SIAM J. NUMER. ANAL.
Vol. 60, No. 4, pp. 2318–2344

A HYBRID HIGH-ORDER METHOD FOR QUASILINEAR ELLIPTIC PROBLEMS OF NONMONOTONE TYPE*

THIRUPATHI GUDI†, GOURANGA MALLIK†, AND TAMAL PRAMANICK‡

Abstract. In this paper, we design and analyze a hybrid high-order approximation for a class of quasilinear elliptic problems of nonmonotone type. The proposed method has several advantages; for instance, it supports an arbitrary order of approximation and general polytopal meshes. The key ingredients involve local reconstruction and high-order stabilization terms. The existence of a unique discrete solution is shown by using Brouwer's fixed point theorem and the contraction principle. A priori error estimation is derived in a discrete energy norm that shows optimal order of convergence. Numerical experiments are performed to substantiate the theoretical results.

Key words. hybrid high-order methods, second-order quasilinear elliptic problems, Brouwer's fixed point theorem, error estimates

MSC codes. 65N12, 65N15, 65N30

DOI. 10.1137/21M1412050

1. Introduction.

We consider here numerical approximation for nonlinear elliptic boundary value problems of the type

$$-\nabla\cdot(a(x,u)\nabla u) = f(x) \quad \text{in} \quad \Omega, \tag{1.1a}$$

$$u(x) = 0 \quad \text{on} \quad \partial\Omega, \tag{1.1b}$$

where Ω is a bounded convex polytopal domain in $\mathbb{R}^d$, $d \in \{2,3\}$, with Lipschitz boundary $\partial\Omega$ and $a : \bar{\Omega}\times\mathbb{R} \to \mathbb{R}$ is a nonlinear function of its arguments. For simplicity, a homogeneous boundary condition is considered. Some additional assumptions are stated in the appropriate section. The main purpose of the article is to devise and analyze hybrid high-order (HHO) approximation for problem (1.1) on general meshes inspired by the HHO methods of [36] for a linear diffusion model problem, [31] for degenerate advection-diffusion-reaction models, and [29] for the nonlinear steady Leray–Lions equation. The proposed method became very famous over the last decade. The method has several advantages, such as HHO discretization supporting general polytopal meshes and allowing arbitrary order of polynomial approximations. The HHO method complies with physics and is robust with respect to the variations of physical coefficients. It focuses on the reproduction of key continuous properties at the discrete level, such as local balances and flux continuity. The computational cost for the HHO method can be reduced by using a compact stencil and static condensation.

Over the last several years, a major focus has been on analyzing discretization methods for partial differential equations (PDEs), which support arbitrary-order discretization on general meshes, including nonmatching interfaces and polytopal cells.

*Received by the editors April 14, 2021; accepted for publication (in revised form) June 6, 2022; published electronically August 30, 2022.
https://doi.org/10.1137/21M1412050
Funding: The work of the first author was supported by the DST MATRICS grant. The work of the second author was supported by the National Board for Higher Mathematics (NBHM) research grant 0204/58/2018/R&D-II/14746 and the DST C.V. Raman grant R(IA)/CVR-PDF/2020.
†Department of Mathematics, Indian Institute of Science, Bangalore 560012, India (gudi@iisc.ac.in, gourangam@iisc.ac.in).
‡Department of Mathematics, National Institute of Technology Calicut, Kozhikode 673601, India (tamal@nitc.ac.in).

A lot of research has been done on the arbitrary-order polytopal method for linear diffusion equations. To name a few, we refer to [2, 5] for the constructions of polyhedral elements for adaptive mesh coarsening, the hybridizable discontinuous Galerkin method of [28, 24], the virtual element method of [6, 7, 16], the high-order mimetic method of [60], the weak Galerkin method of [64, 66, 67], the gradient discretization methods of [46, 44, 33], and the multiscale hybrid mixed method of [3]. Recently, a robust HHO approximation scheme for the Poisson problem on polytopal meshes with small edges/faces was analyzed in [47], where the standard scaling of h_F for the stabilization is replaced by h_T. There are some connections of the HHO method with the hybridizable discontinuous Galerkin (HDG) method, but the choice of stabilization for the HHO method is different from that of HDG to deliver a higher-order convergence rate for HHO discretization. On the other hand, if we consider the nonconforming virtual element methods (ncVEM), the devising viewpoint with HHO is different. ncVEM considers the computable projection of virtual functions instead of a reconstruction operator, and the stabilization for both methods achieves similar convergence rates as HHO but is written differently. The close connections of the HHO method with HDG [28] and ncVEM [4] have been analyzed in [27].

The HHO method in the lowest-order case belongs to the hybrid mixed mimetic family [45], which includes the mixed-hybrid mimetic finite differences [19], the hybrid finite volume [48], and the mixed finite volume [42, 43]. Recently, the HHO method has been bridged in [59] with the virtual element method. Some more related approaches can also be found in [14, 45, 17, 18, 58]. Several works on HHO methods involving linear and nonlinear PDEs can be found in various articles, such as pure diffusion [36], advection-diffusion [31], interface problems [20], the viscosity-dependent Stokes problem [37] for linear PDEs, the elliptic obstacle problem [26], a nonlinear elasticity with infinitesimal deformations [15], steady incompressible Navier–Stokes equations, [38], and Leray–Lions operators [29, 32] for nonlinear PDEs.

The quasilinear problem of nonmonotone type (1.1) can be viewed as a stationary heat problem with variable nonlinear diffusion coefficient. This has many engineering applications, for instance, heat distribution for metal bodies. Finite element approximations for nonlinear problem (1.1) have been studied in [54, 53] for a priori error estimation using discontinuous Galerkin (DG) and *hp*-DG methods and in [11, 9, 12, 65] for various a priori and a posteriori error estimates. For some related works on *hp*-DG methods for strongly nonlinear elliptic problems, we refer to [52, 13] and references therein. We also refer to articles on various nonlinear problems for second-order elliptic PDEs, the weak Galerkin methods of [64], the mimetic finite difference approximation of [1], and the virtual element method of [22]. The finite element approximations for quasilinear problems are studied under various regularity assumptions on the coefficient a and on the solution u, for instance, a in C^2 and $u \in W^{2,2+\epsilon}(\Omega), \epsilon > 0$, or $u \in H^2(\Omega)$; see [68, 54, 10, 11, 41, 61, 63].

In this article, we establish an optimal-order a priori error estimate in a discrete energy norm for the HHO approximation for a quasilinear elliptic problem of nonmonotone type. We assume the solution $u \in H_0^1(\Omega)$ of (1.1) belongs to $W^{2,\infty}(\Omega)$ for $d = 2$ and belongs to $H^3(\Omega) \cap W^{2,\infty}(\Omega)$ for $d = 3$. We use local reconstruction and high-order stabilization in the discrete formulation. First, we establish the existence, uniqueness, and a priori error estimate for an auxiliary second-order non-self-adjoint linear elliptic problem satisfying Gårding-type inequality. The well-posedness of this auxiliary problem helps us to formulate a suitable nonlinear map that possesses a ball to ball mapping and contraction properties. For sufficiently small mesh parameters, the existence, uniqueness, and a priori error estimate for the solution of HHO approximation of quasilinear problems are derived by using Brouwer's fixed point theorem and the contraction principle on a quasiuniform mesh.

The organization of the paper is as follows. Section 1 is introductory in nature, and section 2 is devoted to notation, definition, and preliminaries related to HHO discretization. In section 3, we discuss the HHO approximation for a linear non-self-adjoint elliptic problem and establish error estimates. Section 4 is devoted to the HHO approximation for the solution of quasilinear elliptic problem (4.1). In section 5, numerical experiments are performed to illustrate the theoretical results obtained in this article. Finally, in section 6, we present a summary of the article and describe some possible extensions.

Throughout the paper, standard notation on Lebesgue and Sobolev spaces and their norms are employed. The standard seminorm and norm on $H^s(\Omega)$ (resp., $W^{s,p}(\Omega)$) for $s > 0$ are denoted by $|\bullet|_s$ and $\|\bullet\|_s$ (resp., $|\bullet|_{s,p}$ and $\|\bullet\|_{s,p}$). The positive constants C appearing in the inequalities denote generic constants which do not depend on the meshsize. The notation $a \lesssim b$ means that there exists a generic constant C independent of the mesh parameters such that $a \leq Cb$; $a \approx b$ abbreviates $a \lesssim b \lesssim a$.

2. HHO discretization.

2.1. Discrete setting. We consider a sequence of meshes $(\mathcal{T}_h)_{h>0}$, where the parameter h denotes the meshsize and goes to zero during the refinement process. For all $h > 0$, we assume that the mesh $\mathcal{T}_h$ covers Ω exactly and consists of a finite collection of nonempty disjoint open polyhedral cells T such that $\overline{\Omega} = \cup_{T\in\mathcal{T}_h}\overline{T}$ and $h = \max_{h\in\mathcal{T}_h} h_T$, where h_T is the diameter of T. A closed subset F of Ω is defined to be a mesh face if it is a subset of an affine hyperplane H_F with positive $(d-1)$-dimensional Hausdorff measure and if either of the following two statements holds true: (i) there exist $T_1(F)$ and $T_2(F)$ in $\mathcal{T}_h$ such that $F \subset \partial T_1(F) \cap \partial T_2(F) \cap H_F$; in this case, the face F is called an internal face; (ii) there exists $T(F) \in \mathcal{T}_h$ such that $F \subset \partial T(F) \cap \partial\Omega \cap H_F$; in this case, the face F is called a boundary face. The set of mesh faces is a partition of the mesh skeleton, i.e., $\cup_{T\in\mathcal{T}_h}\partial T = \cup_{F\in\mathcal{F}_h}\bar{F}$, where $\mathcal{F}_h := \mathcal{F}_h^i \cup \mathcal{F}_h^b$ is the collection of all faces that is the union of the set of all the internal faces $\mathcal{F}_h^i$ and the set of all the boundary faces $\mathcal{F}_h^b$. Let h_F denote the diameter of $F \in \mathcal{F}_h$. For each $T \in \mathcal{T}_h$, the set $F_T := \{F \in \mathcal{F}_h \,|\, F \subset \partial T\}$ denotes the collection of all faces contained in ∂T, $\boldsymbol{n}_T$ denotes the unit outward normal to T, and we set $\boldsymbol{n}_{TF} := \boldsymbol{n}_T|_F$ for all $F \in \mathcal{F}_h$. Following [35, Definition 1], we assume that the mesh sequence $(\mathcal{T}_h)_{h>0}$ is admissible in the sense that, for all $h > 0$, $\mathcal{T}_h$ admits a matching simplicial submesh $\mathcal{T}_h$ (i.e., every cell and face of $\mathcal{T}_h$ is a subset of a cell and a face of $\mathcal{T}_h$, respectively) so that the mesh sequence $(\mathcal{T}_h)_{h>0}$ is shape-regular in the usual sense and all the cells and faces of $\mathcal{T}_h$ have a uniformly comparable diameter to the cell and face of $\mathcal{T}_h$ to which they belong. Owing to [34, Lemma 1.42], for $T \in \mathcal{T}_h$ and $F \in \mathcal{F}_T$, h_F is comparable to h_T in the sense that

$$\varrho^2 h_T \leq h_F \leq h_T,$$

where ϱ is the mesh regularity parameter. Moreover, there exists an integer N_∂ depending on ϱ and d such that (see [34, Lemma 1.41]):

$$\max_{T\in\mathcal{T}_h} \operatorname{card}(\mathcal{F}_T) \leq N_\partial.$$

There also exist real numbers C_{tr} and $C_{\mathrm{tr,c}}$ depending on ϱ but independent of h such that the following discrete and continuous trace inequalities hold for all $T \in \mathcal{T}_h$ and $F \in \mathcal{F}_T$ (see [34, Lemmas 1.46 and 1.49]):

$$\|v\|_F \leq C_{\mathrm{tr}} h_F^{-1/2} \|v\|_T \quad \forall v \in \mathbb{P}_d^l(T), \tag{2.1}$$

$$\|v\|_{\partial T} \leq C_{\mathrm{tr,c}} (h_T^{-1} \|v\|_T^2 + h_T \|\nabla v\|_T^2)^{1/2} \quad \forall v \in H^1(T), \tag{2.2}$$

where $\mathbb{P}_d^l(T)$ is the space of polynomial of degree at most l on $T \in \mathcal{T}_h$. There exists a real number C_{app} depending on ϱ and l but independent of h such that, for all $T \in \mathcal{T}_h$, denoting by π_T^l the L^2-orthogonal projector on $\mathbb{P}_d^l(T)$, the following holds (see [34, Lemmas 1.58 and 1.59]): For all $s \in \{1, \ldots, l+1\}$ and all $v \in H^s(T)$,

$$|v - \pi_T^l v|_{H^m(T)} + h_T^{1/2} |v - \pi_T^l v|_{H^m(\partial T)} \leq C_{\mathrm{app}} h_T^{s-m} |v|_{H^s(T)} \quad \forall m \in \{0, \ldots, s-1\}, \tag{2.3}$$

where $|\bullet|_{H^m(\partial T)}$ denotes the facewise H^m-seminorm when the boundary ∂T of an element $T \in \mathcal{T}_h$ is written as a union of faces.

2.2. Discrete spaces. Let $k \geq 0$ be a fixed polynomial degree. For $T \in \mathcal{T}_h$, define the local space of degrees of freedom (DOFs) by

$$\underline{U}_T^k := \mathbb{P}_d^k(T) \times \left\{ \underset{F \in \mathcal{F}_T}{\times} \mathbb{P}_{d-1}^k(F) \right\}, \tag{2.4}$$

where $\mathbb{P}_d^k(T)$ is the space of polynomials of degree at most k on $T \in \mathcal{T}_h$ and $\mathbb{P}_{d-1}^k(F)$ is the space of polynomial of degree at most k on the face $F \in \mathcal{F}_h$. The global space of DOFs is obtained by patching interface values in (2.4) as

$$\underline{U}_h^k := \left\{ \underset{T \in \mathcal{T}_h}{\times} \mathbb{P}_d^k(T) \right\} \times \left\{ \underset{F \in \mathcal{F}_h}{\times} \mathbb{P}_{d-1}^k(F) \right\}.$$

The zero boundary condition can be imposed in the above discrete space $\underline{U}_h^k$ as follows:

$$\underline{U}_{h,0}^k := \left\{ \underline{v}_h = ((v_T)_{T \in \mathcal{T}_h}, (v_F)_{F \in \mathcal{F}_h}) \in \underline{U}_h^k \,|\, v_F \equiv 0 \quad \forall F \in \mathcal{F}_h^b \right\}.$$

For $\underline{v}_h \in \underline{U}_h^k$, we understand $v_h \in L^2(\Omega)$ by $v_h|_T = v_T$. The local interpolation operator $I_T^k : H^1(T) \to \underline{U}_T^k$ is such that, for all $v \in H^1(T)$,

$$I_T^k v := (\pi_T^k v, (\pi_F^k v)_{F \in \mathcal{F}_T}), \tag{2.5}$$

where π_F^k is the L^2-orthogonal projector on $\mathbb{P}_{d-1}^k(F)$. The corresponding global interpolation operator $I_h^k : H^1(\Omega) \to \underline{U}_h^k$ is such that, for all $v \in H^1(\Omega)$,

$$I_h^k v := ((\pi_T^k v)_{T \in \mathcal{T}_h}, (\pi_F^k v)_{F \in \mathcal{F}_h}).$$

When applied to $H_0^1(\Omega)$, I_h^k maps onto $\underline{U}_{h,0}^k$.

Below, we state the Lebesgue embedding result. For proof, we refer to [29, Lemma 5.1].

LEMMA 2.1 (direct and reverse Lebesgue embeddings). *Let $\mathcal{T}_h$ be a regular mesh with $T \in \mathcal{T}_h$. Let $k \in \mathbb{N}$ and $q, m \in [1, \infty]$. Then*

$$\|w\|_{L^q(T)} \approx |T|^{\frac{1}{q} - \frac{1}{m}} \|w\|_{L^m(T)} \quad \forall w \in \mathbb{P}^k(T). \tag{2.6}$$

Define the Sobolev exponent p^* of p by

$$p^* := \begin{cases} \frac{dp}{d-p} & \text{if } p < d, \\ +\infty & \text{if } p \geq d. \end{cases}$$

The following lemma on discrete Sobolev embeddings is stated from [29, Proposition 5.4], and this is used to obtain various boundedness results in the following sections.

Lemma 2.2 (discrete Sobolev embeddings). *Let $(\mathcal{T}_h)_{h>0}$ be an admissible mesh sequence of $\Omega \subset \mathbb{R}^d$. Let $1 \leq q \leq p^*$ if $1 \leq p < d$ and $1 \leq q < \infty$ if $p \geq d$. Then there exists C only depending on $\Omega, \varrho, k, q,$ and p such that*

$$\|v_h\|_{L^q(\Omega)} \leq C\|\underline{v}_h\|_{1,p,h} \quad \forall \underline{v}_h \in \underline{U}_{h,0}^k,$$

where

$$\|\underline{v}_h\|_{1,p,h} := \left(\sum_{T\in\mathcal{T}_h} \|\underline{v}_T\|_{1,p,T} \right)^{1/p}$$

with

$$\|\underline{v}_T\|_{1,p,T} := \left(\|\nabla v_T\|_{L^p(T)^d}^p + \sum_{F\in\mathcal{F}_T} h_F^{1-p} \|v_F - v_T\|_{L^p(T)}^p \right)^{1/p}.$$

In particular,

$$\|v_h\|_{L^6(\Omega)} \leq C\|\underline{v}_h\|_{1,2,h} \quad \forall \underline{v}_h \in \underline{U}_{h,0}^k. \tag{2.7}$$

We use the abbreviation $\|\bullet\|_{1,h}$ for $\|\bullet\|_{1,2,h}$ in the subsequent analysis.

2.3. Local reconstructions and stabilization operators. In this subsection, some essential ingredients related to HHO formulation are defined. For $T \in \mathcal{T}_h$, we define the local reconstruction operator $R_T^{k+1} : \underline{U}_T^k \to \mathbb{P}_d^{k+1}(T)$ such that, for $\underline{v}_T = (v_T, (v_F)_{F\in\mathcal{F}_T})$,

$$(\nabla R_T^{k+1}\underline{v}_T, \nabla w)_T = (\nabla v_T, \nabla w)_T + \sum_{F\in\mathcal{F}_T} (v_F - v_T, \nabla w \cdot \boldsymbol{n}_{TF})_F, \tag{2.8a}$$

$$\left(R_T^{k+1}\underline{v}_T, 1\right)_T = (v_T, 1)_T\,, \tag{2.8b}$$

where (2.8a) is enforced for all $w \in \mathbb{P}_d^{k+1}(T)$. A global reconstruction operator $R_h^{k+1} : \underline{U}_h^k \to \mathbb{P}_d^{k+1}(\mathcal{T}_h)$ is defined by $R_h^{k+1}\underline{v}_h|_T = R_T^{k+1}\underline{v}_T$.

Define the local gradient reconstruction $\boldsymbol{G}_T^k : \underline{U}_T^k \to \mathbb{P}^k(T)^d$ such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$(\boldsymbol{G}_T^k\underline{v}_T, \boldsymbol{\tau})_T = (\nabla v_T, \boldsymbol{\tau})_T + \sum_{F\in\mathcal{F}_T} (v_F - v_T, \boldsymbol{\tau}\cdot\boldsymbol{n}_{TF})_F \quad \forall \boldsymbol{\tau} \in \mathbb{P}^k(T)^d. \tag{2.9}$$

Moreover, from [30, Lemma 4.10], it holds that

$$(\boldsymbol{G}_T^k\underline{v}_T, \boldsymbol{\tau})_T = (\nabla v_T, \boldsymbol{\tau})_T + \sum_{F\in\mathcal{F}_T} (v_F - v_T, (\pi_T^k\boldsymbol{\tau})\cdot\boldsymbol{n}_{TF})_F \quad \forall \boldsymbol{\tau} \in L^1(T)^d. \tag{2.10}$$

The relation between $\boldsymbol{G}_T^k$ and R_T^{k+1} is established by taking $\boldsymbol{\tau} = \nabla w$ with $w \in \mathbb{P}_d^{k+1}(T)$ in (2.8) and comparing with (2.9) as

$$(\boldsymbol{G}_T^k \underline{v}_T - \nabla R_T^{k+1} \underline{v}_T, \nabla w)_T = 0 \quad \forall w \in \mathbb{P}^{k+1}(T). \tag{2.11}$$

In other words, $\nabla R_T^{k+1} \underline{v}_T$ is the L^2-orthogonal projection of $\boldsymbol{G}_T^k \underline{v}_T$ on $\nabla \mathbb{P}^{k+1}(T) \subset P^k(T)^d$ and $\|\nabla R_T^{k+1} \underline{v}_T\|_T \leq \|\boldsymbol{G}_T^k \underline{v}_T\|_T$.

For $F \in \mathcal{F}_T$, define the local stabilization operator $S_F^k : \underline{U}_T^k \to \mathbb{P}_{d-1}^k(F)$ by

$$S_F^k \underline{v}_T := \pi_F^k \left(v_F - v_T - \left(R_T^{k+1} \underline{v}_T - \pi_T^k R_T^{k+1} \underline{v}_T \right) \right). \tag{2.12}$$

The next lemma follows from the property $R_T^{k+1} I_T^k v = \pi_T^{1,k+1} v$ for $v \in W^{1,1}(T)$, where $\pi_T^{1,k+1}$ is the elliptic projector (see [30, Definition 1.39]) and its approximation property [30, Theorem 1.48] is given by the following lemma.

LEMMA 2.3 (approximation properties of $R_T^{k+1} I_T^k$). *There exists a real number $C > 0$ depending on ϱ but independent of h_T such that, for all $v \in H^{s+1}(T)$ for some $s \in \{0, 1, \ldots, k+1\}$,*

$$\begin{aligned} &\|v - R_T^{k+1} I_T^k v\|_T + h_T^{1/2} \|v - R_T^{k+1} I_T^k v\|_{\partial T} + h_T \|\nabla(v - R_T^{k+1} I_T^k v)\|_T \\ &\quad + h_T^{3/2} \|\nabla(v - R_T^{k+1} I_T^k v)\|_{\partial T} \leq C h_T^{s+1} \|u\|_{H^{s+1}(T)}. \end{aligned} \tag{2.13}$$

The property $\boldsymbol{G}_T^k I_T^k v = \pi_T^k(\nabla v)$ for $v \in W^{1,1}(T)$ and the approximation property for L^2 projector π_T^k lead to the next lemma.

LEMMA 2.4 (approximation properties of $\boldsymbol{G}_T^k I_T^k$). [30, Lemma 3.24] *There exists a real number $C > 0$ depending on ϱ but independent of h_T such that, for all $v \in H^{s+1}(T)$ for some $s \in \{0, 1, \ldots, k+1\}$,*

$$\|\nabla v - \boldsymbol{G}_T^k I_T^k v\|_{L^2(T)} \leq C h_T^s |u|_{H^{s+1}(T)}. \tag{2.14}$$

3. Non-self-adjoint linear elliptic problems and error estimate. To analyze the existence and uniqueness of the discrete solution of (1.1), we require a linearized problem which is essentially a non-self-adjoint elliptic problem (the explicit form is described in section 4). This motivates us to consider a general non-self-adjoint elliptic PDE:

$$-\nabla \cdot (a(x) \nabla u) + \vec{b}(x) \cdot \nabla u + a_0(x) u = p(x) \quad \text{in} \quad \Omega, \tag{3.1a}$$

$$u = 0 \quad \text{on} \quad \partial\Omega, \tag{3.1b}$$

where Ω is a bounded convex polytopal domain in $\mathbb{R}^d$, $d \in \{2, 3\}$, with Lipschitz boundary $\partial\Omega$. In this section, we consider the HHO approximation and establish an error estimate for the discrete solution of (3.1). The existence and uniqueness results will be used in the error analysis of the HHO approximation for the quasilinear elliptic problem (1.1). We adopt the following Assumptions A.1–A.4 on the above problem (3.1).

*Assumption A.*1. There exists $\alpha > 0$ such that $\alpha \leq a(x)$ and $a_0(x) \geq 0$ for all $x \in \overline{\Omega}$.

*Assumption A.*2. The functions $a \in C^1(\bar{\Omega})$, $\vec{b} \in \boldsymbol{L}^\infty(\Omega)$ and $a_0 \in L^\infty(\Omega)$ with $M := \max\{\|a\|_{L^\infty(\Omega)}, \|\vec{b}\|_{\boldsymbol{L}^\infty(\Omega)}, \|a_0\|_{L^\infty(\Omega)}\}$.

Assumption A.3. The load function $p \in L^2(\Omega)$.

The model problem (3.1) has a unique solution $u \in H_0^1(\Omega)$ under the above Assumptions A.1–A.3; see [49, Theorem 8.9].

Assumption A.4. The solution $u \in H_0^1(\Omega)$ of (3.1) satisfies the regularity result

$$\|u\|_{H^2(\Omega)} \le C\|p\|_{L^2(\Omega)}. \tag{3.2}$$

It can be observed that under Assumptions A.1–A.3, the last Assumption A.4 on the regularity is verified if the boundary $\partial\Omega$ is of class $C^{1,1}$; see [49, Lemma 9.17] and [50, Theorem 2.4.2.5].

3.1. HHO discretization of a non-self-adjoint linear elliptic problem. Define the discrete counterpart of $(a_T\nabla u, \nabla v)_T$, where $a_T(x) = a(x)|_T$, as the local discrete bilinear form

$$\begin{aligned} B_T(\underline{u}_T, \underline{v}_T) &:= \left(a_T \boldsymbol{G}_T^k \underline{u}_T, \boldsymbol{G}_T^k \underline{v}_T\right)_T + s_T(\underline{u}_T, \underline{v}_T) \\ &\quad + \left(\vec{b}\cdot\nabla R_T^{k+1}(\underline{u}_T), v_T\right)_T + (a_0 u_T, v_T)_T, \end{aligned} \tag{3.3}$$

where the stabilization term $s_T(\underline{u}_T, \underline{v}_T)$ is defined by

$$s_T(\underline{u}_T, \underline{v}_T) := \sum_{F\in\mathcal{F}_T} \frac{a_{TF}^\infty}{h_F}\left(S_F^k \underline{u}_T, S_F^k \underline{v}_T\right)_F \text{ with } a_{TF}^\infty := \|a_T\|_{L^\infty(F)}. \tag{3.4}$$

The discrete formulation of (3.1) seeks $\underline{u}_h \in \underline{U}_{h,0}^k$ such that

$$B_h(\underline{u}_h, \underline{v}_h) = (p, v_h) \quad \forall \underline{v}_h \in \underline{U}_{h,0}^k, \tag{3.5}$$

where $B_h(\underline{u}_h, \underline{v}_h) := \sum_{T\in\mathcal{T}_h} B_T(\underline{u}_T, \underline{v}_T)$.

Define the energy seminorms on $\underline{U}_h^k$ (norms on $\underline{U}_{h,0}^k$ owing to the homogeneous boundary condition) as follows:

$$\|\underline{v}_h\|_{a,h}^2 := \sum_{T\in\mathcal{T}_h} \|\underline{v}_T\|_{a,T}^2 \text{ and } \|\underline{v}_h\|_{1,h} := \|\underline{v}_h\|_{1,2,h} \quad \forall \underline{v}_h \in \underline{U}_h^k, \tag{3.6}$$

where the local contributions are defined as

$$\|\underline{v}_T\|_{a,T}^2 := \|a_T^{1/2}\boldsymbol{G}_T^k \underline{v}_T\|_T^2 + \sum_{F\in\mathcal{F}_T} \frac{a_{TF}^\infty}{h_F}\|v_F - v_T\|_F^2. \tag{3.7}$$

With the help of Assumptions A.1–A.2 and [30, Lemma 3.15], it can be easily shown that the above seminorms $\|\bullet\|_{a,h}$ and $\|\bullet\|_{1,h}$ are equivalent on $\underline{U}_h^k$.

The following boundedness property can be obtained using the definitions of reconstructions $\boldsymbol{G}_T^k, R_T^{k+1}$, the trace inequality, and the Cauchy–Schwarz inequality; see [30, Proposition 3.13 and Lemma 3.15].

LEMMA 3.1 (boundedness). *For $\underline{u}_h, \underline{v}_h \in \underline{U}_h^k$, there exists a constant $C > 0$ independent of mesh parameter h but depending on $\alpha, M, C_{\mathrm{tr}}, C_{\mathrm{app}}, \varrho, N_\partial$ such that*

$$B_h(\underline{u}_h, \underline{v}_h) \le C\left(\|\underline{u}_h\|_{a,h}^2 + \|u_h\|^2\right)^{1/2}\left(\|\underline{v}_h\|_{a,h}^2 + \|v_h\|^2\right)^{1/2}.$$

The following lemma is essential for establishing the existence and uniqueness of the discrete solution (3.5).

LEMMA 3.2 (Gårding-type inequality). *For all $\underline{v}_h \in \underline{U}_h^k$, there exist two real numbers $C_1, C_2 > 0$ independent of h such that*

$$B_h(\underline{v}_h, \underline{v}_h) \geq C_1 \|\underline{v}_h\|_{a,h}^2 - C_2 \|v_h\|^2 \quad \forall \underline{v}_h \in \underline{U}_h^k. \tag{3.8}$$

Proof. Following the proof of [30, Lemma 31.5], the lower bound for $B_h(\bullet, \bullet)$ can be easily obtained as follows:

$$\left(a_T \boldsymbol{G}_T^k \underline{v}_T, \boldsymbol{G}_T^k \underline{v}_T\right)_T + s_T(\underline{v}_T, \underline{v}_T) \geq c_3 \|\underline{v}_h\|_{a,h}^2 \tag{3.9}$$

for some positive constant c_3. Using the Cauchy–Schwarz inequality, we obtain the following estimates for the remaining terms:

(3.10)

$$\sum_{T \in \mathcal{T}_h} \left(\vec{b} \cdot \nabla R_T^{k+1} \underline{v}_T, v_T\right)_T \lesssim \|\vec{b}\|_{L^\infty(\Omega)} \|\underline{v}_h\|_{a,h} \left(\sum_{T \in \mathcal{T}_h} \|v_T\|_T^2\right)^{1/2} \leq c_4 \|\underline{v}_h\|_{a,h} \|v_h\|,$$

$$\sum_{T \in \mathcal{T}_h} (a_0 v_T, v_T)_T \leq \|a_0\|_{L^\infty(\Omega)} \sum_{T \in \mathcal{T}_h} \|v_T\|_T^2 \leq c_5 \|v_h\|^2 \tag{3.11}$$

for some positive constant c_4 and c_5.

Now, we prove the Gårding-type inequality for $B_h(\underline{v}_h, \underline{v}_h)$. The definition of $B_h(\underline{v}_h, \underline{v}_h)$ and a use of the above three estimates (3.9)–(3.11) lead to

$$\begin{aligned} B_h(\underline{v}_h, \underline{v}_h) &\geq c_3 \|\underline{v}_h\|_{a,h}^2 - c_4 \|\underline{v}_h\|_{a,h} \|v_h\| - c_5 \|v_h\|^2 \\ &\geq (c_3 - c_4 \zeta) \|\underline{v}_h\|_{a,h}^2 - \left(\frac{c_4}{\zeta} + c_5\right) \sum_{T \in \mathcal{T}_h} \|v_T\|_T^2 \end{aligned}$$

for any positive ζ. For sufficiently small $\zeta > 0$, the proof of (3.8) follows. ∎

Using Lemma 2.2, we have $\|v_h\| \leq C \|\underline{v}_h\|_{a,h}$. We rewrite the Gårding-type inequality (3.8) as

$$C_1 \|\underline{v}_h\|_{a,h} \leq \sup_{\underline{w}_h \in \underline{U}_h^k,\, \|\underline{w}_h\|_{a,h} = 1} B_h(\underline{v}_h, \underline{w}_h) + C_2 \|v_h\| \quad \forall \underline{v}_h \in \underline{U}_h^k. \tag{3.12}$$

3.2. Existence and uniqueness of the discrete solution of a non-self-adjoint elliptic problem. We now show an a priori result of a discrete auxiliary problem which is used to prove the existence and uniqueness of the discrete solution $\underline{u}_h \in \underline{U}_{h,0}^k$ of (3.5).

LEMMA 3.3. *Let $q \in L^2(\Omega)$. Then for sufficiently small h, there exists a unique solution $\underline{\phi}_h \in \underline{U}_{h,0}^k$ such that*

$$B_h(\underline{v}_h, \underline{\phi}_h) = (q, v_h) \quad \forall \underline{v}_h \in \underline{U}_{h,0}^k. \tag{3.13}$$

Moreover, the solution $\underline{\phi}_h$ satisfies

$$\|\underline{\phi}_h\|_{a,h} \leq C \|q\|. \tag{3.14}$$

Proof. We first prove the a priori bound (3.14). Since (3.13) leads to a finite-dimensional system, the existence and uniqueness of the solution follow from the a priori bound. The Gårding-type inequality (3.8) with $\underline{v}_h = \underline{\phi}_h$ yields

$$C_1 \|\underline{\phi}_h\|_{a,h}^2 - C_2 \|\phi_h\|^2 \le B_h(\underline{\phi}_h, \underline{\phi}_h).$$

Using (3.13) and the Cauchy–Schwarz inequality, we obtain

$$B_h(\underline{\phi}_h, \underline{\phi}_h) = (q, \phi_h) \le \|q\|_{L^2(\Omega)} \|\phi_h\|_{L^2(\Omega)} \le (\|q\|^2 + \|\phi_h\|^2)/2.$$

Combining the above two estimates, we have

$$\|\underline{\phi}_h\|_{a,h} \le C_3 \|q\| + C_4 \|\phi_h\|. \tag{3.15}$$

We apply the Aubin–Nitsche duality argument to estimate $\|\phi_h\|$. For the above $\underline{\phi}_h \in \underline{U}_{h,0}^k$, consider the following auxiliary problem:

$$-\nabla\cdot(a(x)\nabla\psi) + \vec{b}(x)\cdot\nabla\psi + a_0(x)\psi = \phi_h \quad \text{in } \Omega, \tag{3.16a}$$

$$\psi = 0 \quad \text{on } \partial\Omega. \tag{3.16b}$$

Assumptions A.1, A.2, and A.4 ensure the existence and uniqueness of the solution $\psi \in H_0^1(\Omega) \cap H^2(\Omega)$ and the elliptic regularity:

$$\|\psi\|_{H^2(\Omega)} \le C \|\phi_h\|_{L^2(\Omega)}. \tag{3.17}$$

Multiply (3.16) by ϕ_h and integrate over Ω to obtain

$$\|\phi_h\|^2 = -\int_\Omega \nabla\cdot(a\nabla\psi)\phi_h \,\mathrm{dx} + \int_\Omega \vec{b}\cdot\nabla\psi\phi_h \,\mathrm{dx} + \int_\Omega a_0\psi\phi_h \,\mathrm{dx}. \tag{3.18}$$

Applying the integration by parts on the first term of the above equation, we obtain

$$\begin{aligned} -\int_\Omega \nabla\cdot(a\nabla\psi)\phi_h \,\mathrm{dx} &= -\sum_{T\in\mathcal{T}_h} \int_T \nabla\cdot(a_T\nabla\psi)\phi_h \,\mathrm{dx} \\ &= \sum_{T\in\mathcal{T}_h} \left(\int_T a_T \nabla\psi\cdot\nabla\phi_T \,\mathrm{dx} + \sum_{F\in\mathcal{F}_T} \int_F (\phi_F - \phi_T) a_T \nabla\psi\cdot\boldsymbol{n}_{TF} \,\mathrm{ds} \right), \end{aligned} \tag{3.19}$$

where the term related to ϕ_F on the skeleton $\mathcal{F}_h$ is zero owing to the zero boundary condition and [30, Corollary 1.19]. The local term of the above equation is estimated by the definition of gradient reconstruction (2.10) as

$$\begin{aligned} &\int_T a_T \nabla\psi\cdot\nabla\phi_T \,\mathrm{dx} + \sum_{F\in\mathcal{F}_T} \int_F (\phi_F - \phi_T) a_T \nabla\psi\cdot\boldsymbol{n}_{TF} \,\mathrm{ds} \\ &= \int_T a_T \nabla\psi\cdot\boldsymbol{G}_T^k \underline{\phi}_T \,\mathrm{dx} + \sum_{F\in\mathcal{F}_T} \int_F (\phi_F - \phi_T) \left(a_T\nabla\psi - \pi_T^k(a_T\nabla\psi) \right)\cdot\boldsymbol{n}_{TF} \,\mathrm{ds} \\ &= \int_T a_T \boldsymbol{G}_T^k I_T^k \psi\cdot\boldsymbol{G}_T^k \underline{\phi}_T \,\mathrm{dx} + \int_T a_T (\nabla\psi - \boldsymbol{G}_T^k I_T^k\psi)\cdot\boldsymbol{G}_T^k\underline{\phi}_T \,\mathrm{dx} \\ &\quad + \sum_{F\in\mathcal{F}_T} \int_F (\phi_F - \phi_T) \left(a_T\nabla\psi - \pi_T^k(a_T\nabla\psi) \right)\cdot\boldsymbol{n}_{TF} \,\mathrm{ds}. \end{aligned} \tag{3.20}$$

We rewrite the second term of (3.18) as

$$\int_\Omega \vec{b}(x)\cdot\nabla\psi\phi_h\,\mathrm{dx} = \sum_{T\in\mathcal{T}_h}\int_T \vec{b}(x)\cdot\nabla\psi\phi_T\,\mathrm{dx}$$

$$= \sum_{T\in\mathcal{T}_h}\int_T \vec{b}(x)\cdot\nabla(R_T^{k+1}I_T^k\psi)\phi_T\,\mathrm{dx} + \sum_{T\in\mathcal{T}_h}\int_T \vec{b}(x)\cdot\nabla(\psi - R_T^{k+1}I_T^k\psi)\phi_T\,\mathrm{dx}. \tag{3.21}$$

Combining (3.19)–(3.21) in (3.18) and using the definition of $B_h(\bullet,\bullet)$, we obtain

$$\begin{aligned}\|\phi_h\|^2 = B_h(I_h^k\psi,\underline{\phi}_h) - \sum_{T\in\mathcal{T}_h} s_T(I_T^k\psi,\underline{\phi}_T) + \sum_{T\in\mathcal{T}_h}\Big(\int_T a_T(\nabla\psi - \boldsymbol{G}_T^k I_T^k\psi)\cdot\boldsymbol{G}_T^k\underline{\phi}_T\,\mathrm{dx}\\ + \sum_{F\in\mathcal{F}_T}\int_F(\phi_F-\phi_T)\left(a_T\nabla\psi - \pi_T^k(a_T\nabla\psi)\right)\cdot\boldsymbol{n}_{TF}\,\mathrm{ds}\\ + \int_T \vec{b}(x)\cdot\nabla(\psi - R_T^{k+1}I_T^k\psi)\phi_T\,\mathrm{dx} + \int_T a_0(x)(\psi-\pi_T^k\psi)\phi_h\,\mathrm{dx}\Big) =: T_1+T_2+T_3.\end{aligned} \tag{3.22}$$

The Cauchy–Schwarz inequality, trace inequality, and approximation properties of $R_T^{k+1}I_T^k, \boldsymbol{G}_T^k$, and π_T^k yield an estimate for the above last four terms as

$$T_3 \lesssim h\|\psi\|_{H^2(\Omega)}\|\underline{\phi}_h\|_{a,h}. \tag{3.23}$$

Choose $\underline{v}_h = I_h^k\psi$ in (3.13) and use the Cauchy–Schwarz inequality to obtain

$$B_h(I_h^k\psi,\underline{\phi}_h) = \int_\Omega q\pi_h^k\psi\,\mathrm{dx} \le \|q\|\|\pi_h^k\psi\| \lesssim \|q\|\|\psi\|_{H^2(\Omega)}. \tag{3.24}$$

Following [36, equation (46)], we can obtain

$$\sum_{T\in\mathcal{T}_h} s_T(I_T^k\psi,\underline{\phi}_T) \lesssim h\|\psi\|_{H^2(\Omega)}\|\underline{\phi}_h\|_{a,h}. \tag{3.25}$$

Combining the last three estimates (3.23)–(3.25) in (3.22) and using the a priori estimate (3.17) we obtain

$$\|\phi_h\| \le Ch\|\underline{\phi}_h\|_{a,h} + \|q\|. \tag{3.26}$$

Finally, for sufficiently small choice of h, use (3.26) in (3.15) to obtain

$$\|\underline{\phi}_h\|_{a,h} \le C\|q\|.$$

This completes the proof. ∎

Existence and uniqueness of the solution of (3.5). In Lemma 3.3, we proved the existence and uniqueness of the solution of discrete system (3.13) which is the adjoint problem of (3.5). This implies the existence and uniqueness of the solution of discrete system (3.5).

3.3. Error estimate for a non-self-adjoint elliptic problem. In this subsection, we prove the error estimate for the discrete solution (3.5) using Gårding-type inequality and some auxiliary problems.

LEMMA 3.4. *Let $u \in H_0^1(\Omega)$ and $\underline{u}_h \in \underline{U}_{h,0}^k$ be the solutions to the continuous and discrete problems* (3.1) *and* (3.5)*, respectively. Assume $u \in H^{r+2}(\mathcal{T}_h)$ for some $r \in \{0, 1, \dots, k\}$. For sufficiently small h, there exists a real number $C > 0$ independent of h such that*

$$\|I_h^k u - \underline{u}_h\|_{a,h} \le Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)}. \tag{3.27}$$

Proof. From the Gårding-type inequality (3.8), we have

$$C_1\|\underline{v}_h\|_{a,h}^2 - C_2\|v_h\|^2 \le B_h(\underline{v}_h, \underline{v}_h) \quad \forall \underline{v}_h \in \underline{U}_h^k.$$

Set $\underline{\chi}_h := I_h^k u - \underline{u}_h$. Choosing $\underline{v}_h = \underline{\chi}_h$ in the above equation, we obtain

$$C_1\|I_h^k u - \underline{u}_h\|_{a,h}^2 - C_2\|\pi_h^k u - u_h\|^2 \le B_h(I_h^k u - \underline{u}_h, I_h^k u - \underline{u}_h) = B_h(I_h^k u - \underline{u}_h, \underline{\chi}_h). \tag{3.28}$$

Since $\underline{u}_h \in \underline{U}_{h,0}^k$ satisfies (3.5), we have

$$C_1\|I_h^k u - \underline{u}_h\|_{a,h}^2 - C_2\|\pi_h^k u - u_h\|^2 \le B_h(I_h^k u, \underline{\chi}_h) - (p, \chi_h). \tag{3.29}$$

Now, $p = -\nabla\cdot(a\nabla u) + \vec{b}\cdot\nabla u + a_0 u$ of (3.1) leads to

$$(p, \chi_h) = \sum_{T\in\mathcal{T}_h}\int_T p\chi_h \,\mathrm{dx} = \sum_{T\in\mathcal{T}_h}\int_T (-\nabla\cdot(a\nabla u) + \vec{b}\cdot\nabla u + a_0 u)\chi_h \,\mathrm{dx}.$$

Applying the integration by parts in the above equation and following the derivation of (3.18)–(3.22), we obtain

$$\begin{aligned}(p, \chi_h) = B_h(I_h^k u, \underline{\chi}_h) - s_h(I_h^k u, \underline{\chi}_h) + \int_T a_T(\nabla u - \boldsymbol{G}_T^k I_T^k u)\cdot\boldsymbol{G}_T^k\underline{\chi}_T \,\mathrm{dx}\\ + \sum_{F\in\mathcal{F}_T}\int_F (\chi_F - \chi_T)\left(a_T\nabla u - \pi_T^k(a_T\nabla u)\right)\cdot\boldsymbol{n}_{TF}\,\mathrm{ds}\\ + \sum_{T\in\mathcal{T}_h}\int_T \vec{b}(x)\cdot\nabla(u - R_T^{k+1}I_T^k u)\chi_T \,\mathrm{dx} + \int_\Omega a_0(x)(u - \pi_T^k u)\chi_h \,\mathrm{dx}.\end{aligned}$$

The last five terms of the above equation are estimated using similar techniques followed in (3.23)–(3.25) to produce

$$E_h(p; \underline{\chi}_h) := B_h(I_h^k u, \underline{\chi}_h) - (p, \chi_h) \le Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)}\|\underline{\chi}_h\|_{a,h}. \tag{3.30}$$

This leads to an estimation for (3.29) as

$$C_1\|I_h^k u - \underline{u}_h\|_{a,h}^2 - C_2\|\pi_h^k u - u_h\|^2 \le Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)}\|\underline{\chi}_h\|_{a,h}. \tag{3.31}$$

A use of Lemma 2.2 leads the above equation to

$$C_1\|I_h^k u - \underline{u}_h\|_{a,h} \le Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)} + C_2\|\chi_h\|. \tag{3.32}$$

To estimate $\|\chi_h\|$, we use the results of the discrete adjoint problem (3.13) with $q = \chi_h$ and $\underline{v}_h = \underline{\chi}_h$, and this leads to

$$\begin{aligned}\|\chi_h\|^2 &= B_h(\underline{\chi}_h, \underline{\phi}_h) = B_h(I_h^k u - \underline{u}_h, \underline{\phi}_h)\\ &= B_h(I_h^k u, \underline{\phi}_h) - (p, \phi_h) = E_h(p; \underline{\phi}_h).\end{aligned}$$

After using the estimate for $E_h(p; \underline{\phi}_h)$ in (3.30) and the a priori result $\|\underline{\phi}_h\|_{a,h} \leq C\|\chi_h\|$ of (3.13), it holds that

$$\|\chi_h\| \leq Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)}. \tag{3.33}$$

Combining (3.32) and (3.33), we obtain the final estimate

$$\|I_h^k u - \underline{u}_h\|_{a,h} \leq Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)}.$$

This completes the proof. ∎

The triangle inequality, Lemma 3.4, and Lemma 2.4 lead to the error estimate.

THEOREM 3.5 (error estimate). *Let $u \in H_0^1(\Omega)$ and $\underline{u}_h \in \underline{U}_{h,0}^k$ be the solutions to the continuous and discrete problems* (3.1) *and* (3.5)*, respectively. Assume $u \in H^{r+2}(\mathcal{T}_h)$ for some $r \in \{0, 1, \dots, k\}$. For sufficiently small h, it holds that*

$$\left(\sum_{T\in\mathcal{T}_h} \|\nabla u - \boldsymbol{G}_T^k \underline{u}_T\|_T^2\right)^{1/2} \leq Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)} \tag{3.34}$$

for some positive constant C independent of h.

4. Quasilinear elliptic problem.

In this section, we consider the HHO approximation for the quasilinear elliptic boundary value problem:

$$-\nabla\cdot(a(x,u)\nabla u) = f(x) \quad \text{in } \Omega, \tag{4.1a}$$

$$u(x) = 0 \quad \text{on } \partial\Omega, \tag{4.1b}$$

where Ω is a bounded convex polytopal domain in $\mathbb{R}^d$, $d \in \{2, 3\}$ with Lipschitz boundary $\partial\Omega$. We make the following Assumptions Q.1–Q.4 (see [40, 54]) for the problem (4.1).

Assumption Q.1. There exist positive constants α, M such that $0 < \alpha \leq a(x,t) \leq M$, $x \in \bar{\Omega}, t \in \mathbb{R}$.

Assumption Q.2. The coefficient function $a : \bar{\Omega} \times \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function on $\bar{\Omega} \times \mathbb{R}$ such that all derivatives of $a(x,t)$ up to and including second order are bounded in $\bar{\Omega} \times \mathbb{R}$, i.e., $a \in C_b^2(\bar{\Omega} \times \mathbb{R})$.

Henceforth, we understand $a(x,t)$ by $a(t)$ if there is no confusion. For sufficiently smooth data f, problem (4.1) possesses a unique smooth solution u when the boundary is also sufficiently smooth; see [41]. Under the assumption $f \in L^\infty(\Omega)$, Caloz and Rappaz in [21, Theorem 5.1] have proved that problem (4.1) has a unique solution in $W^{2,p}(\Omega)$, $1 \leq p < \infty$, on a domain Ω with smooth boundary $\partial\Omega$. The existence and uniqueness of the weak solution u in $H_0^1(\Omega)$ are established in [56] on a bounded domain with Lipschitz boundary.

Finite element approximation for the above problem (4.1) is proposed and analyzed under various assumptions on the coefficient and data and an assumption on regularity for the solution; see [40, 61, 63, 54, 10, 11, 9, 8]. Following [22, 53, 10, 11], we make the following assumption on the solution of (4.1).

Assumption Q.3. The solution $u \in H_0^1(\Omega)$ of (4.1) belongs to $W^{2,\infty}(\Omega)$ for $d = 2$ and belongs to $H^3(\Omega) \cap W^{2,\infty}(\Omega)$ for $d = 3$.

The linearization of (4.1) around a solution u in the direction ψ is given by an operator $L(u) : H^1_0(\Omega) \to H^{-1}(\Omega)$ defined by

$$L(u)\psi := -\nabla\cdot(a(u)\nabla\psi + a_u(u)\nabla u\,\psi). \tag{4.2}$$

In this article, we consider the numerical approximation of a class of isolated solutions for the quasilinear problem (4.1).

DEFINITION 4.1 (isolated solution). [57, Definition 2.4] *A solution u of* (4.1) *is said to be isolated if $L(u)$ is nonsingular. That is, if $L(u)\psi = 0$, then $\psi = 0$.*

An isolated solution is often named a nonsingular solution or regular solution for the underlying nonlinear problem. We make the following assumption which is used to define a nonlinear map through the Newton method.

Assumption Q.4. The solution $u \in H^1_0(\Omega)$ of (4.1) is isolated.

For the HHO approximation of the solution of (4.1), we consider a linearized problem: find $\psi \in H^1_0(\Omega)$ such that

$$-\nabla\cdot(a(u)\nabla\psi + a_u(u)\nabla u\,\psi) = \phi \quad \text{in} \ \ \Omega, \tag{4.3a}$$

$$\psi = 0 \quad \text{on} \ \ \partial\Omega, \tag{4.3b}$$

for some load function $\phi \in L^2(\Omega)$. The existence and uniqueness of the solution of (4.3) follow from Assumption Q.4. The above problem (4.3) can be converted to an equivalent problem:

$$\begin{aligned} -\Delta\psi &= \tilde{\phi} \quad \text{in } \Omega,\\ \psi &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

with the right-hand-side load function defined by

$$\tilde{\phi} := \big(\phi + \nabla a(u)\cdot\nabla\psi + a_u(u)\nabla u\cdot\nabla\psi + (\nabla a_u(u)\cdot\nabla u + a_u(u)\Delta u)\psi\big)/a(u)$$

which belongs to $L^2(\Omega)$. Since $a \in C^2_b(\mathbb{R})$ and $u \in W^{2,\infty}(\Omega)$ (Assumptions Q.2 and Q.3), the solution satisfies the elliptic regularity $\psi \in H^2(\Omega)$; see [51, Theorem 2.4.3 and Corollary 2.6.8].

In the following sections, we consider an HHO approximation of the above linearized problem (4.3) and analyze the existence and uniqueness of the HHO approximation for the quasilinear problem (4.1).

4.1. HHO approximation for a quasilinear elliptic problem. For $\underline{w}_h, \underline{u}_h$, and $\underline{v}_h$ in $\underline{U}^k_h$, define

$$\mathcal{N}_h(\underline{w}_h; \underline{u}_h, \underline{v}_h) := \sum_{T\in\mathcal{T}_h} \int_T a(R^{k+1}_T \underline{w}_T)\boldsymbol{G}^k_T\underline{u}_T\cdot\boldsymbol{G}^k_T\underline{v}_T \,\mathrm{dx} + s_h(\underline{u}_h, \underline{v}_h), \tag{4.4}$$

where the above stabilization term $s_h(\underline{u}_h, \underline{v}_h) := \sum_{T\in\mathcal{T}_h} s_T(\underline{u}_T, \underline{v}_T)$ with the local stabilization

$$s_T(\underline{u}_T, \underline{v}_T) := \sum_{F\in\mathcal{F}_T} \frac{a^\infty_{TF}}{h_F}\left(S^k_F\underline{u}_T, S^k_F\underline{v}_T\right)_F, \ \text{where } a^\infty_{TF} := \|a_T\|_{L^\infty(F)}.$$

The discrete HHO approximation of (4.1) seeks $\underline{u}_h \in \underline{U}^k_{h,0}$ such that

$$\mathcal{N}_h(\underline{u}_h; \underline{u}_h, \underline{v}_h) = l(\underline{v}_h) \quad \forall \underline{v}_h \in \underline{U}^k_{h,0}, \tag{4.5}$$

where the linear form reads $l(\underline{v}_h) := \sum_{T\in\mathcal{T}_h} \int_T f v_T \,\mathrm{dx}$.

We propose an HHO discretization of the linearized problem (4.3). The discrete linearized problem seeks $\underline{\psi}_h \in \underline{U}^k_{h,0}$ such that

$$\mathcal{N}^{\mathrm{lin}}_h(u; \underline{\psi}_h, \underline{v}_h) = (\phi, v_h) \quad \forall \underline{v}_h \in \underline{U}^k_{h,0}, \tag{4.6}$$

where we considered linearization around the solution u of (4.1) and, for $\underline{\psi}_h, \underline{v}_h \in \underline{U}^k_h$,

$$\begin{aligned}\mathcal{N}^{\mathrm{lin}}_h(u; \underline{\psi}_h, \underline{v}_h) := \sum_{T\in\mathcal{T}_h} \int_T \Big(a(u)\boldsymbol{G}^k_T \underline{\psi}_T \cdot \boldsymbol{G}^k_T \underline{v}_T + a_u(u) R^{k+1}_T \underline{\psi}_T \nabla u \cdot \boldsymbol{G}^k_T \underline{v}_T \Big) \,\mathrm{dx} \\ + s_h(\underline{\psi}_h, \underline{v}_h).\end{aligned} \tag{4.7}$$

Moreover, define a fully discrete version of the linearization term as follows: for $\underline{w}_h, \underline{\psi}_h, \underline{v}_h \in \underline{U}^k_h$,

$$\begin{aligned}\widetilde{\mathcal{N}}^{\mathrm{lin}}_h(\underline{w}_h; \underline{\psi}_h, \underline{v}_h) &:= \sum_{T\in\mathcal{T}_h} \int_T a(R^{k+1}_T \underline{w}_T) \boldsymbol{G}^k_T \underline{\psi}_T \cdot \boldsymbol{G}^k_T \underline{v}_T \,\mathrm{dx} \\ &+ \sum_{T\in\mathcal{T}_h} \int_T a_u(R^{k+1}_T \underline{w}_T) R^{k+1}_T \underline{\psi}_T \boldsymbol{G}^k_T \underline{w}_T \cdot \boldsymbol{G}^k_T \underline{v}_T \,\mathrm{dx} + s_h(\underline{\psi}_h, \underline{v}_h).\end{aligned} \tag{4.8}$$

For $\underline{v}_h \in \underline{U}^k_h$, the definition of reconstruction operator yields $\int_T (R^{k+1}_T \underline{v}_T - v_T)\,\mathrm{dx} = 0$ for each $T \in \mathcal{T}_h$. This leads to the following estimate (see [29, Corollary 5.10]): for $p \le q \le p^*$,

$$\|R^{k+1}_h \underline{v}_h - v_h\|_{L^q(\Omega)} \le C h^{1+\frac{d}{q}-\frac{d}{p}} \|\underline{v}_h\|_{1,p,h}. \tag{4.9}$$

In particular,

$$\|R^{k+1}_h \underline{v}_h\|_{L^6(\Omega)} \le C \|\underline{v}_h\|_{1,h}. \tag{4.10}$$

Using similar arguments as in the proof of Lemma 3.2 and the above estimate, we obtain a Gårding-type inequality for $\mathcal{N}^{\mathrm{lin}}_h(u; \bullet, \bullet)$ as

$$\mathcal{N}^{\mathrm{lin}}_h(u; \underline{v}_h, \underline{v}_h) \ge C_1 \|\underline{v}_h\|^2_{1,h} - C_2 \|v_h\|^2_{L^2(\Omega)} \tag{4.11}$$

for some positive constants C_1 and C_2 independent of h but depending on u and $a(u)$. Then, the existence and uniqueness of the solution $\underline{\psi}_h \in \underline{U}^k_{h,0}$ of (4.6) follow from the existence and uniqueness of the solution of (3.5).

We make the following assumption [54] throughout the section.

Assumption Q.5 (quasiuniformity). We assume the admissible mesh sequence $(\mathcal{T}_h)_{h>0}$ to be quasiuniform; i.e., there exists a constant C_Q independent of h such that

$$\max_{T\in\mathcal{T}_h} h_T \le C_Q \min_{T\in\mathcal{T}_h} h_T.$$

4.2. Fixed point formulation and contraction result. In this section, we prove the existence, local uniqueness, and error estimates for the solution $\underline{u}_h \in \underline{U}_{h,0}^k$ of the above problem (4.5) using fixed point arguments. Following the idea of [62, 23, 54], we define a nonlinear map $\mu : \underline{U}_{h,0}^k \to \underline{U}_{h,0}^k$ which satisfies, for all $\underline{v}_h \in \underline{U}_{h,0}^k$,

$$\mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \mu(\underline{\theta}_h), \underline{v}_h) = \mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \underline{\theta}_h, \underline{v}_h) + \mathcal{N}_h(\underline{\theta}_h; \underline{\theta}_h, \underline{v}_h) - l(\underline{v}_h). \tag{4.12}$$

The map μ is well-defined as (4.6) is well-posed. It can be observed that any fixed point $\underline{\xi}_h$ (say) of μ satisfies $\mathcal{N}_h(\underline{\xi}_h; \underline{\xi}_h, \underline{v}_h) = l(\underline{v}_h)$ for all $\underline{v}_h \in \underline{U}_{h,0}^k$. That is, $\underline{\xi}_h$ is a solution of (4.5). Define a ball of radius R with center at $I_h^k u$ as

$$D(I_h^k u; R) := \left\{ \underline{\theta}_h \in \underline{U}_{h,0}^k \text{ such that } \|I_h^k u - \underline{\theta}_h\|_{1,h} \leq R \right\}.$$

The following results are obtained using the generalized Hölder's inequality, Lemma 2.1, (4.10), discrete Sobolev embedding (2.7), and similar arguments of [54, equation 4.20] as follows.

LEMMA 4.2. *For $\underline{\xi}_h, \underline{\chi}_h \in \underline{U}_h^k$, and $\underline{v}_h \in \underline{U}_{h,0}^k$, the following bounds hold true:*

(a)
$$\begin{aligned}\sum_{T\in\mathcal{T}_h} \int_T |R_T^{k+1} \underline{\xi}_T \boldsymbol{G}_T^k \underline{\chi}_T \cdot \boldsymbol{G}_T^k \underline{v}_T| \,\mathrm{dx} &\leq C\big(\max_{T\in\mathcal{T}_h} h_T^{-d/6}\big) \|R_h^{k+1}\underline{\xi}_h\|_{L^6(\Omega)} \|\underline{\chi}_h\|_{1,h} \|\underline{v}_h\|_{1,h} \\ &\leq C\big(\max_{T\in\mathcal{T}_h} h_T^{-d/6}\big) \|\underline{\xi}_h\|_{1,h} \|\underline{\chi}_h\|_{1,h} \|\underline{v}_h\|_{1,h};\end{aligned}$$

(b)
$$\begin{aligned}\sum_{T\in\mathcal{T}_h} \int_T |(R_T^{k+1} \underline{\xi}_T)^2 \boldsymbol{G}_T^k \underline{\chi}_T \cdot \boldsymbol{G}_T^k \underline{v}_T| \,\mathrm{dx} &\leq C\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \|R_h^{k+1}\underline{\xi}_h\|_{L^6(\Omega)}^2 \|\underline{\chi}_h\|_{1,h} \|\underline{v}_h\|_{1,h} \\ &\leq C\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \|\underline{\xi}_h\|_{1,h}^2 \|\underline{\chi}_h\|_{1,h} \|\underline{v}_h\|_{1,h}.\end{aligned}$$

Now we prove some auxiliary results that will be used in the fixed point theorem.

LEMMA 4.3. *Letting $u \in H^{r+2}(\mathcal{T}_h)$ for $r \in \{0, 1, \ldots, k\}$ and $\underline{\psi}_h, \underline{v}_h \in \underline{U}_h^k$, it holds that*

$$|\mathcal{N}_h^{lin}(u; \underline{\psi}_h, \underline{v}_h) - \widetilde{\mathcal{N}}_h^{lin}(I_h^k u; \underline{\psi}_h, \underline{v}_h)| \leq C h^{r+1-d/6} \|u\|_{H^{r+2}(\mathcal{T}_h)} \|\underline{\psi}_h\|_{1,h} \|\underline{v}_h\|_{1,h}. \tag{4.13}$$

Proof. From the definitions of $\mathcal{N}_h^{\mathrm{lin}}$ and $\widetilde{\mathcal{N}}_h^{\mathrm{lin}}$ in (4.7)–(4.8), we have

$$\begin{aligned}&\mathcal{N}_h^{\mathrm{lin}}(u; \underline{\psi}_h, \underline{v}_h) - \widetilde{\mathcal{N}}_h^{\mathrm{lin}}(I_h^k u; \underline{\psi}_h, \underline{v}_h) \\ &\quad = \sum_{T\in\mathcal{T}_h} \int_T (a(u) - a(R_T^{k+1} I_T^k u)) \boldsymbol{G}_T^k \underline{\psi}_T \cdot \boldsymbol{G}_T^k \underline{v}_T \,\mathrm{dx} \\ &\qquad + \sum_{T\in\mathcal{T}_h} \int_T \Big(a_u(u)\nabla u - a_u(R_T^{k+1} I_T^k u) \boldsymbol{G}_T^k I_T^k u \Big) R_T^{k+1} \underline{\psi}_T \cdot \boldsymbol{G}_T^k \underline{v}_T \,\mathrm{dx}.\end{aligned} \tag{4.14}$$

The Taylor series expansion

$$a(u) = a(w) + \tilde{a}_u(u)(u - w), \tag{4.15}$$

where $\tilde{a}_u(u) = \int_0^1 a_u(u + t(w-u))\,\mathrm{dt}$, Hölder's inequality, inverse inequality, and Lemma 2.3 lead to an estimate for the first term of (4.14) as

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\int_T (a(u) - a(R_T^{k+1}I_T^k u))\boldsymbol{G}_T^k\underline{\psi}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\le C\|\tilde{a}_u\|_{L^\infty(\Omega)}\sum_{T\in\mathcal{T}_h}\|u - R_h^{k+1}I_h^k u\|_{L^2(T)}\|\boldsymbol{G}_h^k\underline{\psi}_h\|_{L^4(T)}\|\boldsymbol{G}_h^k\underline{v}_h\|_{L^4(T)}\\
&\le C\|\tilde{a}_u\|_{L^\infty(\Omega)}h^{r+2-d/2}\|u\|_{H^{r+2}(\mathcal{T}_h)}\|\underline{\psi}_h\|_{1,h}\|\underline{v}_h\|_{1,h}.
\end{aligned}\tag{4.16}$$

The second term of (4.14) is estimated using Taylor series expansion, Hölder's inequality, inverse inequality, quasiuniformity of meshes, Lemmas 2.2–2.3, and Lemma 4.2(a) as

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\int_T\left(a_u(u)\nabla u - a_u(R_T^{k+1}I_T^k u)\boldsymbol{G}_T^k I_T^k u\right)R_T^{k+1}\psi_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&= \sum_{T\in\mathcal{T}_h}\int_T\left(a_u(u) - a_u(R_T^{k+1}I_T^k u)\right)(\nabla u)R_T^{k+1}\psi_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\quad+\sum_{T\in\mathcal{T}_h}\int_T a_u(R_T^{k+1}I_T^k u)\left(\nabla u - \boldsymbol{G}_T^k I_T^k u\right)R_T^{k+1}\psi_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\le C\|\tilde{a}_{uu}\|_{L^\infty(\Omega)}h^{r+2-d/6}\|u\|_{H^{r+2}(\mathcal{T}_h)}\|\nabla u\|_{L^\infty(\Omega)}\|\underline{\psi}_h\|_{1,h}\|\underline{v}_h\|_{1,h}\\
&\quad+C\|\tilde{a}_u\|_{L^\infty(\Omega)}h^{r+1-d/6}\|u\|_{H^{r+2}(\mathcal{T}_h)}\|\underline{\psi}_h\|_{1,h}\|\underline{v}_h\|_{1,h}.
\end{aligned}\tag{4.17}$$

Then, the proof of (4.13) follows from the above two estimations. ∎

THEOREM 4.4 (fixed point result). *Let $u \in H_0^1(\Omega)$ be a solution for (4.1). Assume $u \in H^{r+2}(\mathcal{T}_h)$ for some $r \in \{d-2,\dots,k\}$. For sufficiently small mesh parameter h, there exists $R(h)$ such that the nonlinear map $\mu : \underline{U}_{h,0}^k \to \underline{U}_{h,0}^k$ defined in (4.12) maps from the ball $D(I_h^k u; R(h))$ to itself with radius $R(h) := \tilde{C}h^{r+1}$ for some constant $\tilde{C}$ independent of the mesh parameter. Moreover, μ has a fixed point in $D(I_h^k u; R(h))$.*

Proof. Since $\mathcal{N}_h^{\mathrm{lin}}(u;\bullet,\bullet)$ is associated with the linearized problem, it satisfies the Gårding-type inequality (3.12):

$$C_1\|\underline{w}_h\|_{1,h} \le \sup_{\underline{v}_h\in\underline{U}_{h,0}^k,\,\|\underline{v}_h\|_{a,h}=1}\mathcal{N}_h^{\mathrm{lin}}(u;\underline{w}_h,\underline{v}_h) + C_2\|w_h\| \quad \forall \underline{w}_h \in \underline{U}_{h,0}^k.\tag{4.18}$$

Choose $\underline{w}_h = I_h^k u - \mu(\underline{\theta}_h)$ (we understand $\|\underline{\psi}_h\|_{L^2}$ by $\|\psi_h\|_{L^2(\Omega)}$). Then, there exists $\underline{v}_h \in \underline{U}_{h,0}^k$ with $\|\underline{v}_h\|_{a,h} = 1$ such that

$$C_1\|I_h^k u - \mu(\underline{\theta}_h)\|_{1,h} \le \mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \mu(\underline{\theta}_h),\underline{v}_h) + C_2\|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2}.$$

With this and the definition of μ of (4.12), we obtain

$$\begin{aligned}
&C_1\|I_h^k u - \mu(\underline{\theta}_h)\|_{1,h} \le \mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \mu(\underline{\theta}_h),\underline{v}_h) + C_2\|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2}\\
&= \mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \underline{\theta}_h,\underline{v}_h) + \mathcal{N}_h(\underline{\theta}_h;\underline{\theta}_h,\underline{v}_h) - l(\underline{v}_h) + C_2\|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2}\\
&= \left(\mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \underline{\theta}_h,\underline{v}_h) - \widetilde{\mathcal{N}}_h^{\mathrm{lin}}(I_h^k u; I_h^k u - \underline{\theta}_h,\underline{v}_h)\right)\\
&\quad+\left(\widetilde{\mathcal{N}}_h^{\mathrm{lin}}(I_h^k u; I_h^k u - \underline{\theta}_h,\underline{v}_h) + \mathcal{N}_h(\underline{\theta}_h;\underline{\theta}_h,\underline{v}_h) - l(\underline{v}_h)\right) + C_2\|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2}.
\end{aligned}\tag{4.19}$$

The definition of $l(\underline{v}_h)$ with (4.1) and integration by parts lead to

$$l(\underline{v}_h) = \int_\Omega f v_h \,\mathrm{dx} = -\int_\Omega \nabla\cdot(a(u)\nabla u) v_h \,\mathrm{dx} = -\sum_{T\in\mathcal{T}_h}\int_T \nabla\cdot(a(u)\nabla u) v_h \,\mathrm{dx}$$

$$= \sum_{T\in\mathcal{T}_h}\left(\int_T a(u)\nabla u\cdot\nabla v_T \,\mathrm{dx} + \sum_{F\in\mathcal{F}_T}\int_F (v_F - v_T) a(u)\nabla u\cdot\boldsymbol{n}_{TF}\,\mathrm{ds}\right), \tag{4.20}$$

where the additional term related to v_F on the skeleton $\mathcal{F}_h$ is zero, owing to the zero boundary condition and [30, Corollary 1.19]. Now we rewrite the above terms by adding and subtracting several terms and use the definition (2.10) of gradient reconstruction as follows:

$$\begin{aligned}
l(\underline{v}_h) &= \sum_{T\in\mathcal{T}_h}\left(\int_T a(u)\nabla u\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx} + \sum_{F\in\mathcal{F}_T}\int_F (v_F - v_T)(a(u)\nabla u - \pi_T^k(a(u)\nabla u))\cdot\boldsymbol{n}_{TF}\,\mathrm{ds}\right)\\
&= \sum_{T\in\mathcal{T}_h}\int_T a(u)\boldsymbol{G}_T^k I_T^k u\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx} + \sum_{T\in\mathcal{T}_h}\int_T a(u)(\nabla u - \boldsymbol{G}_T^k I_T^k u)\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx}\\
&\quad + \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T}\int_F (v_F - v_T)(a(u)\nabla u - \pi_T^k(a(u)\nabla u))\cdot\boldsymbol{n}_{TF}\,\mathrm{ds}\\
&= \sum_{T\in\mathcal{T}_h}\int_T a(R_T^{k+1} I_T^k u)\boldsymbol{G}_T^k I_T^k u\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx}\\
&\quad + \sum_{T\in\mathcal{T}_h}\int_T \left(a(u) - a(R_T^{k+1} I_T^k u)\right)\boldsymbol{G}_T^k I_T^k u\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx}\\
&\quad + \sum_{T\in\mathcal{T}_h}\int_T a(u)(\nabla u - \boldsymbol{G}_T^k I_T^k u)\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx}\\
&\quad + \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T}\int_F (v_F - v_T)(a(u)\nabla u - \pi_T^k(a(u)\nabla u))\cdot\boldsymbol{n}_{TF}\,\mathrm{ds}.
\end{aligned} \tag{4.21}$$

The second term of (4.21) is estimated using Taylor series expansion (4.15), the generalized Hölder's inequality, and Lemma 2.3 as

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\int_T \left(a(u) - a(R_T^{k+1} I_T^k u)\right)\boldsymbol{G}_T^k I_T^k u\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx}\\
&\quad\le Ch^{r+2}\|u\|_{H^{r+2}(\mathcal{T}_h)}\|u\|_{W^{1,\infty}(\Omega)}\|\underline{v}_h\|_{1,h}.
\end{aligned} \tag{4.22}$$

The third term of (4.21) is estimated using the Cauchy–Schwarz inequality and Lemma 2.4 as

$$\sum_{T\in\mathcal{T}_h}\int_T a(u)(\nabla u - \boldsymbol{G}_T^k I_T^k u)\cdot\boldsymbol{G}_T^k\underline{v}_T \,\mathrm{dx} \le Ch^{r+1}\|u\|_{H^{r+2}(\mathcal{T}_h)}\|\underline{v}_h\|_{1,h}. \tag{4.23}$$

The last term of (4.21) is estimated by the Cauchy–Schwarz inequality, Lemma 2.3, trace inequality, and Sobolev embedding $H^1(\Omega)\hookrightarrow L^4(\Omega)$ as

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_T}\int_F (v_F - v_T)\left(a(u)\nabla u - \pi_T^k(a(u)\nabla u)\right)\cdot\boldsymbol{n}_{TF}\,\mathrm{ds}\\
&\quad\le Ch^{r+1}\|a(u)\nabla u\|_{H^{r+1}(\mathcal{T}_h)}\|\underline{v}_h\|_{1,h}\\
&\quad\le Ch^{r+1}(\|a'(u)\|_{L^\infty}\|u\|_{H^{r+2}(\mathcal{T}_h)} + \|a(u)\|_{L^\infty})\|u\|_{H^{r+2}(\mathcal{T}_h)}\|\underline{v}_h\|_{1,h}.
\end{aligned} \tag{4.24}$$

For $\underline{\xi}_h, \underline{v}_h \in \underline{U}_h^k$, define

$$\langle \mathcal{F}_h(\underline{\xi}_h), \underline{v}_h\rangle := \sum_{T\in\mathcal{T}_h} \int_T a(R_T^{k+1}\underline{\xi}_T)\boldsymbol{G}_T^k\underline{\xi}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}. \tag{4.25}$$

The definitions of $\mathcal{N}_h(\bullet;\bullet,\bullet)$ and $\mathcal{F}_h$ and the previous estimates (4.24)–(4.22) lead to

$$\mathcal{N}_h(\underline{\theta}_h;\underline{\theta}_h,\underline{v}_h) - l(\underline{v}_h) \le \langle \mathcal{F}_h(\underline{\theta}_h), \underline{v}_h\rangle - \langle \mathcal{F}_h(I_h^k u), \underline{v}_h\rangle + s_h(\underline{\theta}_h,\underline{v}_h) + Ch^{r+1}\|\underline{v}_h\|_{1,h}. \tag{4.26}$$

The Taylor series expansion yields

$$a(w) = a(u) + a_u(u)(w-u) + \tilde{a}_{uu}(w)(w-u)^2, \tag{4.27}$$

where $\tilde{a}_{uu}(w) = \int_0^1 (1-t)a_{uu}(w+t(w-u))\,\mathrm{dt}$. Since $a_u \in C_b^1(\bar\Omega\times\mathbb{R})$ and $a_{uu} \in C_b^0(\bar\Omega\times\mathbb{R})$, we have $\tilde{a}_u \in L^\infty(\Omega\times\mathbb{R})$ and $\tilde{a}_{uu} \in L^\infty(\Omega\times\mathbb{R})$; see [54, equation (4.8)]. We set

$$C_a := \max\left\{\|\tilde{a}_u\|_{L^\infty(\Omega\times\mathbb{R})}, \|\tilde{a}_{uu}\|_{L^\infty(\Omega\times\mathbb{R})}\right\}. \tag{4.28}$$

For $\underline{\xi}_h \in \underline{U}_h^k$ and $\underline{\chi}_h, \underline{v}_h \in \underline{U}_{h,0}^k$, expanding $\mathcal{F}_h(\underline{\xi}_h + \underline{\chi}_h)$ from the above definition (4.25) and using (4.27), we obtain

$$\begin{aligned}
\langle \mathcal{F}_h(\underline{\xi}_h+\underline{\chi}_h), \underline{v}_h\rangle &= \sum_{T\in\mathcal{T}_h}\int_T a(R_T^{k+1}(\underline{\xi}_T+\underline{\chi}_T))\boldsymbol{G}_T^k(\underline{\xi}_T+\underline{\chi}_T)\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&= \sum_{T\in\mathcal{T}_h}\int_T \Big(a(R_T^{k+1}\underline{\xi}_T) + a_u(R_T^{k+1}\underline{\xi}_T)R_T^{k+1}\underline{\chi}_T\Big)\boldsymbol{G}_T^k(\underline{\xi}_T+\underline{\chi}_T)\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\quad + \sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(R_T^{k+1}\underline{\xi}_T)(R_T^{k+1}\underline{\chi}_T)^2\boldsymbol{G}_T^k(\underline{\xi}_T+\underline{\chi}_T)\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&= \langle \mathcal{F}_h(\underline{\xi}_h), \underline{v}_h\rangle + \widetilde{\mathcal{N}}_h^{\mathrm{lin}}(\underline{\xi}_h;\underline{\chi}_h,\underline{v}_h) - s_h(\underline{\chi}_h,\underline{v}_h)\\
&\quad + \sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(R_T^{k+1}\underline{\xi}_T)(R_T^{k+1}\underline{\chi}_T)^2\boldsymbol{G}_T^k\underline{\xi}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\quad + \sum_{T\in\mathcal{T}_h}\int_T a_u(R_T^{k+1}\underline{\xi}_T)R_T^{k+1}\underline{\chi}_T\boldsymbol{G}_T^k\underline{\chi}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\quad + \sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(R_T^{k+1}\underline{\xi}_T)(R_T^{k+1}\underline{\chi}_T)^2\boldsymbol{G}_T^k\underline{\chi}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}.
\end{aligned} \tag{4.29}$$

The fifth term (4.29) is estimated using Lemma 4.2(a) as

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\int_T a_u(R_T^{k+1}\underline{\xi}_T)R_T^{k+1}\underline{\chi}_T\boldsymbol{G}_T^k\underline{\chi}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\le C_aC\big(\max_{T\in\mathcal{T}_h} h_T^{-d/6}\big)\|\underline{\chi}_h\|_{1,h}^2\|\underline{v}_h\|_{1,h}.
\end{aligned} \tag{4.30}$$

The fourth term of (4.29) is estimated using Lemma 4.2(b) as

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(R_T^{k+1}\underline{\xi}_T)(R_T^{k+1}\underline{\chi}_T)^2\boldsymbol{G}_T^k\underline{\xi}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\le C_aC\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big)\|\underline{\chi}_h\|_{1,h}^2\|\underline{\xi}_h\|_{1,h}\|\underline{v}_h\|_{1,h},
\end{aligned} \tag{4.31}$$

and the sixth term of (4.29) is estimated by (4.31) with $\underline{\xi}_h = \underline{\chi}_h$. Combining (4.29)–(4.31) with $\underline{\chi}_h := \underline{\theta}_h - I_h^k u$ and $\underline{\xi}_h = I_h^k u$, we obtain

$$
\begin{aligned}
&\langle \mathcal{F}_h(\underline{\theta}_h), \underline{v}_h\rangle - \langle \mathcal{F}_h(I_h^k u), \underline{v}_h\rangle + s_h(\underline{\theta}_h, \underline{v}_h)\\
&\le \widetilde{\mathcal{N}}_h^{\mathrm{lin}}(I_h^k u; \underline{\theta}_h - I_h^k u, \underline{v}_h) + s_h(I_h^k u, \underline{v}_h)\\
&+ C_a C \max_{T\in\mathcal{T}_h} h_T^{-d/3} \left(\|\underline{\theta}_h - I_h^k u\|_{1,h}^2 + \|\underline{\theta}_h - I_h^k u\|_{1,h}^2 \|I_h^k u\|_{1,h} + \|\underline{\theta}_h - I_h^k u\|_{1,h}^3\right) \|\underline{v}_h\|_{1,h}\\
&\le \widetilde{\mathcal{N}}_h^{\mathrm{lin}}(I_h^k u; \underline{\theta}_h - I_h^k u, \underline{v}_h) + C_a C \max_{T\in\mathcal{T}_h} h_T^{-d/3} \left(\|\underline{\theta}_h - I_h^k u\|_{1,h}^2 + \|\underline{\theta}_h - I_h^k u\|_{1,h}^3\right) \|\underline{v}_h\|_{1,h}\\
&\quad + C h^{r+1} \|\underline{v}_h\|_{1,h},
\end{aligned}
\tag{4.32}
$$

where we used the estimation for the consistency term $s_h(I_h^k u, \underline{v}_h)$; see [36, equation (46)]. Combining (4.26), and (4.32), we obtain

$$
\begin{aligned}
&\widetilde{\mathcal{N}}_h^{\mathrm{lin}}(I_h^k u; I_h^k u - \underline{\theta}_h, \underline{v}_h) + \mathcal{N}_h(\underline{\theta}_h; \underline{\theta}_h, \underline{v}_h) - l(\underline{v}_h)\\
&\quad \le C_a C\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \left(\|\underline{\theta}_h - I_h^k u\|_{1,h}^2 + \|\underline{\theta}_h - I_h^k u\|_{1,h}^3\right) \|\underline{v}_h\|_{1,h} + C h^{r+1} \|\underline{v}_h\|_{1,h}.
\end{aligned}
\tag{4.33}
$$

This implies from (4.19) with Lemma 4.3 and $\|\underline{v}_h\|_{1,h} = 1$ that

$$
\begin{aligned}
C_1 \|I_h^k u - \mu(\underline{\theta}_h)\|_{1,h} \le C\Big(& C_a\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \left(\|\underline{\theta}_h - I_h^k u\|_{1,h}^2 + \|\underline{\theta}_h - I_h^k u\|_{1,h}^3\right)\\
&+ h^{r+1} + h^{r+1-d/6} \|\underline{\theta}_h - I_h^k u\|_{1,h} + \|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2}\Big).
\end{aligned}
\tag{4.34}
$$

Now, we estimate $\|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2}$ using the following dual problem: given $\underline{q}_h := I_h^k u - \mu(\underline{\theta}_h)$, find $\underline{\phi}_h \in \underline{U}_{h,0}^k$ such that

$$
\mathcal{N}_h^{\mathrm{lin}}(u; \underline{v}_h, \underline{\phi}_h) = (q_h, v_h) \quad \forall \underline{v}_h \in \underline{U}_{h,0}^k. \tag{4.35}
$$

Choosing $\underline{v}_h = I_h^k u - \mu(\underline{\theta}_h)$ in the above equation, using the definition (4.12) of μ, the idea of splitting (4.19), and (4.33), we obtain

$$
\begin{aligned}
&\|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2(\Omega)}^2 = \mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \mu(\underline{\theta}_h), \underline{\phi}_h)\\
&= \mathcal{N}_h^{\mathrm{lin}}(u; I_h^k u - \underline{\theta}_h, \underline{\phi}_h) + \mathcal{N}_h(\underline{\theta}_h; \underline{\theta}_h, \underline{\phi}_h) - l(\underline{\phi}_h)\\
&\le C\big(C_a\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \left(\|\underline{\theta}_h - I_h^k u\|_{1,h}^2 + \|\underline{\theta}_h - I_h^k u\|_{1,h}^3\right) \|\underline{\phi}_h\|_{1,h}\\
&\quad + h^{r+1} \|\underline{\phi}_h\|_{1,h} + h^{r+1-d/6} \|\underline{\theta}_h - I_h^k u\|_{1,h} \|\underline{\phi}_h\|_{1,h}\big).
\end{aligned}
$$

Using the a priori bound $\|\underline{\phi}_h\|_{1,h} \lesssim \|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2}$ of (4.35), we obtain

$$
\begin{aligned}
\|I_h^k u - \mu(\underline{\theta}_h)\|_{L^2} \le C\Big(& C_a\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \left(\|\underline{\theta}_h - I_h^k u\|_{1,h}^2 + \|\underline{\theta}_h - I_h^k u\|_{1,h}^3\right)\\
&+ h^{r+1} + h^{r+1-d/6} \|\underline{\theta}_h - I_h^k u\|_{1,h}\Big).
\end{aligned}
\tag{4.36}
$$

Finally, the above estimations (4.34) and (4.36) lead to

$$
\begin{aligned}
\|I_h^k u - \mu(\underline{\theta}_h)\|_{1,h} \le \tilde{C}\Big(& h^{r+1} + h^{r+1-d/6} \|\underline{\theta}_h - I_h^k u\|_{1,h}\\
&+ \big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \left(\|\underline{\theta}_h - I_h^k u\|_{1,h}^2 + \|\underline{\theta}_h - I_h^k u\|_{1,h}^3\right)\Big)
\end{aligned}
\tag{4.37}
$$

for some positive constant $\tilde{C}$ independent of h but depending on u and $a(u)$. Choose h_* such that

$$(1 + 2\tilde{C}h_*^{r+1-d/6} + 4\tilde{C}^2 h_*^{r+1-d/3} + 8\tilde{C}^3 h_*^{2r+2-d/3}) \leq 2.$$

This implies $(1 + 2\tilde{C}h^{r+1-d/6} + 4\tilde{C}^2 h^{r+1-d/3} + 8\tilde{C}^3 h^{2r+2-d/3}) \leq 2$ whenever $h \leq h_*$. Thus if $\|I_h^k u - \underline{\theta}_h\|_{1,h} \leq R(h) := 2\tilde{C}h^{r+1}$, then using Assumption Q.5 of quasi-uniformity, (4.37) yields

$$\begin{aligned}\|I_h^k u - \mu(\underline{\theta}_h)\|_{1,h} &\leq \tilde{C}\left(h^{r+1} + 2\tilde{C}h^{2r+2-d/6} + 4\tilde{C}^2 h^{2r+2-d/3} + 8\tilde{C}^3 h^{3r+3-d/3}\right)\\ &\leq \tilde{C}h^{r+1}\left(1 + 2\tilde{C}h^{r+1-d/6} + 4\tilde{C}^2 h^{r+1-d/3} + 8\tilde{C}^3 h^{2r+2-d/3}\right) \leq \tilde{C}h^{r+1} \times 2 = R(h).\end{aligned}$$

Thus for sufficiently small h ($h \leq h_*$), there exists a ball $D(I_h^k u; R(h))$ of radius $R(h) = 2\tilde{C}h^{r+1}$ with center at $I_h^k u$ such that the following result holds:

$$\|I_h^k u - \underline{\theta}_h\|_{1,h} \leq R(h) \Rightarrow \|I_h^k u - \mu(\underline{\theta}_h)\|_{1,h} \leq R(h).$$

Hence μ is a map from a closed and bounded (compact) convex ball to itself. Therefore, using the Brouwer fixed point theorem, it has a fixed point. This completes the proof. ∎

To prove the unique fixed point of μ, we show the following contraction result.

THEOREM 4.5 (contraction result). *Let $u \in H_0^1(\Omega)$ be a solution for* (4.1). *Assume $u \in H^{r+2}(\mathcal{T}_h)$ for some $r \in \{d-2, \dots, k\}$. Let $\underline{\theta}_1$, $\underline{\theta}_2 \in D(I_h^k u; R(h))$. For sufficiently small h, the following contraction result holds:*

$$\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{1,h} \leq Ch^{r+1-d/3}\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h}.$$

Proof. Let $\underline{\theta}_1$, $\underline{\theta}_2 \in D(I_h^k u; R(h))$; then $\mu(\underline{\theta}_1)$ and $\mu(\underline{\theta}_2)$ satisfy (4.12). That is,

$$\mathcal{N}_h^{\text{lin}}(u; I_h^k u - \mu(\underline{\theta}_1), \underline{v}_h) = \mathcal{N}_h^{\text{lin}}(u; I_h^k u - \underline{\theta}_1, \underline{v}_h) + \mathcal{N}_h(\underline{\theta}_1; \underline{\theta}_1, \underline{v}_h) - l(\underline{v}_h), \tag{4.38}$$

$$\mathcal{N}_h^{\text{lin}}(u; I_h^k u - \mu(\underline{\theta}_2), \underline{v}_h) = \mathcal{N}_h^{\text{lin}}(u; I_h^k u - \underline{\theta}_2, \underline{v}_h) + \mathcal{N}_h(\underline{\theta}_2; \underline{\theta}_2, \underline{v}_h) - l(\underline{v}_h). \tag{4.39}$$

Using Gårding-type inequality (3.12), replacing $\underline{\theta}_h$ by $\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)$ with $\|\underline{v}_h\|_{1,h} = 1$, we have

$$C_1\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{1,h} \leq \mathcal{N}_h^{\text{lin}}(u; \mu(\underline{\theta}_1) - \mu(\underline{\theta}_2), \underline{v}_h) + C_2\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{L^2}, \tag{4.40}$$

where we understand $\|\underline{v}_h\|_{L^2}$ by $\|v_h\|_{L^2(\Omega)}$ in the above term of $\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{L^2}$. From the definition of μ and subtracting (4.39) from (4.38), we get

$$\mathcal{N}_h^{\text{lin}}(u; \mu(\underline{\theta}_2) - \mu(\underline{\theta}_1), \underline{v}_h) = \mathcal{N}_h^{\text{lin}}(u; \underline{\theta}_2 - \underline{\theta}_1, \underline{v}_h) + \mathcal{N}_h(\underline{\theta}_1; \underline{\theta}_1, \underline{v}_h) - \mathcal{N}_h(\underline{\theta}_2; \underline{\theta}_2, \underline{v}_h). \tag{4.41}$$

Using the definition $\mathcal{F}_h$ of (4.25), the last two terms of (4.41) yield

$$\begin{aligned}&\mathcal{N}_h(\underline{\theta}_1; \underline{\theta}_1, \underline{v}_h) - \mathcal{N}_h(\underline{\theta}_2; \underline{\theta}_2, \underline{v}_h)\\ &= \langle \mathcal{F}_h(\underline{\theta}_1), \underline{v}_h\rangle - \langle \mathcal{F}_h(\underline{\theta}_2), \underline{v}_h\rangle + s_h(\underline{\theta}_1 - \underline{\theta}_2, \underline{v}_h)\\ &= \left(\langle \mathcal{F}_h(\underline{\theta}_1), \underline{v}_h\rangle - \langle \mathcal{F}_h(I_h^k u), \underline{v}_h\rangle + s_h(\underline{\theta}_1 - I_h^k u, \underline{v}_h)\right)\\ &\quad - \left(\langle \mathcal{F}_h(\underline{\theta}_2), \underline{v}_h\rangle - \langle \mathcal{F}_h(I_h^k u), \underline{v}_h\rangle + s_h(\underline{\theta}_2 - I_h^k u, \underline{v}_h)\right).\end{aligned} \tag{4.42}$$

Set $\underline{\chi}_{1h} := \underline{\theta}_1 - I_h^k u$, $\underline{\chi}_{2h} := \underline{\theta}_2 - I_h^k u$, and $\check{u}_T := R_T^{k+1} I_T^k u$. From the expansion of $\mathcal{F}_h$ of (4.29), we rewrite (4.42) as

$$\begin{aligned}
&\mathcal{N}_h(\underline{\theta}_1;\underline{\theta}_1,\underline{v}_h) - \mathcal{N}_h(\underline{\theta}_2;\underline{\theta}_2,\underline{v}_h) - \widetilde{\mathcal{N}}_h^{\mathrm{lin}}(I_h^k u;\underline{\theta}_1-\underline{\theta}_2,\underline{v}_h)\\
&= \Bigg(\sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(\check{u}_T)(R_T^{k+1}\underline{\chi}_{1T})^2\nabla\check{u}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\qquad - \sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(\check{u}_T)(R_T^{k+1}\underline{\chi}_{2T})^2\nabla\check{u}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\Bigg)\\
&\quad + \Bigg(\sum_{T\in\mathcal{T}_h}\int_T a_u(\check{u}_T)R_T^{k+1}\underline{\chi}_{1T}\boldsymbol{G}_T^k\underline{\chi}_{1T}\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\qquad - \sum_{T\in\mathcal{T}_h}\int_T a_u(\check{u}_T)R_T^{k+1}\underline{\chi}_{2T}\boldsymbol{G}_T^k\underline{\chi}_{2T}\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\Bigg)\\
&\quad + \Bigg(\sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(\check{u}_T)(R_T^{k+1}\underline{\chi}_{1T})^2\boldsymbol{G}_T^k\underline{\chi}_{1T}\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\qquad - \sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(\check{u}_T)(R_T^{k+1}\underline{\chi}_{2T})^2\boldsymbol{G}_T^k\underline{\chi}_{2T}\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\Bigg) =: T_1+T_2+T_3.
\end{aligned}\tag{4.43}$$

To estimate the above terms, we use the identities

$$a^2-b^2=(a-b)(a+b),\quad a_1b_1-a_2b_2=a_1(b_1-b_2)+(a_1-a_2)b_2,\tag{4.44}$$

$$a_1^2b_1-a_2^2b_2=a_1^2(b_1-b_2)+(a_1-a_2)(a_1+a_2)b_2.\tag{4.45}$$

Using the above identity (4.44) and estimates similar to Lemma 4.2(a), the terms T_1 and T_2 of (4.43) yield

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(\check{u}_T)(R_T^{k+1}\underline{\theta}_{1T}-R_T^{k+1}\underline{\theta}_{2T})(R_T^{k+1}\underline{\theta}_{1T}-\check{u}_T+R_T^{k+1}\underline{\theta}_{2T}-\check{u}_T)\nabla\check{u}_T\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\quad+\sum_{T\in\mathcal{T}_h}\int_T a_u(\check{u}_T)(\check{u}_T-R_T^{k+1}\underline{\theta}_{1T})\boldsymbol{G}_T^k(\underline{\theta}_{1T}-\underline{\theta}_{2T})\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\quad+\sum_{T\in\mathcal{T}_h}\int_T a_u(\check{u}_T)(R_T^{k+1}\underline{\theta}_{1T}-R_T^{k+1}\underline{\theta}_{2T})\boldsymbol{G}_T^k(I_T^k u-\underline{\theta}_{2T})\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\le C_aC\big(\max_{T\in\mathcal{T}_h}h_T^{-d/6}\big)\|\underline{\theta}_1-\underline{\theta}_2\|_{1,h}\left(\|I_h^k u-\underline{\theta}_2\|_{1,h}+\|I_h^k u-\underline{\theta}_1\|_{1,h}\right)\|\underline{v}_h\|_{1,h}.
\end{aligned}$$

The above identity (4.45) and estimates similar to Lemma 4.2 lead to an estimate for the term T_3 of (4.43) as

$$\begin{aligned}
T_3 &= \sum_{T\in\mathcal{T}_h}\int_T \tilde{a}_{uu}(\check{u}_T)\Big((\check{u}_T-R_T^{k+1}\underline{\theta}_{2T})^2\boldsymbol{G}_T^k(\underline{\theta}_{1T}-\underline{\theta}_{2T})+(R_T^{k+1}\underline{\theta}_{1T}-R_T^{k+1}\underline{\theta}_{2T})\times\\
&\qquad(\check{u}_T-R_T^{k+1}\underline{\theta}_{2T}+\check{u}_T-R_T^{k+1}\underline{\theta}_{1T})\boldsymbol{G}_T^k(I_T^k u-\underline{\theta}_{1T})\Big)\cdot\boldsymbol{G}_T^k\underline{v}_T\,\mathrm{dx}\\
&\le C_aC\big(\max_{T\in\mathcal{T}_h}h_T^{-d/3}\big)\|\underline{\theta}_1-\underline{\theta}_2\|_{1,h}\left(\|I_h^k u-\underline{\theta}_2\|_{1,h}^2+\|I_h^k u-\underline{\theta}_1\|_{1,h}^2\right)\|\underline{v}_h\|_{1,h}.
\end{aligned}$$

Combining the estimation for (4.43) in (4.41) with $\|\underline{v}_h\|_{1,h} = 1$, yields

$$\begin{aligned} \mathcal{N}_h^{\text{lin}}(u; \mu(\underline{\theta}_2) - \mu(\underline{\theta}_1), \underline{v}_h) &= \Big(\mathcal{N}_h^{\text{lin}}(u; \underline{\theta}_2 - \underline{\theta}_1, \underline{v}_h) - \widetilde{\mathcal{N}}_h^{\text{lin}}(I_h^k u; \underline{\theta}_2 - \underline{\theta}_1, \underline{v}_h)\Big) \\ &+ \Big(\mathcal{N}_h(\underline{\theta}_1; \underline{\theta}_1, \underline{v}_h) - \mathcal{N}_h(\underline{\theta}_2; \underline{\theta}_2, \underline{v}_h) - \widetilde{\mathcal{N}}_h^{\text{lin}}(I_h^k u; \underline{\theta}_1 - \underline{\theta}_2, \underline{v}_h)\Big) \\ &\le Ch^{r+1-d/6}\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h} + C_a C\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big)\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h}\Big(\|I_h^k u - \underline{\theta}_1\|_{1,h} \\ &+ \|I_h^k u - \underline{\theta}_1\|_{1,h}^2 + \|I_h^k u - \underline{\theta}_2\|_{1,h} + \|I_h^k u - \underline{\theta}_2\|_{1,h}^2\Big). \end{aligned} \tag{4.46}$$

To obtain the estimation for the L^2-term $\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{L^2}$ of (4.40), we consider the following dual linear problem: given $\underline{q}_h := \mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)$, find $\underline{\phi}_h \in \underline{U}_{h,0}^k$ such that

$$\mathcal{N}_h^{\text{lin}}(u; \underline{v}_h, \underline{\phi}_h) = (q_h, v_h) \quad \forall \underline{v}_h \in \underline{U}_{h,0}^k. \tag{4.47}$$

Choose $\underline{v}_h = \mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)$ in (4.47) and use (4.46) to obtain

$$\begin{aligned} &\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{L^2}^2 = \mathcal{N}_h^{\text{lin}}(u; \mu(\underline{\theta}_1) - \mu(\underline{\theta}_2), \underline{\phi}_h) \\ &\le Ch^{r+1-d/6}\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h}\|\underline{\phi}_h\|_{1,h} + C_a C\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big)\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h} \\ &\quad \times \Big(\|I_h^k u - \underline{\theta}_1\|_{1,h} + \|I_h^k u - \underline{\theta}_1\|_{1,h}^2 + \|I_h^k u - \underline{\theta}_2\|_{1,h} + \|I_h^k u - \underline{\theta}_2\|_{1,h}^2\Big)\|\underline{\phi}_h\|_{1,h}. \end{aligned}$$

The a priori bound $\|\underline{\phi}_h\|_{1,h} \le C\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{L^2}$ of (4.47) leads to

$$\begin{aligned} \|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{L^2} &\le Ch^{r+1-d/6}\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h} + C_a C\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h}\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \\ &\times \Big(\|I_h^k u - \underline{\theta}_1\|_{1,h} + \|I_h^k u - \underline{\theta}_1\|_{1,h}^2 + \|I_h^k u - \underline{\theta}_2\|_{1,h} + \|I_h^k u - \underline{\theta}_2\|_{1,h}^2\Big). \end{aligned} \tag{4.48}$$

Using (4.46) and (4.48), we obtain from (4.40) that

$$\begin{aligned} \|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{1,h} &\le Ch^{r+1-d/6}\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h} + C_a C\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h}\big(\max_{T\in\mathcal{T}_h} h_T^{-d/3}\big) \\ &\times \Big(\|I_h^k u - \underline{\theta}_1\|_{1,h} + \|I_h^k u - \underline{\theta}_1\|_{1,h}^2 + \|I_h^k u - \underline{\theta}_2\|_{1,h} + \|I_h^k u - \underline{\theta}_2\|_{1,h}^2\Big). \end{aligned}$$

Since $\underline{\theta}_1$, $\underline{\theta}_2 \in D(I_h^k u; R(h))$ with $R(h) = 2\tilde{C}h^{r+1}$, we have

$$\|I_h^k u - \underline{\theta}_1\|_{1,h} \le 2\tilde{C}h^{r+1} \quad \text{and} \quad \|I_h^k u - \underline{\theta}_2\|_{1,h} \le 2\tilde{C}h^{r+1}.$$

This implies

$$\|\mu(\underline{\theta}_1) - \mu(\underline{\theta}_2)\|_{1,h} \le Ch^{r+1-d/3}\|\underline{\theta}_1 - \underline{\theta}_2\|_{1,h}$$

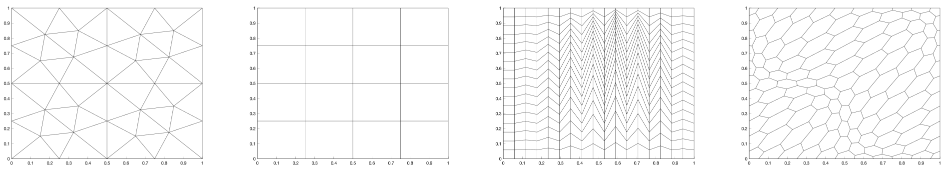
for sufficiently small mesh parameter h. This completes the proof. ∎

For sufficiently small h, Theorem 4.5 proves the local uniqueness of the fixed point of μ and hence the local uniqueness of the solution of (4.5).

Error estimate for a quasilinear problem. Adding and subtracting $\boldsymbol{G}_h^k I_h^k u$, using the triangle inequality and Theorem 4.4, we have the following error estimate.

THEOREM 4.6 (error estimate). *Let $u \in H_0^1(\Omega)$ be the solution of the nonlinear problem* (4.1) *and $\underline{u}_h \in \underline{U}_{h,0}^k$ be the solution of the discrete problem* (4.5). *Assume $u \in H^{r+2}(\mathcal{T}_h)$ for some $r \in \{d-2, \dots, k\}$. Then for sufficiently small h, we have*

$$\|\nabla u - \boldsymbol{G}_h^k \underline{u}_h\| \le Ch^{r+1}. \tag{4.49}$$



(a) (b) (c) (d)

FIG. 1. (a) *Triangular,* (b) *Cartesian,* (c) *Kershaw, and* (d) *hexagonal initial meshes.*

5. Numerical experiments. In this section, numerical experiments are performed using the HHO approximation (4.5) for the quasilinear problem (4.1). We consider the following nonlinear problem [54]:

$$\begin{aligned} -\nabla\cdot((1+u)\nabla u) &= f \quad \text{in} \;\; \Omega, \\ u &= 0 \quad \text{on} \;\; \partial\Omega, \end{aligned}$$

where $\Omega := (0,1)\times(0,1) \subset \mathbb{R}^2$ and the source term f is taken in such a way that the exact solution reads $u(x,y) = x(1-x)y(1-y)$. We rewrite the nonlinear map (4.12) in order to obtain practical iterative solutions. In the computation, we do not demand the exact solution u, and this is replaced by the previous step's computed (initial guess) solution. We start with an initial guess $u_h^0 \in \underline{U}_{h,0}^k$ obtained from solving the Dirichlet Poisson problem $-\Delta u = f$ with the same load function f as defined above. The $(n+1)$th iteration is given by, for all $\underline{v}_h \in \underline{U}_{h,0}^k$,

$$\widetilde{\mathcal{N}}_h^{\mathrm{lin}}(\underline{u}_h^n; \underline{u}_h^{n+1}, \underline{v}_h) = \widetilde{\mathcal{N}}_h^{\mathrm{lin}}(\underline{u}_h^n; \underline{u}_h^n, \underline{v}_h) - \mathcal{N}_h(\underline{u}_h^n; \underline{u}_h^n, \underline{v}_h) + l(\underline{v}_h) \;\; \text{for } n = 0,1,2,\ldots,$$

where the linearized $\widetilde{\mathcal{N}}_h^{\mathrm{lin}}$ and nonlinear $\mathcal{N}_h$ forms are as defined in (4.7) and (4.4), respectively. The stopping criterion is prescribed by a tolerance 10^{-10} for the difference of two successive iterative solutions as $\|\boldsymbol{G}_h^k(\underline{u}_h^{n+1} - \underline{u}_h^n)\|/\|\boldsymbol{G}_h^k \underline{u}_h^{n+1}\| \leq 10^{-10}$. We consider the triangular, Cartesian, Kershaw, and hexagonal mesh families for numerical experiments which are depicted in Figure 1. The triangular, Cartesian, and Kershaw mesh families are discussed in [55], and the hexagonal mesh family is introduced in [39]. The experiments are performed in MATLAB. Some of the basic implementation methodologies for HHO methods are adopted from [30, 25, 36]. It has been observed that the iterative step terminates within 4 steps using the above stopping criterion. The experimental rate of convergence is computed as

$$\texttt{rate}(\ell) := \log\left(e_{h_\ell}/e_{h_{\ell-1}}\right)/\log\left(h_\ell/h_{\ell-1}\right) \text{ for } \ell = 1,2,3,\ldots,$$

where e_{h_ℓ} and $e_{h_{\ell-1}}$ are the errors associated to the two consecutive meshsizes h_ℓ and $h_{\ell-1}$, respectively. In Figure 2, we have plotted the convergence histories for the relative reconstructed gradient error $e_h = \|\nabla u - \boldsymbol{G}_h^k \underline{u}_h\|/\|\nabla u\|$ as a function of mesh parameter h on the sequence of triangular, Cartesian, Kershaw, and hexagonal meshes for the polynomial degree $k = 0,1,2$. The convergence rates for the polynomial degree $k = 0,1,2$ are, respectively, close to $1,2,3$ for each mesh. The convergence rates are in line with the theoretical convergence found in Theorem 4.6.

6. Conclusions. In this paper, we have discussed an HHO approximation for the second-order quasilinear elliptic problem of nonmonotone type defined on a polytopal domain in $\mathbb{R}^d, d = 2,3$. First, we have deduced the existence, uniqueness, and error

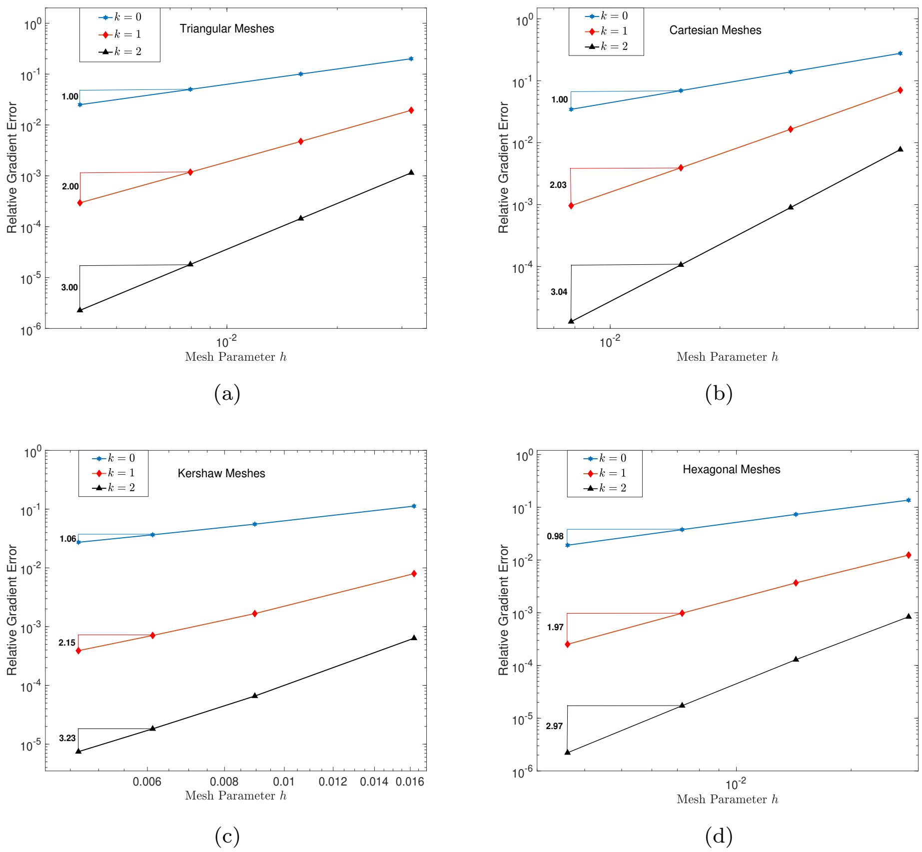


FIG. 2. *Convergence histories for the relative gradient error on* (a) *triangular,* (b) *Cartesian,* (c) *Kershaw, and* (d) *hexagonal meshes.*

estimate for the discrete solution of a general second-order non-self-adjoint problem. This has helped us to construct a nonlinear map that satisfies the contraction property over a small ball. The discrete solution of the nonlinear problem is essentially a fixed point of the nonlinear map. The existence, local uniqueness, and error estimation for the HHO approximation of the nonlinear problem are established. The analysis does not require any user-specified large penalty parameter, unlike the DG method of [54]. The analysis also supports lowest-order ($k = 0$ when $d = 2$) polynomial approximation with linear order of convergence. It is possible to extend our analysis without much difficulty to the more general nonlinear problem of the type $\nabla\cdot(a(x,u)\nabla u)+f(x,u) = 0$ with $f \in C_b^2(\bar{\Omega}\times\mathbb{R})$ and when $a(x,u)$ is a uniformly bounded positive-definite matrix.

REFERENCES

[1] P. F. Antonietti, N. Bigoni, and M. Verani, *Mimetic finite difference approximation of quasilinear elliptic problems*, Calcolo, 52 (2015), pp. 45–67.

[2] P. F. Antonietti, S. Giani, and P. Houston, *hp-version composite discontinuous Galerkin methods for elliptic problems on complicated domains*, SIAM J. Sci. Comput., 35 (2013), pp. A1417–A1439.

[3] R. Araya, C. Harder, D. Paredes, and F. Valentin, *Multiscale hybrid-mixed method*, SIAM J. Numer. Anal., 51 (2013), pp. 3505–3531.

[4] B. Ayuso de Dios, K. Lipnikov, and G. Manzini, *The nonconforming virtual element method*, ESAIM Math. Model. Numer. Anal., 50 (2016), pp. 879–904.
[5] F. Bassi, L. Botti, A. Colombo, D. A. Di Pietro, and P. Tesini, *On the flexibility of agglomeration based physical space discontinuous Galerkin discretizations*, J. Comput. Phys., 231 (2012), pp. 45–65.
[6] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo, *Basic principles of virtual element methods*, Math. Models Methods Appl. Sci., 23 (2013), pp. 199–214.
[7] L. Beirão da Veiga, F. Brezzi, and L. D. Marini, *Virtual elements for linear elasticity problems*, SIAM J. Numer. Anal., 51 (2013), pp. 794–812.
[8] C. Bi and V. Ginting, *Two-grid finite volume element method for linear and nonlinear elliptic problems*, Numer. Math., 108 (2007), pp. 177–198.
[9] C. Bi and V. Ginting, *A residual-type a posteriori error estimate of finite volume element method for a quasi-linear elliptic problem*, Numer. Math., 114 (2009), pp. 107–132.
[10] C. Bi and V. Ginting, *Two-grid discontinuous Galerkin method for quasi-linear elliptic problems*, J. Sci. Comput., 49 (2011), pp. 311–331.
[11] C. Bi and V. Ginting, *A posteriori error estimates of discontinuous Galerkin method for nonmonotone quasi-linear elliptic problems*, J. Sci. Comput., 55 (2013), pp. 659–687.
[12] C. Bi and M. Liu, *A discontinuous finite volume element method for second-order elliptic problems*, Numer. Methods Partial Differential Equations, 28 (2012), pp. 425–440.
[13] C. Bi, C. Wang, and Y. Lin, *A posteriori error estimates of two-grid finite element methods for nonlinear elliptic problems*, J. Sci. Comput., 74 (2018), pp. 23–48.
[14] J. Bonelle and A. Ern, *Analysis of compatible discrete operator schemes for elliptic problems on polyhedral meshes*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 553–581.
[15] M. Botti, D. A. Di Pietro, and P. Sochala, *A hybrid high-order method for nonlinear elasticity*, SIAM J. Numer. Anal., 55 (2017), pp. 2687–2717.
[16] F. Brezzi, R. S. Falk, and L. D. Marini, *Basic principles of mixed virtual element methods*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 1227–1240.
[17] F. Brezzi, K. Lipnikov, and M. Shashkov, *Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes*, SIAM J. Numer. Anal., 43 (2005), pp. 1872–1896.
[18] F. Brezzi, K. Lipnikov, M. Shashkov, and V. Simoncini, *A new discretization methodology for diffusion problems on generalized polyhedral meshes*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 3682–3692.
[19] F. Brezzi, K. Lipnikov, and V. Simoncini, *A family of mimetic finite difference methods on polygonal and polyhedral meshes*, Math. Models Methods Appl. Sci., 15 (2005), pp. 1533–1551.
[20] E. Burman and A. Ern, *An unfitted hybrid high-order method for elliptic interface problems*, SIAM J. Numer. Anal., 56 (2018), pp. 1525–1546.
[21] G. Caloz and J. Rappaz, *Numerical analysis for nonlinear and bifurcation problems*, in Techniques of Scientific Computing (Part 2), P. G. Ciarlet and J. L. Lions, eds., Handb. Numer. Anal. V, North-Holland, Amsterdam, 1997, pp. 487–637.
[22] A. Cangiani, P. Chatzipantelidis, G. Diwan, and E. H. Georgoulis, *Virtual element method for quasilinear elliptic problems*, IMA J. Numer. Anal., 40 (2020), pp. 2450–2472.
[23] C. Carstensen, G. Mallik, and N. Nataraj, *A priori and a posteriori error control of discontinuous Galerkin finite element methods for the von Kármán equations*, IMA J. Numer. Anal., 39 (2019), pp. 167–200.
[24] P. Castillo, B. Cockburn, I. Perugia, and D. Schötzau, *An a priori error analysis of the local discontinuous Galerkin method for elliptic problems*, SIAM J. Numer. Anal., 38 (2000), pp. 1676–1706.
[25] M. Cicuttin, D. A. Di Pietro, and A. Ern, *Implementation of discontinuous skeletal methods on arbitrary-dimensional, polytopal meshes using generic programming*, J. Comput. Appl. Math., 344 (2018), pp. 852–874.
[26] M. Cicuttin, A. Ern, and T. Gudi, *Hybrid high-order methods for the elliptic obstacle problem*, J. Sci. Comput., 83 (2020), 8.
[27] B. Cockburn, D. A. Di Pietro, and A. Ern, *Bridging the hybrid high-order and hybridizable discontinuous Galerkin methods*, ESAIM Math. Model. Numer. Anal., 50 (2016), pp. 635–650.
[28] B. Cockburn, J. Gopalakrishnan, and R. Lazarov, *Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems*, SIAM J. Numer. Anal., 47 (2009), pp. 1319–1365.

[29] D. A. Di Pietro and J. Droniou, *A hybrid high-order method for Leray-Lions elliptic equations on general meshes*, Math. Comp., 86 (2017), pp. 2159–2191.
[30] D. A. Di Pietro and J. Droniou, *The Hybrid High-Order Method for Polytopal Meshes: Design, Analysis, and Applications*, Springer, Cham, 2020.
[31] D. A. Di Pietro, J. Droniou, and A. Ern, *A discontinuous-skeletal method for advection-diffusion-reaction on general meshes*, SIAM J. Numer. Anal., 53 (2015), pp. 2135–2157.
[32] D. A. Di Pietro, J. Droniou, and A. Harnist, *Improved error estimates for hybrid high-order discretizations of Leray-Lions problems*, Calcolo, 58 (2021), 19.
[33] D. A. Di Pietro, J. Droniou, and G. Manzini, *Discontinuous skeletal gradient discretisation methods on polytopal meshes*, J. Comput. Phys., 355 (2018), pp. 397–425.
[34] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Math. Appl. (Berlin) 69, Springer, Heidelberg, 2012.
[35] D. A. Di Pietro and A. Ern, *A hybrid high-order locking-free method for linear elasticity on general meshes*, Comput. Methods Appl. Mech. Engrg., 283 (2015), pp. 1–21.
[36] D. A. Di Pietro, A. Ern, and S. Lemaire, *An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators*, Comput. Methods Appl. Math., 14 (2014), pp. 461–472.
[37] D. A. Di Pietro, A. Ern, A. Linke, and F. Schieweck, *A discontinuous skeletal method for the viscosity-dependent Stokes problem*, Comput. Methods Appl. Mech. Engrg., 306 (2016), pp. 175–195.
[38] D. A. Di Pietro and S. Krell, *A hybrid high-order method for the steady incompressible Navier-Stokes problem*, J. Sci. Comput., 74 (2018), pp. 1677–1705.
[39] D. A. Di Pietro and S. Lemaire, *An extension of the Crouzeix-Raviart space to general meshes with application to quasi-incompressible linear elasticity and Stokes flow*, Math. Comp., 84 (2015), pp. 1–31.
[40] J. Douglas, Jr., and T. Dupont, *A Galerkin method for a nonlinear Dirichlet problem*, Math. Comp., 29 (1975), pp. 689–696.
[41] J. Douglas, Jr., T. Dupont, and J. Serrin, *Uniqueness and comparison theorems for nonlinear elliptic equations in divergence form*, Arch. Ration. Mech. Anal., 42 (1971), pp. 157–168.
[42] J. Droniou, *Finite volume schemes for diffusion equations: Introduction to and review of modern methods*, Math. Models Methods Appl. Sci., 24 (2014), pp. 1575–1619.
[43] J. Droniou and R. Eymard, *A mixed finite volume scheme for anisotropic diffusion problems on any grid*, Numer. Math., 105 (2006), pp. 35–71.
[44] J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin, *The Gradient Discretisation Method*, Math. Appl. (Berlin) 82, Springer, Cham, 2018.
[45] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin, *A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods*, Math. Models Methods Appl. Sci., 20 (2010), pp. 265–295.
[46] J. Droniou, R. Eymard, and R. Herbin, *Gradient schemes: Generic tools for the numerical analysis of diffusion equations*, ESAIM Math. Model. Numer. Anal., 50 (2016), pp. 749–781.
[47] J. Droniou and L. Yemm, *Robust hybrid high-order method on polytopal meshes with small faces*, Comput. Methods Appl. Math., 22 (2022), pp. 44–71.
[48] R. Eymard, T. Gallouët, and R. Herbin, *Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: A scheme using stabilization and hybrid interfaces*, IMA J. Numer. Anal., 30 (2010), pp. 1009–1043.
[49] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics Math., Springer, Berlin, 2001. Reprint of the 1998 edition.
[50] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Monogr. Stud. Math. 24, Pitman, Boston, MA, 1985.
[51] P. Grisvard, *Singularities in Boundary Value Problems*, Rech. Math. Appl. 22, Masson, Paris, 1992.
[52] T. Gudi, N. Nataraj, and A. K. Pani, *hp-discontinuous Galerkin methods for strongly nonlinear elliptic boundary value problems*, Numer. Math., 109 (2008), pp. 233–268.
[53] T. Gudi, N. Nataraj, and A. K. Pani, *An hp-local discontinuous Galerkin method for some quasilinear elliptic boundary value problems of nonmonotone type*, Math. Comp., 77 (2008), pp. 731–756.
[54] T. Gudi and A. K. Pani, *Discontinuous Galerkin methods for quasi-linear elliptic problems of nonmonotone type*, SIAM J. Numer. Anal., 45 (2007), pp. 163–192.
[55] R. Herbin and F. Hubert, *Benchmark on discretization schemes for anisotropic diffusion problems on general grids*, in Finite Volumes for Complex Applications V, John Wiley & Sons, 2008, pp. 659–692.

[56] I. Hlaváček, M. Křížek, and J. Malý, *On Galerkin approximations of a quasilinear nonpotential elliptic problem of a nonmonotone type*, J. Math. Anal. Appl., 184 (1994), pp. 168–189.

[57] H. B. Keller, *Approximation methods for nonlinear problems with application to two-point boundary value problems*, Math. Comp., 29 (1975), pp. 464–474.

[58] Y. Kuznetsov, K. Lipnikov, and M. Shashkov, *The mimetic finite difference method on polygonal meshes for diffusion-type problems*, Comput. Geosci., 8 (2004), pp. 301–324.

[59] S. Lemaire, *Bridging the hybrid high-order and virtual element methods*, IMA J. Numer. Anal., 41 (2021), pp. 549–593.

[60] K. Lipnikov and G. Manzini, *A high-order mimetic method on unstructured polyhedral meshes for the diffusion equation*, J. Comput. Phys., 272 (2014), pp. 360–385.

[61] L. Liu, M. Křížek, and P. Neittaanmäki, *Higher order finite element approximation of a quasilinear elliptic boundary value problem of a non-monotone type*, Appl. Math., 41 (1996), pp. 467–478.

[62] G. Mallik and N. Nataraj, *A nonconforming finite element approximation for the von Karman equations*, ESAIM Math. Model. Numer. Anal., 50 (2016), pp. 433–454.

[63] F. A. Milner, *Mixed finite element methods for quasilinear second-order elliptic problems*, Math. Comp., 44 (1985), pp. 303–320.

[64] L. Mu, J. Wang, and X. Ye, *Weak Galerkin finite element methods on polytopal meshes*, Int. J. Numer. Anal. Model., 12 (2015), pp. 31–53.

[65] L. Song and Z. Zhang, *Superconvergence property of an over-penalized discontinuous Galerkin finite element gradient recovery method*, J. Comput. Phys., 299 (2015), pp. 1004–1020.

[66] J. Wang and X. Ye, *A weak Galerkin finite element method for second-order elliptic problems*, J. Comput. Appl. Math., 241 (2013), pp. 103–115.

[67] J. Wang and X. Ye, *A weak Galerkin mixed finite element method for second order elliptic problems*, Math. Comp., 83 (2014), pp. 2101–2126.

[68] J. Xu, *Two-grid discretization techniques for linear and nonlinear PDEs*, SIAM J. Numer. Anal., 33 (1996), pp. 1759–1777.