A note on energy contraction and optimal convergence of adaptive iterative linearized finite element methods

P. Heid, D. Praetorius, and T.P. Wihler
Most recent ASC Reports

20/2020  M. Faustmann, J.M. Melenk, M. Parvizi
Caccioppoli-type estimates and \(H\)-Matrix approximations to inverses for FEM-BEM couplings

19/2020  G. Di Fratta, A. Fiorenza, and V. Slastikov
An estimate of the blow-up of Lebesgue norms in the non-tempered case

18/2020  A. Bespalov, D. Praetorius, and M. Ruggeri
Two-level a posteriori error estimation for adaptive multilevel stochastic Galerkin FEM

17/2020  G. Di Fratta
Micromagnetics of curved thin films

15/2020  T.S. Gutleb, N.J. Mauser, M. Ruggeri, H.-P. Stimming
A time splitting method for the three-dimensional linear Pauli equation

14/2020  N. Angleitner, M. Faustmann, J.M. Melenk
Approximating inverse FEM matrices on non-uniform meshes with \(H\)-matrices

13/2020  F. Achleitner, C. Kuehn, J.M. Melenk, A. Rieder
Metastable speeds in the fractional Allen-Cahn equation

12/2020  A. Haberl, D. Praetorius, S. Schimanko, and M. Vohralik
Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver

11/2020  L. Banjai, J.M. Melenk, and C. Schwab
hp-FEM for reaction-diffusion equations
II. Robust exponential convergence for multiple length scales in corner domains

10/2020  G. Gantner and D. Praetorius
Adaptive BEM for elliptic PDE systems,
Part I: Abstract framework for weakly-singular integral equations
A NOTE ON ENERGY CONTRACTION AND OPTIMAL CONVERGENCE OF ADAPTIVE ITERATIVE LINEARIZED FINITE ELEMENT METHODS

PASCAL HEID\(^1\), DIRK PRAETORIUS\(^2\), AND THOMAS P. WIHLER\(^3\)

Abstract. In this note, we revisit a unified methodology for the iterative solution of nonlinear equations in Hilbert spaces. Our key observation is that the general approach from [HW\(20\), HW\(20\)a] satisfies an energy contraction property in the context of (abstract) strongly monotone problems. This property, in turn, is the crucial ingredient in the recent convergence analysis in [GHPS\(20\)]. In particular, we deduce that adaptive iterative linearized finite element methods (AILFEMs) lead to linear convergence with optimal algebraic rates with respect to the degrees of freedom as well as the total computational time.

1. Introduction

On a real Hilbert space \(X\) with inner product \((\cdot, \cdot)_X\) and induced norm \(\|\cdot\|_X\), we consider a nonlinear operator \(F : X \to X^*\), where \(X^*\) denotes the dual space of \(X\). We aim to generate optimally converging (with respect to the number of degrees of freedom and the total computational time) finite element approximations of the following nonlinear operator equation:

\[
F(u) = 0 \quad \text{in } X^*. \tag{1}
\]

In weak form, this problem reads:

Find \(u \in X\) such that \(\langle F(u), v \rangle = 0 \quad \forall v \in X; \tag{2}\)

here, \(\langle \cdot, \cdot \rangle\) signifies the duality pairing on \(X^* \times X\). For the purpose of this work, we suppose that \(F\) satisfies the following conditions:

(F1) Lipschitz continuity: There exists a constant \(L_F > 0\) such that

\[
|\langle F(u) - F(v), w \rangle| \leq L_F \|u - v\|_X \|w\|_X \quad \forall u, v, w \in X.
\]

(F2) Strong monotonicity: There exists a constant \(\nu > 0\) such that

\[
\nu \|u - v\|_X^2 \leq \langle F(u) - F(v), u - v \rangle \quad \forall u, v \in X.
\]

(F3) The operator \(F\) possesses a potential, i.e., there exists a Gâteaux differentiable (energy) functional \(E : X \to \mathbb{R}\) such that \(E' = F\).

Assuming (F1) and (F2), the main theorem of strongly monotone operators guarantees that (1) (or equivalently (2)) has a unique solution \(u^* \in X\); see, e.g., [Neč86, §3.3] or [Zei90, Thm. 25.B].
1.1. Iterative linearization. Following the recent approach [HW20a], we apply a general fixed point iteration scheme for (1). For given \( v \in X \), we consider a linear and invertible preconditioning operator \( A[v] : X \to X^* \). Then, the nonlinear problem (1) is equivalent to \( A[u]u = f(u) := A[u]u - F(u) \). For any suitable initial guess \( u^0 \in X \), this leads to the following iterative scheme:

\[
\text{Find } u^{n+1} \in X \text{ such that } A[u^n]u^{n+1} = f(u^n) \text{ in } X^* \quad \forall n \geq 0. \quad (3)
\]

Note that the above iteration is a linear equation for \( u^{n+1} \), thereby rendering (3) an iterative linearization scheme for (1). Given \( u^n \in X \), the weak form of (3) is based on the bilinear form \( a(u^n; v, w) := (A[u^n]v, w) \), for \( v, w \in X \), and the solution \( u^{n+1} \in X \) of (3) can be obtained from

\[
a(u^n; u^{n+1}, w) = (f(u^n), w) \quad \forall w \in X. \quad (4)
\]

Throughout, for any \( u \in X \), we suppose that the bilinear form \( a(\cdot, \cdot) \) is uniformly coercive and bounded, i.e., there are two constants \( \alpha, \beta > 0 \) independent of \( u \in X \) such that

\[
a(u; v, v) \geq \alpha \|v\|_X^2 \quad \forall v \in X,
\]

and

\[
a(u; v, w) \leq \beta \|v\|_X \|w\|_X \quad \forall v, w \in X. \quad (5)
\]

For any given \( u^n \in X \), owing to the Lax–Milgram theorem, these two properties imply that the linear equation (4) admits a unique solution \( u^{n+1} \in X \).

1.2. Iterative linearized Galerkin approach (ILG). In order to cast (4) into a computational framework, we consider closed (and mainly finite dimensional) subspaces \( Y \subseteq X \), endowed with the inner product and norm on \( X \). Then, ILG [HW20a] is based on restricting the weak iteration scheme (4) to \( Y \). Specifically, for a prescribed initial guess \( u_Y^0 \in Y \), define a sequence \( \{u_Y^n\}_{n \geq 0} \subset Y \) inductively by

\[
a(u^n_Y; u_Y^{n+1}, w) = (f(u^n_Y), w) \quad \forall w \in Y. \quad (7)
\]

Note that (7) has a unique solution, since the conditions (5) and (6) above remain valid for the restriction to \( Y \subseteq X \). For the purpose of the convergence results in this paper, we introduce a monotonicity condition on the functional \( E \) from (F3).

(F4) There exists a (uniform) constant \( C_E > 0 \) such that for any closed subspace \( Y \subseteq X \), the sequence defined by (7) fulfills the bound

\[
E(u_Y^n) - E(u_Y^{n+1}) \geq C_E \|u_Y^n - u_Y^{n+1}\|_X^2 \quad \forall n \geq 0, \quad (8)
\]

where \( E \) is the potential of \( F \) (restricted to \( Y \subseteq X \)) introduced in (F3).

If \( F \) satisfies (F1)–(F3) and if the ILG bilinear form \( a(\cdot, \cdot) \) from (4) is coercive (5) with coercivity constant \( \alpha > \frac{1}{\beta N/2} \), then (F4) is fulfilled; see [HW20b, Prop. 2.1]. Moreover, upon imposing alternative conditions, we may still be able to satisfy (8) even when \( \alpha \leq \frac{1}{\beta N/2} \). For instance, [HW20a, Rem. 2.8] proposed an a posteriori step size strategy that guarantees the bound (8) in the context of the damped Newton method. This argument can be generalized to other methods containing a damping parameter. Moreover, the Kačanov linearization approach (see (22) below) satisfies (8) with \( C_E = \alpha/2 \) without requiring \( \alpha > \frac{1}{\beta N/2} \); see, e.g., [HW20a, Eq. (18)].

1.3. Adaptive iterative linearized finite element method (AILFEM). We restrict ILG (7) to a sequence of Galerkin subspaces \( X_0 \subset X_1 \subset X_2 \subset \ldots \subset X_N \subset \ldots \subset X \) with corresponding sequences \( \{u_Y^n\}_{n \geq 0} \subset X_N \) for \( N \geq 0 \) given by

\[
a(u_N^N; u_N^{N+1}, v) = (f(u_N^N), v) \quad \forall v \in X_N. \quad (9)
\]

The spaces \( X_N \) will be conforming finite element spaces associated to admissible triangulations \( T_N \) of an underlying bounded Lipschitz domain \( \Omega \subset \mathbb{R}^d \), with \( d \geq 2 \). AILFEM (Algorithm 1 below) exploits an interplay of adaptive mesh refinements and ILG (9) and chooses the initial guesses \( u_0^0 \in X_0 \) and \( u_{N+1}^0 = u_N^{n(N)} \in X_N \subset X_{N+1} \) for all \( N \geq 0 \) and appropriate indices \( n(N) \geq 1 \).
1.4. Goal of the paper. The work [GHPS18] has recently analyzed the AILFEM for the solution of (1) in the specific context of a Zarantonello linearization approach originally proposed in [CW17]. The results of [GHPS18] have been extended to contractive iterative solvers in [GHPS20]. In particular, it has been shown that an optimal convergence rate with respect to the number of degrees of freedom as well as with respect to the overall computational cost can be achieved. The purpose of the present note is to show that these optimality results from [GHPS20] also apply to AILFEM beyond the Zarantonello linearization approach (cf. [HW20b] for some convergence results in abstract spaces). Especially, we will exploit (F4) to establish an energy contraction property for the fixed point iteration (3), which constitutes the crucial ingredient in the analysis of [GHPS20]. In particular, we will revisit a few example AILFEM schemes from [HW20a], and verify our theoretical findings numerically.

2. Energy contraction

In this section, we will establish an energy contraction result for ILG (7). To this end, let $Y \subseteq X$ be a closed linear subspace of $X$ as in §1.2. Moreover, denote by $u^*_Y \in Y$ the unique solution of the equation

$$(F(u^*_Y), v) = 0 \quad \forall v \in Y,$$  

(10)

where existence and uniqueness of $u^*_Y$ follow again from the main theorem of strongly monotone operators.

**Proposition 2.1.** Consider the sequence $\{u^n_Y\}_{n \geq 0} \subset Y$ generated by the iteration (7). If (F1)–(F4) are satisfied and $a(u; \cdot, \cdot)$ fulfils (5)–(6) for any $u \in Y$, then there holds the energy contraction property

$$0 \leq E(u^{n+1}_Y) - E(u^*_Y) \leq q_{12}^2 [E(u^n_Y) - E(u^*_Y)] \quad \forall n \geq 0,$$  

(11)

with a contraction constant

$$0 \leq q_{12} := (1 - 2C \nu^2 \beta^{-2} L_T^{-1})^{1/2} < 1$$  

(12)

independent of the subspace $Y$ and of the iteration number $n$. In addition, the sequence $\{u^n_Y\}_{n \geq 0}$ converges to the unique solution $u^*_Y \in Y$ of (10).

**Proof.** With (F2) and since $u^*_Y$ is the (unique) solution of (10), for $n \geq 0$, we first observe that

$$\nu \|u^*_Y - u^n_Y\|^2_X \leq \langle F(u^*_Y) - F(u^n_Y), u^*_Y - u^n_Y \rangle \overset{(F2)}{=} \langle F(u^n_Y), u^*_Y - u^n_Y \rangle.$$  

Recalling that $f(u) = A[u]u - F(u)$, (7) and (6) prove that

$$\langle F(u^n_Y), u^n_Y - u^*_Y \rangle \overset{(7)}{=} a(u^n_Y; u^n_Y - u^*_Y, u^n_Y - u^*_Y) \leq \beta \|u^{n+1}_Y - u^n_Y\|_X \|u^n_Y - u^*_Y\|_X.$$  

Altogether, we derive the a posteriori error estimate

$$\|u^*_Y - u^n_Y\|_X \leq \beta \nu^{-1} \|u^n_Y - u^{n+1}_Y\|_X \quad \forall n \geq 0.$$  

(13)

Next, we exploit the structural assumptions (F1)–(F3) to obtain the well-known inequalities

$$\frac{\nu}{2} \|u^*_Y - v\|^2_X \leq E(v) - E(u^*_Y) \leq \frac{L_T}{2} \|u^*_Y - v\|^2_X \quad \forall v \in Y;$$  

(14)

see, e.g., [HW20b, Lem. 2.3] or [GHPS18, Lem. 5.1]. For any $n \geq 0$, we thus infer that

$$0 \overset{(14)}{\leq} E(u^{n+1}_Y) - E(u^*_Y) = E(u^*_Y) - E(u^n_Y) - [E(u^n_Y) - E(u^{n+1}_Y)]$$  

$$\overset{(8)}{\leq} E(u^n_Y) - E(u^*_Y) - C \|u^n_Y - u^*_Y\|^2_X$$  

$$\overset{(13)}{\leq} E(u^n_Y) - E(u^*_Y) - \frac{C \nu^2}{\beta^2} \|u^*_Y - u^n_Y\|^2_X$$  

$$\overset{(14)}{\leq} E(u^*_Y) - E(u^n_Y) - \frac{2C \nu^2}{\beta^2 L_T} [E(u^*_Y) - E(u^n_Y)] = q_{12}^2 [E(u^*_Y) - E(u^n_Y)].$$


This proves, in particular, that $0 \leq q_{12}^2 < 1$. Finally, iterating this inequality, we obtain that
\[
0 \leq \frac{\nu}{2} \|u_Y^n - u_Y^n\|_{L^2}^2 \leq \mathbb{E}(u_Y^n) - \mathbb{E}(u_Y^n) \leq q_{12}^{2n} \|\mathbb{E}(u_Y^n) - \mathbb{E}(u_Y^n)\|_{L^2} \quad \forall n \geq 0.
\]
In particular, we conclude that $u_Y^n \to u_Y^f$ in $Y$ as $n \to \infty$. □

3. Adaptive finite element discretization

In this section, we thoroughly formulate AILFEM. For each $N \geq 0$, the space $X_N$ in (9) will be a conforming finite element space that is associated to an admissible triangulation $\mathcal{T}_N$ of an underlying bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, with $d \geq 2$. Following the recent approach [GHPS20], we will present an adaptive iterative linearization finite element algorithm that exploits the interplay of adaptive mesh refinements and the iterative scheme (9). Throughout this section, we will always assume that $\mathcal{F}$ satisfies (F1)–(F4) and that (5) and (6) hold true.

3.1. Mesh refinements. We adopt the framework from [GHPS20, §2.2–2.4], with slightly modified notation. Consider a shape-regular mesh refinement strategy $\text{refine}(\cdot)$ such as, e.g., newest vertex bisection [Mit91]. For a subset $\mathcal{M}_N$ of marked elements in a regular triangulation $\mathcal{T}_N$, let $\text{refine}(\mathcal{T}_N, \mathcal{M}_N)$ be the coarsest regular refinement of $\mathcal{T}_N$ such that all elements $\mathcal{M}_N$ have been refined. Specifically, we write $\text{refine}(\mathcal{T}_N)$ for the set of all possible meshes that can be generated from $\mathcal{T}_N$ by (repeated) use of $\text{refine}(\cdot)$. For a mesh $\mathcal{T}_N^{ref} \in \text{refine}(\mathcal{T}_N)$, we assume the nestedness of the corresponding finite element spaces $X^{ref}_{\mathcal{N}}$ and $X_N$, respectively, i.e., $X_N \subseteq X^{ref}_{\mathcal{N}}$. In the sequel, starting from a given initial triangulation $\mathcal{T}_0$ of $\Omega$, we let $\mathcal{T} := \text{refine}(\mathcal{T}_0)$ be the set of all possible refinements of $\mathcal{T}_0$.

With regards to the optimal convergence rate of the algorithm with respect to the overall computational cost (see the next section §4), a few assumptions on the mesh refinement strategy are required, cf. [GHPS20, §2.8]. These are satisfied, in particular, for newest vertex bisection.

(R1) Splitting property: Each refined element is split into at least two and at most $C_{\text{ref}}$ many subelements, where $C_{\text{ref}} \geq 2$ is a generic constant. In particular, for all triangulations $\mathcal{T} \in \mathcal{T}$, and for any $\mathcal{M} \subseteq \mathcal{T}$, the (one-level) refinement $\mathcal{T}' := \text{refine}(\mathcal{T}, \mathcal{M})$ satisfies
\[
\#(\mathcal{T} \setminus \mathcal{T}') + \#(\mathcal{T}' \setminus \mathcal{T}) \leq C_{\text{ref}} \#(\mathcal{T} \setminus \mathcal{T}') + \#(\mathcal{T} \cap \mathcal{T}').
\]
Here, $\mathcal{T} \setminus \mathcal{T}'$ is the set of all elements in $\mathcal{T}$ which have been refined in $\mathcal{T}'$, and $\mathcal{T} \cap \mathcal{T}'$ comprises all unrefined elements.

(R2) Overlay estimate: For all meshes $\mathcal{T} \in \mathcal{T}$, and refinements $\mathcal{T}_1^{ref}, \mathcal{T}_2^{ref} \in \text{refine}(\mathcal{T})$ there exists a common refinement denoted by $\mathcal{T}_1^{ref} \oplus \mathcal{T}_2^{ref} \in \text{refine}(\mathcal{T}_1^{ref} \cap \text{refine}(\mathcal{T}_2^{ref})) \subseteq \text{refine}(\mathcal{T})$, which satisfies $\#(\mathcal{T}_1^{ref} \oplus \mathcal{T}_2^{ref}) \leq \#(\mathcal{T}_1^{ref}) + \#(\mathcal{T}_2^{ref}) - \#(\mathcal{T})$.

(R3) Mesh-closure estimate: There exists a constant $C_{\text{mesh}} > 0$ such that, for each sequence $\{\mathcal{T}_N\}_{N \geq 1}$ of successively refined meshes, i.e., $\mathcal{T}_{N+1} := \text{refine}(\mathcal{T}_N, \mathcal{M}_N)$ for some $\mathcal{M}_N \subseteq \mathcal{T}_N$, it holds that
\[
\#(\mathcal{T}_N - \mathcal{T}_0) \leq C_{\text{mesh}} \sum_{j=0}^{N-1} \#(\mathcal{M}_j) \quad \forall N \in \mathbb{N}.
\]

3.2. Error estimators. For a mesh $\mathcal{T}_N \in \mathcal{T}$ associated to a discrete space $X_N$, suppose that there exists a computable local refinement indicator $\eta_N : \mathcal{T}_N \times X_N \to \mathbb{R}$, with $\eta_N(T, v) \geq 0$ for all $T \in \mathcal{T}_N$ and $v \in X_N$. Then, for any $v \in X_N$ and $\mathcal{U}_N \subseteq \mathcal{T}_N$, let
\[
\eta_N(\mathcal{U}_N, v) := \left( \sum_{T \in \mathcal{U}_N} \eta_N(T, v)^2 \right)^{1/2} \quad \text{and} \quad \eta_N(v) := \eta_N(\mathcal{T}_N, v).
\]
We recall the following axioms of adaptivity from [CFPP14] for the refinement indicators: There are fixed constants $C_{\text{stab}}, C_{\text{cl}} \geq 1$ and $0 < q_{\text{red}} < 1$ such that, for all $\mathcal{T}_N \in \mathcal{T}$ and $\mathcal{T}_N^{ref} \in \text{refine}(\mathcal{T}_N)$, with associated refinement indicators $\eta_N$ and $\eta_N^{ref}$, the following properties hold, cf. [GHPS20, §2.8]:
η \in \text{the specific context of our application in section } §E \text{ approach in [HW20a]. The key idea is the same in all algorithms: On a given discrete space, we it-
eralization [GHPS20]), which was studied in the specific context of finite element discretizations
from [GHPS20, Prop. 3] and [GHPS18, Prop. 4.5 & 4.6]: For any \( \nu \in X \text{ of } (1) \) and the exact discrete
solution \( u_N^\ast \in X \text{ of } (10) \), we have the a posteriori error estimate \( \| u^\ast - u_N^\ast \|_X \leq C_{\text{rel}} \eta_N( u_N^\ast ) \).
(A4) Discrete reliability: \( \| u_N^{\text{ref}} - u_N^\ast \|_X \leq C_{\text{rel}} \eta_N( T_N \setminus T_N^{\text{ref}} , u_N^\ast ) \), where \( u_N^{\text{ref}} \in X_N^{\text{ref}} \) is the
solution of \( (10) \) on the discrete space \( X_N^{\text{ref}} \) associated with \( T_N^{\text{ref}} \).

We emphasize that (A1)–(A4) are satisfied for the usual \( h \)-weighted residual error estimators
in the specific context of our application in section §5, cf. (23).

3.3. AILFEM algorithm. We recall the adaptive algorithm from [GHPS18] (and its generalization [GHPS20]), which was studied in the specific context of finite element discretizations
of the Zanartunello iteration. These algorithms are closely related to the general adaptive ILG
approach in [HW20a]. The key idea is the same in all algorithms: On a given discrete space, we it-
erate the linearization scheme (9) as long as the linearization error estimator dominates. Once the
ratio of the linearization error estimator \( |E(u_N^\ast) - E(u_N^{n-1})|^{1/2} \) and the a posteriori error estimator
\( \eta_N(u_N^\ast) \) falls below a prescribed tolerance, the discrete space is refined appropriately.

Algorithm 1 AILFEM algorithm

\begin{enumerate}
\item Prescribe adaptivity parameters \( \lambda > 0, \theta < 1, \) and \( C_{\text{mark}} \geq 1 \). Moreover, set \( N := 0 \)
and \( n := 0 \). Start with an initial triangulation \( T_0 \), a corresponding finite element space \( X_0 \),
and an arbitrary initial guess \( u_0^0 \in X_0 \).
\item While true do
\item \hspace{0.5cm} Repeat with \( n \leftarrow 0 \)
\item \hspace{1cm} Perform a single iterative linearization step (9) on \( X_N \) to obtain \( u_N^{n+1} \) from \( u_N^n \).
\item \hspace{1cm} Update \( n \leftarrow n + 1 \).
\item \hspace{1cm} Until \( |E(u_N^\ast) - E(u_N^{n-1})|^{1/2} \leq \lambda \eta_N(u_N^\ast) \)
\item \hspace{1cm} Determine a marking set \( M_N \subseteq T_N \) with minimal cardinality (up to the multiplicative con-
stant \( C_{\text{mark}} \geq 1 \)) such that \( \theta \eta_N(u_N^\ast) \leq \eta_N(M_N, u_N^\ast) \), and set \( T_{N+1} := \text{refine}(T_N, M_N) \).
\item \hspace{1cm} Set \( n(N) := n \) and define \( u_N^{n+1} := u_N^{n(N)} \) by inclusion \( X_{N+1} \supseteq X_N \).
\item \hspace{1cm} Update \( N \leftarrow N + 1 \).
\end{enumerate}

Remark 3.1. Under the conditions (A1) and (A3), we notice some facts about Algorithm 1
from [GHPS20, Prop. 3] and [GHPS18, Prop. 4.5 & 4.6]: For any \( N \geq 0 \), there holds the a posteriori
error estimate
\[
\| u^\ast - u_N^{n(N)} \|_X \leq C_{16} \eta_N(u_N^{n(N)}) ,
\]
where \( C_{16} \) depends only on \( \nu, \beta, \lambda, L_F, C_{\text{rel}}, C_{\text{stab}} \), and \( C_{E} \). In particular, if the repeat loop termi-
nates with \( \eta_N(u_N^{n(N)}) = 0 \) for some \( N \geq 0 \), then \( u_N^{n(N)} = u^\ast \), i.e., the exact solution is obtained.
Moreover, in the non-generic case that the repeat loop in Algorithm 1 does not terminate after
finitely many steps, for some \( N \geq 0 \), the generated sequence \( \{ u_N^0 \}_{n \geq 0} \) converges to \( u_N^\ast = u^\ast \) (in
particular, the solution \( u^\ast \) is discrete).

4. Optimal convergence

We are now ready to outline the linear convergence of the sequence of approximations from
Algorithm 1 to the unique solution of (1). In addition, the rate optimality with respect to the
overall computational costs will be discussed.
4.1. Step counting. Following [GHPS20], we introduce an ordered index set
\[ Q := \{(N, n) \in \mathbb{N}^2_0 : \text{index pair } (N, n) \text{ occurs in Algorithm 1 and } n < n(N)\}, \]
where \( n(N) \geq 1 \) counts the number of steps in the repeat loop for each \( N \). We exclude the pair \((N, n(N))\) from \( Q \), since either \( u_N^0 = u_{n(N)}^N \) and \((N + 1, 0) \in Q\) or even \( n(N) := \infty \) if the loop does not terminate after finitely many steps; see also Remark 3.1. Observing that Algorithm 1 is sequential, the index set \( Q \) is naturally ordered: For \((N, n), (N', n') \in Q\), we write \((N', n') < (N, n)\) if and only if \((N', n')\) appears earlier in Algorithm 1 than \((N, n)\). With this order, we can define the total step counter
\[ |(N, n)| := \#\{(N', n') \in Q : (N', n') < (N, n)\} = n + \sum_{N'=0}^{N-1} n(N'), \]
which provides the total number of solver steps up to the computation of \( u_N^N \). Finally, we introduce the notation \( \mathfrak{T} := \sup \{N \in \mathbb{N} : (N, 0) \in Q\} \).

4.2. Linear convergence. Based on the notation used in Algorithm 1, let us introduce the quasi-error
\[ \Delta_N^n := \|u^* - u_N^N\|_X + \eta_N(u_N^N) \quad \forall (N, n) \in Q := Q \cup \{(N, n(N)) : n(N) < \infty\}, \quad (17) \]
and suppose that the estimator \( \eta_N \) satisfies (A1)–(A3). Then, appealing to the energy contraction property from Proposition 2.1, we make the crucial observation that ILG (7) satisfies the contraction property (C1) from [GHPS20]. Consequently, [GHPS20, Thm. 4] directly applies to our setting:

**Theorem 4.1** ([GHPS20, Thm. 4]). Suppose (A1)–(A3). Then, for all \( 0 < \theta \leq 1 \) and \( 0 < \lambda < \infty \), there exist constants \( C_{\text{lin}} \geq 1 \) and \( 0 < q_{\text{lin}} < 1 \) such that the quasi-error (17) is linearly convergent in the sense of
\[ \Delta_N^n \leq C_{\text{lin}} q_{\text{lin}}^{(N, n)-(N', n')} \Delta_N^{n'} \quad \forall (N, n), (N', n') \in Q \text{ with } (N', n') < (N, n). \]
The constants \( C_{\text{lin}} \) and \( q_{\text{lin}} \) depend only on \( L_F, \nu, \beta, C_E, C_{\text{stb}}, q_{\text{red}}, C_{\text{rel}} \), and the adaptivity parameters \( \theta \) and \( \lambda \).

4.3. Optimal convergence rate and computational work. Furthermore, we address the optimal convergence rate of the quasi-error (17) with respect to the degrees of freedom as well as the computational work. As before, we can directly apply a result from [GHPS20] owing to the energy contraction from Proposition 2.1. For its statement, we need further notation: First, for \( L \in \mathbb{N}_0 \), let \( T(L) \) be the set of all refinements \( T \) of \( T_0 \) with \# \( T \) = \# \( T_0 \) \leq L. Next, for \( s > 0 \), define
\[ \|u^*\|_{A_s} := \sup_{L \in \mathbb{N}} (L + 1)^s \inf_{T_{\text{opt}} \in T(L)} \{ \|u^* - u^*_s\|_X + \eta_{T_{\text{opt}}} (u^*_s) \} \in \mathbb{R} \geq 0 \cup \{\infty\}, \]
where \( u^*_s \) is the discrete solution (10) on the finite element space related to an optimal (in terms of the above infimum) mesh \( T_{\text{opt}} \). For \( s > 0 \), we note that \( \|u^*\|_{A_s} < \infty \) if and only if the quasi-error converges at least with rate \( s \) along a sequence of optimal meshes.

**Theorem 4.2** ([GHPS20, Thm. 7]). Suppose (R1)–(R3) and (A1)–(A4), and define
\[ \lambda_{\text{opt}} := \frac{1 - q_{\text{rel}}^2}{q_{\text{rel}}^2} \sqrt{s/2}. \]
Let \( 0 < \theta \leq 1 \) and \( 0 < \lambda < \lambda_{\text{opt}} \theta \) such that
\[ 0 < \theta' := \frac{\theta + \sqrt{s/2}}{1 - \sqrt{s/2}} \left(1 + C_{\text{stb}}^2 C_{\text{rel}}^2\right)^{-1/2}. \]
Then, for any \( s > 0 \), there exist positive constants \( c_{\text{opt}},C_{\text{opt}} \) such that
\[
\begin{align*}
&c_{\text{opt}} \|u^*\|_{\mathcal{H}_s} \leq \sup_{(N',n') \in \mathcal{Q}} (\#T_{N'} - \#T_0 + 1)^s \Delta_{N'}^s, \\
&\quad \leq \sup_{(N',n') \in \mathcal{Q}} \left( \sum_{(N,n) \in \mathcal{Q}} \#T_N \right)^s \Delta_{N'}^s \leq C_{\text{opt}} \max\{\|u^*\|_{\mathcal{H}_s}, \Delta_{0}^s\}. \tag{18}
\end{align*}
\]

Here, the constant \( c_{\text{opt}} > 0 \) depends only on \( \nu, L_F, C_{\text{ref}}, C_{\text{stab}}, C_{\text{rel}}, \#T_0, \) and \( s \), and additionally on \( \mathcal{R} \) and \( N_0 \), respectively, if \( \mathcal{R} < \infty \) or \( \eta_{N_0}(u_{N_0}) = 0 \) for some \( (N_0 + 1,0) \in \mathcal{Q} \), respectively; moreover, the constant \( C_{\text{opt}} > 0 \) depends only on \( C_{16}, \nu, C_{\text{stab}}, q_{\text{red}}, C_{\text{rel}}, C_{\text{mesh}}, 1 - \lambda/\lambda_{\text{opt}}, C_{\text{mark}}, C_{\text{lin}}, q_{\text{lin}}, \#T_0, \) and \( s \).

The significance of (18) is that the quasi-error \( \Delta_{N'}^s \) from (17) decays at rate \( s \) (with respect to the number of elements or, equivalently, the number of degrees of freedom) if and only if rate \( s \) is achievable for the discrete solutions on optimal meshes (with respect to the number of elements). If, in addition, all of the (single) steps in Algorithm 1 can be performed at linear cost, \( \mathcal{O}(\#T_N) \), then the quasi-error even decays with rate \( s \) with respect to the overall computational cost if and only if rate \( s \) is attainable with respect to the number of elements. We note that the total computational cost is proportional to the total computational time, which is therefore monitored in the subsequent numerical experiment.

5. Numerical experiment

In this section, we test Algorithm 1 with a numerical example.

5.1. Model problem. On an open, bounded, and polygonal domain \( \Omega \subset \mathbb{R}^2 \) with Lipschitz boundary \( \Gamma = \partial \Omega \), we consider the quasi-linear second-order elliptic boundary value problem:
\[
\begin{align*}
\text{Find } u \in X \text{ such that } \quad & F(u) := -\nabla \cdot \left\{ \mu(|\nabla u|^2) \nabla u \right\} - g = 0 \quad \text{in } X^*. \tag{19}
\end{align*}
\]

Here, \( X := H^1_0(\Omega) \) is the standard Sobolev space of \( H^1 \)-functions on \( \Omega \) with zero trace along \( \Gamma \). For \( u,v \in X \), the inner product and norm on \( X \) are defined by \( (u,v)_X := (\nabla u, \nabla v)_{L^2(\Omega)} \) and \( \|u\|_X := \|\nabla u\|_{L^2(\Omega)} \), respectively. We suppose that \( g \in X^* = H^{-1}(\Omega) \) in (19) is given, and the diffusion parameter \( \mu \in C^1([0,\infty)) \) fulfills the monotonicity property
\[
\begin{align*}
m_\mu(t-s) \leq \mu(t^2) - \mu(s^2) s \leq M_\mu(t-s) \quad \forall t \geq s \geq 0, \tag{20}
\end{align*}
\]
with constants \( M_\mu \geq m_\mu > 0 \). Under this condition, the nonlinear operator \( F : H^1_0(\Omega) \to H^{-1}(\Omega) \) from (19) satisfies (F1) and (F2) with \( \nu = m_\mu \) and \( L_F = 3M_\mu \); see [Zei90, Prop. 25.26]. Moreover, \( F \) has a potential \( E : X \to \mathbb{R} \) given by
\[
\begin{align*}
E(u) := \int_{\Omega} \psi(|\nabla u|^2) \, dx - \langle g, u \rangle \quad \forall u \in X,
\end{align*}
\]
where \( \psi(s) := 1/2 \int_0^s \mu(t) \, dt \), \( s \geq 0 \). The weak form of the boundary value problem (19) reads:
\[
\begin{align*}
\text{Find } u \in X \text{ such that } \quad & \int_{\Omega} \mu(|\nabla u|^2) \nabla u \cdot \nabla v \, dx = \langle g, v \rangle \quad \forall v \in X. \tag{21}
\end{align*}
\]

For the nonlinear boundary value problem (19), [HW20a, §5.1] proves the convergence of a few particular iterative linearization schemes, which can be cast into the general approach (4). These include the following:

(i) Zarantonello (or Picard) iteration, for \( \delta Z \in (0,2/m_\mu) \):
\[
-\Delta u^{n+1} = -\Delta u^n - \delta_Z F(u^n), \quad n \geq 0;
\]
cf. Zarantonello’s original report [Zar60] or the monographs [Neč86, §3.3] and [Zei90, §25.4]. We further point to [HW20b, Rem. 1] for the extended domain of the damping parameter \( \delta Z \).
(ii) Kačanov iteration:
\[- \nabla \cdot \left\{ \mu (|\nabla u|^2) \nabla u^{n+1} \right\} - g = 0, \quad n \geq 0; \tag{22}\]
cf. [Zeit90, §25.13].
(iii) Newton iteration, for a damping parameter \(0 < \delta_{\min} \leq \delta(u^n) \leq \delta_{\max} < 2m_M/M_\mu\):
\[F'(u^n)u^{n+1} = F'(u^n)u^n - \delta(u^n)F(u^n), \quad n \geq 0;\]

Here, for \(u \in X\), the Gâteaux derivative of \(F\) is given through
\[
\langle F'(u)v, w \rangle = \int_{\Omega} 2\mu(|\nabla u|^2)(\nabla u \cdot \nabla v)(\nabla u \cdot \nabla w) \, dx + \int_{\Omega} \mu(|\nabla u|^2)\nabla v \cdot \nabla w \, dx, \quad v, w \in X.
\]

5.2. Discretization and local refinement indicator. AILFEM for (21) is based on regular
triangular triangulations \(\{T_N\}_{N \geq 0}\) that partition the domain \(\Omega\) into open and disjoint triangles \(T \in T_N\) such that \(\Omega = \bigcup_{T \in T_N} T\). Moreover, we consider associated finite element spaces \(X_N := \{ v \in H^1_0(\Omega) : v|_T \in P_1(T) \; \forall T \in T_N \}\), where we signify by \(P_1(T)\) the space of all affine functions
on \(T \in T_N\). The mesh refinement strategy \(\text{refine}()\) in Algorithm 1 is given by newest vertex
bisection [Mit91]. Moreover, for any \(v \in X_N\) and any \(T \in T_N\), we define the local refinement
indicator respectively the global error indicator from (15) by
\[
\eta_N(T, v^2) := h_T^2 \|g\|^2_{L^2(T)} + h_T \left\| \left[ \mu(|\nabla v|^2) \nabla v \right] \right\|^2_{L^2(\partial T \setminus T)} ,
\]
\[
\eta_N(v) := \left( \sum_{T \in T_N} \eta_N(T, v^2) \right)^{1/2}, \tag{23}\]
where \(\|\cdot\|\) signifies the normal jump across element faces, and \(h_T := |T|^{1/2}\) is equivalent to the
diameter of \(T \in T\). This error estimator satisfies the assumptions (A1)–(A4) for the problem
under consideration; see, e.g., [GMZ12, §3.2] or [CFPP14, §10.1].

5.3. Computational example. Consider the L-shaped domain \(\Omega = (-1, 1)^2 \setminus ([0, 1] \times [-1, 0])\),
and the nonlinear diffusion parameter \(\mu(t) = 1 + e^{-t}\), which satisfies (20) with \(m_\mu = 1 - 2\exp(-\frac{1}{2})\)
and \(M_\mu = 2\). Moreover, we choose \(g\) such that the analytical solution of (19) is given by
\[
u^*(r, \varphi) = r^{3/5} \sin \left( \frac{2\varphi}{3} \right) \left( 1 - r \cos \varphi \right) \left( 1 + r \cos \varphi \right) \left( 1 - r \sin \varphi \right) \left( 1 + r \sin \varphi \right) \cos \varphi,
\]
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 \(\nu^*\) is unbounded at the origin.

In all our experiments below we set the adaptive mesh refinement parameters to \(\theta = 0.5\), and
\(C_{\text{mark}} = 1\). The computations employ an initial mesh \(T_0\) consisting of 192 uniform triangles, and
with the starting guess \(u^0_0 \equiv 0\). Then, the procedure is run until the number of elements exceeds
\(10^6\). We always choose the damping parameter \(\delta = 1\) for the Newton iteration, and vary the
damping parameter \(\delta_Z\) for the Zarantonello iteration, as well as the adaptivity parameter \(\lambda\), cf.
line 6 in Algorithm 1, throughout the experiments. Our implementation is based on the MATLAB
package [FPW11] with the necessary modifications.

In general, for the Newton scheme, we note that choosing the damping parameter to be \(\delta = 1\)
(potentially resulting in quadratic convergence of the iterative linearization close to the solution)
might lead to a divergent iteration for the given boundary value problem (cf. [AW14]). Our
numerical computations illustrate, however, that this is not of concern in the current experiments.
Indeed, for \(\delta = 1\), the bound (8) from (F4) remains satisfied in each iteration. Otherwise, a
prediction and correction strategy which obeys the bound (8), could be employed (see [HW20a,
Rem. 2.8]). This would guarantee the convergence of the (damped) Newton method.

(1) \(\delta_Z = 0.1\) and \(\lambda = 0.5\): In Figure 1, we display the performance of Algorithm 1 with respect to
both the number of elements and the total computational time. We clearly see a convergence
rate of \(-1/2\) for the Kačanov and Newton method, which is optimal for linear finite elements.
Moreover, the Zarantonello iteration has a pre-asymptotic phase of reduced convergence, which
Figure 1. $\delta_Z = 0.1$ and $\lambda = 0.5$: Performance plot for adaptively refined meshes with respect to the number of elements (left) and the total computational time (right). The solid and dashed lines correspond to the estimator and the error, respectively. The dashed lines without any markers indicate the optimal convergence order of $-1/2$ for linear finite elements.

Figure 2. $\delta_Z = 0.1$ and $\lambda = 0.5$: The contraction factor $\kappa_N$ (left, see (24)) and the number of iterations (right) on each finite element space.

becomes optimal for finer meshes. In Figure 2 (left) we observe that the energy contraction factor given by

$$\kappa_N := \frac{E(u_N^{n(N)}) - E(u^*)}{E(u_0^N) - E(u^*)}$$

is inferior for the Zarantonello iteration in the initial phase (compared to the Kačanov and Newton methods), which might explain the reduced convergence. This contraction factor becomes better for an increased number of iterations, see Figure 2 (right), thereby leading to the asymptotically optimal convergence rate for the Zarantonello iteration. Finally, in Figure 3 (left), we display the quotient

$$\kappa_N := \frac{E(u_N^{n(N)} - u_N^{n(N)-1})}{\|u_N^{n(N)} - u_N^{n(N)-1}\|_X}$$

which experimentally verifies the assumption (F4).

(2) $\delta_Z = 0.3$ and $\lambda = 0.1$: As before, in Figure 4, we display the performance of Algorithm 1 with respect to both the number of elements and the total computational time. We clearly observe
Figure 3. The quotient $\kappa_N$ from (25) on each finite element space for $\delta_Z = 0.1$ and $\lambda = 0.5$ (left) and $\delta_Z = 0.3$ and $\lambda = 0.1$ (right), respectively.

Figure 4. $\delta_Z = 0.3$ and $\lambda = 0.1$: Performance plot for adaptively refined meshes with respect to the number of elements (left) and the total computational time (right). The solid and dashed lines correspond to the estimator and the error, respectively. The dashed lines without any markers indicate the optimal convergence order of $-1/2$ for linear finite elements.

The optimal convergence rate of $-1/2$ for all of the three iteration schemes presented above from the initial mesh onwards. In contrast to the experiment before, the energy contraction factor $\kappa_N$ from (24) is now of comparable size for all the iteration schemes, as we can see from Figure 5 (left). Moreover, the number of iterations does not significantly differ for the three iterative methods. Again, we plot in Figure 3 (right) the quotient (25) for the numerical evidence of the assumption (F4).

(3) $\delta_Z = 0.3$ and $\lambda = 0.01$: Once more, we observe optimal convergence rate for all our three iteration schemes with respect to both the number of elements and the total computational time, see Figure 6. The total computational times obtained, however, differ noticeably, see Figure 6 (right), as a consequence of the varying number of iterative linearization steps of the three methods, see Figure 7 (right). In contrast, the energy contraction factor $\kappa_N$ from (24) almost coincides for the different iteration schemes, see Figure 7 (left), which is due to the small adaptivity parameter $\lambda = 0.01$. 
**Figure 5.** $\delta_{\mathcal{Z}} = 0.3$ and $\lambda = 0.1$: The contraction factor $\kappa_N$ (left, see (24)) and the number of iterations (right) on each finite element space.

**Figure 6.** $\delta_{\mathcal{Z}} = 0.3$ and $\lambda = 0.01$: Performance plot for adaptively refined meshes with respect to the number of elements (left) and the total computational time (right). The solid and dashed lines correspond to the estimator and the error, respectively. The dashed lines without any markers indicate the optimal convergence order of $-1/2$ for linear finite elements.

**Figure 7.** $\delta_{\mathcal{Z}} = 0.3$ and $\lambda = 0.01$: The contraction factor $\kappa_N$ (left, see (24)) and the number of iterations (right) on each finite element space.
6. Conclusions

We have established an energy contraction property for the iterative linearization Galerkin method (ILG, see (7)). This result is the decisive prerequisite to apply the convergence theorems from [GHPS20]. In particular, the sequence generated by the adaptive iterative linearized finite element method (AILFEM, see Algorithm 1) converges linearly and, under some additional assumptions, even with optimal rate with respect to the overall computational cost. This is confirmed by our numerical test in the context of quasi-linear elliptic problems, where we also underline that the theoretical constraints on the adaptivity parameters $\lambda > 0$ and $0 < \theta \leq 1$ are less restrictive in practice.

References


