ADAPTIVE ITERATIVE LINEARIZATION GALERKIN METHODS FOR NONLINEAR PROBLEMS

PASCAL HEID AND THOMAS P. WHLER

Abstract. Fixed point iterations are widely used for the analysis and numerical treatment of nonlinear problems. In many cases, such schemes can be interpreted as iterative local linearization methods, which, as will be shown, can be obtained by applying a suitable preconditioning operator to the original (nonlinear) equation. Based on this observation, we will derive a unified abstract framework which recovers some prominent iterative schemes. In particular, for Lipschitz continuous and strongly monotone operators, we derive a general convergence analysis. Furthermore, in order to solve nonlinear problems numerically, we propose a combination of the iterative linearization approach and the classical Galerkin discretization method, thereby giving rise to the so-called iterative linearization Galerkin (ILG) methodology. Moreover, still on an abstract level, based on two different elliptic reconstruction techniques, we derive a posteriori error estimates which separately take into account the discretization and linearization errors. Furthermore, we propose an adaptive algorithm, which provides an efficient interplay between these two effects. Finally, our abstract theory and the performance of the adaptive ILG approach are illustrated by means of a finite element discretization scheme for various examples of quasilinear stationary conservation laws.

1. Introduction

The aim of this paper is to establish a general (adaptive) iterative linearization Galerkin (ILG) framework for the numerical solution of nonlinear problems, with application to stationary conservation laws. To set the stage, we consider two continuously and densely embedded real Hilbert spaces $X \hookrightarrow H$, with inner products $(\cdot,\cdot)_X$ and $(\cdot,\cdot)_H$, and induced norms denoted by $\|\cdot\|_X$ and $\|\cdot\|_H$, respectively. Then, given a nonlinear operator $F : X \to X^*$, we focus on the equation

$$u \in X : \quad F(u) = 0 \quad \text{in } X^*,$$

where $X^*$ denotes the dual space of $X$. In weak form, this problem reads

$$u \in X : \quad (F(u),v)_X \times X = 0 \quad \text{for all } v \in X,$$

with $(\cdot,\cdot)_{X^* \times X}$ signifying the duality pairing in $X^* \times X$.

Iterative linearization. The development of an iterative linearization scheme for (1) is based on applying a suitable preconditioning operation. More precisely, for given $v \in X$, we introduce a linear and invertible preconditioning operator

$$A[v] : X \to X^*,$$

which allows to transform (1) into $A[v]^{-1}F(u) = 0$. This in turn gives rise to a fixed point iteration

$$u^{n+1} = u^n - A[u^n]^{-1}F(u^n), \quad n \geq 0,$$

for an initial guess $u^0 \in X$, or equivalently,

$$u^{n+1} \in X : \quad A[u^n]u^{n+1} = A[u^n]u^n - F(u^n), \quad n \geq 0.$$
Letting
\[ f : X \to X^*, \quad f(u) := A[u]u - F(u), \]
the fixed point iteration (4) takes the form of the following *iterative linearization scheme*:
\[ A[u^n]u^{n+1} = f(u^n), \quad n \geq 0. \]

We emphasize that, given \( u^n \in X \), this is a *linear* problem for \( u^{n+1} \in X \).

The general iteration scheme (6) recovers some of the widely used fixed point iterations occurring in the literature. These include, for instance, the Zarantonello iteration, the Kačanov scheme, and the Newton method. In the context of the Zarantonello iteration, the interested reader is referred to the original work [26] (cf. also [7] for a generalization), or the monographs [22, §3.3] and [29, §25.4]. Incidentally, the latter two references also deal with the Kačanov approach, see [22, §4.5] or [29, §25.14]. For the (damped and adaptive) Newton method we refer to [10] for an extensive overview, or the recent works on adaptive Newton schemes [3, 4, 18, 23, 24].

**Iterative linearized Galerkin approach.** The iteration (6) generates a sequence \( \{u^n\}_{n \geq 0} \) which potentially converges to a solution \( u^* \in X \) of (1). In general, however, the computation of this sequence is not feasible if \( X \) is infinite- or high-dimensional. Therefore, in order to cast the iterative linearization approach described above into a computational framework, we will consider Galerkin discretizations of (6) in terms of finite-dimensional conforming subspaces \( X \). Letting the discrete system
\[ u^{n+1}_N \in X_N : \quad \langle A[u^n_N]u^{n+1}_N, v \rangle_{X_N \times X_N} = \langle f(u^n_N), v \rangle_{X_N \times X_N} \quad \forall v \in X_N, \quad n \geq 0. \]

For the resulting sequence \( \{u^n_N\}_{n \geq 0} \subset X_N \) of discrete solutions it is possible, under certain conditions, to obtain general a posteriori estimates for the difference to the exact solution, \( u^* \in X \), i.e. for \( \|u^* - u^{n+1}_N\|_X, n \geq 0 \). The emphasis of such bounds is that they enable the individual identification of different sources of error in the approximation process, such as, e.g., the linearization and discretization errors (further errors, not to be considered here, may result, for instance, from a linear solver iteration, see, e.g., [13], or from quadrature). This can be accomplished by means of two conceptually different techniques, both of which will be presented in this work:

(a) The first approach is based on the assumption that a computable bound for the residual of the *linear* Galerkin discretization of the form (7) is available. Then, applying an elliptic reconstruction technique (see, e.g., [19, 21]) yields a computable a posteriori error estimate for the error \( \|u^* - u^{n+1}_N\|_X \), which can be expressed in terms of a discretization and linearization contribution. We note that this approach has been applied previously in [9] in the specific context of the Zarantonello iteration scheme.

(b) Alternatively, we may consider, for \( n \geq 0 \), a *nonlinear* discrete problem which, on the one hand, features the *nonlinear* operator \( F \) from (1), and, on the other hand, possesses the same solution, \( u^{n+1}_N \in X_N \), as the *linear* Galerkin formulation (7). Assuming that it exists a computable bound for the residual of the discrete solution to a suitably reconstructed nonlinear problem, our analysis will show that such a bound can be exploited for the purpose of deriving an a posteriori error estimator.

A posteriori error estimates as outlined above constitute an essential building block in the development of adaptive ILG schemes for nonlinear problems (1). Indeed, recalling that such bounds allow to distinguish the different sources of error in the approximation process, the key idea of the fully adaptive ILG methodology is to provide an appropriate *interplay* between the fixed point linearization iteration and possible Galerkin space enrichments (e.g., mesh refinements for finite elements) depending on whether the discretization error or the linearization error is dominant. In this way, the goal of the adaptive ILG approach is to keep the number of fixed point iterations at a minimum in the sense that no unnecessary iterations are performed if they are not expected to contribute a substantial reduction of the error on the actual Galerkin space.

The simultaneous control of different sources of error in the context of adaptive finite element methods for monotone problems has been presented in a number of earlier papers. For instance, in the work [8], the authors have considered general linearizations of strongly monotone operators,
and have derived computable a posteriori estimators for the total error (consisting of the linearization error and the Galerkin error) with identifiable components for each of the error sources. Moreover, in the more specific context of the Newton linearization scheme, we refer to the papers [12] (see also [13]) on second-order monotone quasilinear diffusion, where a posteriori error estimates including—in addition to the discretization and linearization errors—also the iterative linear solver error have been derived (more precisely, an inexact Newton iteration scheme has been presented); moreover, the authors have proposed an adaptive iterative procedure, which takes into account all components of the numerical scheme in each refinement step. Furthermore, first a posteriori error estimates in the framework of the Kačanov iteration for quasilinear diffusion problems in divergence form have been presented in [17]. Later on, an adaptive iterative linearized Galerkin type approach has been introduced and discussed in [16]; indeed, the convergence of the Kačanov-Galerkin iteration is proved therein. Moreover, for semilinear second-order elliptic problems, two different linearization schemes of Kačanov type have been analyzed in [6]. Just recently, based on the ILG approach in [9], the convergence of an adaptive Zarantonello-Galerkin iterative scheme for monotone elliptic PDE has been proved in [15]. Finally, we point to the fact that the ILG methodology has been applied also to high-order (so-called hp) [2] and discontinuous Galerkin [18] finite element discretizations, as well as to nonlinear parabolic problems [5].

Outline of the paper. In Section 2 we state and prove a global convergence result for the unified iteration scheme (6). In particular, in order to provide a few examples, we apply our result to the Zarantonello, Kačanov, and (damped) Newton methods, thereby recovering some of the well-known convergence results from the literature. Furthermore, still on an abstract level, in Section 3 we discuss conforming Galerkin discretizations of (6), and present general a posteriori error estimates based on the two approaches outlined in (a) and (b) above. On that account, we propose in Section 4 a fully adaptive algorithm based on the a posteriori error estimates. More specifically, in Section 5, we derive computable error bounds for a second-order stationary diffusion-reaction conservation law; finally, in Section 5.3, these theoretical estimates are employed within a series of numerical experiments in the framework of the fully adaptive ILG approach.

2. Iterative linearization

The goal of this section is to prove a general convergence result for the iterative linearization iteration (6) under the condition that \( F \) in (1) is a Lipschitz continuous and strongly monotone operator. Furthermore, we will review a few classical examples.

2.1. Abstract framework. For the purpose of this work, we restrict ourselves to Lipschitz continuous, strongly monotone operators \( F \):

(F1) The operator \( F \) is Lipschitz continuous, i.e. it exists a constant \( L_F > 0 \) such that
\[
|\langle F(u) - F(v), w \rangle_{\mathcal{X} \times \mathcal{X}}| \leq L_F \|u - v\|_{\mathcal{X}} \|w\|_{\mathcal{X}},
\]
for all \( u, v, w \in \mathcal{X} \).

(F2) The operator \( F \) is strongly monotone, i.e. there is a constant \( \nu > 0 \) such that
\[
\nu \|u - v\|^2_{\mathcal{X}} \leq \langle F(u) - F(v), u - v \rangle_{\mathcal{X} \times \mathcal{X}},
\]
for all \( u, v \in \mathcal{X} \).

Under these conditions, the theory of strongly monotone operators implies that (1) possesses a unique solution \( u^* \in \mathcal{X} \); see, e.g., [22, §3.3] or [29, §25.4].

Furthermore, for given \( u \in \mathcal{X} \), we introduce the bilinear form
\[
a(u; v, w) := \langle A[u]v, w \rangle_{\mathcal{X} \times \mathcal{X}}, \quad v, w \in \mathcal{X},
\]
where \( A[\cdot] \) is the preconditioning operator from (3). Then, we can write (6) in weak form: given \( u^n \in \mathcal{X} \), find \( u^{n+1} \in \mathcal{X} \) such that
\[
a(u^n; u^{n+1}, w) = \langle f(u^n), w \rangle_{\mathcal{X} \times \mathcal{X}} \quad \forall w \in \mathcal{X}.
\]
Throughout this paper, for any \( u \in X \), we assume that the bilinear form \( a(u; \cdot, \cdot) \) is symmetric, and uniformly coercive and bounded. The latter two assumptions refer to the fact that there are two constants \( \alpha, \beta > 0 \) independent of \( u \in X \), such that
\[
a(u; v, v) \geq \alpha \|v\|_X^2 \quad \forall v \in X, \tag{10}
\]
and
\[
a(u; v, w) \leq \beta \|v\|_X \|w\|_X \quad \forall v, w \in X, \tag{11}
\]
respectively. In particular, owing to the Lax-Milgram Theorem, these properties imply the well-posedness of the solution \( u^{n+1} \in X \) of the linear equation (6), for any given \( u^n \in X \).

2.2. A global convergence result. Given the framework introduced in the previous Section 2.1, the ensuing proposition is an abstract global convergence result for the iteration scheme (6).

**Proposition 2.1.** Suppose that (F2), (10) and (11) are satisfied, and \( u \mapsto a(u; u, \cdot) \) and \( u \mapsto f(u) \) are continuous mappings from \( X \) into its dual space \( X^* \) with respect to the weak topology on \( X^* \). If the sequence \( \{u^n\}_{n \geq 0} \) defined by (6) satisfies \( \|u^{n+1} - u^n\|_X \to 0 \) as \( n \to \infty \), then it converges to the unique solution \( u^* \in X \) of (1).

**Proof.** We begin by showing that \( \{u^n\}_{n \geq 0} \) is a Cauchy sequence. Indeed, by virtue of (F2) and (5), for any \( m \geq n \geq 0 \), it holds that
\[
\nu \|u^m - u^n\|_X^2 \leq \langle F(u^m) - F(u^n), u^m - u^n \rangle_{X^* \times X} = \langle A[u^m]u^m - f(u^m), u^m - u^n \rangle_{X^* \times X} - \langle A[u^n]u^n - f(u^n), u^m - u^n \rangle_{X^* \times X}.
\]
Hence, involving (8) and (9) gives
\[
\nu \|u^m - u^n\|_X^2 = a(u^m; u^m - u^{m+1}, u^m - u^n) - a(u^n; u^n - u^{n+1}, u^m - u^n).
\]
Furthermore, (11) implies that
\[
\|u^m - u^n\|_X \leq \frac{\beta}{\nu} \left( \|u^{m+1} - u^m\|_X^2 + \|u^{n+1} - u^n\|_X^2 \right) \to 0,
\]
for \( n, m \to \infty \). Hence, \( \{u^n\}_{n \geq 0} \) is a Cauchy sequence, and, therefore, converges to some limit \( u^* \in X \). Next, we show that \( u^* \) is the unique solution of (1). Owing to (9), we notice the identity
\[
a(u^n; u^n, v) - \langle f(u^n), v \rangle_{X^* \times X} + a(u^n; u^{n+1} - u^n, v) = 0 \quad \forall v \in X,
\]
for all \( n \geq 0 \). Here, due to (11), and because \( \|u^{n+1} - u^n\|_X \) is a null sequence, we observe that
\[
a(u^n; u^n, v) - \langle f(u^n), v \rangle_{X^* \times X} \to 0 \quad \forall v \in X,
\]
for \( n \to \infty \). Thence, by continuity of \( a \) and \( f \), we deduce that
\[
a(u^*; u^*, v) = \langle f(u^*), v \rangle_{X^* \times X} \quad \forall v \in X,
\]
i.e. \( u^* \) is a solution of (1). It remains to show that \( u^* \) is the only solution of (1). In fact, if \( u^\Box \in X \) is any other solution, then (F2) leads to
\[
\nu \|u^* - u^\Box\|_X^2 \leq \langle F(u^*), F(u^\Box) - F(u^*), u^* - u^\Box \rangle_{X^* \times X} = 0,
\]
i.e. \( u^* = u^\Box \). \( \square \)

2.3. Applications. In the ensuing section we will discuss the general Proposition 2.1 in the context of the Zarantonello, Kačanov, and Newton iterations.
2.3.1. Zarantonello iteration. A most simple choice for the preconditioning operator from (3) is $A[v]u := (\delta^{-1}u, \cdot)_X$, where $\delta > 0$ is a fixed constant; in particular, here, $A = A[v]$ is independent of $v$. In this case, the iterative linearization scheme (6) turns out to be

$$(u^{n+1}, \cdot)_X = (u^n, \cdot)_X - \delta \langle F(u^n), \cdot \rangle_{X^* \times X}. \quad (12)$$

**Theorem 2.2** (Convergence of the Zarantonello iteration). Assuming (F1) and (F2), the Zarantonello iteration (12) converges to the unique solution $u^*$ of (1) for any $\delta \in [0, 2\nu/L]$. Proof. We verify the assumptions required for Proposition 2.1 to hold. For $a(u, v) = (\delta^{-1}u, v)_X$, $u, v \in X$, we note that (10) and (11) are satisfied with

$$\alpha = \beta = \delta^{-1} > 0. \quad (13)$$

Moreover, both $u \mapsto a(u, \cdot) = (\delta^{-1}u, \cdot)_X$ and $u \mapsto \langle f(u), \cdot \rangle_{X^* \times X} = (\delta^{-1}u, \cdot)_X - \langle F(u), \cdot \rangle_{X^* \times X}$ are continuous on $X$. It remains to show that $\|u^{n+1} - u^n\|^2_X$ is a null sequence. For that purpose, we denote by $J : X \to X^*$ the Riesz-Frèchet isometry. The iteration (12) can then be written, in strong form, as $u^{n+1} = T(u^n)$, where $T(u) := u - \delta J^{-1}F(u)$. This leads to

$$\|u^{n+1} - u^n\|^2_X = \|T(u^n) - T(u^{n-1})\|^2_X$$

$$= \|u^n - u^{n-1}\|^2_X - 2\delta \langle F(u^n) - F(u^{n-1}), u^n - u^{n-1}\rangle_{X^* \times X}$$

$$+ \delta^2 \|J^{-1}(F(u^n) - F(u^{n-1}))\|^2_X,$$

where we have used the linearity of $J^{-1}$. Invoking (F1) and (F2), together with the fact that $J^{-1}$ is isometric, we further get

$$\|u^{n+1} - u^n\|^2_X \leq (1 - 2\delta \nu + \delta^2 L^2_F) \|u^n - u^{n-1}\|^2_X.$$ We note that

$$\gamma := (1 - 2\delta \nu + \delta^2 L^2_F) < 1 \quad (14)$$

if and only if $\delta \in [0, 2\nu/L]$. Hence, by induction,

$$\|u^{n+1} - u^n\|^2_X \leq \gamma^n \|u^1 - u^0\|^2_X,$$

which shows that $\|u^{n+1} - u^n\|^2_X \to 0$ as $n \to 0$. \qed

**Remark 2.3.** We notice that the contraction factor $\gamma$ from (14) is minimal for the choice $\delta = \nu/L_F$.  

2.3.2. Kačanov iteration. Here we assume that the nonlinear operator $F$ from (1) takes the form $F(u) = A[u]u - g$, where $A[u] : X \to X^*$ is linear (for given $u \in X$), and $g = F(0) \in X^*$ is fixed. Then, the Kačanov iteration is defined by

$$A[u^n]u^{n+1} = g, \quad n \geq 0. \quad (15)$$

Note that this iteration can be cast into the setting of (6), where $A[u^n]$ takes the role of the preconditioning operator, and $f(u^n) = A[u^n]u^n - F[u^n] = g$ is constant. We make the assumption that there exists a Gâteaux differentiable functional $G : X \to \mathbb{R}$ which, for any $u, v \in X$, satisfies the following properties:

(K1) $G'(u) = a(u; u, \cdot)$ on $X$, and $G'$ is continuous and strongly monotone, i.e. there exists a real number $c_0 > 0$ such that, for any $u, v \in X$, it holds

$$\langle G'(u) - G'(v), u - v \rangle_{X^* \times X} \geq c_0 \|u - v\|^2_X; \quad (16)$$

(K2) $G(u) \geq h(u) + h(v)$ for all $u, v \in X$, where $h$ is a convex function.

In order to be able to apply Proposition 2.1, we need an auxiliary result, which will also be crucial in the analysis of the Newton method in Section 2.3.3 below.

**Lemma 2.4.** If $H : X \to \mathbb{R}$ is Gâteaux differentiable with $H'$ continuous and strongly monotone, then $H$ is bounded from below.
Proof. For fixed \( v \in X \), and \( t \in [0,1] \), we define the function \( \varphi(t) := H(tv) \). We note that \( \varphi'(t) = \langle H'(tv), v \rangle_{X^* \times X} \), and, invoking the fundamental theorem of calculus, we find that

\[
H(v) - H(0) = \int_0^1 \langle H'(tv), v \rangle_{X^* \times X} \, dt = \int_0^1 \langle H'(tv) - H'(0), v \rangle_{X^* \times X} \, dt + \langle H'(0), v \rangle_{X^* \times X}.
\] (17)

Since \( H' \) is strongly monotone, there exists a constant \( \gamma > 0 \) such that

\[
\langle H'(tv) - H'(0), v \rangle_{X^* \times X} = \frac{1}{t} \langle H'(tv) - H'(0), tv \rangle_{X^* \times X} \geq \gamma t \| v \|^2_X,
\]

for any \( t \in [0,1] \). Inserting this bound into (17), integrating with respect to \( t \), and using the sub-multiplicativity of the operator norm, yields

\[
H(v) \geq \frac{\gamma}{2} \| v \|^2_X - \| H(0) \|_X \| v \|_X + H(0).
\]

It is elementary to verify that the right-hand side is minimal for \( \| v \|_X = (H'(0))_X^{-1} \| H'(0) \|_X \). With this choice we arrive at \( H(v) \geq H(0) - \frac{1}{2} \gamma \| H'(0) \|_X^2 \) for all \( v \in X \), i.e., \( H \) is bounded from below. \( \square \)

**Theorem 2.5 (Convergence of the Kacanov iteration).** Suppose that \((K1)\) and \((K2)\) hold. Furthermore, assume that \( A \) satisfies the properties from Section 2.1. Then the sequence \( \{u^n\}_{n \geq 0} \) defined by (15) converges to the unique solution \( u^* \) of (1).

**Proof.** Because of \((K1)\) it follows that \( u \mapsto a(u; u, \cdot) = G'(u) \) is continuous; moreover, \( u \mapsto f(u) \) is constant, and thus continuous as well. We show that \( \| u^{n+1} - u^n \|_X \) is a null sequence.

To this end, we follow closely along the lines of the proof of [29, Theorem 25.1]. Let us introduce the functional \( H(u) := G(u) - \langle g, u \rangle_{X \times X} \). We note that \( H'(u) = G'(u) - g = A[u]u - g = F(u) \), i.e., \( H \) is the potential of \( F \). Moreover, by virtue of \((K1)\), the derivative \( H' = G' - g \) is continuous and strongly monotone, and thus \( F \) satisfies \((F2)\). In particular, with the aid of Lemma 2.4, we deduce that \( H \) is bounded from below. Next, we will verify that \( \{H(u^n)\}_{n \geq 0} \) is a monotone decreasing sequence. Indeed, noticing that \( a(u^n; u^{n+1}, u^{n+1} - u^n) = \langle g, u^{n+1} - u^n \rangle_{X^* \times X} \), and employing \((K2)\), yields:

\[
H(u^n) - H(u^{n+1}) = \langle g, u^{n+1} - u^n \rangle_{X^* \times X} + G(u^n) - G(u^{n+1})
\]

\[
\geq a(u^n; u^{n+1}, u^{n+1} - u^n) + \frac{1}{2} a(u^n; u^n) - \frac{1}{2} a(u^n; u^{n+1}, u^{n+1})
\]

\[
\geq \frac{1}{2} a(u^n; u^n, u^n) - a(u^n; u^{n+1}, u^n) + \frac{1}{2} a(u^n; u^{n+1}, u^{n+1}),
\]

for any \( n \geq 0 \). Then, employing the symmetry of \( a(u^n; \cdot, \cdot) \), and involving \((10)\), we obtain

\[
H(u^n) - H(u^{n+1}) \geq \frac{1}{2} a(u^n; u^{n+1} - u^n, u^{n+1} - u^n) \geq \frac{\alpha}{2} \| u^{n+1} - u^n \|_X^2 \geq 0,
\]

which shows that \( \{H(u^n)\}_{n \geq 0} \) is monotone decreasing. Then, recalling the boundedness from below, we conclude that \( H(u^n) - H(u^{n+1}) \to 0 \) as \( n \to \infty \). Hence, exploiting \((18)\), it follows that \( \| u^{n+1} - u^n \|_X \) is a null sequence, and the proof is complete. \( \square \)

**2.3.3. Newton iteration.** For the Newton iteration the preconditioning operator in (4) is selected to be \( A[v] = \delta(v)^{-1} F'(v) \), \( v \in X \), where \( \delta(v) > 0 \) is a (damping) parameter, and \( F' \) signifies the Gâateaux derivative of \( F \). Then, the (damped) Newton iteration is given by

\[
F'(u^n)u^{n+1} = F'(u^n)u^n - \delta(u^n)F(u^n), \quad n \geq 0.
\] (19)

For the purpose of applying Proposition 2.1, we make the following assumptions:

\( \text{(N1)} \) The operator \( F \) is Gâateaux differentiable, with

\[
\langle F'(u)v, w \rangle_{X^* \times X} = \langle F'(u)w, v \rangle_{X^* \times X} \quad \forall u, v, w \in X.
\] (20)

Moreover, \( F' \) is coercive and bounded in the sense that, for any given \( u \in X \), it holds

\[
\langle F'(u)v, v \rangle_{X^* \times X} \geq \alpha_F \| v \|^2_X \quad \forall v \in X,
\] (21)

and

\[
\langle F'(u)v, w \rangle_{X^* \times X} \leq \beta_F \| v \|_X \| w \|_X \quad \forall v, w \in X,
\] (22)
where \( \alpha_F, \beta_F > 0 \) are independent of \( u \).

(N2) It exists a Gâteaux differentiable functional \( G : X \to \mathbb{R} \) such that \( G'(u) = F'(u)u \) in \( X^* \) for any \( u \in X \), and \( G' \) is continuous when \( X^* \) is endowed with the weak topology.

(N3) It exists a Gâteaux differentiable functional \( H : X \to \mathbb{R} \) such that \( H' = F \). We note that, if \( F \) is Fréchet differentiable, then the existence of \( H \) follows from (N1); see, e.g., [27, Proposition 41.5].

(N4) There are some constants \( 0 < \delta_{\min} \leq \delta_{\max} < \infty \) such that \( \delta : X \to [\delta_{\min}, \delta_{\max}] \) is a continuous functional.

**Theorem 2.6** (Convergence of the damped Newton iteration). Assume (F1) and (F2), as well as (N1)-(N4). Then, for \( \delta_{\max} < 2\alpha_F/L_F \) in (N4) the damped Newton iteration (19) converges to the unique solution \( u^* \in X \) of (1).

**Proof.** We aim at employing Proposition 2.1 as before. To do so, notice first that \( a(u; v, w) = \delta(u)^{-1} (F'(u)u, F'(u)w)_{X^*, X} \), for \( u, v, w \in X \). Hence, by (20), we deduce the symmetry property of the bilinear form \( a(u, \cdot, \cdot) \), for \( u \in X \). Furthermore, by virtue of (21), (22), and (N4), we obtain

\[
a(u; v, v) \geq \alpha_F \delta^{-1}_{\max} \|v\|_X^2, \quad u, v \in X,
\]

and

\[
a(u; v, w) \leq \beta_F \delta^{-1}_{\min} \|v\|_X \|w\|_X, \quad u, v, w \in X,
\]

which are the coercivity and boundedness conditions (10) and (11), with

\[
\alpha = \alpha_F/\delta_{\min}, \quad \beta = \beta_F/\delta_{\min},
\]

respectively. Next, we remark that the maps \( u \mapsto a(u; u, \cdot) = \delta(u)^{-1} F'(u)u \) and \( u \mapsto f(u) = \delta(u)^{-1} F(u) - F(u) \) are both continuous, when \( X^* \) is endowed with the weak topology, by (N2) and (N4), and by (N2), (N4) and (F1), respectively. Therefore, by the same arguments as in the proof of Theorem 2.5, it suffices to show that it exists a constant \( C > 0 \) such that

\[
H(u^n) - H(u^{n+1}) \geq C \|u^{n+1} - u^n\|_X^2, \quad n \geq 0.
\]

(24)

To this end, we define the function \( \varphi(t) := H(u^n + t(u^{n+1} - u^n)) \), \( t \in [0, 1] \), and observe that

\[
\varphi'(t) = \langle H'(u^n + t(u^{n+1} - u^n)), u^{n+1} - u^n \rangle_{X^* \times X} = \langle F(u^n + t(u^{n+1} - u^n)), u^{n+1} - u^n \rangle_{X^* \times X}.
\]

Then, the fundamental theorem of calculus implies that

\[
H(u^n) - H(u^{n+1}) = - \int_0^1 \langle F(u^n + t(u^{n+1} - u^n)), u^{n+1} - u^n \rangle_{X^* \times X} \, dt
\]

\[
= - \int_0^1 \langle F(u^n + t(u^{n+1} - u^n)) - F(u^n), u^{n+1} - u^n \rangle_{X^* \times X} \, dt
\]

\[
- \langle F(u^n), u^{n+1} - u^n \rangle_{X^* \times X}.
\]

By the definition of the Newton iteration (19), it holds that \( F(u^n) = \delta(u^n)^{-1} F'(u^n)(u^n - u^n) \), \( n \geq 0 \). Thus, with the aid of (F1) and (10), it follows that

\[
H(u^n) - H(u^{n+1}) \geq - L_E \int_0^1 \|u^{n+1} - u^n\|_X^2 \, dt + \delta(u^n)^{-1} \langle F'(u^n)(u^{n+1} - u^n), u^{n+1} - u^n \rangle_{X^* \times X}
\]

\[
\geq - \frac{L_E}{2} \|u^{n+1} - u^n\|_X^2 + \alpha_F \delta(u^n)^{-1} \|u^{n+1} - u^n\|_X^2.
\]

If \( \delta(u^n) \leq \delta_{\max} < 2\alpha_F/L_E \), then

\[
\frac{\alpha_F}{\delta(u^n)} = \frac{L_E}{2} \geq \frac{\alpha_F}{\delta_{\max}} - \frac{L_E}{2} =: C > 0, \quad n \geq 0.
\]

(25)

We conclude that (24) is satisfied. \( \square \)
Remark 2.7 (Classical Newton scheme). Recalling (21) with \( u = u^0 \), and applying [22, Theorem 3.3.23], we deduce the bound \( \| F'(u^0)^{-1}v \|_X \leq \alpha_{F'}^{-1} \| v \|_X \), for all \( v \in X^* \). Furthermore, assume that \( F \) is Fréchet differentiable and \( F' \) is Lipschitz continuous, i.e. there exists a constant \( L_{F'} > 0 \) such that
\[
\| F'(v) - F'(w) \|_X \leq L_{F'} \| v - w \|_X \quad \forall v, w \in X.
\]
This leads to
\[
\| F'(u^0)^{-1}(F'(u) - F'(v)) \|_X \leq L_{F'} \| u - v \|_X \quad \forall u, v \in X.
\]
Moreover, if the initial guess \( u^0 \in X \) in (19) is sufficiently close to the solution \( u^* \in X \) of (1) in the sense that \( \| F(u^0) \|_X < \alpha_{F'} / 2L_{F'} \), then we infer that
\[
\| F'(u^0)^{-1}F(u^0) \|_X \leq \frac{1}{\alpha_{F'}} \| F(u^0) \|_X < \frac{\alpha_{F'}}{2L_{F'}}.
\]
Referring to [10, Theorem 2.1], it follows that the classical Newton iteration with \( \delta(u^n) = 1 \) in (19) is well-defined, converges to a solution of (1), and converges quadratically.

Remark 2.8. The proof of Theorem 2.6 is crucially based on (24). We emphasize that this bound may be satisfied even if the damping parameter \( \delta(u^n) \) in (19) is larger than \( 2\alpha_{F'}/L_{F'} \). This is particularly important when \( 2\alpha_{F'}/L_{F'} \leq 1 \) and the choice \( \delta(u^n) = 1 \) (leading to local quadratic convergence, cf. Remark 2.7) is not admissible a priori. In this case, we may fix \( \epsilon > 0 \) small, and aim to a posteriori attain the bound, for \( n \geq 0 \),
\[
H(u^n) - H(u^{n+1}) \geq \epsilon \left\| u^{n+1} - u^n \right\|_X^2.
\]
To this end, we may pursue, for instance, the adaptive damping parameter selection approach proposed in [10, §3.1]. More precisely, in each iterative step, we define an initial value for \( \delta(u^n) \) by the following prediction strategy:
\[
\delta^{n,0} = \begin{cases} 
\min \left( \frac{\delta(u^{n-1})}{\kappa}, 1 \right) & \text{if } \delta(u^{n-2}) \leq \delta(u^{n-1}), \\
\delta(u^{n-1}) & \text{else},
\end{cases}
\]
where \( 0 < \kappa < 1 \) is a fixed (correction) factor. Here, we set \( \delta(u^{-2}) = \delta(u^{-1}) = \delta^0 \), with \( \delta^0 \) an initial choice. If \( u^{n+1} \) is obtained by the damped Newton method with damping parameter \( \delta^{n,i} \), for some \( i \geq 0 \), then we need to verify whether or not (26) is satisfied. If not, then we adjust the damping parameter according to the correction strategy
\[
\delta^{n,i+1} = \max \left( \alpha_{F'} (\epsilon + L_{F'}/2)^{-1}, \kappa \delta^{n,i} \right), \quad i \geq 0.
\]
Subsequently, we will compute \( u^{n+1} \) for the new choice \( \delta^{n,i+1} \). This process is repeated until (26) is true, say after \( i^n \) iterations of (27). At this point, we let \( \delta(u^n) := \delta^{n,i^n} \). Evidently, in view of (25), we remark that (26) will certainly hold once \( \delta^{n,i} \leq \alpha_{F'} (\epsilon + L_{F'}/2)^{-1} \).

3. Galerkin approach and a posteriori error analysis

The numerical solution of (1) is based on a finite-dimensional subspace \( X_N \subset X \), and on the iterative linearization Galerkin (ILG) formulation (7), with a given initial guess \( u^0_N \in X_N \). Since \( X_N \subset X \), the assumptions in Section 2.1 guarantee the existence of \( u^{n+1}_N \in X_N \) in each iteration step.

In this section, we will pursue two different strategies for the derivation of a posteriori error estimates for \( \| u^* - u^{n+1}_N \|_X \), where \( u^* \in X \) is the unique solution of (1). In both approaches an elliptic reconstruction technique, cf. [19, 21], will be employed. In the first method we use an elliptic reconstruction for the solution of the linear problem (7), and the second strategy is based on applying a similar idea for a nonlinear discrete problem equivalent to (7). We will refer to this methods as the linear and nonlinear elliptic reconstruction, respectively.
3.1. A posteriori error analysis based on a linear elliptic reconstruction. For the sake of a general a posteriori error analysis, using a linear elliptic reconstruction, we suppose that it exists a computable bound $\eta(u_{n+1}^\star, u_N)$ for the residual

$$\sup_{v \in X} \| a(u_N^\star; u_{n+1}^n, v) - (f(u_N^\star), v) \|_{X^* \times X} \leq \eta(u_{n+1}^\star, u_N^n).$$

(28)

We remark that, in the context of finite element methods for linear elliptic problems, there is a large body of literature focusing on the development of such estimates; see, e.g., [1, 25].

Let us point out that the above result permits to bound the error $\| u^\star - u_{n+1}^n \|_X$ separately in terms of the discretization error estimator $\beta/\alpha \eta(u_{n+1}^\star, u_N^n)$ and of the linearization error estimator $\eta(u_{n+1}^\star - u_{n+1}^n, u_N^n)$.

**Theorem 3.1.** Suppose that (F1) and (F2), as well as (10) and (11) hold true. Then, we have the a posteriori error bound

$$\| u^\star - u_{n+1}^n \|_X \leq \frac{\beta}{\alpha \nu} \eta(u_{n+1}^\star, u_N^n) + \frac{\beta + L_F}{\nu} \| u_{n+1}^n - u_N^n \|_X,$$

where $u^\star$ is the unique solution of (1).

**Proof.** Due to (10) and (11) it exists a unique $\tilde{u}_{n+1}^\star \in X$ such that

$$a(u_N^n; \tilde{u}_{n+1}^\star, v) = (f(u_N^n), v)_{X^* \times X} \quad \forall v \in X.$$

(29)

We note that $\tilde{u}_{n+1}^\star$ is a reconstruction in the sense that $u_{n+1}^\star \in X_N$ is the Galerkin projection of $\tilde{u}_{n+1}^\star$. By using the assumption (F2), we find that

$$\nu \| u^\star - u_{n+1}^n \|_X^2 \leq \langle F(u^\star) - F(u_{n+1}^\star), u^\star - u_{n+1}^n \rangle_{X^* \times X} = - \langle F(u_{n+1}^\star), u^\star - u_{n+1}^n \rangle_{X^* \times X}.$$

since $u^\star \in X$ is the solution of (1). Hence,

$$\nu \| u^\star - u_{n+1}^n \|_X^2 \leq - a(u_N^n; u_{n+1}^\star, u^\star - u_{n+1}^n) + \langle f(u_N^n), u^\star - u_{n+1}^n \rangle_{X^* \times X} + a(u_N^n; u_{n+1}^n - u_N^n, u^\star - u_{n+1}^n) + a(u_N^n; u_N^n, u^\star - u_{n+1}^n) - \langle f(u_N^n), u^\star - u_{n+1}^n \rangle_{X^* \times X}.$$

Using (29) and (5), this estimate transforms into

$$\nu \| u^\star - u_{n+1}^n \|_X^2 \leq a(u_N^n; u_{n+1}^n - u_N^n, u^\star - u_{n+1}^n) + a(u_N^n; u_{n+1}^n - u_N^n, u^\star - u_{n+1}^n) - \langle F(u_{n+1}^\star), u^\star - u_{n+1}^n \rangle_{X^* \times X}.$$

Applying (11) and (F1), we find that

$$\nu \| u^\star - u_{n+1}^n \|_X^2 \leq \beta \| u_{n+1}^n - u_N^n \|_X \| u^\star - u_{n+1}^n \|_X + \beta \| u_{n+1}^n - u_N^n \|_X \| u^\star - u_{n+1}^n \|_X.$$

Dividing by $\| u^\star - u_{n+1}^n \|_X$ yields

$$\nu \| u^\star - u_{n+1}^n \|_X \leq \beta \| u_{n+1}^n - u_N^n \|_X + (\beta + L_F) \| u_{n+1}^n - u_N^n \|_X.$$

(30)

Moreover, by the coercivity property (10), for $u_{n+1}^n \neq \tilde{u}_{n+1}^\star$, we note that

$$\alpha \| u_{n+1}^n - \tilde{u}_{n+1}^\star \|_X \leq \sup_{v \in X} \frac{a(u_N^n; u_{n+1}^n - \tilde{u}_{n+1}^\star, v)}{\| u_{n+1}^n - \tilde{u}_{n+1}^\star \|_X} \leq \sup_{v \in X} \frac{a(u_N^n; u_{n+1}^n, v)}{\| u_{n+1}^n \|_X} \leq \| u_{n+1}^n \|_X.$$

Involving (29) and (28), we arrive at

$$\alpha \| u_{n+1}^n - \tilde{u}_{n+1}^\star \|_X \leq \sup_{v \in X} \frac{a(u_N^n; u_{n+1}^n, v) - \langle f(u_N^n), v \rangle_{X^* \times X}}{\| u_{n+1}^n \|_X} \leq \eta(u_{n+1}^n, u_N^n).$$

Inserting this estimate into (30), finishes the proof. □
Remark 3.2. We mention that the constants for the estimator in Theorem 3.1 for the Zarantonello iteration can be slightly improved. This is due to the fact that the preconditioning operator A is constant in this case; cf. [9, Proposition 2.2].

3.2. A posteriori error analysis based on a nonlinear elliptic reconstruction. In this section we devise an a posteriori error estimate for the linear Galerkin iteration (7) based on applying the reconstruction technique to a nonlinear discrete problem equivalent to (7). We underline that our approach does not require the computation of a solution to a nonlinear problem in practice.

We define an operator \( \psi_N : X \to X_N \), where, for fixed \( w \in X \), we let \( \psi_N(w) \) to be the Riesz representative of \( F(w) \) with respect to the inner product in \( H \), i.e.

\[
(\psi_N(w), v)_H = (F(w), v)_{X^* \times X} \quad \forall v \in X_N.
\]

(31)

Note that, if \( u_N \) is a solution of the nonlinear Galerkin approximation of (2) with respect to the discrete space \( X_N \), i.e.

\[
u \in X_N : \quad (F(u_N), v)_{X^* \times X} = 0 \quad \forall v \in X_N,
\]

(32)

then it holds that \( \psi_N(u_N) = 0 \).

For each \( n \geq 0 \), we define the nonlinear elliptic reconstruction of the solution \( u_{N+1} \in X_N \) of (7) by

\[
(F(\tilde{u}_{N+1}^n), v)_{X^* \times X} = (\psi_N(u_{N+1}^n), v)_H \quad \forall v \in X.
\]

(33)

By construction of the operator \( \psi_N \), it holds that \( u_{N+1}^n \) is the Galerkin approximation of (33), i.e.

\[
(F(\tilde{u}_{N+1}^n) - F(u_{N+1}^n), v)_{X^* \times X} = 0 \quad \forall v \in X_N.
\]

Then, with the aid of (F2), we infer that

\[
\nu \left\| \tilde{u}_{N+1}^n - u_{N+1}^n \right\|_X^2 \leq \left( F(\tilde{u}_{N+1}^n) - F(u_{N+1}^n) \right)_{X^* \times X} - u_{N+1}^n - \frac{1}{X^* \times X}.
\]

Hence,

\[
\nu \left\| \tilde{u}_{N+1}^n - u_{N+1}^n \right\|_X \leq \sup_{u \in X} \left( F(\tilde{u}_{N+1}^n) - F(u_{N+1}^n), w \right)_{X^* \times X}
\]

\[
\leq \sup_{u \in X} \left( (\psi_N(u_{N+1}^n), w)_H - (F(u_{N+1}^n), w)_{X^* \times X} \right).
\]

Now, suppose that it exists a computable bound \( \eta(u_{N+1}^n) \) such that

\[
\sup_{||w||_X = 1} \left\{ (\psi_N(u_{N+1}^n), w)_H - (F(u_{N+1}^n), w)_{X^* \times X} \right\} \leq \eta(u_{N+1}^n).
\]

Then,

\[
\left\| \tilde{u}_{N+1}^n - u_{N+1}^n \right\|_X \leq \nu^{-1} \eta(u_{N+1}^n).
\]

(34)

Theorem 3.3. Given (F1) and (F2), there holds the a posteriori error bound

\[
\left\| u^* - u_{N+1}^n \right\|_X \leq \frac{1}{\nu} \eta(u_{N+1}^n) + \frac{C_E}{\nu} \left\| \psi_N(u_{N+1}^n) \right\|_H,
\]

where \( u^* \) is the exact solution of (1), and \( C_E \) is the embedding constant of \( X \to H \).

We note that, in the bound above, \( \nu^{-1} \eta(u_{N+1}^n) \) is an estimator for the discretization error, and \( C_E \nu^{-1} \left\| \psi_N(u_{N+1}^n) \right\|_H \) controls the linearization error. Indeed, since \( \eta_N(u_N) = 0 \) and \( \{u_N^n\}_{n \geq 0} \) converges to the solution \( u_N \) of (32) by our analysis in Section 2, we see that \( \left\| \psi_N(u_{N+1}^n) \right\|_H \to 0 \) as \( n \to \infty \).
By invoking the triangle inequality and (34), we find that
\[ \| u^* - u_N^{n+1} \|_X \leq \nu^{-1} \eta(u_N^{n+1}) + \| u^* - \tilde{u}_N^{n+1} \|_X. \] (35)
Moreover, due to (F2), we observe that
\[ \nu \| u^* - \tilde{u}_N^{n+1} \|_X^2 \leq \langle F(u^*) - F(\tilde{u}_N^{n+1}), u^* - \tilde{u}_N^{n+1} \rangle_{X \times X} = \langle F(\tilde{u}_N^{n+1}), \tilde{u}_N^{n+1} - u^* \rangle_{X \times X}. \]
By using (33), and upon applying the Cauchy-Schwarz inequality, this leads to
\[ \nu \| u^* - \tilde{u}_N^{n+1} \|_X^2 \leq \| \psi_N(u_N^{n+1}) \|_H \| \tilde{u}_N^{n+1} - u^* \|_H. \] (36)
Noting the continuous embedding \( X \hookrightarrow H \) with
\[ \| v \|_H \leq C_E \| v \|_X \quad \forall v \in X, \] (37)
the inequality (36) transforms into
\[ \| u^* - \tilde{u}_N^{n+1} \|_X \leq \frac{C_E}{\nu} \| \psi_N(u_N^{n+1}) \|_H. \] (38)
This yields the claim. \( \square \)

We comment that the application of the continuous embedding (37) might cause the bound (38) to be pessimistic. At some additional computational cost, however, an improved estimate can be derived. To this end, we refer to the approach pursued in [20, Lemma 3.9].

Let \( v \in X_N \) be fixed, and consider the elements \( \xi \in X \) and \( \Xi \in X_N \) defined by
\[ \langle \xi, w \rangle_X = (v, w)_H \quad \forall w \in X, \] (39)
\[ \langle \Xi, w \rangle_X = (v, w)_H \quad \forall w \in X_N. \] (40)
Then
\[ \| v \|_X^2 = \| \xi \|_X^2 + \| \Xi - \xi \|_X^2. \]
Proof. We begin by observing the Gelfand triple \( X \hookrightarrow H \hookrightarrow X^* \) if we consider \( v \in X_N \subset X \) as an element of the dual space \( X^* \), by isometry, we have that \( \| v \|_X^2 = \| \xi \|_X^2 \). From (39) and (40) we further get that \( \langle \xi - \Xi, w \rangle_X = 0 \) for all \( w \in X_N \), i.e. \( \xi - \Xi \) is orthogonal to \( X_N \) with respect to \( \langle \cdot, \cdot \rangle_X \). Hence, applying the Pythagorean theorem completes the proof. \( \square \)

Corollary 3.5. Given the same assumptions as in Theorem 3.3, it holds that
\[ \| u^* - u_N^{n+1} \|_X \leq \frac{1}{\nu} \eta(u_N^{n+1}) + \frac{1}{\nu} \left( \| \Xi_N^{n+1} \|_X^2 + \| \xi_N^{n+1} - \Xi_N^{n+1} \|_X^2 \right)^{1/2}. \] (41)
where \( \xi_N^{n+1} \in X \) and \( \Xi_N^{n+1} \in X_N \) are such that, for any \( n \geq 0 \),
\[ \langle \xi_N^{n+1}, w \rangle_X = (\psi_N(u_N^{n+1}), w)_H, \quad \forall w \in X, \] (42)
\[ \langle \Xi_N^{n+1}, w \rangle_X = (\psi_N(u_N^{n+1}), w)_H, \quad \forall w \in X_N. \] (43)
Proof. Revisiting (36), and applying the boundedness of the duality pairing, gives
\[ \nu \| u^* - \tilde{u}_N^{n+1} \|_X^2 \leq \| \psi_N(u_N^{n+1}) \|_X \| \tilde{u}_N^{n+1} - u^* \|_X, \]
and thus
\[ \| u^* - \tilde{u}_N^{n+1} \|_X \leq \frac{1}{\nu} \| \psi_N(u_N^{n+1}) \|_X. \]
Moreover, Lemma 3.4 implies that
\[ \| \psi_N(u_N^{n+1}) \|_X^2 = \| \xi_N^{n+1} \|_X^2 = \| \Xi_N^{n+1} \|_X^2 + \| \xi_N^{n+1} - \Xi_N^{n+1} \|_X^2. \]
Inserting this estimate into (35) completes the proof. \( \square \)

Remark 3.6. We note that the term \( \| \Xi_N^{n+1} \|_X \) occurring in Corollary 3.5 is computable while \( \| \xi_N^{n+1} - \Xi_N^{n+1} \|_X \) is not. We may suppose, however, that there exists a computable a posteriori estimator \( \Lambda \) such that \( \| \xi_N^{n+1} - \Xi_N^{n+1} \|_X \leq \Lambda(\psi_N(u_N^{n+1})). \)
Remark 3.7. When dealing with second-order elliptic PDE, we typically use the standard Sobolev spaces $X := H^1_0(\Omega)$ and $H := L^2(\Omega)$. Here, the strong formulations of (42) and (43), respectively, are given by $-\Delta_N u^{n+1}_N = \psi_N(u^{n+1}_N)$ in $X^*$ and $-\Delta_N \Xi^{n+1}_N = \psi_N(u^{n+1}_N)$ in $X_N$, respectively, where $\Delta_N : X_N \to X_N$ is the discrete Laplacian. In particular, in this special case, the computation of $\|\Xi^{n+1}_N\|_X$ requires a discrete Laplace solve, and $\Lambda$ from Remark 3.6 is any a posteriori error estimator for the Poisson equation.

4. An abstract ILG procedure

The estimates from Theorems 3.1 and 3.3 allow to control the error between the solution of (1) and the discrete system (7) with respect to two individual terms, one of which expresses the error of the linearization, and will be denoted by $E_{\text{Linear},N}$, and the other, which we signify by $E_{\text{Galerkin},N}$, bounds the Galerkin discretization error. In a finite element context, the latter error will typically be composed of local contributions for each element; this, in turn, enables to refine the mesh locally. The algorithm, which will be presented below, uses an adaptive interplay between those two controlling terms. More precisely, on a given Galerkin space, we iterate as long as the linearization error dominates and, in addition, until it is, in a certain way, smaller than a given bound depending on $N$. Once the linearization error is small enough and is up to a factor $\vartheta$ less than the one arising from the Galerkin method, we enrich the Galerkin space according to the local error indicators in order to attain a smaller discretization error. Subsequently, we will perform the linearization on the enriched space. In this way, the goal of the ILG algorithm is to compute an approximation of the solution of (1) which, on the one hand, is sufficiently accurate, and, on the other hand, is attained from a minimal number of iterations.

4.1. Adaptive ILG algorithm. For the purpose of this section, we assume that our ILG Algorithm 4.1, to be presented below, generates a sequence of hierarchically enriched Galerkin spaces, $X_0 \subset X_1 \subset X_2 \subset \ldots$, on each of which we perform at least one iterative step. Furthermore, we will make use of a prescribed positive function $\sigma : \mathbb{N} \to (0, \infty)$ which satisfies

$$\sigma(N) \to 0 \quad \text{for} \quad N \to \infty. \quad (44)$$

Its role is to ensure that the linearization error tends to zero for an increasing number $N$ of Galerkin space enrichments. Recall that, for any fixed $N > 0$, our theory in Section 2 guarantees, under certain conditions, that the difference $\|u_N^N - u_N^{n-1}\|_X$ tends to zero for increasing $n$; in particular, it can be made smaller than $\sigma(N)$ for $n$ large enough.

An adaptive ILG procedure for the interactive reduction of discretization and linearization errors is proposed as follows.

Algorithm 4.1. Start with an initial Galerkin space $X_0$ and an initial guess $u^0_0$.
Until stopping criterion is fulfilled
While $E_{\text{Galerkin},N}^n \leq \vartheta E_{\text{Linear},N}^n$ or $\|u_N^n - u_N^{n-1}\|_X > \sigma(N)$,
Perform a single iterative linearization step to obtain $u_N^{n+1}$ from $u_N^n$; cf. (7).
Estimate the linearization error $E_{\text{Linear},N}^n$ and the Galerkin error indicator $E_{\text{Galerkin},N}^n$.
Update $n \leftarrow n + 1$.
EndWhile
Enrich the Galerkin space $X_N$ appropriately based on the error indicator $E_{\text{Galerkin},N}^n$.
Define $u_N^{n+1} \leftarrow u_N^n$ by inclusion $X_{N+1} \leftarrow X_N$.
End Until

We note that the above abstract algorithm can be performed with any of the iterative procedures from Section 2.3, and with any of the error estimators obtained from the linear and nonlinear elliptic reconstructions from Section 3. The input and output arguments as well as the components of the implemented algorithm may, of course, depend on the error estimator and the specific iterative linearization scheme applied.
4.2. A remark on convergence. Given a Galerkin space $X_N$, we recall the solution $u_N \in X_N$ of (32). Furthermore, we let $u_N := u_N^*$ be the final approximation (i.e. before the Galerkin space is enriched) obtained from the while loop in Algorithm 4.1. We establish the convergence of $u_N$ to the unique solution $u^*$ of (1) under the following assumption:

(AG) The hierarchically enriched Galerkin spaces $X_0 \subset X_1 \subset X_2 \subset \ldots$ generated by Algorithm 4.1 are such that the iterative Galerkin approximations $u_N \in X_N$ from (32) converge to the exact solution $u^* \in X$ of (1) for $N \to \infty$.

**Proposition 4.2.** If $F$ from (1) fulfills (F1) and (F2), and the Galerkin method satisfies (AG), then Algorithm 4.1 generates a sequence of Galerkin solutions which converges to the unique solution $u^* \in X$ of (1).

**Proof.** Using (F2) and involving (32), it holds that

$$
\nu \left\| u_N - u_N^{n-1} \right\|^2_X \leq \left\langle F(u_N) - F(u_N^{n-1}), u_N - u_N^{n-1} \right\rangle_{X^* \times X} = \left\langle F(u_N^*), u_N^* - u_N \right\rangle_{X^* \times X}.
$$

Invoking (5), (7), and (11), this leads to

$$
\nu \left\| u_N - u_N^{n-1} \right\|^2_X \leq \nu \left\| u_N^* - u_N \right\|_{X^*} \left\| u_N - u_N^{n-1} \right\|_X,
$$

and thus

$$
\left\| u_N - u_N^{n-1} \right\|_X \leq \frac{\nu \left\| u_N^* - u_N \right\|_{X^*}}{\nu} \left\| u_N - u_N^{n-1} \right\|_X.
$$

By the triangle inequality, this leads to

$$
\left\| u_N - u_N^{n} \right\|_X \leq \left\| u_N - u_N^{n-1} \right\|_X + \left\| u_N^n - u_N^{n-1} \right\|_X \leq \left( \frac{\beta + 1}{\nu} \right) \left\| u_N^n - u_N^{n-1} \right\|_X.
$$

Notice that the stopping criterion for the while loop in Algorithm 4.1 implies that

$$
\left\| u_N^n - u_N^{n-1} \right\|_X \leq \sigma(N) \quad \forall N \geq 0,
$$

with $\sigma$ satisfying (44). Then, invoking the triangle inequality, as well as (45) and (46), yields

$$
\left\| u^* - u_N \right\|_X \leq \left\| u^* - u \right\|_X + \left\| u_N - u_N^n \right\|_X \leq \left\| u^* - u \right\|_X + \left( \frac{\beta + 1}{\nu} \right) \sigma(N).
$$

The first term on the right-hand side tends to zero for $N \to \infty$ by virtue of (AG), and the same holds true for the second term due to (44). We deduce that $u_N \to u^*$ as $N \to \infty$. \qed

We conclude our remarks on convergence by pointing to the fact that, for the Zarantonello iteration (12) based on a conforming finite element method, condition (AG) can be verified under reasonable assumptions on the refinement indicators; cf. [15, §4.4 and (A1)-(A4)]. Here, it is crucial to ensure that the refinement indicators for $u_N$ and $u_N$ are asymptotically equivalent in some suitable sense. Evidently, in general, the verification of (AG), under sensible conditions on the Galerkin method, will require a sophisticated analysis that reaches beyond the scope of this work.

5. Application to stationary conservation laws

On an open, bounded and polygonal domain $\Omega \subset \mathbb{R}^2$, with Lipschitz boundary $\Gamma = \partial \Omega$, let us consider the second-order elliptic partial differential equation

$$
u \left\| u_N - u_N^{n-1} \right\|^2_X \leq \left\langle F(u_N) - F(u_N^{n-1}), u_N - u_N^{n-1} \right\rangle_{X^* \times X} = \left\langle F(u_N^*), u_N^* - u_N \right\rangle_{X^* \times X}.
$$

Here, we choose $X := H^s(\Omega)$ to be the standard Sobolev space of $H^{s}$-functions on $\Omega$ with zero trace along $\Gamma$; the inner product and norm on $X$ are defined, respectively, by $(u, v)_X := \langle \nabla u, \nabla v \rangle_{L^2(\Omega)}$ and $\|u\|_X := \|\nabla u\|_{L^2(\Omega)}$, for $u, v \in X$. By virtue of the Poincaré-Friedrichs inequality, for $H := L^2(\Omega)$, we note the continuous and dense embedding $X \hookrightarrow H$, with

$$
\|u\|_H \leq C_\Omega \|u\|_X \quad \forall u \in X,
$$

(48)
where $C_\Omega > 0$ is a constant only depending on $\Omega$. We suppose that $g \in X^* = H^{-1}(\Omega)$ in (47) is given, and $\mu \in C^1([0, \infty))$ fulfills
\begin{equation}
m_\mu(t - s) \leq \mu(t^2) t - \mu(s^2) s \leq M_\mu(t - s), \quad t \geq s \geq 0,
\end{equation}
with constants $m_\mu, M_\mu > 0$. In particular, upon setting $s = 0$, we observe that
\begin{equation}
m_\mu \leq \mu(t) \leq M_\mu \quad \forall t \geq 0.
\end{equation}
Under condition (49) it can be shown that the nonlinear operator $F$ from (47) satisfies the properties (F1) and (F2) with
\begin{equation}
v = m_\mu, \quad L_F = 3M_\mu;
\end{equation}
see [29, Proposition 25.26].

We note the weak form of the boundary value problem (47) in $X$:
\begin{equation}u \in X : \int_\Omega \mu \left( |\nabla u|^2 \right) \nabla u \cdot \nabla v \, dx = \langle g, v \rangle_{X^*,X} \quad \forall v \in X.
\end{equation}

5.1. Convergence of iterative linearizations. In the sequel, we will investigate the convergence of the various iteration schemes from Section 2.3 as applied to the PDE (47). The convergence of the Zarantonello iteration follows immediately from Theorem 2.2.

**Proposition 5.1.** If $\mu$ satisfies (49) and $F$ is given by (47), then the Zarantonello iteration (12), i.e.
\begin{equation}u^{n+1} \in X : \quad -\Delta u^{n+1} = -\Delta u^n + \delta \nabla \cdot \left\{ \mu \left( |\nabla u^n|^2 \right) \nabla u^n \right\} + \delta g, \quad n \geq 0,
\end{equation}
converges to the unique solution of (47) for any $\delta \in [0, 2m_\mu/3M_\mu^2]$.

In order to study the Kačanov iteration method for (47), let us define, for $u \in X$, the linear preconditioning operator
\begin{equation}A[u]v := -\nabla \cdot \left\{ \mu \left( |\nabla u|^2 \right) \nabla v \right\}, \quad v \in X.
\end{equation}
In addition to (49), we assume that $\mu$ is monotone decreasing, i.e.
\begin{equation}\mu'(t) \leq 0 \quad \forall t \geq 0.
\end{equation}

**Proposition 5.2.** Let $\mu$ satisfy (49) and (54). Then, the Kačanov iteration (15), i.e.
\begin{equation}u^{n+1} \in X : \quad -\nabla \cdot \left\{ \mu \left( |\nabla u^n|^2 \right) \nabla u^{n+1} \right\} = g, \quad n \geq 0,
\end{equation}
converges to the unique solution of (47).

**Proof.** We will show that the assumptions of Theorem 2.5 are satisfied. To this end, for $A$ from (53), and any $u \in X$, we define the bilinear form $a(u; v, w) := \langle A[u]v, w \rangle_{X^*,X}$, for $v, w \in X$. Then, using (50) in combination with the Cauchy-Schwarz inequality shows the coercivity and continuity properties (10) and (11) with
\begin{equation}\alpha = m_\mu, \quad \beta = M_\mu,
\end{equation}
respectively. Furthermore, we introduce the potential $G : X \to \mathbb{R}$ by
\begin{equation}G(u) := \int_\Omega \psi \left( |\nabla u|^2 \right) \, dx, \quad \text{with} \quad \psi(s) := \frac{1}{2} \int_0^s \mu(t) \, dt.
\end{equation}
For $u \in X$, taking the Gâteaux derivative of $G$, we find that
\begin{equation}\langle G'(u), v \rangle_{X^*,X} = \int_\Omega 2\psi' \left( |\nabla u|^2 \right) \nabla u \cdot \nabla v \, dx = \int_\Omega \mu \left( |\nabla u|^2 \right) \nabla u \cdot \nabla v \, dx = a(u; u, v),
\end{equation}
for any $v \in X$. Thus, we infer that $G'(u) = a(u; u, \cdot) = F(u) + g$. Recalling (F2), this implies the strong monotonicity property (16) with $c_0 = m_\mu$, and we conclude that (K1) holds true. In addition, due to (54), for any $t \geq s \geq 0$, it holds that
\begin{equation}\psi(t) - \psi(s) = \frac{1}{2} \int_s^t \mu(\tau) \, d\tau \geq \frac{1}{2}(t - s)\mu(t),
\end{equation}
and similarly for $s \geq t \geq 0$,\[\psi(t) - \psi(s) = -\frac{1}{2} \int_t^s \mu(\tau) \, d\tau \geq -\frac{1}{2} (s - t) \mu(t) = \frac{1}{2} (t - s) \mu(t).\]

Hence, for any $u, v \in X$, we have\[G(u) - G(v) \geq \frac{1}{2} \int_{\Omega} \mu \left( (\nabla u)^2 - (\nabla v)^2 \right) \, dx = \frac{1}{2} \left( a(u; u, u) - a(u; v, v) \right),\]
which shows (K2).

Finally, we turn our attention to the damped Newton iteration.

**Proposition 5.3.** Let $\mu$ satisfy (49) and (54). Moreover, suppose that the damping parameter $\delta : X \to [\delta_{\min}, \delta_{\max}]$ is a continuous functional, for some constants $\delta_{\min}, \delta_{\max}$, with $0 < \delta_{\min} \leq \delta_{\max} < 2^m v^3 M_{\mu}$. Then, the damped Newton iteration (19) for the nonlinear PDE (47) converges to its unique solution in $X$.

We will prove this proposition by showing that the assumptions of Theorem 2.6 are satisfied. For this purpose we require the following auxiliary result.

**Lemma 5.4.** If $\mu$ satisfies (49), then the operator $u \mapsto F'(u)u$ is continuous from $X$ to $X^*$ with respect to the weak topology on $X^*$.

**Proof.** By taking the limit $s \searrow t$ in (49), we infer that $m_\mu \leq \frac{d}{dt} (\mu(t)^2) t \leq M_\mu$, and, thereby,\[m_\mu \leq 2\mu'(t^2)t^2 + \mu(t^2) \leq M_\mu \quad \forall t \geq 0. \quad (57)\]

Moreover, a simple but lengthy calculation shows that\[\langle F'(u)v, w \rangle_{X^*, X} = \int_{\Omega} 2\mu'(|\nabla u|^2)(\nabla u \cdot \nabla v)(\nabla u \cdot \nabla w) \, dx + \int_{\Omega} \mu(|\nabla u|^2) \nabla v \cdot \nabla w \, dx, \quad (58)\]
for any $u, v, w \in X$. Consider a sequence $\{u^k\}_{k \geq 0} \subset X$ which converges to a limit $u \in X$, i.e.\[\|u - u^k\|_X \to 0, \quad k \to \infty. \quad (59)\]

Proceeding as in the proof of [29, Proposition 26.6], there is a subsequence such that\[\nabla u^k \to \nabla u \quad \text{a.e. in } \Omega \quad \text{for } k' \to \infty. \quad (60)\]

Hence, defining the function $\omega'(t) := 2\mu'(t) t + \mu(t)$, $t \geq 0$, it holds\[\langle F'(u)u - F'(u^k)u^k, w \rangle_{X^*, X} = \int_{\Omega} \omega(|\nabla u|^2) - \omega(|\nabla u^k|^2) \nabla u \cdot \nabla w \, dx\]
\[+ \int_{\Omega} \omega(|\nabla u^k|^2) \nabla (u - u^k) \cdot \nabla w \, dx.\]

We note that both terms on the right-hand side tend to 0 as $k' \to \infty$: Indeed, for the first integral this follows from the continuity of $\omega$, (60), (57), and the dominated convergence theorem; for the second integral, we recall (57) and (59). Finally, referring to [28, Proposition 10.13(2)], we conclude the weak convergence of the entire sequence, i.e. $F'(u^k)u^k \to F'(u)u$ as $k \to \infty$. This finishes the proof. \qed

**Proof of Proposition 5.3.** Notice that (58) is symmetric with respect to $v$ and $w$. Moreover, for $v = w$, we have that\[\langle F'(u)v, v \rangle_{X^*, X} = \int_{\Omega} 2\mu'(|\nabla u|^2)|\nabla u \cdot \nabla v|^2 + \int_{\Omega} \mu(|\nabla u|^2)|\nabla v|^2 \, dx.\]

Exploiting (54), and using the Cauchy-Schwarz inequality, we notice that\[2\mu'(|\nabla u|^2)|\nabla u \cdot \nabla v|^2 \geq 2\mu'(|\nabla u|^2)|\nabla u|^2 |\nabla v|^2.

It follows that\[\langle F'(u)v, v \rangle_{X^*, X} \geq \int_{\Omega} \left( 2\mu'(|\nabla u|^2)|\nabla u|^2 + \mu \left(|\nabla u|^2\right) \right) |\nabla v|^2 \, dx.\]
Applying (57) implies that $\langle F'(u)v, v \rangle_{X^\times X} \geq m_\mu \|v\|_X^2$ for any $u, v \in X$; this shows (21) with
\[ \alpha_F = m_\mu. \]

In addition, in view of (51), we observe that $2\alpha_F/\mu_F = 2m_\mu/3M_\mu > \delta_{\text{max}}$, as required in Theorem 2.6. Furthermore, application of the Cauchy-Schwarz inequality, and involving (54), yields
\[
\langle F'(u)v, v \rangle_{X^\times X} \leq \int_\Omega \left( 2\mu'(|\nabla u|^2) |\nabla u|^2 + \mu \left( |\nabla u|^2 \right) \right) |\nabla v||\nabla w| \, dx \\
- \int_\Omega \left( 2\mu'(|\nabla u|^2) |\nabla u|^2 + \mu \left( |\nabla u|^2 \right) \right) |\nabla v||\nabla w| \, dx \\
+ 2\int_\Omega \mu \left( |\nabla u|^2 \right) |\nabla v||\nabla w| \, dx.
\]

Employing (57) and (50), this leads to
\[
\langle F'(u)v, w \rangle_{X^\times X} \leq (2M_\mu - m_\mu) \int_\Omega |\nabla v||\nabla w| \, dx \leq (2M_\mu - m_\mu) \|v\|_X \|w\|_X,
\]
which gives (22) with
\[ \beta_F = 2M_\mu - m_\mu. \]

In order to prove (N3), let us define the functional $H : X \to \mathbb{R}$ by
\[
H(u) := \int_\Omega \psi \left( |\nabla u|^2 \right) \, dx - \langle g, u \rangle_{X^\times X}, \quad u \in X,
\]
with $\psi$ as in (56). It holds that
\[
\langle H'(u), v \rangle_{X^\times X} = \int_\Omega \mu \left( |\nabla u|^2 \right) \nabla u \cdot \nabla v \, dx - \langle g, v \rangle_{X^\times X} = \langle F(u), v \rangle_{X^\times X},
\]
for all $v \in X$. Finally, to establish (N2), we introduce the functional $G : X \to \mathbb{R}$ by $G(u) := F(u) - H(u)$, where $F(u) := \langle F(u), u \rangle_{X^\times X}, u \in X$. For $u \in X$, the Gâteaux derivative of $F$ is given by
\[
\langle F'(u), v \rangle_{X^\times X} = \langle F'(u)u, v \rangle_{X^\times X} + \langle F(u), v \rangle_{X^\times X}, \quad \forall v \in X.
\]
It follows that $G'(u) = F'(u) - H'(u) = F'(u)u + F(u) - F(u) = F'(u)u$. Finally, due to Lemma 5.4 the mapping $u \mapsto G'(u) = F'(u)u$ is continuous with respect to the weak topology on $X^\times$.

5.2. Iterative linearized FEM. For the sake of discretizing (52), and thereby, of obtaining an ILG formulation for (47), we will use a conforming finite element framework. We consider regular and shape-regular meshes $\mathcal{T}_h$ that partition the domain $\Omega$ into open and disjoint triangles $K \in \mathcal{T}_h$ such that $\Omega = \bigcup_{K \in \mathcal{T}_h} K$. We denote by $h_K := \text{diam}(K)$ the diameter of $K \in \mathcal{T}_h$, and let $h := \max_{K \in \mathcal{T}_h} h_K$. Moreover, we consider the finite element space
\[ X_h := \{ v \in H^1_0(\Omega) : v|_K \in \mathcal{P}_p(K) \ \forall K \in \mathcal{T}_h \}, \]
where, for fixed $p \in \mathbb{N}$, we signify by $\mathcal{P}_p(K)$ the space of all polynomials of total degree at most $p \geq 1$ on $K \in \mathcal{T}_h$.

Within the adaptive ILG framework, we will consider a sequence of meshes $\{\mathcal{T}_N\}_{N \geq 0}$, whereby we start with an initial conforming triangulation $\mathcal{T}_0$ of $\Omega$. All subsequent meshes are obtained by refinement, i.e. for $N \geq 0$, the mesh $\mathcal{T}_{N+1}$ is a hierarchical refinement of $\mathcal{T}_N$. Moreover, we will denote by $X_N$ the finite element space associated to the mesh $\mathcal{T}_N$.

For an edge $e \subset \partial K^+ \cap \partial K^-$, which is the intersection of two neighbouring elements $K^\pm \in \mathcal{T}_N$, we signify by $[v]|_e = v^+|_e \cdot n_{K^+} + v^-|_e \cdot n_{K^-}$ the jump of a (vector-valued) function $v$ along $e$, where $v^\pm|_e$ denote the traces of the function $v$ on the edge $e$ taken from the interior of $K^\pm$, respectively, and $n_{K^\pm}$ are the unit outward normal vectors on $\partial K^\pm$, respectively.
5.2.1. A posteriori error analysis via linear elliptic reconstruction. In this section, we discuss the a posteriori error estimate from Theorem 3.1 in the specific context of the nonlinear PDE (47) and the finite element framework presented above. Introducing the residual
\[
R(u; v, w) := a(u; v, w) - (f(u), w)_{X^* \times X}, \quad u, v \in X_N, w \in X,
\]
it is fairly straightforward to verify that, for all of the three iterative linearization schemes from Section 5.1, and for \( g \in L^2(\Omega) \) in (47), it holds the special form
\[
R(u_N^n; u_{N+1}^n, w) = -\int_\Omega q_N^n \cdot \nabla w \, dx + \int_\Omega p_N^n w \, dx \quad \forall w \in X,
\]
with some \( p_N^n \in H = L^2(\Omega) \) and \( q_N^n \in H^1(\Omega)^2 \), which can be represented explicitly. Then, recalling (7), we may conclude that
\[
R(u_N^n; u_{N+1}^n, w) = R(u_N^n; u_N^n, w - w_N) = -\int_\Omega q_N^n \cdot \nabla (w - w_N) \, dx + \int_\Omega p_N^n (w - w_N) \, dx,
\]
for any \( w_N \in X_N \). Therefore, choosing \( w_N \) to be a quasi-interpolant of \( w \), and pursuing a standard residual-based a posteriori error analysis (see, e.g., [25]), we deduce the upper bound
\[
\sup_{w \in X} R(u_N^n; u_{N+1}^n, w) \leq C_1 \left( \sum_{K \in T_N} \eta_K^2 \right)^{1/2},
\]
where \( C_1 > 0 \) is an interpolation constant (only depending on the polynomial degree \( p \) and on the shape-regularity of the mesh), and
\[
\eta_K^2 = h_K^2 \| \nabla \cdot q_N^n + p_N^n \|_{L^2(K)}^2 + \frac{1}{2} h_K \| \|q_N^n\|_{L^2(\partial K \setminus T)}^2, \quad K \in T_N,
\]
is a computable error indicator.

**Theorem 5.5.** Let \( F \) be defined by (47) with \( \mu \) fulfilling (49) and (54), and let \( X_N \subset H^1_0(\Omega) \) be a conforming finite element space as in (63) on a mesh \( T_N \). If \( u^* \) is the unique solution of (47), and \( \{u_N^n\}_{n \geq 0} \) is a sequence of ILG solutions obtained by any of the iterative linearization procedures from Section 5.1 on \( X_N \), then it holds the a posteriori estimate
\[
\| u^* - u_{N+1} \|_X \leq \frac{\beta C_1}{\alpha m_\mu} \left( \sum_{K \in T_N} \eta_K^2 \right) + \frac{\beta}{m_\mu} \| u_{N+1}^n - u_N^n \|_X,
\]
where \( C_1 > 0 \) is a constant, and
\[
(\alpha, \beta) = \begin{cases} 
(\delta^{-1}, \delta^{-1}) \quad & \text{for the Zarantonello iteration, cf. (13),} \\
(m_\mu, M_\mu) \quad & \text{for the Kačanov iteration, cf. (55),} \\
M_\mu / \delta_{\max} (2M_\mu - m_\mu) / \delta_{\min} \quad & \text{for the Newton iteration, cf. (23), (61), and (62),}
\end{cases}
\]
and \( \eta_K \), for \( K \in T_N \), is defined in (64).

**Proof.** The result follows from Theorem 3.1, whereby we replace the constants \( \nu \) and \( L_F \) from (51), and insert the values of \( \alpha \) and \( \beta \) from (10) and (11) for the respective iterative schemes from Section 5.1. \( \square \)

5.2.2. Error estimator via nonlinear elliptic reconstruction. Following our abstract analysis in Section 3.2, we consider the residual
\[
R(u_N^n+1) := \sup_{u \in X \atop \|u\|_X = 1} \left\{ (\psi_N(u_{N+1}^n), w)_H - \langle F(u_N^n+1), w \rangle_{X^* \times X} \right\}.
\]
Noticing (31), for any \( w_N \in X_N \), we have
\[
R(u_N^n+1) := \sup_{w \in X \atop \|w\|_X = 1} \left\{ (\psi_N(u_{N+1}^n), w - w_N)_H - \langle F(u_N^n+1), w - w_N \rangle_{X^* \times X} \right\}.
\]
Then, for \( g \in L^2(\Omega) \) in (47), and \( w_N \in X_N \) an appropriate quasi-interpolant of \( w \in H_0^1(\Omega) \), we employ a standard residual-based a posteriori error analysis (see, e.g., [25]) to infer the upper bound

\[
R(u_n^{n+1}) \leq C_1 \left( \sum_{K \in T_N} \eta_N^2 \right)^{1/2},
\]

where \( C_1 \) is a quasi-interpolation constant, and

\[
\eta_N^2 = h^2_N \left\| \psi_N(u_n^{n+1}) + g + \nabla \cdot \left\{ \mu \left( \left| \nabla u_n^{n+1} \right|^2 \right) \nabla u_n^{n+1} \right\} \right\|_{L^2(K)}^2 \\
+ \frac{1}{2} h_K \left\| \mu \left( \left| \nabla u_n^{n+1} \right|^2 \right) \nabla u_n^{n+1} \right\|_{L^2(\partial K \setminus \Gamma)}^2,
\]

for any \( K \in T_N \). Then, invoking Theorem 3.3 and recalling (51), we obtain the following result.

**Theorem 5.6.** Given the same assumptions as in Theorem 5.5, then it holds the a posteriori error estimate

\[
\| u^* - u_n^{n+1} \|_X \leq \frac{C_1}{m_\mu} \left( \sum_{K \in T_N} \eta_N^2 \right)^{1/2} + \frac{C_\Omega}{m_\mu} \| \psi_N(u_n^{n+1}) \|_{L^2(\Omega)},
\]

where \( u^* \) is the unique solution of (1), \( C_1 \) is a constant, \( C_\Omega \) is from (48), and \( \eta_K \), for \( K \in T_N \), is given in (65).

Furthermore, revisiting Corollary 3.5, the ensuing result can be established.

**Theorem 5.7.** Let \( \Xi_n^{n+1} \in X_N \) be the unique solution of the finite element formulation

\[
(\nabla \Xi_n^{n+1}, \nabla w)_{L^2(\Omega)} = (\psi_N(u_n^{n+1}), w)_{L^2(\Omega)} \quad \forall w \in X_N.
\]

Then, given the same assumptions as in Theorem 5.5, there holds the a posteriori error estimate

\[
\| u^* - u_n^{n+1} \|_X \leq \frac{C_1}{m_\mu} \left( \sum_{K \in T_N} \eta_N^2 \right)^{1/2} + \frac{1}{m_\mu} \left( \| \Xi_n^{n+1} \|_X^2 + C'_1 \sum_{K \in T_N} \zeta_K^2 \right)^{1/2},
\]

where \( u^* \) is the unique solution of (1), \( C_1, C'_1 > 0 \) are constants depending only on the shape regularity of the mesh, and \( \eta_K, K \in T_N \), is given by (65). Furthermore,

\[
\zeta_K^2 = h^2_K \left\| \psi_N(u_n^{n+1}) + \Delta \Xi_n^{n+1} \right\|_{L^2(K)}^2 + \frac{1}{2} h_K \left\| \nabla \Xi_n^{n+1} \right\|_{L^2(\partial K \setminus \Gamma)}^2,
\]

for any \( K \in T_N \).

**Proof.** Applying Remark 3.7, and using a standard a posteriori error estimate for the Poisson equation (see, e.g., [1,25]), we have that

\[
\| \nabla (\xi_n^{n+1} - \Xi_n^{n+1}) \|_{L^2(\Omega)} \leq C'_1 \left( \sum_{K \in T_N} \zeta_K^2 \right)^{1/2},
\]

where

\[
\zeta_K^2 = h^2_K \left\| \psi_N(u_n^{n+1}) + \Delta \Xi_n^{n+1} \right\|_{L^2(K)}^2 + \frac{1}{2} h_K \left\| \nabla \Xi_n^{n+1} \right\|_{L^2(\partial K \setminus \Gamma)}^2,
\]

and \( C'_1 > 0 \) is a constant. Inserting this bound into (41), and using (51), yields the desired bound. \( \Box \)
5.3. Numerical Experiments. In this section, we test the adaptive ILG Algorithm 4.1 in the context of the iterative linearized FEM for stationary conservation laws discussed in Section 5. We perform a series of numerical experiments to compare the various iterative linearization procedures from Section 2.3 and to validate the \textit{a posteriori} error estimators from Section 5.2. For all our experiments, we consider the L-shaped domain $\Omega = (-1,1)^2 \setminus ([0,1] \times [-1,0])$, and an initial mesh consisting of 192 uniform triangles. Moreover, we will always choose the initial guess to be $u^0 \equiv 0$, and run the algorithm until the number of elements exceeds $10^6$. On a given mesh, we perform at least one iterative linearization step, and continue until the linearization error is at most half as large as the discretization error, i.e. we let $\vartheta = 2$ in Algorithm 4.1, and $\sigma(N) := (N+1)^{-1/2} \|u^0_N\|_X$, for $N \geq 0$, which is in line with the expected convergence rate of $O(N^{-1/2})$. We set the constant factors for the discretization and linearization estimators appearing in the right-hand sides of the \textit{a posteriori} error bounds to 1 (cf. Theorems 5.5, 5.6, and 5.7). In the adaptive process, we mark the elements for refinement by use of the Dörfler marking strategy, see [11], and process them by the newest vertex bisection method. The true error $\|u^* - u^N\|_X$ and the error estimator will be displayed each time before a mesh refinement is undertaken. Our implementation is based on the matlab package [14], with the necessary modifications.

We remark that, even though we have derived two different error estimators for the nonlinear elliptic reconstruction, see Theorems 5.6 and 5.7, we will focus on the estimator from Theorem 5.7. Indeed, as indicated earlier in the context of the possibly pessimistic bound (38), our computational studies (not presented here) have revealed that the bound for the linearization error in Theorem 5.6 is fairly rough. As a consequence, a higher number of iterative linearization steps is required, which, in turn, leads to longer overall running times of the algorithm. We note that, for the damped Newton method, however, the situation is better as for the Kačanov and Zarantonello iterations. This might be due to the local quadratic convergence property which may compensate, to some extent, for the rough estimate of the linearization error. For any iterative procedure in the Experiments 5.3.1–5.3.3 below, with either the \textit{a posteriori} error estimators from Theorem 5.5 or Theorem 5.7, the algorithm performs between one and three iterative steps on the initial mesh, and then only one linearization is needed on each refined mesh. Hence, our algorithm is highly efficient for the proposed examples.

5.3.1. Smooth solution. We consider the nonlinear diffusion coefficient $\mu(t) = (t + 1)^{-1} + 1/2$, for $t \geq 0$, and select $g$ in (47) such that the analytical solution of (52) is given by the smooth function $u^*(x,y) = \sin(\pi x)\sin(\pi y)$. It is straightforward to verify that $\mu$ fulfills the requirements (49) and (54) from Section 5, so that the convergence of the three iterative procedures from Section 2.3 is guaranteed. The parameter $\delta$ in the Zarantonello iteration (12) is chosen to be 0.85 as this seems to be close to optimal. The initial damping parameter on the initial mesh for the damped Newton method is chosen to be $\delta^0 = 1$ in Remark 2.8; moreover, throughout all our experiments, the factor $\kappa$ for the correction and prediction strategy of the damping parameter is set to be $1/2$.

In Figure 1, for each of the three iterative linearization schemes presented in Section 5.1, we plot the error $\|u^* - u^N_N\|_X$ and both error estimators from Theorems 5.5 and 5.7 against the number $N$ of elements in the mesh. In addition, we display the effectiveness indices for each experiment, i.e. the ratio of the error estimator and the true error; we see that they are roughly bounded between 2 and 4. Furthermore, we notice that (nearly) optimal convergence rates $O\left(N^{-1/2}\right)$ are achieved in all plots.

5.3.2. Nonsmooth solution. In our second experiment, we consider the nonlinear diffusion parameter $\mu(t) = 1 + e^{-t}$, for $t \geq 0$. Again, it is easily seen that $\mu$ satisfies the assumptions (49) and (54). We choose $g$ in (47) such that the analytical solution is given by

$$u^*(r, \varphi) = r^{2/3} \sin\left(\frac{2\varphi}{3}\right) \left(1 - r \cos(\varphi)\right)(1 + r \cos(\varphi))(1 - r \sin(\varphi))(1 + r \sin(\varphi)) \cos(\varphi),$$

where $r$ and $\varphi$ are polar coordinates. This is the prototype singularity for (linear) second-order elliptic problems with homogeneous Dirichlet boundary conditions in the L-shaped domain; in particular, we note that the gradient of $u^*$ is unbounded at the origin. As before, in Figure 2, we plot the error $\|u^* - u^N_N\|_X$, the error estimators from Theorems 5.5 and 5.7, as well as the effectiveness indices versus the number $N$ of elements in the mesh for each of the three iterative
linearization schemes from Section 5.2. We let $\delta = 0.5$ for the Zarantonello iteration, and use the initial damping parameter $\delta^0 = 1$ for the Newton method as in Experiment 5.3.1. As before, we observe that optimal rates of convergence are attained in all six cases.

5.3.3. Nonsmooth solution with monotone increasing diffusion. Finally, we consider the nonlinear diffusivity function $\mu(t) = 2 - e^{-t}$, for $t \geq 0$. Again, we choose $g$ in (47) such that the analytical solution is given by the nonsmooth function (66). Since $\mu$ is monotone increasing, it does not have the property (54), which is needed to guarantee the convergence of the Kačanov iteration and of the damped Newton method. It still fulfills, however, the assumption (49), which, in turn, is sufficient to guarantee the convergence of the Zarantonello method. In this experiment, we choose the damping parameter for the Zarantonello method to be $\delta = 0.4$, and the initial damping parameter in the Newton method to be $\delta^0 = 1$. We see from the plots in Figure 3 that the Kačanov and damped Newton methods converge, even with optimal order, which indicates that the property (54) does not seem to be necessary for the current example and the initial setup chosen here. We emphasize that this observation for the Kačanov method was already made in [16].

**References**

1. M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Series in Computational and Applied Mathematics, Elsevier, 1996.
2. M. Amrein, J. M. Melenk, and T. P. Wihler, *An hp-adaptive Newton-Galerkin finite element procedure for semilinear boundary value problems*, Math. Methods Appl. Sci. 40 (2017), no. 6, 1973–1985.
3. M. Amrein and T. P. Wihler, *An adaptive Newton-method based on a dynamical systems approach*, Commun. Nonlinear Sci. Numer. Simul. 19 (2014), no. 9, 2958–2973.
4. M. Amrein and T. P. Wihler, *Fully adaptive Newton-Galerkin methods for semilinear elliptic partial differential equations*, SIAM J. Sci. Comput. 37 (2015), no. 4, A1637–A1657.
5. M. Amrein and T. P. Wihler, *An adaptive space-time Newton-Galerkin approach for semilinear singularly perturbed parabolic evolution equations*, IMA J. Numer. Anal. 37 (2017), no. 4, 2004–2019.
6. C. Bernardi, J. Dákrub, G. Mansour, and T. Sayah, *A posteriori analysis of iterative algorithms for a nonlinear problem*, J. Sci. Comput. 65 (2015), no. 2, 672–697.
7. F. E. Browder, *Remarks on nonlinear functional equations. II, III*, Illinois J. Math. 9 (1965) 608–616; ibid. 9 (1965), 617–622.
8. A. L. Chaillou and M. Suri, *A posteriori estimation of the linearization error for strongly monotone nonlinear operators*, J. Comput. Appl. Math. 205 (2007), no. 1, 72–87.
9. S. Congreve and T. P. Wihler, *Iterative Galerkin discretizations for strongly monotone problems*, Journal of Computational and Applied Mathematics 311 (2017), 457–472.
10. P. Deuflhard, *Neutron methods for nonlinear problems*, Springer Series in Computational Mathematics, vol. 35, Springer-Verlag, Berlin, 2004, Affine invariance and adaptive algorithms.
11. W. Dörfler, *A convergent adaptive algorithm for Poisson’s equation*, SINUM 33 (1996), 1106–1124.
12. L. El Alaoui, A. Ern, and M. Vohralík, *Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems*, Comput. Methods Appl. Mech. Engrg. 200 (2011), no. 37–40, 2782–2795.
13. A. Ern and M. Vohralík, *Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs*, SIAM J. Sci. Comput. 35 (2013), no. 4, A1761–A1791.
14. S. Funken, D. Praetorius, and P. Wissgott, *Efficient Implementation of Adaptive P1-FEM in Matlab*, Computational Methods in Applied Mathematics 11 (2011), no. 4, 460–490. MR 2875100
15. G. Gantner, A. Haberl, D. Praetorius, and B. Stiftner, *Rate optimal adaptive fem with inexact solver for nonlinear operators*, IMA Journal of Numerical Analysis (2017), drx050.
16. E. M. Garau, P. Morin, and C. Zuppa, *Convergence of an adaptive Kačanov FEM for quasi-linear problems*, Appl. Numer. Math. 61 (2011), no. 4, 512–529.
17. W. Han, S. Jensen, and I. Shimansky, *The Kačanov method for some nonlinear problems*, Appl. Numer. Meth. 24 (1997), 57–79.
18. P. Houston and T. P. Wihler, *An hp-adaptive neutron-discontinuous-galerkin finite element approach for semilinear elliptic boundary value problems*, Math. Comp. (2018), in press.
19. O. Lakkis and C. Makridakis, *Elliptic reconstruction and a posteriori error estimates for fully discrete linear parabolic problems*, Math. Comp. 75 (2006), no. 256, 1627–1658.
20. O. Lakkis and T. Pryer, *Gradient recovery in adaptive finite-element methods for parabolic problems*, IMA J. Numer. Anal. 32 (2012), no. 1, 246–278.
21. C. Makridakis and R. H. Nochetto, *Elliptic reconstruction and a posteriori error estimates for parabolic problems*, SIAM Journal on Numerical Analysis 41 (2003), no. 4, 1585–1594.
22. J. Nečas, *Introduction to the theory of nonlinear elliptic equations*, John Wiley and Sons, 1986.
23. A. Potschka, *Backward step control for global newton-type methods*, SIAM J. Numer. Anal. 54 (2016), no. 1, 361–387.
24. H. R. Schneebeli and T. P. Wihler, *The Newton-Raphson method and adaptive ODE solvers*, Fractals. Complex Geometry, Patterns, and Scaling in Nature and Society 19 (2011), no. 1, 87–99.
25. R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.
26. E. H. Zarantonello, *The closure of the numerical range contains the spectrum*, Bull. Amer. Math. Soc. 70 (1964), 781–787.
27. E. Zeidler, *Nonlinear functional analysis and its applications. III: Variational methods and optimization.*, Springer-Verlag, New York, 1985.
28. E. Zeidler, *Nonlinear functional analysis and its applications. I*, Springer-Verlag, New York, 1986, Fixed-point theorems.
29. ——, *Nonlinear functional analysis and its applications. II/B*, Springer-Verlag, New York, 1990.
Figure 1. Experiment 5.3.1: Performance data for the error estimators from Theorem 5.5 (left) and Theorem 5.7 (right) for the Zarantonello, Kačanov and Newton iterations.
Figure 2. Experiment 5.3.2: Performance data for the error estimators from Theorem 5.5 (left) and Theorem 5.7 (right) for the Zarantonello, Kačanov and Newton iterations.
Figure 3. Experiment 5.3.3: Performance data for the error estimators from Theorem 5.5 (left) and Theorem 5.7 (right) for the Zarantonello, Kačanov and Newton iterations.