smoothing Newton Method for Complementarity and Variational Inequality Problems. *Journal of Optimization Theory and Applications 113*, 1 (Apr. 2002), 121–147.

[37] Rao, M. M., and Ren, Z. D. *Theory of Orlicz Spaces*, vol. 146 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, 1991.

[38] Repin, S. I. A posteriori error estimates for approximate solutions of variational problems with functionals of power growth. *Journal of Mathematical Sciences 101*, 5 (2000), 3531–3538.

[39] Repin, S. I. A posteriori error estimation for variational problems with uniformly convex functionals. *Mathematics of Computation 69*, 230 (2000), 481–500.

[40] Repin, S. I. *A Posteriori Estimates for Partial Differntial Equations*. De Gruyter, 2008.

[41] Smears, I., and Vohralík, M. Simple and robust equilibrated flux a posteriori estimates for singularly perturbed reaction–diffusion problems. *ESAIM: Mathematical Modelling and Numerical Analysis 54*, 6 (Nov. 2020), 1951–1973.

[42] Stein, E. M., and Shakarchi, R. *Real Analysis*, vol. 3 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2005.

[43] Verdugo, F., and Badia, S. The software design of Gridap: A Finite Element package based on the Julia JIT compiler. *Computer Physics Communications 276* (2022), 108341.

[44] Verfürth, R. *A Posteriori Error Estimation Techniques for Finite Element Methods*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2013.

[45] Viorel Barbu, T. P. *Convexity and Optimization in Banach Spaces*. Springer Netherlands, 2014.

[46] Wan, Z., Li, H., and Huang, S. A smoothing inexact Newton method for nonlinear complementarity problems. *Abstract and Applied Analysis 2015* (2015), e731026.

[47] Zeidler, E. *Nonlinear Functional Analysis and Its Applications. II/B*. Springer-Verlag, New York, 1990.

Appendix B Performance study for the estimator

In this section we discuss the cost associated to the evaluation of the estimators in (7.1). We constrast this with the cost of solving the linearization step (3.5). In particular, we consider the cost of assembling the flux reconstruction given in (6.3) since this is by far the most expensive part of calculating the estimators. The patchwise problems are mutually independent and therefore can be solved in parallel. We use the threading model in Julia to possibly take advantage of this parallelism in the results presented below. We also use routines from the `Gridap.jl` library [2, 43] in the assembly. The tests in this section were carried out on a cluster with 20 dual socket Cascade Lake Intel Xeon 5218 nodes and 192GB of 2667 MHz RAM. We do not discuss any gains due to significantly reducing the number of linearization iterations due to our adaptive stopping criteria.

For the setting of §10.1, we plot the time to assemble the flux via (6.3) versus the time required to solve the linearization step (3.5) by a direct LU solver in Figure 16 for polynomial degrees $p = 1, 2, 3, 4$. We use differently-sized meshes for the various polynomial degrees to have a roughly constant number of DOFs. We first remark that the assembly time for (6.3) is linear in the number of DOFs for sufficiently many DOFs. This is in contrast to the time to solve the linearization step (3.5), which is superlinear as we use a direct LU solver. Next, we observe that the total time to assemble the flux decreases with more processors for sufficiently many DOFs, and that this is more pronounced for $p \geq 2$.

We now consider Table 1, where we tabulate the percentage of the total runtime, i.e.,

$$P_{\mathrm{f}} := \frac{T_{\mathrm{f}}}{T_{\mathrm{f}} + T_{\mathrm{N}}}, \tag{B.1}$$

where T_{f} is the time for assembling the flux and T_{N} is the time for solving the linearization step. We calculate this quantity for the largest number of DOFs that we consider, approximately 1.25e6 DOFs for the various polynomial degrees. We see that this value is less than 0.5 (meaning the flux assembly is cheaper than solving the linearization step) for $p \geq 2$ and 8 or more processors. The table also reflects a monotone decrease of P_{f} as we increase the number of processors, which is more pronounced for $p \geq 2$. We conclude that the cost of the estimator can be effectively reduced by adding computational resources, at least for problems of a sufficient size.

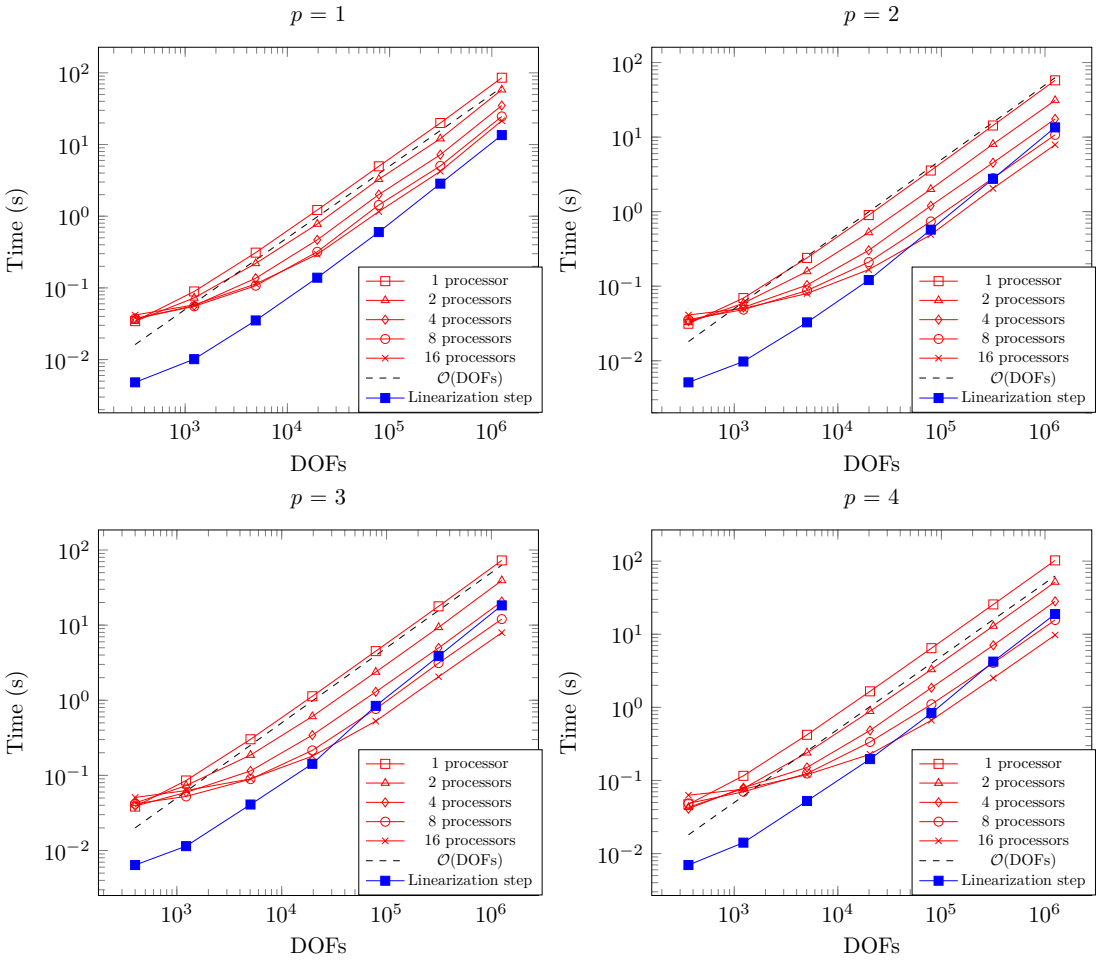

Figure 16 – Comparison between solving linearization step (3.5) and assembling the equilibrated flux via (6.3) for different polynomial degrees p and different numbers of processors.

Table 1 – Percentage of total runtime (B.1) for 1.25e6 DOFs.

p	Processors				
	1	2	4	8	16
1	0.87	0.81	0.74	0.66	0.62
2	0.81	0.70	0.58	0.45	0.39
3	0.81	0.68	0.54	0.41	0.28
4	0.85	0.74	0.61	0.45	0.33