Analysis of the Staggered DG Method for the Quasi-Newtonian Stokes flows

Liu, Jingyu; Liu, Yang; Zhao, Lina

Published in:
Journal of Scientific Computing

Published: 01/01/2025

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

Published version (DOI):
10.1007/s10915-024-02741-9

Publication details:
Liu, J., Liu, Y., & Zhao, L. (2025). Analysis of the Staggered DG Method for the Quasi-Newtonian Stokes flows. *Journal of Scientific Computing*, *102*(1), Article 14. https://doi.org/10.1007/s10915-024-02741-9

Journal of Scientific Computing (2025) 102:14
https://doi.org/10.1007/s10915-024-02741-9

Analysis of the Staggered DG Method for the Quasi-Newtonian Stokes flows

Jingyu Liu[1] · Yang Liu[2] · Lina Zhao[1]

Received: 26 November 2023 / Revised: 30 October 2024 / Accepted: 8 November 2024

Abstract
This paper introduces and analyzes a staggered discontinuous Galerkin (DG) method for quasi-Newtonian Stokes flow problems on polytopal meshes. The method introduces the flux and tensor gradient of the velocity as additional unknowns and eliminates the pressure variable via the incompressibility condition. Thanks to the subtle construction of the finite element spaces used in our staggered DG method, no additional numerical flux or stabilization terms are needed. Based on the abstract theory for the non-linear twofold saddle point problems, we prove the well-posedness of our scheme. A priori error analysis for all the involved unknowns is also provided. In addition, the proposed scheme can be hybridizable and the global problem only involves the trace variables, rendering the method computationally attractive. Finally, several numerical experiments are carried out to illustrate the performance of our scheme.

Keywords Discontinuous Galerkin methods · Quasi-Newtonian Stokes flow · Polygonal mesh · Hybridization

1 Introduction

This paper addresses the quasi-Newtonian Stokes flow problems on a bounded polygonal or polyhedral domain $\Omega \subset \mathbb{R}^d$, $d = 2,\ 3$, with a Lipschitz boundary $\Gamma = \partial\Omega$

$$-\nabla \cdot \{\mu(\boldsymbol{x}, |\nabla \boldsymbol{u}|)\nabla \boldsymbol{u}\} + \nabla p = \boldsymbol{f} \qquad \text{in } \Omega, \tag{1.1}$$

$$\nabla \cdot \boldsymbol{u} = 0 \qquad \text{in } \Omega, \tag{1.2}$$

$$\boldsymbol{u} = 0 \qquad \text{on } \Gamma, \tag{1.3}$$

✉ Lina Zhao
linazha@cityu.edu.hk

Jingyu Liu
jingyliu28-c@my.cityu.edu.hk

Yang Liu
liuyang@dhu.edu.cn

1 Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China

2 College of Information Science and Technology, Donghua University, 2999 North Renmin Road, Songjiang District, Shanghai 201-620, China

Published online: 20 November 2024

where $\boldsymbol{f} \in L^2(\Omega)^d$ is a given source term, $\boldsymbol{u}$ is the velocity vector, and p is the pressure satisfying the condition $\int_\Omega p\, dx = 0$.
Furthermore, $|\nabla \boldsymbol{u}|$ denotes the Frobenius norm of $\nabla \boldsymbol{u}$. The coefficient function μ is assumed to satisfy the following hypothesis.

Hypothesis 1 *The function $\mu(\boldsymbol{x}, t)$ satisfies the following conditions:*

(A1) $\mu \in C(\bar{\Omega} \times [0, \infty))$
(A2) There exist constants $M_\mu \geq m_\mu > 0$, such that

$$m_\mu(t - s) \leq \mu(\boldsymbol{x}, t)t - \mu(\boldsymbol{x}, s)s \leq M_\mu(t - s), \ \forall t \geq s \geq 0, \ \forall \boldsymbol{x} \in \bar{\Omega}.$$

Designing suitable numerical schemes for quasi-Newtonian Stokes flow problems has become an active research area in the past few decades. The abstract solvability and Galerkin approximation of non-linear twofold saddle point problems have been provided in [6]. The studies on the classical continuous Galerkin method for quasi-Newtonian Stokes flow problems are quite complete (readers can refer to [16]). With the development of numerical schemes for non-linear problems, such as quasi-linear elliptic problems(for example, [23–25, 30]) and non-linear diffusion problems (for example, [3, 7, 8, 22]), various DG methods on the quasi-Newtonian Stokes flow problems have been proposed. In [1], a posteriori error estimate is proposed for the quasi-Newtonian flow problem, where the mixed finite element discretizations are employed. A low-order mixed finite element method for the non-linear Stokes model is studied in [17], where the artificial flux and tensor corresponding to the velocity gradient are introduced as auxiliary unknowns. A local DG method for the non-linear model arising from the quasi-Newtonian Stokes flow with high-order polynomials is proposed in [9]. By introducing both tensors for velocity and pseudo-stress for incompressible non-linear Stokes problems, [26] suggested a dual-mixed finite element scheme for linear and non-linear Stokes model, and [19] considered an augmented mixed method with a priori and posterior estimate for the quasi-Newtonian Stokes flow problems. By introducing an extra term arising from the constitutive equation relating the pseudo-stress and the velocity, [10] suggested a mixed virtual element method. [15] suggested a hp-version interior penalty DG method with a priori, a posterior error analysis, and an automatic hp-adaptive refinement algorithm. The augmented hybridizable DG method for quasi-Newtonian Stokes flows is studied in [19–21]. In addition, some studies on solving the quasi-Newtonian Stokes flow problems on the unfitted meshes(for example, [29]) and coupled problems(for example, [2]) have been published recently.

The staggered DG method was initially proposed for wave propagation problems in [11, 12]. Since then, this method has been widely applied in solving partial differential equations (for example, [27, 28, 33–36]). Different from other DG methods, there is no need to introduce additional numerical flux or stabilization terms in the staggered DG method thanks to the construction of carefully balanced spaces. The key idea of the methods lies in the construction of the subtriangulations that are generated via connecting the interior point to all the vertices. Then, finite element spaces earning partial continuity properties are defined. We emphasize that the construction of the spaces needs to be carefully balanced to ensure the inf-sup condition, which is crucial in guaranteeing the unique solvability of the method. Thanks to this particular construction, many physical properties, such as the local conservation law, can be satisfied. Also, the staggered DG method can be flexibly applied in coupled problems (for example, [33, 36]). It is worth mentioning that even though only partial continuity properties are enforced on the finite element spaces, we can prove that the obtained velocity in our scheme is $H(\text{div}, \Omega)$-conforming and satisfies the pointwise divergence-free condition.

In this paper, we aim to design a staggered DG method for the quasi-Newtonian Stokes flow problems applicable to polytopal meshes. In our scheme, the incompressible condition is adopted, and the pressure variable can be eliminated by introducing additional tensor and pseudo-stress variables. Then, the second-order partial differential equation system (1.1)–(1.3) can be recast into a first-order partial differential equation system. Besides, the staggered continuity in our scheme is defined in the following sense: The functions in the finite element spaces for the pseudo-stress are normal continuous over the primal faces and normal tangential continuous over the dual faces; the functions in the finite element spaces for velocity are normal continuous over the dual faces. For the velocity gradient, we only enforce the trace-free condition on the corresponding finite element space, which arises from the incompressible condition. The proposed scheme yields divergence-free fluid velocity, which makes it well-suited for solving incompressible flow problems. Moreover, the proposed scheme also allows hanging nodes, which can be treated in a natural way by viewing them as additional vertices. This makes it particularly attractive for adaptive mesh refinement. The proposed scheme is hybridizable and the resulting global system only involves the trace variables, which makes our scheme computationally attractive. Moreover, our numerical experiments indicate that our scheme still works well even if the coefficient function μ is small, which is consistent with the study in [31].

The remainder of this paper is organized as follows. In Sect. 2, we introduce the quasi-Newtonian Stokes flow problem, detailing the derivation of the weak form and providing a proof of its well-posedness. In Sect. 3, we describe the construction of the staggered DG method for these flow problems. In Sect. 4, we show the inf-sup condition of the corresponding bilinear forms, which are fundamental for proving the well-posedness. In Sect. 5, we prove the convergence error estimates for all the involved variables measured in the suitable norms. To reduce the computational costs, the implementation with hybridization is introduced in Sect. 6. Finally, several numerical experiments are presented in Sect. 7 to demonstrate the performance of the proposed scheme.

2 Quasi-Newtonian Stokes Flows Problem

In this section, we shall derive the variational form of the quasi-Newtonian Stokes flow problem (1.1)–(1.3), and prove the well-posedness of it.

To start with, we introduce two additional unknowns: $\underline{L} = \nabla \boldsymbol{u}$ and $\underline{G} = \mu(|\underline{L}|)\underline{L} - p\underline{I}$, where $\underline{I}$ represents the $d \times d$ identity matrix. This enables us to reformulate the quasi-Newtonian Stokes flow problems (1.1)–(1.3) into the following first-order system:

$$\underline{G} = \mu(\boldsymbol{x}, |\underline{L}|)\underline{L} - p\underline{I} \quad \text{in } \Omega, \tag{2.1}$$

$$\underline{L} = \nabla \boldsymbol{u} \quad \text{in } \Omega, \tag{2.2}$$

$$-\nabla \cdot \underline{G} = \boldsymbol{f} \quad \text{in } \Omega, \tag{2.3}$$

$$\nabla \cdot \boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{2.4}$$

$$\boldsymbol{u} = \boldsymbol{0} \quad \text{on } \Gamma, \tag{2.5}$$

with the restriction $\int_\Omega p \, dx = 0$. Due to the divergence-free condition (2.4), we can infer that

$$\operatorname{tr}(\underline{L}) = \operatorname{tr}(\nabla \boldsymbol{u}) = \nabla \cdot \boldsymbol{u} = 0,$$
$$\operatorname{tr}(\underline{G}) = \operatorname{tr}(\mu(\boldsymbol{x}, |\underline{L}|)\underline{L} - p\underline{I}) = \mu(\boldsymbol{x}, |\underline{L}|)\operatorname{tr}(\underline{L}) - dp = -dp.$$

Inspired by this, we shall introduce an operator $\mathcal{A} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ defined by $\mathcal{A}\underline{\sigma} = \underline{\sigma} - \frac{\text{tr}(\underline{\sigma})}{d}\underline{I}$ to recast the system (2.1)–(2.5) into the following system:

$$\mathcal{A}\underline{G} = \mu(\boldsymbol{x}, |\underline{L}|)\underline{L} \quad \text{in } \Omega, \tag{2.6}$$

$$\mathcal{A}\underline{L} = \nabla\boldsymbol{u} \quad \text{in } \Omega, \tag{2.7}$$

$$-\nabla\cdot\underline{G} = \boldsymbol{f} \quad \text{in } \Omega, \tag{2.8}$$

$$\boldsymbol{u} = \boldsymbol{0} \quad \text{on } \Gamma, \tag{2.9}$$

where we apply the fact $\mathcal{A}\underline{L} = \underline{L} - \frac{\text{tr}(\underline{L})}{d}\underline{I} = \underline{L}$ in (2.7).

Before deriving the variational form of (2.6)–(2.9), we shall introduce some definitions and notations applied throughout this paper.

For a set $D \subset \mathbb{R}^d$, the notation $(\cdot,\cdot)_D$ represents the inner product in $L^2(D)$, specifically $(p,q)_D := \int_D pq\, dx$ for $p,\ q \in L^2(D)$. The same symbol is used for inner products in $L^2(D)^d$ and $L^2(D)^{d\times d}$. For example, for $\boldsymbol{u},\ \boldsymbol{v} \in L^2(D)^d$, $(\boldsymbol{u},\boldsymbol{v})_D := \sum_{i=1}^{d}(u_i, v_i)_D$ and for $\underline{P},\ \underline{Q} \in L^2(D)^{d\times d}$, $(\underline{P},\underline{Q})_D := \sum_{i=1}^{d}\sum_{j=1}^{d}(p_{ij}, q_{ij})_D$. $C_c^\infty(D)$ represents the set of functions defined on D with compact support and differentiable for all degrees of differentiation. For any integer $m \geq 0$, the scalar-value Sobolev spaces are denoted as $H^m(D) = W^{m,2}(D)$ with norm $\|\cdot\|_{m,D}$ and seminorm $|\cdot|_{m,D}$. Additionally, in unambiguous case, $\|\cdot\|_{m,D}$ and $|\cdot|_{m,D}$ are also used to denote the Sobolev norm and seminorm of functions in $H^m(D)^d$ or $H^m(D)^{d\times d}$, respectively. When D coincides with Ω, the subscript Ω will be omitted for convenience. Similarly, for $F \subset \mathbb{R}^{d-1}$ (or duality pair), $(\cdot,\cdot)_F$ denotes the inner product of functions in $L^2(F)$, $L^2(F)^d$, or $L^2(F)^{d\times d}$, and $\|\cdot\|_{0,F}$ denotes the corresponding L^2-norm on F.

Now, we will introduce the spaces of test functions for deriving the continuous variational form. For (2.6)–(2.9), we seek the weak solution $(\boldsymbol{u}, \underline{L}, \underline{G}) \in \boldsymbol{U} \times \underline{X} \times \underline{W}_{tr}$, where

$$\boldsymbol{U} = H_0^1(\Omega)^d,$$

$$\underline{X} = \{\underline{Q} \in L^2(\Omega)^{d\times d} : \text{tr}(\underline{Q}) = 0\},$$

$$\underline{W}_{tr} = \{\underline{H} \in L^2(\Omega)^{d\times d} : \int_\Omega \text{tr}(\underline{H})dx = 0\}.$$

Poincare inequality implies that $\|\nabla\boldsymbol{v}\|_0$ is a norm defined on $\boldsymbol{U}$. Multiplying $\underline{Q} \in \underline{X}$ on the both side of (2.6), we can reach to

$$(\mathcal{A}\underline{G}, \underline{Q}) = (\mu(|\underline{L}|)\underline{L}, \underline{Q}).$$

After multiplying $\underline{H} \in \underline{W}_{tr}$ on both sides of (2.7), we can obtain that

$$(\mathcal{A}\underline{L}, \underline{H}) = (\nabla\boldsymbol{u}, \underline{H}).$$

Finally, by multiplying $\boldsymbol{v} \in \boldsymbol{U}$ on both sides of (2.9) and conducting integration by part, we can infer that

$$(\underline{G}, \nabla\boldsymbol{v}) = (\boldsymbol{f}, \boldsymbol{v}).$$

After defining the following forms

$$\begin{aligned} A(\underline{L}, \underline{Q}) &= (\mu(|\underline{L}|\underline{L}, \underline{Q}), \ \forall \underline{L}, \underline{Q} \in \underline{X}, \\ B(\underline{H}, \boldsymbol{v}) &= (\underline{H}, \nabla \boldsymbol{v}), \forall \underline{H} \in \underline{W}_{tr}, \ \boldsymbol{v} \in \boldsymbol{U}, \\ C^*(\underline{H}, \underline{Q}) &= (\mathcal{A}\underline{H}, \underline{Q}), \ \forall \underline{H} \in \underline{W}_{tr}, \ \underline{Q} \in \underline{X}, \end{aligned} \tag{2.10}$$

the weak formulation can be introduced as follows: Find $(\underline{G}, \underline{L}, \boldsymbol{u}) \in \underline{W}_{tr} \times \underline{X} \times \boldsymbol{U}$, such that, the following equations hold for all $(\underline{H}, \underline{Q}, \underline{v}) \in \underline{W}_{tr} \times \underline{X} \times \boldsymbol{U}$:

$$\begin{aligned} A(\underline{L}, \underline{Q}) - C^*(\underline{G}, \underline{Q}) &= 0, \\ C^*(\underline{H}, \underline{L}) - B(\underline{H}, \boldsymbol{u}) &= 0, \\ B(\underline{G}, \boldsymbol{v}) &= (\boldsymbol{f}, \boldsymbol{v}). \end{aligned} \tag{2.11}$$

In order to prove the well-posedness of (2.11), we shall introduce the following theorem:

Theorem 2.1 [18, Theorem 2.1] *Let $\tilde{\underline{W}}_{tr}$ be the kernel space of B, that is*

$$\tilde{\underline{W}}_{tr} := \{\underline{H} \in \underline{W}_{tr} : B(\underline{H}, \boldsymbol{v}) = 0, \forall \boldsymbol{v} \in \boldsymbol{U}\}.$$

Assume that:

(i) The nonlinear form $A : \underline{X} \times \underline{X} \to \mathbb{R}$ is Lipschitz continuous and strongly monotone on the first variable, that is, there exist $C_{c,1}, C_{c,2} > 0$, such that

$$\sup_{\underline{0} \neq \underline{Q} \in \underline{X}} \frac{A(\underline{Q}_1, \underline{Q}) - A(\underline{Q}_2, \underline{Q})}{\|\underline{Q}\|_0} \leq C_{c,1}\|\underline{Q}_1 - \underline{Q}_2\|_0, \ \forall \underline{Q}_1, \ \underline{Q}_2, \ \underline{Q} \in \underline{X},$$

and

$$A(\underline{Q}_1, \underline{Q}_1 - \underline{Q}_2) - A(\underline{Q}_2, \underline{Q}_1 - \underline{Q}_2) \geq C_{c,2}\|\underline{Q}_1 - \underline{Q}_2\|_0^2, \ \forall \underline{Q}_1, \ \underline{Q}_2 \in \underline{X}.$$

(ii) The bilinear form B satisfies a uniform inf-sup condition on $\boldsymbol{U} \times \underline{W}_{tr}$, that is, there exists a constant $C_{c,3} > 0$, such that for any $\boldsymbol{v} \in \boldsymbol{U}$

$$\sup_{\underline{0} \neq \underline{H} \in \underline{W}_{tr}} \frac{B(\underline{H}, \boldsymbol{v})}{\|\underline{H}\|_0} \geq C_{c,3}\|\nabla \boldsymbol{v}\|_0.$$

(iii) The bilinear form C^* satisfies an inf-sup condition on $\tilde{\underline{W}}_{tr} \times \underline{X}$, that is, there exists a constant $C_{c,4} > 0$, such that

$$\sup_{\underline{0} \neq \underline{Q} \in \underline{X}} \frac{C^*(\underline{H}, \underline{Q})}{\|\underline{Q}\|_0} \geq C_{c,4}\|\underline{H}\|_0.$$

Then, for any given $\boldsymbol{f} \in L^2(\Omega)^d$, there exists a unique $(\underline{G}, \underline{L}, \boldsymbol{u}) \in \underline{W}_{tr} \times \underline{X} \times \boldsymbol{U}$ satisfying (2.11). In addition, there exists a constant $C > 0$, only depending on $C_{c,1}, C_{c,2}, C_{c,3}, C_{c,4}$, such that

$$\|\underline{G}\|_0 + \|\underline{L}\|_0 + \|\nabla \boldsymbol{u}\|_0 \leq C\|\boldsymbol{f}\|_0.$$

According to Theorem 2.1, it suffices to verify conditions (i)–(iii). To this end, we shall first introduce two lemmas.

Lemma 2.1 [5, Lemma 2.1] *Under the assumption that μ satisfies the Hypothesis 1., there exist constants $C_1,\ C_2 > 0$, such that for any $\underline{W},\ \underline{V} \in \mathbb{R}^{d\times d}$ and any $\boldsymbol{x} \in \bar{\Omega}$, the following inequalities hold*

$$|\mu(\boldsymbol{x}, |\underline{W}|)\underline{W} - \mu(\boldsymbol{x}, |\underline{V}|)\underline{V}| \le C_1|\underline{W} - \underline{V}|, \tag{2.12}$$
$$C_2|\underline{W} - \underline{V}|^2 \le (\mu(\boldsymbol{x}, |\underline{W}|)\underline{W} - \mu(\boldsymbol{x}, |\underline{V}|)\underline{V}) : (\underline{W} - \underline{V}). \tag{2.13}$$

where $|\cdot|$ means the Frobenius norm in $\mathbb{R}^{d\times d}$, which is for any $\underline{A} = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix}$,
$|\underline{A}|^2 = \sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}^2.$

Lemma 2.2 (Energy norm defined on $\underline{\tilde{W}}_{tr}$) *In the space $\underline{\tilde{W}}_{tr}$, a non-negative operator is defined as:*

$$\begin{aligned} \|\cdot\|_E : \underline{\tilde{W}}_{tr} &\to \mathbb{R}, \\ \underline{H} &\to \|\underline{H}\|_E = \|\mathcal{A}\underline{H}\|_0. \end{aligned} \tag{2.14}$$

Then, (2.14) defines a norm on $\underline{\tilde{W}}_{tr}$. In addition, $\|\cdot\|_E$ is equivalent to $\|\cdot\|_0$.

Proof First, we shall prove that $\|\cdot\|_E$ is a norm defined on $\underline{\tilde{W}}_{tr}$.

(i) For the semi-linear property, for any $\underline{H_1},\ \underline{H_2} \in \underline{\tilde{W}}_{tr}$ and $\lambda,\ \nu \in \mathbb{R}$, thanks to the semi-linear property of L^2-norm and the linear property of $\mathcal{A}$, we can derive that

$$\begin{aligned} \|\lambda\underline{H_1} + \nu\underline{H_2}\|_E &= \|\lambda\mathcal{A}(\underline{H_1}) + \nu\mathcal{A}(\underline{H_2})\|_0 \\ &\le |\lambda|\|\mathcal{A}(\underline{H_1})\|_0 + |\nu|\|\mathcal{A}(\underline{H_2})\|_0 = |\lambda|\|\underline{H_1}\|_E + |\nu|\|\underline{H_2}\|_E. \end{aligned}$$

(ii) For positiveness, the non-negativity can be shown by

$$\|\underline{H}\|_E = \|\mathcal{A}\underline{H}\|_0 \ge 0,\ \forall\underline{H} \in W_{tr}.$$

(iii) It remains to prove that for any $\underline{H} \in \underline{\tilde{W}}_{tr}$, $\|\underline{H}\|_E = 0$ can lead to $\underline{H} = \underline{0}$. It suffices to show that there exists a constant $C > 0$ such that

$$\|\underline{H}\|_0 \le C\|\underline{H}\|_E$$

holds for all $\underline{H} \in \underline{\tilde{W}}_{tr}$.

Since $\mathrm{tr}\underline{H} \in L_0^2(\Omega)$, according to [4, Lemma 5.4.2], there exists $\boldsymbol{w} \in H_0^1(\Omega)^d$ such that $\nabla\cdot\boldsymbol{w} = \mathrm{tr}\underline{H}$. In addition, there exists a constant $C > 0$, only depending on Ω, such that $\|\boldsymbol{w}\|_1 \le C\|\mathrm{tr}\underline{H}\|_0$. Then, we can infer that

$$\|\mathrm{tr}\underline{H}\|_0^2 = (\mathrm{tr}\underline{H}, \nabla\cdot\boldsymbol{w}) = ((\mathrm{tr}\underline{H})\underline{I}, \nabla\boldsymbol{w}) = d(\underline{H} - \mathcal{A}\underline{H}, \nabla\boldsymbol{w}) = -d(\mathcal{A}\underline{H}, \nabla\boldsymbol{w}) \le d\|\underline{H}\|_E\|\boldsymbol{w}\|_1,$$

leading to the fact that $\|\mathrm{tr}\underline{H}\|_0 \le dC\|\underline{H}\|_E$. Thus, simple calculation implies

$$\|\underline{H}\|_0 \le \|\underline{H}\|_E + \|\mathrm{tr}\underline{H}\|_0 \le (dC + 1)\|\underline{H}\|_E, \tag{2.15}$$

holds for all $\underline{H} \in \underline{\tilde{W}}_{tr}$, where $C > 0$ only depends on Ω.

So far, we have proved that $\|\cdot\|_E$ is a norm defined on $\underline{\tilde{W}}_{tr}$. Furthermore, a straightforward calculation reveals that, for any $\underline{O} \in L^2(\Omega)^{d\times d}$,

$$(\mathcal{A}\underline{O}, \operatorname{tr}(\underline{O})\underline{I}) = 0.$$

This orthogonal condition combined with the non-negativity of $L^2(\Omega)$ norm indicates that

$$\|\underline{H}\|_E^2 \le \|\underline{H}\|_E^2 + \|\frac{\operatorname{tr}(H)}{d}\underline{I}\|_0^2 = \|\underline{H}\|_0^2. \tag{2.16}$$

Inequalities (2.15) and (2.16) imply that $\|\cdot\|_0$ is equivalent to $\|\cdot\|_E$ on $\underline{\tilde{W}}_{tr}$. □

Now, we are ready to prove the following theorem, which will lead to the well-posedness of continuous problem (2.11).

Theorem 2.2 *The forms defined in* (2.10) *satisfy the condition (i)–(iii) in Theorem* 2.1.

Proof (i) According to (2.12) and (2.13) in Lemma 2.1, it is easy to prove the condition (i) in Theorem 2.1.

(ii) For any $\boldsymbol{v} \in \boldsymbol{U}$ and constant $c \in \mathbb{R}$, we can observe that

$$B(c\underline{I}, \boldsymbol{v}) = (c\underline{I}, \nabla\boldsymbol{v}) = -(\boldsymbol{v}, c\nabla\cdot\underline{I}) = 0. \tag{2.17}$$

In addition, for any $\underline{R} \in L^2(\Omega)^{d\times d}$ with $\int_\Omega \operatorname{tr}(\underline{R})dx = c$, simple calculation implies that

$$\|\underline{R}\|_0^2 = \|\underline{R} - \frac{c}{d}\underline{I}\|_0^2 + \frac{c^2}{d}|\Omega| \ge \|\underline{R} - \frac{c}{d}\underline{I}\|_0^2. \tag{2.18}$$

By taking $\underline{H} = \nabla\boldsymbol{v} - \frac{\int_\Omega \nabla\cdot\boldsymbol{v}\,dx}{d|\Omega|}\underline{I} \in \underline{W}_{tr}$, with the help of equations (2.17)–(2.18) and the linear property of bilinear form B, we can conclude that

$$\frac{B(\underline{H}, \boldsymbol{v})}{\|\underline{H}\|_0} = \frac{\|\nabla\boldsymbol{v}\|_0^2}{\|\nabla\boldsymbol{v} - \frac{\int_\Omega \nabla\cdot\boldsymbol{v}\,dx}{d|\Omega|}\underline{I}\|_0} \ge \|\nabla\boldsymbol{v}\|_0. \tag{2.19}$$

(iii) For any $\underline{H} \in \underline{\tilde{W}}_{tr}$, there is $\operatorname{tr}(\mathcal{A}\underline{H}) = 0$, leading to the fact that $\mathcal{A}\underline{H} \in \underline{X}$. By taking $\underline{Q} = \mathcal{A}\underline{H}$ and taking the norm equivalence obtained in Lemma 2.2 into account, the inf-sup condition can be obtained by

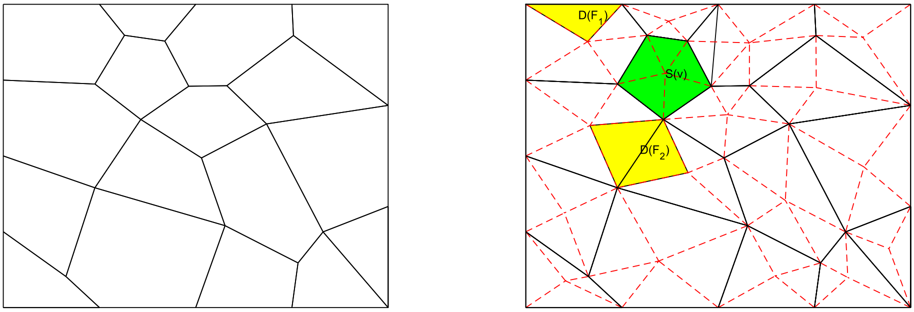
$$\begin{aligned} C^*(\underline{H}, \underline{Q}) &= (\mathcal{A}\underline{H}, \underline{Q}) \\ &= (\mathcal{A}\underline{H}, \mathcal{A}\underline{H}) = (\|\underline{H}\|_E)^2 \ge C^{-1}\|\underline{H}\|_0\|\underline{H}\|_E = C^{-1}\|\underline{H}\|_0\|\underline{Q}\|_0. \end{aligned} \tag{2.20}$$

□

3 Description of the Staggered DG Scheme

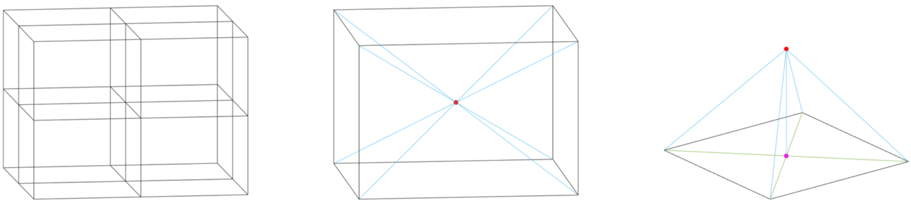
In this section, we present our staggered DG scheme for numerically solving the quasi-Newtonian Stokes flow problems. We first introduce some notation that will be used throughout the paper. Following [31], we let $\mathcal{T}_u$ represent the general polygonal or polyhedral mesh of domain Ω, consisting of nonempty, connected, disjoint subsets of Ω:

$$\Omega = \cup_{M\in\mathcal{T}_u} M.$$



(a) Mesh before subdivision (b) Mesh after subdivision

Fig. 1 Schematic of $D(F_1)$, $D(F_2)$ and $S(v)$ in two-dimensional case



(a) primal mesh (b) first division (c) second division

Fig. 2 subdivision in three-dimensional case

We now proceed to subdivide the primal mesh. In the two-dimensional case ($d = 2$), for each element $M \in \mathcal{T}_u$, by connecting the midpoint v_M of M, to all the vertices of M, we obtain the submesh $\mathcal{T}_h$ of $\mathcal{T}_u$ (see Fig. 1).

In the three-dimensional case ($d = 3$), we propose an innovative approach to subdivide the primal meshes, building upon the methodology presented in [12], originally developed for tetrahedral meshes, we extend it to general polyhedral meshes. To begin with, for each convex polyhedral element $M \in \mathcal{T}_u$, we define $\mathcal{F}_M$ as the set of all faces of M, and $\mathcal{N}_M$ as the set of all vertices of M. We then connect the midpoint v_M of M to each vertex $v \in \mathcal{N}_M$. This process divides M into $|\mathcal{F}_M|$ disjoint pyramids. Next, within each pyramid M_k, where $k = 1, \cdots, |\mathcal{F}_M|$, there exists exactly one polygonal face $F_{M,k} \in \mathcal{F}_M$, excluding v_M from its vertex set. All the remaining faces in this pyramid are triangles sharing v_M as their common vertex. We designate $v_{M,k}$ as the midpoint of $F_{M,k}$ and connect it to all the vertices of M_k. Through this two-step subdivision process, we successfully obtain a tetrahedral submesh $\mathcal{T}_h$, which adheres to the assumptions outlined in [12, Lemma 2.1]. To facilitate better understanding, we provide an illustrative example using a hexahedral mesh (Fig. 2a). In this example, each element M (Fig. 2b) has a red point representing the midpoint v_M. We generate six disjoint pyramids by connecting v_M to all the vertices of M (indicated by the blue lines). Within each pyramid M_k (Fig. 2c), the pink point denotes the midpoint $v_{M,k}$ of the face $F_{M,k}$, which excludes v_M (red point) from its vertex set. Finally, by connecting $v_{M,k}$ to all the vertices of this pyramid (represented by the green lines), we achieve the desired tetrahedral submesh.

In the subsequent sections, we introduce various notations and assumptions to facilitate the analysis. Firstly, we denote the set of midpoints of primal elements as $\mathcal{N}$, specifically defined as

$$\mathcal{N} := \{v : v \text{ is the midpoint of } M, M \in \mathcal{T}_u\}.$$

Furthermore, we represent the set of all faces in $\mathcal{T}_h$ as $\mathcal{F}$. The set of all dual faces, where each dual face has a vertex $v \in \mathcal{N}$, is denoted as $\mathcal{F}_{dl}$, while the set of primal faces is denoted as $\mathcal{F}_{pr} := \mathcal{F} \backslash \mathcal{F}_{dl}$. For later use, we partition $\mathcal{F}_{pr}$ into $\mathcal{F}_{pr}^0$ and $\mathcal{F}_{pr}^b$, where $\mathcal{F}_{pr}^b := \{F \in \mathcal{F} : |F \cap \partial\Omega| > 0\}$ denotes the set of boundary primal faces, and $\mathcal{F}_{pr}^0 := \mathcal{F}_{pr} \backslash \mathcal{F}_{pr}^b$ denotes the set of interior primal faces. For example, in Fig. 1, the red dotted lines represent the dual faces, and the black solid lines represent the primal faces. In Fig. 2c, faces with two blue edges and one black edge are dual faces, and faces with two green edges and one black edge are primal faces.

We assume that the triangulation $\mathcal{T}_h$ satisfies the standard regularity assumption as outlined in [13]. For each element τ in $\mathcal{T}_h$, we use h_τ to denote the diameter of τ and define $h = \max\{h_\tau : \tau \in \mathcal{T}_h\}$. Furthermore, the diameter of face $F \in \mathcal{F}$ is denoted as h_F. Throughout this paper, C represents a generic positive constant independent of mesh size h, which may have different values in different contexts.

For each interior point $v \in \mathcal{N}$, let $S(v)$ denote the union of elements in $\mathcal{T}_h$ sharing the common vertex v. We assume that $S(v)$ is star-shaped with respect to a ball of radius $\rho h_{S(v)}$, where ρ is a positive constant. Additionally, for each $F \in \partial S(v)$, we assume that $h_F \geq C h_{S(v)}$. For each boundary primal face $F_1 \subset \mathcal{F}_{pr}^b$, $D(F_1)$ denotes the element that includes F_1 as one of its faces. For each interior primal face $F_2 \in \mathcal{F}_{pr}^0$, $D(F_2)$ represents the two elements sharing the common interior primal face F_2, as illustrated in Fig. 1. For each face $F \in \mathcal{F}$, a unit normal vector $\boldsymbol{n}_F$ is specified in the following manner: If $F \in \mathcal{F}_{pr}^b$, $\boldsymbol{n}_F$ refers to the outward unit normal vector of F. Otherwise, $\boldsymbol{n}_F$ is selected as one of the two possible unit normal vectors of F. The $d-1$ dimensional linear space consisting of vectors that are parallel to F is denoted as

$$\boldsymbol{T}_F := \{\boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{a} \cdot \boldsymbol{n}_F = 0\}.$$

Furthermore, to maintain consistency in our analysis for both two-dimensional and three-dimensional cases, we introduce the tangential component of vector $\boldsymbol{v}$ on $F \in \mathcal{F}$ as:

$$(\boldsymbol{v})^{\boldsymbol{t}_F} := \boldsymbol{v} - (\boldsymbol{v} \cdot \boldsymbol{n}_F)\boldsymbol{n}_F.$$

Let $k \geq 0$, $P^k(\tau)$ and $P^k(F)$ represent the polynomials of degree up to k on $\tau \in \mathcal{T}_h$ and $F \in \mathcal{F}$, respectively. Furthermore, the jump and average of a scalar, vector, or tensor function v on interior face $F \in \mathcal{F}^0$ are defined as

$$[v]_{|F} := v_1 - v_2, \quad \{v\}_{|F} := \frac{v_1 + v_2}{2},$$

where τ_1 and τ_2 denote the two elements sharing the face F, and $v_i = (v_{|\tau_i})_{|F}$, $i = 1, 2$. For $F \in \mathcal{F}_{pr}^b$, the jump and average of v on F are defined as

$$[v]_{|F} := v_1, \quad \{v\}_{|F} := v_1,$$

where $v_1 = (v_{|\tau_1})_{|F}$. When there is no ambiguity about which face is under consideration, the subscript of $\boldsymbol{n}_F$, $(\cdot)^{\boldsymbol{t}_F}$, $[\cdot]_{|F}$, and $\{\cdot\}_{|F}$ will be omitted for brevity. Also, in the following sections, the elementwise gradient and divergence are denoted as ∇_h and $\nabla_h\cdot$, respectively.

Then, we introduce the following finite element spaces for later use. To begin with, we define the locally $H(\text{div};\Omega)$-conforming finite element space for velocity $\boldsymbol{u}$

$$\boldsymbol{U}^h := \{\boldsymbol{v}_h : (\boldsymbol{v}_h)_{|\tau} \in P^k(\tau)^d;\ \tau \in \mathcal{T}_h;\ [\boldsymbol{v}_h \cdot \boldsymbol{n}]_{|F} = 0, \forall F \in \mathcal{F}_{dl}\},$$

equipped with the norm

$$\|\boldsymbol{v}_h\|^2_{\boldsymbol{U}^h} = \|\nabla_h \boldsymbol{v}_h\|_0^2 + \sum_{F\in\mathcal{F}_{pr}} h_F^{-1}\|[\boldsymbol{v}_h]\|^2_{0,F} + \sum_{F\in\mathcal{F}_{dl}} h_F^{-1}\|[(\boldsymbol{v}_h)^{\boldsymbol{t}}]\|^2_{0,F}.$$

Secondly, we define the trace-free finite element space for velocity gradient $\underline{L}$

$$\underline{X}^h := \{\underline{Q}_h : (\underline{Q}_h)_{|\tau} \in P^k(\tau)^{d\times d};\ \text{tr}(\underline{Q}_h) = 0\},$$

equipped with L^2-norm.
Finally, we define the finite element space for pseudo-stress $\underline{G}$. To start with, we introduce the finite element space given in [31]

$$\underline{W}^h := \{\underline{R}_h : (\underline{R}_h)_{|\tau} \in P^k(\tau)^{d\times d};\ [\underline{R}_h\boldsymbol{n}]_{|F} = \boldsymbol{0},\ \forall F \in \mathcal{F}^0_{pr};\ [(\underline{R}_h\boldsymbol{n})^{\boldsymbol{t}}]_{|F} = 0,\ \forall F \in \mathcal{F}_{dl}\},$$

equipped with the norm

$$\|\underline{R}_h\|^2_{\underline{W}^h} = \|\underline{R}_h\|_0^2 + \sum_{F\in\mathcal{F}_{pr}} h_F\|\underline{R}_h\boldsymbol{n}\|^2_{0,F} + \sum_{F\in\mathcal{F}_{dl}} h_F\|(\underline{R}_h\boldsymbol{n})^{\boldsymbol{t}}\|^2_{0,F}.$$

We define the finite element space for pseudo-stress $\underline{G}$

$$\underline{W}^h_{tr} := \{\underline{H}_h : \underline{H}_h \in \underline{W}^h,\ \text{tr}(\underline{H}_h) \in L^2_0(\Omega)\},$$

equipped with the norm $\|\cdot\|_{\underline{W}^h}$. The scaling arguments imply that there exists a positive constant C, independent of mesh size h, such that

$$\|\underline{H}_h\|_0 \le \|\underline{H}_h\|_{\underline{W}^h} \le C\|\underline{H}_h\|_0,\ \forall \underline{H}_h \in \underline{W}^h_{tr}. \tag{3.1}$$

Now, we are ready to derive our staggered DG scheme for the quasi-Newtonian Stokes flow problems (2.6)–(2.9). Firstly, by multiplying a test function $\underline{Q}_h \in \underline{X}^h$ on both sides of (2.6) and performing integration over the whole domain Ω, we obtain

$$(\mathcal{A}\underline{G}, \underline{Q}_h) = (\mu(|\underline{L}|)\underline{L}, \underline{Q}_h).$$

Next, by multiplying a test function $\underline{H}_h \in \underline{W}^h_{tr}$ on both sides of (2.7) and performing integration over $D(F)$ for $F \in \mathcal{F}_{pr}$, we can infer

$$(\mathcal{A}\underline{L}, \underline{H}_h)_{D(F)} = -(\boldsymbol{u}, \nabla_h \cdot \underline{H}_h)_{D(F)} + \sum_{F\in\mathcal{F}_{dl}\bigcap\partial D(F)} \int_F (\boldsymbol{u}\cdot\boldsymbol{n})\boldsymbol{n}\cdot\underline{H}_h\boldsymbol{n}\, ds,$$

where we use the decomposition for $\boldsymbol{u}$ on $F \in \mathcal{F}_{dl}$, i.e., $\boldsymbol{u} = (\boldsymbol{u}\cdot\boldsymbol{n})\boldsymbol{n} + (\boldsymbol{u})^{\boldsymbol{t}}$ and the facts $\boldsymbol{u}_{|F} = 0$ for any $F \in \mathcal{F}^b_{pr}$, $[\underline{H}_h\boldsymbol{n}]_{|F} = 0$ for any $\mathcal{F}^0_{pr}$ and $[(\underline{H}_h\boldsymbol{n})^{\boldsymbol{t}}]_{|F} = \boldsymbol{0}$, $(\boldsymbol{u})^{\boldsymbol{t}}\cdot\boldsymbol{n} = (\underline{H}_h\boldsymbol{n})^{\boldsymbol{t}}\cdot\boldsymbol{n} = 0$ for any $F \in \mathcal{F}_{dl}$.
Finally, by multiplying a test function $\boldsymbol{v}_h \in \boldsymbol{U}^h$ on both sides of (2.8) and performing integration over $S(\nu)$ for $\nu \in \mathcal{N}$, we can deduce

$$(\boldsymbol{f}, \boldsymbol{v}_h)_{S(\nu)} = (\underline{G}, \nabla_h\boldsymbol{v}_h)_{S(\nu)} - \sum_{F\in\mathcal{F}_{dl}\cap\partial S(\nu)} \int_F \underline{G}\boldsymbol{n}\cdot[(\boldsymbol{v}_h)^{\boldsymbol{t}}]\, ds - \sum_{F\in\mathcal{F}_{pr}\cap\partial S(\nu)} \int_F \underline{G}\boldsymbol{n}\cdot\boldsymbol{v}_h\, ds,$$

where we exploit the fact $[\boldsymbol{v}_h\cdot\boldsymbol{n}]_{|F} = 0$ for any $F \in \mathcal{F}_{dl}$ here.

After defining the forms

$$
\begin{aligned}
A_h(\underline{W}_h, \underline{Q}_h) &:= (\mu(|\underline{W}_h|)\underline{W}_h, \underline{Q}_h),\ \forall(\underline{W}_h, \underline{Q}_h) \in \underline{X}^h \times \underline{X}^h, \\
B_h^*(\boldsymbol{v}_h, \underline{H}_h) &:= -(\boldsymbol{v}_h, \nabla_h \cdot \underline{H}_h) \\
&\quad + \sum_{F\in\mathcal{F}_{dl}} \int_F (\boldsymbol{v}_h \cdot \boldsymbol{n})\boldsymbol{n} \cdot [\underline{H}_h \boldsymbol{n}]\, ds,\ \forall(\boldsymbol{v}_h, \underline{H}_h) \in \boldsymbol{U}^h \times \underline{W}_{tr}^h, \\
B_h(\underline{H}_h, \boldsymbol{v}_h) &:= (\underline{H}_h, \nabla_h \boldsymbol{v}_h) - \sum_{F\in\mathcal{F}_{pr}} \int_F [\boldsymbol{v}_h] \cdot (\underline{H}_h \boldsymbol{n})\, ds \\
&\quad - \sum_{F\in\mathcal{F}_{dl}} \int_F [(\boldsymbol{v}_h)^{\boldsymbol{t}} \cdot (\underline{H}_h \boldsymbol{n})^{\boldsymbol{t}}]\, ds,\ \forall(\underline{H}_h, \boldsymbol{v}_h) \in \underline{W}_{tr}^h \times \boldsymbol{U}^h, \\
C_h(\underline{H}_h, \underline{Q}_h) &:= (\mathcal{A}\underline{H}_h, \underline{Q}_h),\ \forall(\underline{H}_h, \underline{Q}_h) \in \underline{W}_{tr}^h \times \underline{X}^h,
\end{aligned}
\tag{3.2}
$$

we can formulate our scheme for quasi-Newtonian Stokes flow problems: Find $(\underline{G}_h,\ \underline{L}_h,\ \boldsymbol{u}_h) \in \underline{W}_{tr}^h \times \underline{X}^h \times \boldsymbol{U}^h$ such that

$$A_h(\underline{L}_h, \underline{Q}_h) - C_h(\underline{G}_h, \underline{Q}_h) = 0, \tag{3.3}$$

$$B_h^*(\boldsymbol{u}_h, \underline{H}_h) - C_h(\underline{H}_h, \underline{L}_h) = 0, \tag{3.4}$$

$$B_h(\underline{G}_h, \boldsymbol{v}_h) = (\boldsymbol{f}, \boldsymbol{v}_h), \tag{3.5}$$

for all $(\underline{H}_h, \underline{Q}_h, \boldsymbol{v}_h) \in \underline{W}_{tr}^h \times \underline{X}^h \times \boldsymbol{U}^h$. Here, we apply the fact that for any $\underline{W}_h,\ \underline{H}_h \in P^k(\tau)^{d\times d}$ and $\tau \in \mathcal{T}_h$, the equality $(\mathcal{A}\underline{W}_h, \underline{H}_h)_\tau = (\underline{W}_h, \mathcal{A}\underline{H}_h)_\tau$ holds.

Before closing this section, we would like to show that, even though only the normal continuity on the dual faces is defined on the finite element space $\boldsymbol{U}^h$, the fluid velocity $\boldsymbol{u}_h$ obtained from our scheme (3.3)–(3.5) is normal continuous over both primal faces and dual faces, and satisfies the divergence-free condition, which makes it a good candidate for solving incompressible fluid equations. Before we prove this property, we first introduce the following space

$$P^h := \{q_h : (q_h)_{|\tau} \in P^k(\tau),\ \forall \tau \in \mathcal{T}_h;\ [q_h]_{|F} = 0,\ \forall F \in \mathcal{F}_{pr}^0\}. \tag{3.6}$$

The degrees of freedom for the space P^h are defined in the following lemma from [11, Lemma 2.2], [12, Lemma 2.2] and [31].

Lemma 3.1 *Any function $q_h \in P^h$ is uniquely determined by the following degrees of freedom (VD1)–(VD2):*

(VD1) For any $F \in \mathcal{F}_{pr}$,

$$\Phi_F(q_h) = \int_F q_h p_k\, ds,\ \forall p_k \in P^k(F),$$

(VD2) For any $\tau \in \mathcal{T}_h$,

$$\Phi_\tau(q_h) = \int_\tau q_h p_{k-1}\, dx,\ \forall p_{k-1} \in P^{k-1}(\tau).$$

Now, we are ready to prove the divergence-free property for velocity.

Lemma 3.2 (Divergence-free fluid velocity) *For any $\boldsymbol{v}_h \in \boldsymbol{U}^h$, if there exists $\underline{W}_h \in \underline{X}^h$, such that*

$$C_h(\underline{H}_h, \underline{W}_h) = B_h^*(\boldsymbol{v}_h, \underline{H}_h),\ \forall \underline{H}_h \in \underline{W}_{tr}^h, \tag{3.7}$$

then, $\boldsymbol{v}_h \in H(div, \Omega)$ *and* $\nabla \cdot \boldsymbol{v}_h = 0$.

Proof To start with, we observe that for any constant c, Eq. (3.7) still holds if $\underline{H}_h = c\underline{I}$. Thus, (3.7) is equivalent to

$$C_h(\underline{H}_h, \underline{W}_h) = B_h^*(\boldsymbol{v}_h, \underline{H}_h), \ \forall \underline{H}_h \in \underline{W}^h. \tag{3.8}$$

We can discover that for any $q_h \in P^h$, there is $q_h\underline{I} \in \underline{W}^h$, where we use the fact that

$$\left(\left(\underline{I}\boldsymbol{n}\right)^{\boldsymbol{t}}\right)_{|F} = 0, \forall F \in \mathcal{F}.$$

Further calculation can lead to

$$C_h(q_h\underline{I}, \underline{W}_h) = 0, \tag{3.9}$$

$$B_h^*(\boldsymbol{v}_h, q_h\underline{I}) = (\nabla_h \cdot \boldsymbol{v}_h, q_h) - \sum_{F\in\mathcal{F}_{pr}} \int_F q_h[\boldsymbol{v}_h \cdot \boldsymbol{n}]\, ds. \tag{3.10}$$

According to Lemma 3.1, we can define $q_h \in P^h$ satisfying

$$(q_h, p_{k-1})_\tau = (\nabla \cdot \boldsymbol{v}_h, p_{k-1})_\tau, \ \forall p_{k-1} \in P^{k-1}(\tau), \ \forall \tau \in \mathcal{T}_h,$$
$$(q_h, p_k)_F = -([\boldsymbol{v}_h \cdot \boldsymbol{n}], p_k)_F, \ \forall p_k \in P^k(F), \ \forall F \in \mathcal{F}_{pr}.$$

We can obtain that

$$0 = C_h(q_h\underline{I}, \underline{W}_h) = B_h^*(\boldsymbol{v}_h, q_h\underline{I}) = \|\nabla_h \cdot \boldsymbol{v}_h\|_0^2 + \sum_{F\in\mathcal{F}_{pr}} \|[\boldsymbol{v}_h \cdot \boldsymbol{n}]\|_{0,F}^2.$$

which shows that

$$[\boldsymbol{v}_h \cdot \boldsymbol{n}]_{|F} = 0, \ \forall F \in \mathcal{F}_{pr},$$
$$\nabla_h \cdot \boldsymbol{v}_h = 0.$$

Thus, the proof is completed. □

4 Well-Posedness

This section aims to show that the non-linear system (3.3)–(3.5) admits a unique solution $(\underline{G}_h, \underline{L}_h, \boldsymbol{u}_h) \in \underline{W}_{tr}^h \times \underline{X}^h \times \boldsymbol{U}^h$. To achieve this goal, we shall first introduce the kernel space $\underline{\tilde{W}}_{tr}^h \subset \underline{W}_{tr}^h$ associated with the bilinear form B_h as

$$\underline{\tilde{W}}_{tr}^h := \{\underline{H}_1 \in \underline{W}_{tr}^h : B_h(\underline{H}_1, \boldsymbol{v}_h) = 0, \ \forall \boldsymbol{v}_h \in \boldsymbol{U}^h\}.$$

We recall the following theorem (cf. [18, Theorem 3]), which will play an important role in proving the well-posedness of our discrete scheme.

Theorem 4.1 *Assume that*

(i) The operator $A_1 : \underline{X}^h \to \underline{X}^{h'}$ *is Lipschitz continuous and strongly monotone on the first variable, that is, there exist constants* $C_1, \ C_2 > 0$*, such that for any* $\underline{Q}_1, \ \underline{Q}_2 \in \underline{X}^h$

$$\sup_{0\neq\underline{Q}_h\in\underline{X}^h} \frac{A_h(\underline{Q}_1 - \underline{Q}_2, \underline{Q}_h)}{\|\underline{Q}_h\|_0} \le C_1\|\underline{Q}_1 - \underline{Q}_2\|_0, \tag{4.1}$$

and

$$A_h(\underline{Q}_1, \underline{Q}_1 - \underline{Q}_2) - A_h(\underline{Q}_2, \underline{Q}_1 - \underline{Q}_2) \geq C_2 \|\underline{Q}_1 - \underline{Q}_2\|_0^2. \tag{4.2}$$

(ii) The bilinear form B_h satisfies a uniform inf-sup condition on $\boldsymbol{U}^h \times \underline{W}^h_{tr}$, that is, there exists a constant $C_3 > 0$, independent of mesh size h, such that for any $\boldsymbol{v}_h \in \boldsymbol{U}^h$

$$\sup_{0 \neq \underline{H}_h \in \underline{W}^h_{tr}} \frac{B_h(\boldsymbol{v}_h, \underline{H}_h)}{\|\underline{H}_h\|_{\underline{W}^h}} \geq C_3 \|\boldsymbol{v}_h\|_{\boldsymbol{U}^h}. \tag{4.3}$$

(iii) The bilinear operator C_h satisfies a uniform inf-sup condition on $\underline{X}^h \times \tilde{\underline{W}}^h_{tr}$, that is, there exists a constant $C_4 > 0$, independent of mesh size h, such that for any $\underline{W}_h \in \tilde{\underline{W}}^h_{tr}$

$$\sup_{0 \neq \underline{Q}_h \in \underline{X}^h} \frac{C_h(\underline{Q}_h, \underline{W}_h)}{\|\underline{Q}_h\|_0} \geq C_4 \|\underline{W}_h\|_{\underline{W}^h}. \tag{4.4}$$

Then, the system (3.3)–(3.5) admits a unique solution $(\underline{G}_h, \underline{L}_h, \boldsymbol{u}_h) \in (\underline{W}^h_{tr}, \underline{X}^h, \boldsymbol{U}^h)$. Moreover, there exists a constant $C > 0$, depending on $C_1,\ C_2,\ C_3,\ C_4$, such that

$$\|\underline{G}_h\|_{\underline{W}^h} + \|\underline{L}_h\|_0 + \|\boldsymbol{u}_h\|_{\boldsymbol{U}^h} \leq C\|\boldsymbol{f}\|_0.$$

In this section, we will prove conditions (i-iii) sequentially in the following subsections, where the Lipschitz continuous and strongly monotone property of A_1 can be proved easily with the help of Lemma 2.1 introduced in Sect. 2.

4.1 Uniform Inf–Sup Condition of B_h

In this subsection, we shall prove the condition (ii) in Theorem 4.1. Prior to working on the proof, we shall introduce the following lemma from [12, Lemma 2.3] and [31].

Lemma 4.1 *Any function $\boldsymbol{v}_h \in \boldsymbol{U}^h$ is uniquely determined by the following degrees of freedom (WD1)–(WD2):*

(WD1) For any $F \in \mathcal{F}_{dl}$,

$$\Phi_F(\boldsymbol{v}_h) = \int_F \boldsymbol{v_h} \cdot \boldsymbol{n} p_k \, ds, \ \forall p_k \in P^k(F),$$

(WD2) For any $\tau \in \mathcal{T}_h$,

$$\Phi_F(\boldsymbol{v}_h) = \int_\tau \boldsymbol{v_h} \cdot \boldsymbol{p}_{k-1} \, dx, \ \forall \boldsymbol{p}_{k-1} \in P^{k-1}(\tau)^d.$$

With the help of this Lemma, the degrees of freedom of $\underline{W}^h$ can be described in the following Lemma.

Lemma 4.2 *Any function $\underline{H}_h \in \underline{W}^h$ is uniquely determined by the following degrees of freedom (UD1)–(UD3):*
(UD1) For any $F_1 \in \mathcal{F}_{dl}$,

$$\Phi_{F_1}(\underline{H}_h) = \int_{F_1} (\underline{H}_h \boldsymbol{n})^{\boldsymbol{t}} \cdot \boldsymbol{p}_k \, ds, \ \forall \boldsymbol{p}_{k,F_1} \in \boldsymbol{P}^{k,\boldsymbol{T}}(F_1),$$

where $\boldsymbol{P}^{k,\boldsymbol{T}}(F_1) := \{\boldsymbol{p}_k \in P^k(F_1)^d : \boldsymbol{p}_k \cdot \boldsymbol{n} = 0\}$.

(UD2) For any $F_2 \in \mathcal{F}_{pr}$,

$$\Phi_{F_2}(\underline{H}_h) = \int_{F_2} (\underline{H}_h \boldsymbol{n}) \cdot \boldsymbol{p}_k \, ds, \ \forall \boldsymbol{p}_k \in P^k(F_2)^d.$$

(UD3) For any $\tau \in \mathcal{T}_h$,

$$\Phi_\tau(\underline{H}_h) = \int_\tau \underline{H}_h : \underline{p}_{k-1} \, dx, \ \forall \underline{p}_{k-1} \in P^{k-1}(\tau)^{d\times d}.$$

Proof First, we define $P^{k-1}(\tau) = \emptyset$, for all $\tau \in \mathcal{T}_h$ to ensure consistent notation. Additionally, we denote $|UD|$ as the total number of degrees of freedom associated with $(UD1) - -(UD3)$.
For any $F \in \mathcal{F}, \ \tau \in \mathcal{T}_h$, the dimensions of polynomial spaces are given by

$$\dim\left(P^k(F)\right) = \frac{1}{(d-1)!}\prod_{i=1}^{d-1}(k+i),$$
$$\dim\left(P^k(\tau)\right) = \frac{1}{d!}\prod_{i=1}^{d}(k+i).$$

We observe that, for any $\underline{R}_h \in \underline{W}^h$, the condition $[(\underline{R}_h\boldsymbol{n})^{\boldsymbol{t}}]_{|F} = 0, \forall F \in \mathcal{F}_{dl}$ can be equivalently expressed as

$$[(\underline{R}_h\boldsymbol{n})\cdot\boldsymbol{a}]_{|F} = 0, \ \forall \boldsymbol{a} \in \boldsymbol{T}_F, \ \forall F \in \mathcal{F}_{dl},$$

where $\dim(\boldsymbol{T}_F) = d-1$. Thus, the degrees of freedom associated with $(UD1)$ are

$$|UD1| = \sum_{F\in\mathcal{F}_{dl}} (d-1)\dim\left(P^k(F)\right).$$

Therefore, the dimensions of $\underline{W}^h$ and $|UD|$ can be described as follows:

$$\begin{aligned}
\dim(\underline{W}^h) &= d^2\sum_{\tau\in\mathcal{T}_h}\dim\left(P^k(\tau)\right) - (d-1)\sum_{F\in\mathcal{F}_{dl}}\dim(P^k(F)) - d\sum_{F\in\mathcal{F}_{pr}^0}\dim\left(P^k(F)\right)\\
&= d^2|\mathcal{T}_h|\frac{1}{d!}\prod_{i=1}^{d}(k+i) - (d|\mathcal{F}_{pr}^0| + (d-1)|\mathcal{F}_{dl}|)\frac{1}{(d-1)!}\prod_{i=1}^{d-1}(k+i),\\
|UD| &= d^2\sum_{\tau\in\mathcal{T}_h}\dim\left(P^{k-1}(\tau)\right) + (d-1)\sum_{F\in\mathcal{F}_{dl}}\dim(P^k(F)) + d\sum_{F\in\mathcal{F}_{pr}}\dim\left(P^k(F)\right)\\
&= d^2|\mathcal{T}_h|\frac{1}{d!}\prod_{i=1}^{d}(k-1+i) + (d|\mathcal{F}_{pr}| + (d-1)|\mathcal{F}_{dl}|)\frac{1}{(d-1)!}\prod_{i=1}^{d-1}(k+i).
\end{aligned}$$

Subtracting these two equations, we find that

$$|UD| - \dim(\underline{W}^h) = \frac{1}{(d-1)!}\prod_{i=1}^{d-1}(k+i)(d^2|\mathcal{T}_h| - d(|\mathcal{F}_{pr}| + |\mathcal{F}_{pr}^0|) - 2(d-1)|\mathcal{F}_{dl}|).$$

It is important to note that, each face in $\mathcal{F}_{pr}^0 \cup \mathcal{F}_{dl}$ is associated with two elements in $\mathcal{T}_h$, and each face in $\mathcal{F}_{pr}^b$ is associated with one element in $\mathcal{T}_h$, while each element has exactly one face in $\mathcal{F}_{pr}$ and d faces in $\mathcal{F}_{dl}$, $d = 2, 3$. Thus, we can establish the equations

$$|\mathcal{F}_{pr}| + |\mathcal{F}_{pr}^0| = |\mathcal{T}_h|,$$
$$2|\mathcal{F}_{dl}| = d|\mathcal{T}_h|.$$

Thanks to this observation, we successfully prove that $\dim(\underline{W}^h) = |UD|$. In the following, it suffices to prove uniqueness. Let $\underline{H}_h \in \underline{W}^h$ be defined by

$$\Phi_{F_1}(\underline{H}_h) = 0, \forall F_1 \in \mathcal{F}_{dl}, \tag{4.5}$$
$$\Phi_{F_2}(\underline{H}_h) = 0, \forall F_2 \in \mathcal{F}_{pr}, \tag{4.6}$$
$$\Phi_{\tau}(\underline{H}_h) = 0, \forall \tau \in \mathcal{T}_h. \tag{4.7}$$

In the following, we will demonstrate that $\underline{H}_h = \underline{0}$.
Since $\left((\underline{H}_h \boldsymbol{n})^{\boldsymbol{t}}\right)_{|F_1} \in \boldsymbol{P}^{k,\boldsymbol{T}}(F_1)$, $\forall F_1 \in \mathcal{F}_{dl}$ and $(\underline{H}_h \boldsymbol{n})_{|F_2} \in P^k(F_2)^d$, $\forall F_2 \in \mathcal{F}_{pr}$, conditions (4.5) and (4.6) imply that

$$\left((\underline{H}_h \boldsymbol{n})^{\boldsymbol{t}}\right)_{|F_1} = \boldsymbol{0},\ \forall F_1 \in \mathcal{F}_{dl}, \tag{4.8}$$
$$(\underline{H}_h \boldsymbol{n})_{|F_2} = \boldsymbol{0},\ \forall F_2 \in \mathcal{F}_{pr}. \tag{4.9}$$

For each $\tau \in \mathcal{T}_h$, let $F^{(i)}, i = 1, \cdots, d+1$ be the faces of τ, where $F^{(1)} \in \mathcal{F}_{pr}$ and $F^{(i)} \in \mathcal{F}_{dl}, i = 2 \cdots, d+1$. We denote $\boldsymbol{n}^{(i)}$ as the unit outward normal vector of $F^{(i)}, i = 1 \cdots, d+1$. It can be proven that $\boldsymbol{n}^{(i)},\ i = 2 \cdots, d+1$ are d linearly independent vectors. According to Lemma 4.1, we can find $\boldsymbol{v}_\tau \in \boldsymbol{U}^h$, such that

$$(\boldsymbol{v}_\tau, \boldsymbol{p}_{k-1})_\tau = -(\nabla \cdot \underline{H}_h, \boldsymbol{p}_{k-1})_\tau,\ \forall \boldsymbol{p}_{k-1} \in \boldsymbol{P}^{k-1}(\tau)^d,$$
$$(\boldsymbol{v}_\tau \cdot \boldsymbol{n}^{(i)}, p_k)_{F^{(i)}} = \left((\underline{H}_h)_{|\tau} \boldsymbol{n}^{(i)} \cdot \boldsymbol{n}^{(i)}, p_k\right)_{F^{(i)}},\ \forall p_k \in P^k(F^{(i)}),\ i = 2, \cdots, d+1.$$

These special choices of $\underline{H}_h$ and $\boldsymbol{v}_\tau$ reveal that

$$\begin{aligned} 0 &= (\underline{H}_h, \nabla_h \boldsymbol{v}_\tau)_\tau - \left(\left((\underline{H}_h)_{|\tau} \boldsymbol{n}^{(1)}\right)^{\boldsymbol{t}}, (\boldsymbol{v}_\tau)^{\boldsymbol{t}}\right)_{F^{(1)}} - \sum_{i=2}^{d+1} \left((\underline{H}_h)_{|\tau} \boldsymbol{n}^{(i)}, \boldsymbol{v}_\tau\right)_{F^{(i)}} \\ &= -(\nabla_h \cdot \underline{H}_h, \boldsymbol{v}_\tau)_\tau + \sum_{i=2}^{d+1} \left((\underline{H}_h)_{|\tau} \boldsymbol{n}^{(i)} \cdot \boldsymbol{n}^{(i)}, (\boldsymbol{v}_\tau \cdot \boldsymbol{n}^{(i)})\right)_{F^{(i)}} \\ &= \|\nabla_h \cdot \underline{H}\|_\tau^2 + \sum_{i=2}^{d+1} \|(\underline{H}_h)_{|\tau} \boldsymbol{n}^{(i)} \cdot \boldsymbol{n}^{(i)}\|_{F^{(i)}}^2. \end{aligned} \tag{4.10}$$

Combining (4.8) and (4.10), we can infer that

$$\left((\underline{H}_h)_{|\tau} \boldsymbol{n}^{(i)}\right)_{|F^{(i)}} = \boldsymbol{0},\ i = 2, \cdots, d+1. \tag{4.11}$$

The analytical approach diverges subtly when comparing the cases of $k = 0$ and $k > 0$.
Case 1: $k = 0$
In this case, every entry of $(\underline{H}_h)_{|\tau}$ is constant. Thus, we can infer from (4.11) that

$$(\underline{H}_h)_{|\tau} \boldsymbol{n}^{(i)} = \boldsymbol{0},\ i = 2, \cdots, d+1. \tag{4.12}$$

The linear independence of $\boldsymbol{n}^{(i)},\ i=2,\cdots,d+1$ implies that (4.12) constitutes a well-posed d^2-dimensional system of linear equations. Thus, we can infer that

$$(\underline{H}_h)_{|\tau} = \underline{0}.$$

Applying this procedure to all $\tau \in \mathcal{T}_h$, we can conclude

$$\underline{H}_h = \underline{0}.$$

Case 2: $k \geq 1$

This case is more complex than **Case 1**. Here, we define v_i as the vertex of τ that lies outside $F^{(i)}$. Let λ_i be the affine function whose value equals $\delta_{i,j}$ at $v_j,\ i,j=1,2,3,4$, where

$$\delta_{i,j} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$$

Thus, there exist $\boldsymbol{q}_\tau^{(i)} \in P^{k-1}(\tau)^d,\ i=2,\cdots,d+1$, such that

$$(\underline{H}_h)_{|\tau}\boldsymbol{n}^{(i)} = \lambda_i \boldsymbol{q}_\tau^{(i)},\ i=2,\cdots,d+1. \tag{4.13}$$

By defining

$$\underline{\Lambda}_\tau = \begin{bmatrix} \lambda_2 \underline{I}_d & & \\ & \ddots & \\ & & \lambda_{d+1}\underline{I}_d \end{bmatrix} \in P^1(\tau)^{d^2\times d^2},$$

where $\underline{I}_d$ is the $d\times d$ identity matrix, a $d^2\times d^2$ matrix $\underline{A}_\tau$ can be defined from (4.13), such that

$$\underline{A}_\tau (\underline{H}_h)_{|\tau} = \underline{\Lambda} \begin{bmatrix} \boldsymbol{q}_\tau^{(2)} \\ \vdots \\ \boldsymbol{q}_\tau^{(d+1)} \end{bmatrix} := \underline{\Lambda}\boldsymbol{q}_\tau \tag{4.14}$$

The linear independence of $\boldsymbol{n}^{(i)},\ i=2,\cdots,d+1$, ensures that $\underline{A}_\tau$ is invertible. Consequently, we can define $\boldsymbol{p}_{\tau,k-1} = \underline{A}_\tau^{-1}\boldsymbol{q}_\tau \in P^{k-1}(\tau)^{d^2}$. Substituting this into (4.7), we derive

$$\begin{aligned} 0 &= \left((\boldsymbol{H}_h)_{|\tau}, \boldsymbol{p}_{\tau,k-1}\right)_\tau = \int_\tau (\underline{\Lambda}_\tau \boldsymbol{q}_\tau)^T \underline{A}_\tau^{-T}\underline{A}_\tau^{-1}\boldsymbol{q}_\tau dx \\ &\geq C\int_\tau \boldsymbol{q}_\tau^T \underline{A}_\tau^{-T}\underline{A}_\tau^{-1}\boldsymbol{q}_\tau dx = C\int_\tau |\boldsymbol{p}_{\tau,k-1}|^2 dx \geq 0, \end{aligned}$$

leading to the conclusion that $(\underline{H}_h)_{|\tau} = \underline{0}$. Applying this procedure to all $\tau \in \mathcal{T}_h$, we can conclude that

$$\underline{H}_h = \underline{0}.$$

□

Remark When $d=2$, for any $F \in \mathcal{T}_F$, there is only one unit basis function $\boldsymbol{t}$ in $\boldsymbol{T}_F$, which makes the definition of $(UD1)--(UD2)$ consistent with the definition of $(WD1)$ given in [31].

Now, we are ready to prove the validity of condition (ii) in Theorem 4.1.

Theorem 4.2 *The bilinear form B_h satisfies a uniform inf-sup condition on $\boldsymbol{U}^h \times \underline{W}_{tr}^h$, that is, there exists a constant $C > 0$, independent of mesh size h, such that*

$$\inf_{\mathbf{0}\neq \boldsymbol{v}\in \boldsymbol{U}^h} \sup_{0\neq \underline{H}\in \underline{W}_{tr}^h} B_h(\boldsymbol{v}, \underline{H}) \geq C\|\boldsymbol{v}\|_{\boldsymbol{U}^h}\|\underline{H}\|_{\underline{W}^h}.$$

Proof It suffices to prove that for any $\boldsymbol{v} \in \boldsymbol{U}^h$, there exists $\underline{H} \in \underline{W}_{tr}^h$, such that

$$B_h(\underline{H}, \boldsymbol{v}) = \|\boldsymbol{v}\|_{\boldsymbol{U}^h}^2 \quad \text{and} \quad \|\underline{H}\|_{\underline{W}^h} \leq C\|\boldsymbol{v}\|_{\boldsymbol{U}^h}. \tag{4.15}$$

In view of Lemma 4.2, for any $\boldsymbol{v} \in \boldsymbol{U}^h$, we can define $\underline{\tilde{H}} \in \underline{W}^h$ in the following sense:

$$\begin{aligned}
(\underline{\tilde{H}}, \underline{p}_{k-1})_\tau &= (\nabla \boldsymbol{v}, \underline{p}_{k-1})_\tau, \ \forall \underline{p}_{k-1} \in P^{k-1}(\tau)^{d\times d}, \ \forall \tau \in \mathcal{T}_h,\\
(\underline{\tilde{H}}\boldsymbol{n}, \boldsymbol{p}_k)_F &= h_F^{-1}([\boldsymbol{v}]_{|F}, \boldsymbol{p}_k)_F, \ \forall \boldsymbol{p}_k \in P^k(F)^d, \ \forall F \in \mathcal{F}_{pr},\\
((\underline{\tilde{H}}\boldsymbol{n})^{\boldsymbol{t}}, \boldsymbol{p}_k)_F &= h_F^{-1}([(\boldsymbol{v})^{\boldsymbol{t}}]_{|F}, \boldsymbol{p}_k)_F, \ \forall \boldsymbol{p}_k \in \boldsymbol{P}^{k,\boldsymbol{T}}(F), \ \forall F \in \mathcal{F}_{dl}.
\end{aligned}$$

This special choice implies

$$\begin{aligned}
B_h(\underline{\tilde{H}}, \boldsymbol{v}) &= \|\boldsymbol{v}\|_{\boldsymbol{U}^h}^2,\\
\|\underline{\tilde{H}}\|_{\underline{W}^h} &\leq C\|\boldsymbol{v}\|_{\boldsymbol{U}^h}.
\end{aligned}$$

We define a function $\underline{H} \in \underline{W}_{tr}^h$ as $\underline{H} = \underline{\tilde{H}} - \frac{1}{d|\Omega|}\left(\int_\Omega \mathrm{tr}\left(\underline{\tilde{H}}\right)\, dx\right)\underline{I}$. We will show that $\underline{H}$ fulfills the condition (4.15). First, the triangle inequality yields

$$\begin{aligned}
\|\underline{H}\|_{\underline{W}^h} &= \|\underline{\tilde{H}} - \frac{1}{d|\Omega|}\left(\int_\Omega \mathrm{tr}\left(\underline{\tilde{H}}\right)\, dx\right)\underline{I}\|_{\underline{W}^h}\\
&\leq \|\underline{\tilde{H}}\|_{\underline{W}^h} + \frac{1}{d|\Omega|}\int_\Omega \|\mathrm{tr}(\underline{\tilde{H}})\underline{I}\|_{\underline{W}^h}\, dx \leq C\|\underline{\tilde{H}}\|_{\underline{W}^h} \leq C\|\boldsymbol{v}\|_{\boldsymbol{U}^h}.
\end{aligned}$$

Second, by the linear property of jump, we can observe that, for any $F \in \mathcal{F}_{dl}$

$$\begin{aligned}
&\int_F (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}\cdot[\underline{H}\boldsymbol{n}]_{|F}\, ds\\
=&\int_F (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}\cdot\left[\left(\underline{\tilde{H}} - \frac{1}{d|\Omega|}\left(\int_\Omega \mathrm{tr}\left(\underline{\tilde{H}}\right)\, dx\right)\underline{I}\right)\boldsymbol{n}\right]_{|F}\, ds\\
=&\int_F (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}\cdot\left(\left[\underline{\tilde{H}}\boldsymbol{n}\right]_{|F} - \left[\frac{1}{d|\Omega|}\left(\int_\Omega \mathrm{tr}\left(\underline{\tilde{H}}\right)\, dx\right)\underline{I}\boldsymbol{n}\right]_{|F}\right)\, ds\\
=&\int_F (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}\cdot[\underline{\tilde{H}}\boldsymbol{n}]_{|F}\, ds.
\end{aligned} \tag{4.16}$$

Moreover, we notice that $\nabla_h \cdot \underline{H} = \nabla_h \cdot \underline{\tilde{H}}$. Combining with (4.16), we can obtain that

$$\begin{aligned}
B_h(\boldsymbol{v}, \underline{H}) &= B_h^*(\underline{H}, \boldsymbol{v})\\
&= -(\boldsymbol{v}, \nabla_h\cdot\underline{H}) + \sum_{F\in\mathcal{F}_{dl}}\int_F (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}\cdot[\underline{H}\boldsymbol{n}]_{|F}\, ds\\
&= -(\boldsymbol{v}, \nabla_h\cdot\underline{\tilde{H}}) + \sum_{F\in\mathcal{F}_{dl}}\int_F (\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}\cdot[\underline{\tilde{H}}\boldsymbol{n}]_{|F}\, ds\\
&= B_h^*(\underline{\tilde{H}}, \boldsymbol{v}) = B_h(\boldsymbol{v}, \underline{\tilde{H}}) = \|\boldsymbol{v}\|_{\boldsymbol{U}^h}^2.
\end{aligned}$$

Therefore, the proof is completed. □

4.2 Uniform Inf–Sup Condition of C_h

In this subsection, we will consider condition (iii) in Theorem 4.1. In addition, we can describe $\underline{\tilde{W}}^h_{tr}$ in the following lemma.

Lemma 4.3 *$\underline{\tilde{W}}^h_{tr}$ is a subset of $\underline{\tilde{W}}_{tr}$.*

Proof For any $\underline{H} \in \underline{\tilde{W}}^h_{tr}$, according to Lemma 4.1, we can take $\boldsymbol{w}_h \in \boldsymbol{U}^h$ satisfying

$$\int_F \boldsymbol{w}_h \cdot \boldsymbol{n} p_k \, ds = \int_F [\underline{H}\boldsymbol{n}] \cdot \boldsymbol{n} p_k \, ds, \forall p_k \in P^k(F), \forall F \in \mathcal{F}_{dl},$$
$$\int_\tau \boldsymbol{w}_h \cdot \boldsymbol{p}_{k-1} \, ds = -\int_\tau \nabla \cdot \underline{H} \cdot \boldsymbol{p}_{k-1} \, ds, \forall \boldsymbol{p}_{k-1} \in P^{k-1}(\tau)^d, \forall \tau \in \mathcal{T}_h.$$

Then, simple calculation implies that

$$0 = B_h(\underline{H}, \boldsymbol{w}_h) = B_h^*(\boldsymbol{w}_h, \underline{H}) = \sum_{\tau \in \mathcal{T}_h} \|\nabla \cdot \underline{H}\|^2_{\tau,0} + \sum_{F \in \mathcal{F}_{dl}} \|[\underline{H}\boldsymbol{n}] \cdot \boldsymbol{n}\|^2_{F,0}.$$

Thus, we can infer that $[\underline{H}\boldsymbol{n} \cdot \boldsymbol{n}]_{|F} = 0$, for all $F \in \mathcal{F}_{dl}$. Combining with the definition of $\underline{W}^h_{tr}$, we can induce that $\underline{H}\boldsymbol{n}$ is continuous over all $F \in \mathcal{F}_{dl} \cup \mathcal{F}_{pr}$, leading to the fact that $\underline{H} \in H(\text{div}, \Omega)^d$ with $\nabla \cdot \underline{H} = \boldsymbol{0}$. Such divergence-free condition demonstrates that $\underline{\tilde{W}}^h_{tr} \subset \underline{\tilde{W}}_{tr}$. □

Thanks to Lemma 4.3, we can conlcude that $\|\cdot\|_E$ is a norm defined on $\underline{\tilde{W}}^h_{tr}$. Furthermore, there exists a constant $C_4 > 0$, independent of mesh size h, such that

$$C_4\|\underline{H}\|_0 \le \|\underline{H}\|_E \le 2\|\underline{H}\|_0, \ \forall \underline{H} \in \underline{\tilde{W}}^h_{tr}. \tag{4.17}$$

Now we are ready to verify the condition (iii) in Theorem 4.1.

Theorem 4.3 *The bilinear form C_h satisfies a uniform inf-sup condition on $\underline{X}^h \times \underline{\tilde{W}}^h_{tr}$, that is, there exists a constant $C > 0$, independent of mesh size h, such that for any $\underline{Q} \in \underline{X}^h$, there holds*

$$\inf_{0 \ne \underline{H} \in \underline{\tilde{W}}^h_{tr}} \sup_{0 \ne \underline{Q} \in \underline{X}^h} \frac{C_h(\underline{Q}, \underline{H})}{\|\underline{Q}\|_0 \|\underline{H}\|_{\underline{W}^h}} \ge C > 0.$$

Proof For given $\underline{H} \in \underline{\tilde{W}}^h_{tr}$, it holds $\text{tr}(\mathcal{A}\underline{H}) = 0$, we can conclude $\mathcal{A}\underline{H} \in \underline{X}^h$. In the following proof, we will take $\underline{Q} = \mathcal{A}\underline{H}$. With inequality (4.17) and the fact that $(\mathcal{A}\underline{Q}, \underline{H}) = (\underline{Q}, \mathcal{A}\underline{H})$, it can be observed that

$$C_h(\underline{Q}, \underline{H}) = (\mathcal{A}\underline{H}, \mathcal{A}\underline{H}) = \|\underline{H}\|^2_E \ge C_4\|\underline{H}\|_0\|\mathcal{A}\underline{H}\|_0 = C_4\|\underline{H}\|_0\|\underline{Q}\|_0. \tag{4.18}$$

Combining (3.1) and (4.18), we can obtain

$$C_h(\underline{Q}, \underline{H}) \ge C_4\|\underline{H}\|_{\underline{W}^h}\|\underline{Q}\|_0.$$

□

Combining the above arguments, we have proved that our scheme satisfies the conditions (i)–(iii) given in Theorem 4.1, which implies the well-posedness.

5 Convergence Estimates

In this section, we will prove the convergence estimates of the scheme (3.3)–(3.5). Before we start our analysis, we would like to introduce some projections associated with the bilinear forms defined in (2.10) and (3.2).
Let $\Pi_1 : \underline{X} \to \underline{X}^h$ be the L_2 projection from $\underline{X}$ to $\underline{X}^h$

$$(\Pi_1 \underline{W} - \underline{W}, \underline{Q}_h)_\tau = 0,\ \forall \underline{Q}_h \in \underline{X}^h,\ \forall \tau \in \mathcal{T}_h. \tag{5.1}$$

Since $\underline{X}^h$ is a closed convex Hilbert space, this projection is well-defined and unique, which shows that this is a polynomial preserving operator (cf. [14]).
Combining (5.1) and the fact that for any $\underline{H}_h \in \underline{W}_{tr}^h$, there is $\mathcal{A}\underline{H}_h \in \underline{X}^h$, we notice that for any $\underline{W} \in \underline{X}$

$$(\underline{W}, \mathcal{A}\underline{H}_h)_\tau = (\Pi_1 \underline{W}, \mathcal{A}\underline{H}_h)_\tau,\ \forall \underline{H}_h \in \underline{W}_{tr}^h,\ \forall \tau \in \mathcal{T}_h. \tag{5.2}$$

Second, Lemma 4.2 implies that we can define an operator $\Pi_2 : H^1(\Omega)^{d\times d} \to \underline{W}^h$ satisfying

$$\begin{aligned}
(\Pi_2 \underline{H} - \underline{H}, \underline{p}_{k-1})_\tau &= 0,\ \forall \underline{p}_{k-1} \in P^{k-1}(\tau)^{d\times d},\ \forall \tau \in \mathcal{T}_h,\\
((\Pi_2 \underline{H} - \underline{H})\boldsymbol{n}, \boldsymbol{p}_k)_F &= 0,\ \forall \boldsymbol{p}_k \in P^k(F)^d,\ \forall F \in \mathcal{F}_{pr}^0,\\
(((\Pi_2 \underline{H} - \underline{H})\boldsymbol{n})^{\boldsymbol{t}}, \boldsymbol{p}_k)_F &= 0,\ \forall \boldsymbol{p}_k \in \boldsymbol{P}^{k,\boldsymbol{T}}(F),\ \forall F \in \mathcal{F}_{dl}.
\end{aligned} \tag{5.3}$$

Furthermore, if $\underline{H} \in H^1(\Omega)^{d\times d} \cap \underline{W}_{tr}$, we can infer that $\Pi_2 \underline{H} \in \underline{W}_{tr}^h$.
In view of Lemma 4.1, we can define a projection $I_h : H^1(\Omega)^d \to \boldsymbol{U}^h$ satisfying the following properties for any $\boldsymbol{v} \in H^1(\Omega)^2$

$$\begin{aligned}
((I_h \boldsymbol{v} - \boldsymbol{v}) \cdot \boldsymbol{n}, p_k)_F &= 0,\ \forall p_k \in P^k(F),\ \forall F \in \mathcal{F}_{dl},\\
((I_h \boldsymbol{v} - \boldsymbol{v}), \boldsymbol{p}_{k-1})_\tau &= 0,\ \forall \boldsymbol{p}_{k-1} \in P^{k-1}(\tau)^d,\ \forall \tau \in \mathcal{T}_h.
\end{aligned} \tag{5.4}$$

It is easy to see that Π_2 and I_h are well-defined polynomial preserving operators. The definitions (5.2), (5.3), and (5.4) imply that for any $\underline{H}_1 \in H^1(\Omega)^{d\times d}$, $\underline{Q}_1 \in L^2(\Omega)^{d\times d}$, $\boldsymbol{v}_1 \in H_0^1(\Omega)^d$, we have

$$B_h(\underline{H}_1 - \Pi_2 \underline{H}_1, \boldsymbol{v}_h) = 0,\ \forall \boldsymbol{v}_h \in \boldsymbol{U}^h, \tag{5.5}$$

$$C_h(\underline{H}_h, \underline{Q}_1 - \Pi_1 \underline{Q}_1) = 0,\ \forall \underline{H}_h \in \underline{W}_{tr}^h, \tag{5.6}$$

$$B_h^*(\boldsymbol{v}_1 - I_h \boldsymbol{v}_1, \underline{H}_h) = 0,\ \forall \underline{H}_h \in \underline{W}_{tr}^h. \tag{5.7}$$

By standard theory for polynomial preserving operators and the trace inequalities (cf. [12, 13] [31, Lemma 3.1]), the following error estimates hold for any $(\underline{H}_1, \underline{Q}_1, \boldsymbol{v}_1) \in H^{k+1}(\Omega)^{d\times d} \times H^{k+1}(\Omega)^{d\times d} \times H^{k+1}(\Omega)^d$

$$\begin{aligned}
\|\underline{Q}_1 - \Pi_1 \underline{Q}_1\|_0 &\le Ch^{k+1} |\underline{Q}_1|_{k+1},\\
\|\underline{H}_1 - \Pi_2 \underline{H}_1\|_{\underline{W}^h} &\le Ch^{k+1} |\underline{H}_1|_{k+1},\\
\|\boldsymbol{v}_1 - I_h \boldsymbol{v}_1\|_{\boldsymbol{U}^h} &\le Ch^k |\boldsymbol{v}_1|_{k+1}.
\end{aligned} \tag{5.8}$$

Now we are ready to prove the convergence estimate in the following theorem.

Theorem 5.1 *Let $(\underline{L}, \underline{G}, \boldsymbol{u}) \in \underline{W}_{tr} \times \underline{X} \times \boldsymbol{U}$ be the solution of the continuous variational form (2.11) and let $(\underline{L}_h, \underline{G}_h, \boldsymbol{u}_h) \in \underline{X}^h \times \underline{W}_{tr}^h \times \boldsymbol{U}^h$ be the discrete solution of (3.3)–(3.5).*

Assume that $(\underline{L}, \underline{G}, \boldsymbol{u}) \in H^{k+1}(\Omega)^{d\times d} \times H^{k+1}(\Omega)^{d\times d} \times H^{k+1}(\Omega)^2$, *then, there exists a positive constant independent of mesh size h such that*

$$
\begin{aligned}
\|\underline{L} - \underline{L}_h\|_0 &\le Ch^{k+1}|\underline{L}|_{k+1},\\
\|\underline{G} - \underline{G}_h\|_{\underline{W}^h} &\le Ch^{k+1}(|\underline{G}|_{k+1} + |\underline{L}|_{k+1}),\\
\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\boldsymbol{U}^h} &\le Ch^{k}(|\boldsymbol{u}|_{k+1} + |\underline{L}|_{k+1}).
\end{aligned}
$$

In addition, the superconvergence holds in the following sense:

$$\|I_h\boldsymbol{u} - \boldsymbol{u}_h\|_{\boldsymbol{U}^h} \le Ch^{k+1}|\underline{L}|_{k+1}.$$

Proof Performing integration by parts, we can get the following error equations:

$$A_h(\underline{L}, \underline{Q}_h) - A_h(\underline{L}_h, \underline{Q}_h) - C_h(\underline{G} - \underline{G}_h, \underline{Q}_h) = 0,\ \forall \underline{Q}_h \in \underline{X}^h, \tag{5.9}$$

$$B_h^*(\boldsymbol{u} - \boldsymbol{u}_h, \underline{H}_h) - C_h(\underline{H}_h, \underline{L} - \underline{L}_h) = 0,\ \forall \underline{H}_h \in \underline{W}_{tr}^h, \tag{5.10}$$

$$B_h(\underline{G} - \underline{G}_h, \boldsymbol{v}_h) = 0,\ \forall \boldsymbol{v}_h \in \boldsymbol{U}^h. \tag{5.11}$$

It follows from (5.6) and (5.11)

$$B_h(\Pi_2\underline{G} - \underline{G}_h, \boldsymbol{v}_h) = 0,\ \forall \boldsymbol{v}_h \in \boldsymbol{U}^h,$$

which shows that $\Pi_2\underline{G} - \underline{G}_h \in \tilde{\underline{W}}_{tr}^h$. Taking $\underline{Q}_h = \Pi_1\underline{L} - \underline{L}_h,\ \underline{H}_h = \Pi_2\underline{G} - \underline{G}_h,\ \boldsymbol{v}_h = I_h\boldsymbol{u} - \boldsymbol{u}_h$ into (5.9)–(5.11), we can obtain

$$
\begin{aligned}
A_h(\underline{L}, \Pi_1\underline{L} - \underline{L}_h) - A_h(\underline{L}_h, \Pi_1\underline{L} - \underline{L}_h) - C_h(\underline{G} - \underline{G}_h, \Pi_1\underline{L} - \underline{L}_h) &= 0,\\
B_h^*(\boldsymbol{u} - \boldsymbol{u}_h, \Pi_2\underline{G} - \underline{G}_h) - C_h(\Pi_2\underline{G} - \underline{G}_h, \underline{L} - \underline{L}_h) &= 0,\\
B_h(\underline{G} - \underline{G}_h, I_h\boldsymbol{u} - \boldsymbol{u}_h) &= 0.
\end{aligned}
$$

Summing these equations together and applying the adjoint properties of bilinear forms yields to

$$A_h(\underline{L}, \Pi_1\underline{L} - \underline{L}_h) - A_h(\underline{L}_h, \Pi_1\underline{L} - \underline{L}_h) = 0.$$

Due to the linear property of A_h in the second variable, we can infer that

$$A_h(\underline{L}, \underline{L} - \underline{L}_h) - A_h(\underline{L}_h, \underline{L} - \underline{L}_h) = A_h(\underline{L}, \underline{L} - \Pi_1\underline{L}) - A_h(\underline{L}_h, \underline{L} - \Pi_1\underline{L}).$$

Hence, the Lipschitz continuous property (4.1) and the strongly monotone property (4.2) of A_1 lead to

$$\|\underline{L} - \underline{L}_h\|_0 \le C\|\underline{L} - \Pi_1\underline{L}\|_0 \le Ch^{k+1}|\underline{L}|_{k+1}.$$

On the other hand, with the inf-sup condition (4.4) of C_h, we deduce that

$$
\begin{aligned}
\|\Pi_2\underline{G} - \underline{G}_h\|_{\underline{W}^h} &\le C \sup_{0\ne\underline{Q}_h\in\underline{X}^h} \frac{C_h(\Pi_2\underline{G} - \underline{G}_h, \underline{Q}_h)}{\|\underline{Q}_h\|_0}\\
&= C \sup_{0\ne\underline{Q}_h\in\underline{X}^h} \frac{A_h(\underline{L}, \underline{Q}_h) - A_h(\underline{L}_h, \underline{Q}_h)}{\|\underline{Q}_h\|_0} \le C\|\underline{L} - \underline{L}_h\|_0.
\end{aligned}
$$

The inf-sup condition (4.3) of B_h and (5.10) lead to

$$\begin{aligned}\|I_h\boldsymbol{u}-\boldsymbol{u}_h\|_{\boldsymbol{U}^h} &\le C\sup_{0\neq\underline{H}_h\in\underline{W}^h_{tr}}\frac{B_h^*(I_h\boldsymbol{u}-\boldsymbol{u}_h,\underline{H}_h)}{\|\underline{H}_h\|_{\underline{W}^h}}\\ &= C\sup_{0\neq\underline{H}_h\in\underline{W}^h_{tr}}\frac{C_h(\underline{H}_h,\underline{L}-\underline{L}_h)}{\|\underline{H}_h\|_{\underline{W}^h}}\\ &= C\sup_{0\neq\underline{H}_h\in\underline{W}^h_{tr}}\frac{(\underline{L}-\underline{L}_h,\mathcal{A}\underline{H}_h)}{\|\underline{H}_h\|_{\underline{W}^h}}\\ &\le C\sup_{0\neq\underline{H}_h\in\underline{W}^h_{tr}}\frac{\|\underline{L}-\underline{L}_h\|_0\|\mathcal{A}\underline{H}_h\|_0}{\|\underline{H}_h\|_{\underline{W}^h}}\\ &\le C\|\underline{L}-\underline{L}_h\|_0,\end{aligned}$$

where in the last inequality we use the fact that $\|\mathcal{A}\underline{H}\|_0 \le 2\|\underline{H}\|_{\underline{W}^h}$ from (2.16). Combining the above estimates, the proof is completed. □

6 Implementation with the Hybridization

This section is split into two parts. The first part introduces how we use the iterative method to solve the nonlinear problem, and the second part introduces the hybridization of the proposed scheme and shows that the resulting global system only involves the trace variables, rendering the method computationally efficient.

6.1 Iterative Method for Nonlinear System

First, we shall introduce a form N_h associated with the non-linear operator A_h defined in Sect. 3

$$N_h(\underline{Q}_1,\underline{Q}_2,\underline{Q}_3) := (\mu(\underline{Q}_1)\underline{Q}_2,\underline{Q}_3),\ \forall(\underline{Q}_1,\underline{Q}_2,\underline{Q}_3)\in L^2(\Omega)^{d\times d}\times L^2(\Omega)^{d\times d}\times L^2(\Omega)^{d\times d},$$

where there is $A_h(\underline{Q}_1,\underline{Q}_2) = N_h(\underline{Q}_1,\underline{Q}_1,\underline{Q}_2)$ when $\underline{Q}_1,\ \underline{Q}_2\in\underline{X}^h$.
In order to solve the nonlinear system, we adopt the iterative method as follows: For given data $\underline{L}_h^n\in\underline{X}^h$, $\boldsymbol{f}\in L^2(\Omega)^d$, find $(\underline{L}_h^{n+1},\underline{G}_h^{n+1},\boldsymbol{u}_h^{n+1})\in\underline{X}^h\times\underline{W}^h_{tr}\times\boldsymbol{U}^h$ such that

$$N_h(\underline{L}_h^n,\underline{L}_h^{n+1},\underline{Q}_h) - C_h(\underline{G}_h^{n+1},\underline{Q}_h) = 0, \tag{6.1}$$

$$B_h^*(\boldsymbol{u}_h^{n+1},\underline{H}_h) - C_h(\underline{H}_h,\underline{L}_h^{n+1}) = 0, \tag{6.2}$$

$$B_h(\underline{G}_h^{n+1},\boldsymbol{v}_h) = (\boldsymbol{f},\boldsymbol{v}_h) \tag{6.3}$$

for all $(\underline{Q}_h,\underline{H}_h,\boldsymbol{v}_h)\in\underline{X}^h\times\underline{W}^h_{tr}\times\boldsymbol{U}^h$. Here the bilinear forms B_h, B_h^*, C_h, C_h^* are defined in Sect. 3 and $\underline{L}_h^n$ is obtained from the previous iteration. For each step, (6.1)–(6.3) is a linear system. Concerning the hypothesis on μ, the well-posedness of scheme (6.1)–(6.3) and the boundness of sequence $\{\underline{L}_h^n\}_{n=1}^\infty$ in $L^2(\Omega)^{d\times d}$ can be easily obtained. According to Theorem 4.1 and the fact that $\underline{X}^h$ is finite dimensional, $\{\underline{L}_h^n\}_{n=1}^\infty$ will converge to the solution of (3.3)–(3.5).

6.2 Implement with Hybridization

In this subsection, we introduce the hybridization of the proposed scheme and show that the resulting global system only involves the trace variables, rendering the method computationally efficient.

To start with, we define the piecewise polynomial finite element spaces as follows:

$$\begin{aligned}\underline{\Sigma} &:= \{\underline{w} : \underline{w}_{|\tau} \in P^k(\tau)^{d\times d}, \tau \in \mathcal{T}_h; \mathrm{tr}(\underline{\sigma}) = 0\},\\ \widetilde{\boldsymbol{U}} &:= \{\boldsymbol{\omega} : \boldsymbol{\omega}_{|\tau} \in P^k(\tau)^d, \tau \in \mathcal{T}_h\}.\end{aligned}$$

Then, in order to enforce the staggered continuity, piecewise polynomial spaces on dual faces and primal faces are introduced, respectively, in the following sense:

$$\begin{aligned}\boldsymbol{\Lambda}_{pr} &= \{\boldsymbol{\mu} : \boldsymbol{\mu}_{|F} \in P^k(F)^d, F \in \mathcal{F}^0_{pr}\},\\ \Lambda_{dl,1} &= \{\mu : \mu_{|F} \in P^k(F), F \in \mathcal{F}_{dl}\},\\ \boldsymbol{\Lambda}_{dl,2} &= \{\boldsymbol{\mu} : \boldsymbol{\mu}_{|F} \in \boldsymbol{P}^{k,\boldsymbol{T}}(F), F \in \mathcal{F}_{dl}\}.\end{aligned}$$

In addition, we can introduce the following piecewise polynomial spaces:

$$\begin{aligned}\widehat{W} &:= \{\underline{w} : \underline{w}_{|\tau} \in P^k(\tau)^{d\times d},\ (\mathrm{tr}(\underline{w}), 1)_\tau = 0\ \forall \tau \in \mathcal{T}_h\},\\ \overline{W} &:= \{w : w_{|\tau} \in P_0(\tau),\ \forall \tau \in \mathcal{T}_h; \int_\Omega w\, dx = 0\},\\ \widetilde{W} &:= \{\underline{w} : \underline{w} = \underline{w_1} + w_2\underline{I},\ \underline{w_1} \in \widehat{W},\ w_2 \in \overline{W}\}.\end{aligned}$$

Also, we introduce the following bilinear forms:

$$\begin{aligned}D_h(\boldsymbol{\sigma}, \omega) &= \sum_{F\in\mathcal{F}_{dl}} ([\boldsymbol{\sigma}\cdot\boldsymbol{n}]_{|F}, \omega)_F,\ \forall(\boldsymbol{\sigma}, \omega) \in \widetilde{\boldsymbol{U}} \times \Lambda_{dl,1},\\ E_h(\underline{\sigma}, \omega) &= \sum_{F\in\mathcal{F}_{dl}} ([(\underline{\sigma}\boldsymbol{n})^{\boldsymbol{t}}]_{|F}, \boldsymbol{\omega})_F,\ \forall(\underline{\sigma}, \boldsymbol{\omega}) \in \widehat{W} \times \boldsymbol{\Lambda}_{dl,2},\\ F_h(\underline{\sigma}, \boldsymbol{\omega}) &= \sum_{F\in\mathcal{F}^0_{pr}} ([\underline{\sigma}\boldsymbol{n}]_{|F}, \boldsymbol{\omega})_F,\ \forall(\underline{\sigma}, \boldsymbol{\omega}) \in \widetilde{W} \times \boldsymbol{\Lambda}_{pr}.\end{aligned}$$

Notice that, since $\left((\underline{I}\boldsymbol{n})^{\boldsymbol{t}}\right)_{|F} = \boldsymbol{0}, \forall F \in \mathcal{F}$, the domain of bilinear form E_h can be extended to $\widetilde{W} \times \Lambda_{dl}$ with $E_h(\underline{\sigma}, \omega) = E_h(\underline{\sigma} + \sigma\underline{I}, \omega),\ \forall\sigma \in \overline{W}$. Moreover, we define

$$\begin{aligned}\widehat{B}^*_h(\boldsymbol{v}, \underline{H}) &= -(\boldsymbol{v}, \nabla_h\cdot\underline{H}) + \sum_{\tau\in\mathcal{T}_h} (\boldsymbol{v}\cdot\boldsymbol{n}, \underline{H}\boldsymbol{n}\cdot\boldsymbol{n})_{\partial\tau\cap\mathcal{F}_{dl}},\ \forall(\boldsymbol{v}, \underline{H}) \in \widetilde{\boldsymbol{U}} \times \widetilde{W},\\ \widehat{B}_h(\underline{H}, \boldsymbol{v}) &= (\underline{H}, \nabla_h\boldsymbol{v}) - \sum_{\tau\in\mathcal{T}_h} (\boldsymbol{v}, \underline{H}\boldsymbol{n})_{\partial\tau\cap\mathcal{F}_{pr}} - \sum_{\tau\in\mathcal{T}_h} ((\boldsymbol{v})^{\boldsymbol{t}}, (\underline{H}\boldsymbol{n})^{\boldsymbol{t}})_{\partial\tau\cap\mathcal{F}_{dl}},\ \forall(\boldsymbol{v}, \underline{H}) \in \widetilde{\boldsymbol{U}} \times \widetilde{W}.\end{aligned}$$

After defining

$$\psi_0 := \int_\tau \frac{1}{d}\mathrm{tr}(\underline{\psi})dx, \quad \underline{\widetilde{\psi}} := \underline{\psi} - \psi_0\underline{I},$$

we have

$$\mathcal{A}\underline{\psi} = \mathcal{A}\underline{\widetilde{\psi}}.$$

Now, we can propose the scheme with Lagrange multiplier corresponding to the scheme developed in Sect. 3: For given $\underline{\omega}_h \in \underline{X}^h$, find $(\underline{\widetilde{G}}_h, G_0, \underline{\widehat{L}}_h, \widehat{\boldsymbol{u}}_h, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \lambda_u) \in \underline{\widehat{W}} \times \overline{W} \times \underline{\Sigma} \times \widetilde{\boldsymbol{U}} \times \boldsymbol{\Lambda}_{dl,2} \times \boldsymbol{\Lambda}_{pr} \times \Lambda_{dl,1}$, such that

$$
\begin{aligned}
N_h(\underline{\omega}_h, \underline{\widehat{L}}_h, \underline{Q}) - C_h(\underline{\widetilde{G}}_h, \underline{Q}) &= 0,\\
\widehat{B}_h^*(\widehat{\boldsymbol{u}}_h, \underline{\widetilde{H}} + H_0\underline{I}) - C_h(\underline{\widetilde{H}}, \underline{L}_h) + E_h(\underline{\widetilde{H}}, \boldsymbol{\lambda}_1) + F_h(\underline{\widetilde{H}} + H_0\underline{I}, \boldsymbol{\lambda}_2) &= 0,\\
\widehat{B}_h(\underline{\widetilde{G}}_h + G_0\underline{I}, \boldsymbol{v}) + D_h(\boldsymbol{v}, \lambda_u) &= (\boldsymbol{f}, \boldsymbol{v}),\\
D_h(\widehat{\boldsymbol{u}}_h, \mu_u) &= 0,\\
E_h(\underline{\widetilde{G}}_h, \boldsymbol{\mu}_1) &= 0,\\
F_h(\underline{\widetilde{G}}_h + G_0\underline{I}, \boldsymbol{\mu}_2) &= 0
\end{aligned}
$$

for all $(\underline{\widetilde{H}}, H_0, \underline{Q}, \boldsymbol{v}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \lambda_u) \in \underline{\widehat{W}} \times \overline{W} \times \underline{\Sigma} \times \widetilde{\boldsymbol{U}} \times \boldsymbol{\Lambda}_{dl,2} \times \boldsymbol{\Lambda}_{pr} \times \Lambda_{dl,1}$.
Let $\underline{\Sigma}^\tau$, $\underline{\widehat{W}}^\tau$, $\widetilde{\boldsymbol{U}}^\tau$ respectively represent the restriction of $\underline{\Sigma}$, $\underline{\widehat{W}}$, $\widetilde{\boldsymbol{U}}$ to each element τ for $\tau \in \mathcal{T}_h$. The corresponding local problem is defined by: Given $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \lambda_u) \in \boldsymbol{\Lambda}_{dl,2} \times \boldsymbol{\Lambda}_{pr} \times \Lambda_{dl,1}$, find $(\underline{\widehat{L}}_h, \underline{\widetilde{G}}_h, \widehat{\boldsymbol{u}}_h) \in \underline{\Sigma}^\tau \times \underline{\widehat{W}}^\tau \times \widetilde{\boldsymbol{U}}^\tau$ such that

$$
\begin{aligned}
&(\mu(|\underline{\omega}_h|)\underline{\widehat{L}}_h, \underline{Q})_\tau - (\mathcal{A}\underline{\widetilde{G}}_h, \underline{Q})_\tau = 0 \quad \forall \underline{Q} \in \underline{\Sigma}^\tau,\\
&\widehat{B}_h^{*,\tau}(\boldsymbol{u}_h^*, \underline{H}) - (\mathcal{A}\underline{\widehat{L}}_h, \underline{H})_\tau = -(\boldsymbol{\lambda}_1, [(\underline{H}\boldsymbol{n})^{\boldsymbol{t}}])_{\partial\tau\cap\mathcal{F}_{dl}} - (\boldsymbol{\lambda}_2, \underline{H}\boldsymbol{n})_{\partial\tau\cap\mathcal{F}_{pr}^0} \quad \forall \underline{H} \in \underline{\widehat{W}}^\tau,\\
&\widehat{B}_h^\tau(\underline{\widetilde{G}}_h, \boldsymbol{v}) = -(\lambda_u, \boldsymbol{v}\cdot\boldsymbol{n})_{\partial\tau\cap\mathcal{F}_{dl}} + (\boldsymbol{f}, \boldsymbol{v})_\tau \quad \forall \boldsymbol{v} \in \widetilde{\boldsymbol{U}}^\tau.
\end{aligned}
\tag{6.4}
$$

The solution to the local problem can be written as

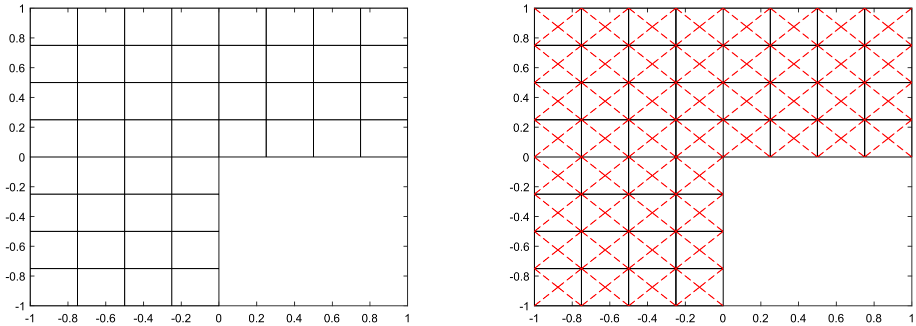
$$
(\underline{\widehat{L}}_h, \underline{\widetilde{G}}_h, \widehat{\boldsymbol{u}}_h) = (\underline{\widehat{L}}_h^{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)}, \underline{\widetilde{G}}_h^{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)}, \widehat{\boldsymbol{u}}_h^{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)}) + (L_h^{\boldsymbol{f}}, \underline{\widetilde{G}}_h^{\boldsymbol{f}}, \widehat{\boldsymbol{u}}_h^{\boldsymbol{f}})
$$

by considering separately the influence of $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \lambda_u)$ and $\boldsymbol{f}$. Indeed, $(\underline{\widehat{L}}_h^{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)}, \underline{\widetilde{G}}_h^{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)}, \widehat{\boldsymbol{u}}_h^{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)})$ is the solution to (6.4) by setting $\boldsymbol{f} = \boldsymbol{0}$.

The global (hybrid) problem is to find $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \lambda_u, G_0) \in \boldsymbol{\Lambda}_{dl,2} \times \boldsymbol{\Lambda}_{pr} \times \Lambda_{dl,1} \times \overline{W}$ such that

$$
\begin{aligned}
\sum_{F\in\mathcal{F}_{dl}} ([(G_{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)}\boldsymbol{n})^{\boldsymbol{t}}], \boldsymbol{\mu}_1)_F &= -\sum_{F\in\mathcal{F}_{dl}} ([(G_{(\boldsymbol{f})}\boldsymbol{n})^{\boldsymbol{t}}], \boldsymbol{\mu}_1)_F,\\
\sum_{F\in\mathcal{F}_{pr}^0} ([(G_{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)} + G_0\underline{I})\boldsymbol{n}], \boldsymbol{\mu}_2) &= -\sum_{F\in\mathcal{F}_{pr}^0} ([G_{(\boldsymbol{f})}\boldsymbol{n}], \boldsymbol{\mu}_2),\\
\sum_{F\in\mathcal{F}_{dl}} ([\boldsymbol{u}_{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_u)}\cdot\boldsymbol{n}], \mu_u) &= -\sum_{F\in\mathcal{F}_{dl}} ([\boldsymbol{u}_{\boldsymbol{f}}\cdot\boldsymbol{n}], \mu_u),\\
\sum_{F\in\mathcal{F}_{pr}^0} (\boldsymbol{\lambda}_2, [\psi_0\underline{I}\boldsymbol{n}])_F &= 0.
\end{aligned}
\tag{6.5}
$$

for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \mu_u, \psi_0) \in \boldsymbol{\Lambda}_{dl,2} \times \boldsymbol{\Lambda}_{pr} \times \Lambda_{dl,1} \times \overline{W}$.
Therefore, at each iteration, we first solve the local problem (6.4) and then the global problem (6.5), where the global problem only involves the trace variables.



(a) primal mesh (b) mesh after staggered DG division

Fig. 3 Example mesh for Examples 7.1 and 7.3 when $h^{-1} = 4$

7 Numerical Experiments

This section presents numerical experiments to verify the convergence estimates outlined in Sect. 5. Examples given in Examples 7.1–7.3 are executed in two dimensions, whereas Example 7.4 is carried out in three dimensions. The numerical results from Examples 7.1 and 7.2 demonstrate the robust performance of our scheme under varying values of μ including scenarios where μ is exceedingly small. Furthermore, Example 7.3 incorporates a numerical test featuring a singular solution, which deviates from the regularity assumptions outlined in Sect. 5. In such cases, the convergence rates are observed to be limited by the solution's regularity. Additionally, to extend our analysis to three-dimensional scenarios, we have included a three-dimensional numerical experiment in Example 7.4.

7.1 Example 1. Power's Law

The first example considers the quasi-Newtonian Stokes flow problems with μ in Power's law from [15]. We let Ω be the L-shape $(-1, 1)^2 \backslash [0, 1) \times (-1, 0]$ domain and consider the non-linear coefficient

$$\mu(\boldsymbol{x}, t) = k_0 + \frac{k_0 - k_\infty}{1 + t^2}.$$

We choose the source term $\boldsymbol{f}$ with the analytic solution given by

$$\begin{aligned} u_1(x, y) =& - e^x (y \cos(y) + \sin(y)), \\ u_2(x, y) =& e^x y \sin(y), \\ p(x, y) =& 2e^x \sin(y) - \frac{2(1-e)(\cos(1) - 1)}{3}. \end{aligned}$$

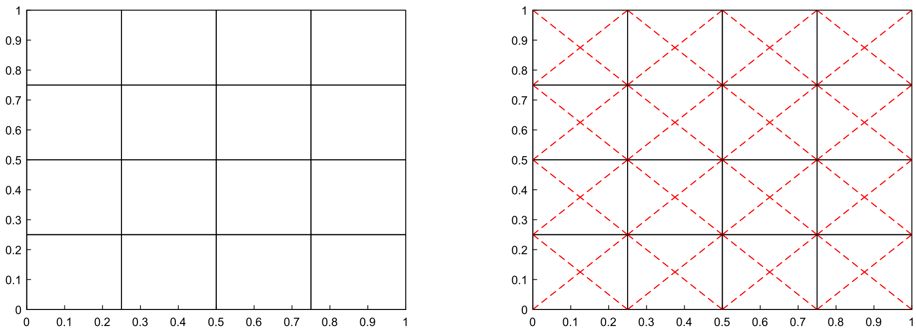
In this case, we use uniform rectangle mesh before subdivision. The mesh size h^{-1} represents the face length of rectangles. Readers can refer to Fig. 3, which shows the primal mesh and mesh after subdivision when $h^{-1} = 4$. Additionally, we consider the cases where $k_0 = 2, k_\infty = 1$ and $k_0 = 2e-4, k_\infty = 1e-4$. The Tables 1 and 2 display the convergence history for polynomial orders $k = 0, 1, 2$. It can be observed that the piecewise polynomials of order k yield optimal convergence rates of $k + 1$.

Table 1 convergence history of Example 7.1 for $k_0 = 2, k_\infty = 1$

k	mesh h^{-1}	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_0$ Error	Rate	$\|G - G_h\|_0$ Error	Rate	$\|L - L_h\|_0$ Error	Rate
0	2	3.0969e−01	NA	1.0607e+00	NA	6.7103e−01	NA
	4	1.5480e−01	1.0004	5.1701e−01	1.0367	3.5203e−01	0.9306
	8	7.7454e−02	0.9989	2.5084e−01	1.0434	1.7957e−01	0.9712
	16	3.8737e−02	0.9996	1.2340e−01	1.0234	9.0439e−02	0.9895
	32	1.9370e−02	0.9999	6.1298e−02	1.0095	4.5329e−02	0.9965
	64	9.6852e−03	0.9999	3.0579e−02	1.0033	2.2682e−02	0.9989
1	2	6.4878e−02	NA	1.4011e−01	NA	1.1117e−01	NA
	4	1.5977e−02	2.0218	3.4234e−02	2.0330	2.7267e−02	2.0275
	8	3.9782e−03	2.0058	8.5001e−03	2.0099	6.7858e−03	2.0066
	16	9.9355e−04	2.0015	2.1205e−03	2.0031	1.6950e−03	2.0012
	32	2.4832e−04	2.0004	5.2972e−04	2.0011	4.2372e−04	2.0001
2	2	3.0601e−03	NA	5.9453e−03	NA	5.0604e−03	NA
	4	3.8112e−04	3.0053	7.4082e−04	3.0045	6.2996e−04	3.0059
	8	4.7593e−05	3.0014	9.2499e−05	3.0016	7.8622e−05	3.0023
	16	5.9476e−06	3.0004	1.1557e−05	3.0007	9.8220e−06	3.0008
	32	7.4340e−07	3.0001	1.4443e−06	3.0003	1.2275e−06	3.0003

Table 2 convergence history of Example 7.1 for $k_0 = 2e-4, k_\infty = 1e-4$

k	mesh h^{-1}	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_0$ Error	Rate	$\|G - G_h\|_0$ Error	Rate	$\|L - L_h\|_0$ Error	Rate
0	2	9.6139e−01	NA	3.3088e−01	NA	4.3444e+00	NA
	4	1.6733e−01	2.5225	1.6724e−01	0.9843	4.2904e−01	3.3400
	8	7.7809e−02	1.1047	8.3853e−02	0.9960	1.7961e−01	1.2562
	16	3.8748e−02	1.0058	4.1956e−02	0.9990	9.0420e−02	0.9901
	32	1.9370e−02	1.0003	2.0982e−02	0.9997	4.5329e−02	0.9962
1	2	3.1341e+00	NA	4.0053e−02	NA	3.2388e+01	NA
	4	2.1155e−01	3.8889	9.6364e−03	2.0553	4.0195e+00	3.0104
	8	1.4610e−02	3.8560	2.3822e−03	2.0162	4.9178e−01	3.0309
	16	1.4730e−03	3.3101	5.9377e−04	2.0043	6.1719e−02	2.9942
	32	2.6937e−04	2.4511	1.4833e−04	2.0011	7.7345e−03	2.9963
2	2	1.0878e−01	NA	2.2194e−03	NA	1.3217e+00	NA
	4	3.6249e−03	4.9073	2.7667e−04	3.0039	8.4208e−02	3.9723
	8	1.3834e−04	4.7116	3.4569e−05	3.0006	5.3072e−03	3.9879
	16	8.2305e−06	4.0711	4.3208e−06	3.0001	3.3348e−04	3.9923
	32	8.0380e−07	3.3561	5.4009e−07	3.0000	2.0928e−05	3.9941



(a) primal mesh (b) mesh after staggered DG division

Fig. 4 Example mesh for Example 7.2 when $h^{-1} = 4$

7.2 Example 2. Cavity Problem

In this example, we consider the cavity problem from [6] with non-linear coefficient μ in Carreau-Law with $\theta = 1.2,\ \lambda = 1$ given by

$$\mu(t) = k_\infty + (k_0 - k_\infty)(1 - \lambda t^2)^{\frac{\theta-2}{2}}.$$

The domain Ω is assumed to be the unit square $(0, 1)^2 \subset \mathbb{R}^2$ and select the source term $\boldsymbol{f}$ with the analytic solution is given by

$$\begin{aligned}
u_1(x, y) &= (1 - \cos(2\pi \frac{e^{\theta x} - 1}{e^\theta - 1})) \sin(2\pi y),\\
u_2(x, y) &= -\theta e^{\theta x} \sin(2\pi \frac{e^{\theta x} - 1}{e^\theta - 1}) \frac{1 - \cos(2\pi y)}{e^\theta - 1},\\
p(x, y) &= 2\pi\theta e^{\theta x} \sin(2\pi \frac{e^{\theta x} - 1}{e^\theta - 1}) \frac{\sin(2\pi y)}{e^\theta - 1}.
\end{aligned}$$

In this case, we still use uniform rectangle mesh before subdivision. The mesh size h^{-1} here denotes the face length of rectangles. Readers can also refer to Fig. 4 for the case $h^{-1} = 4$, where the black solid lines represent primal faces and red dots lines represent dual faces. In the following Tables 3 and 4, we will still displace the convergence history of polynomial order $k = 0, 1, 2$ with coefficient $k_0 = 1, k_\infty = 0.5$ and $k_0 = 1e - 4, k_\infty = 5e - 5$. Similar to Example 7.1, we can observe that the convergence rates of polynomial order k are $k + 1$ for all variables as expected.

7.3 Example 3. Singular Solution

In this example, we shall test a nonlinear version of the classical singular solution presented in [15] using the nonlinearity

$$\mu(\boldsymbol{x}, t) = 1 + e^{-t}.$$

Table 3 convergence history of Example 7.2 for $k_0 = 1, k_\infty = 0.5$

k	mesh h^{-1}	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_0$ Error	Rate	$\|G - G_h\|_0$ Error	Rate	$\|L - L_h\|_0$ Error	Rate
0	2	1.5176e+00	NA	1.0371e+01	NA	1.0529e+01	NA
	4	6.4992e−01	1.2235	5.4963e+00	0.9160	6.6520e+00	0.6625
	8	2.3762e−01	1.4516	2.6494e+00	1.0528	3.3545e+00	0.9877
	16	1.0660e−01	1.1565	1.3247e+00	0.9999	1.6944e+00	0.9853
	32	5.1568e−02	1.0477	6.6222e−01	1.0003	8.4866e−01	0.9975
	64	2.5548e−02	1.0133	3.3098e−01	1.0005	4.2452e−01	0.9993
1	2	4.5446e−01	NA	6.8757e+00	NA	7.0231e+00	NA
	4	1.6101e−01	1.4970	1.7182e+00	2.0006	2.2282e+00	1.6562
	8	4.1435e−02	1.9582	3.6599e−01	2.2310	5.3708e−01	2.0527
	16	1.0660e−02	1.9587	8.9950e−02	2.0246	1.3866e−01	1.9536
	32	2.6722e−03	1.9961	2.2471e−02	2.0010	3.5142e−02	1.9803
	64	6.6106e−04	2.0152	5.6099e−03	2.0020	8.6917e−03	2.0155
2	2	1.3483e−01	NA	2.4142e+00	NA	1.9369e+00	NA
	4	2.1471e−02	2.6507	3.1524e−01	2.9370	3.8193e−01	2.3424
	8	3.2563e−03	2.7211	4.0538e−02	2.9591	5.7577e−02	2.7298
	16	5.0015e−04	2.7028	6.8517e−03	2.5647	8.1445e−03	2.8216
	32	5.5964e−05	3.1598	8.6299e−04	2.9891	1.0188e−03	2.9990
	64	7.0266e−06	2.9936	1.1164e−04	2.9505	1.3102e−04	2.9590

Table 4 convergence history of Example 7.2 for $k_0 = 1e-4, k_\infty = 5e-5$

k	mesh h^{-1}	$\|\|\boldsymbol{u} - \boldsymbol{u}_h\|\|_0$ Error	Rate	$\|\|G - G_h\|\|_0$ Error	Rate	$\|\|L - L_h\|\|_0$ Error	Rate
0	2	2.4380e+02	NA	4.3344e+00	NA	1.2582e+03	NA
	4	8.5355e+01	1.5142	1.6109e+00	1.4280	8.4109e+02	0.5810
	8	2.2086e+00	5.2723	7.1349e−01	1.1749	4.5718e+01	4.2014
	16	1.1699e−01	4.2387	3.5296e−01	1.0154	3.2881e+00	3.7974
	32	5.1200e−02	1.1921	1.7619e−01	1.0024	8.6734e−01	1.9226
	64	2.5535e−02	1.0037	8.8062e−02	1.0005	4.2470e−01	1.0301
1	2	2.0294e+03	NA	4.0718e+00	NA	1.5941e+04	NA
	4	1.0710e+02	4.2440	8.2327e−01	2.3062	1.8132e+03	3.1361
	8	6.6049e+00	4.0192	1.3685e−01	2.5888	2.3469e+02	2.9497
	16	3.7603e−01	4.1346	2.7341e−02	2.3235	2.9594e+01	2.9874
	32	2.2050e−02	4.0920	6.3080e−03	2.1158	3.4980e+00	3.0807
	64	1.4636e−03	3.9132	1.5418e−03	2.0325	4.2871e−01	3.0285
2	2	6.4899e+02	NA	1.4856e+00	NA	5.1059e+03	NA
	4	1.1769e+01	5.7851	1.5888e−01	3.2250	2.6289e+02	4.2796
	8	4.4214e−01	4.7343	1.9351e−02	3.0375	2.0525e+01	3.6790
	16	1.1310e−02	5.2889	2.4189e−03	3.0000	1.2278e+00	4.0632
	32	3.3135e−04	5.0931	3.0231e−04	3.0002	7.6825e−02	3.9984
	64	1.2400e−05	4.7399	3.7788e−05	3.0000	4.8228e−03	3.9936

Table 5 convergence history of Example 7.3

k	mesh h^{-1}	$\|\boldsymbol{u}-\boldsymbol{u}_h\|_0$ Error	Rate	$\|G-G_h\|_0$ Error	Rate	$\|L-L_h\|_0$ Error	Rate
0	1	9.2120e−01	NA	5.4845e+00	NA	1.6657e+00	NA
	2	4.8369e−01	0.9294	4.0093e+00	0.4520	1.3827e+00	0.2686
	4	2.5189e−01	0.9412	2.5516e+00	0.6519	1.0744e+00	0.3639
	8	1.2846e−01	0.9715	1.5800e+00	0.6914	7.8725e−01	0.4486
	16	6.4533e−02	0.9931	1.0095e+00	0.6462	5.5766e−01	0.4974
	32	3.2244e−02	1.0010	6.7341e−01	0.5841	3.8865e−01	0.5209
	64	1.6217e−02	0.9914	4.6931e−01	0.5209	2.6898e−01	0.5309
1	1	5.4632e−01	NA	4.8146e+00	NA	7.8248e+00	NA
	2	2.0767e−01	1.3954	3.3005e+00	0.5447	5.2959e+00	0.5631
	4	7.6698e−02	1.4371	2.2630e+00	0.5444	3.6131e+00	0.5516
	8	2.8269e−02	1.4400	1.5522e+00	0.5439	2.4784e+00	0.5438
	16	1.0876e−02	1.3781	1.0650e+00	0.5434	1.7067e+00	0.5382
	32	5.0215e−03	1.1149	7.3095e−01	0.5430	1.1806e+00	0.5317

In this example, we consider the L-shape domain $\Omega = (-1,1)^2 \backslash [0,1) \times (-1,0]$. The source term $\boldsymbol{f}$ is chosen such that the solutions, expressed in polar system (r,ϕ), are given by

$$\boldsymbol{u}(r,\phi) = r^{\lambda} \begin{pmatrix} (1+\lambda)\sin(\phi)\psi(\phi) + \cos(\phi)\psi'(\phi) \\ \sin(\phi)\psi'(\phi) - (1+\lambda)\cos(\phi)\psi(\phi) \end{pmatrix},$$

$$p(r,\phi) = -r^{\lambda-1} \frac{(1+\lambda)^2\psi'(\phi) + \psi'''(\phi)}{1-\lambda},$$

where

$$\psi(\phi) = \frac{\sin((1+\lambda)\phi)\cos(\lambda\omega)}{1+\lambda} - \cos((1+\lambda)\phi) - \frac{\sin((1-\lambda)\phi)\cos(\lambda\omega)}{1-\lambda} + \cos((1-\lambda)\phi).$$

Here, $\omega = \frac{3}{2}\pi$, and λ is determined as the smallest positive solution of $\sin(\lambda\omega) + \lambda\sin(\omega) = 0$, approximately $\lambda \approx 0.54448373678246$. It is noteworthy that $\boldsymbol{u}$ and p are analytic in $\Omega \backslash \{\mathbf{0}\}$, but $\boldsymbol{u} \notin H^2(\Omega)^2$ and $p \notin H^1(\Omega)$. In this case, we still use uniform rectangle mesh before subdivision. The mesh size h^{-1} here denotes the face length of rectangles. Readers can also refer to Fig. 3 for the case $h^{-1} = 4$, where the black solid lines represent primal faces and red dots lines represent dual faces.
The convergence history of this example is depicted in Table 5. Unlike the results in Examples 7.1 and 7.2, the convergence rate is not constrained by the polynomial degree but by the regularity of the solution, except when $k = 0$. Despite our theoretical framework not accommodating this case, the expected convergence rates reflecting the regularity of the solution can be observed. Furthermore, the results exhibit similar behavior to the theoretical results provided in [32], where $\|\boldsymbol{u}-\boldsymbol{u}_h\|_0 = \mathcal{O}(h^{\min\{k+1,2\alpha\}})$.

Table 6 convergence history of Example 7.4 in Power's law

k	mesh h^{-1}	$\|\boldsymbol{u}-\boldsymbol{u}_h\|_0$ Error	Rate	$\|G-G_h\|_0$ Error	Rate	$\|L-L_h\|_0$ Error	Rate
0	1	5.6016e−01	NA	3.3538e+00	NA	9.4506e−01	NA
	2	2.7811e−01	1.0102	1.7639e+00	0.9270	5.3737e−01	0.8144
	4	1.3874e−01	1.0033	8.5161e−01	1.0505	2.8716e−01	0.9040
	8	6.9328e−02	1.0009	4.0154e−01	1.0846	1.4781e−01	0.9581
	16	3.4658e−02	1.0003	1.9347e−01	1.0535	7.4748e−02	0.9836
1	1	2.4974e−01	NA	1.0837e+00	NA	4.7155e−01	NA
	2	6.0583e−02	2.0435	2.4417e−01	2.1500	1.1306e−01	2.0604
	4	1.4974e−02	2.0165	5.9169e−02	2.0450	2.7950e−02	2.0161
	8	3.7319e−03	2.0044	1.4680e−02	2.0110	6.9813e−03	2.0013

7.4 Example 4. Three-Dimensional Case

In this example, we consider the three-dimensional case with μ in Power's law and Carreau-Law, respectively. We select $\Omega=(0,1)^3$, and choose the source term $\boldsymbol{f}$ with analytic solution given by

$$\begin{aligned} u_1(x,y,z) &= -e^x(\sin(y)+\sin(z)+y\cos(y)+z\cos(z)),\\ u_2(x,y,z) &= e^x(y\sin(y)+z\cos(z)),\\ u_3(x,y,z) &= e^x(y\cos(y)+z\sin(z)),\\ p(x,y,z) &= 2e^x(\sin(y)+\sin(z))-4\frac{e-1}{1-\cos(1)}, \end{aligned}$$

For Power's law, we select the coefficient as:

$$\mu(\boldsymbol{x},t)=2+\frac{1}{1+t^2}, \tag{7.1}$$

and for Carreau-Law, we select the coefficient as

$$\mu(\boldsymbol{x},t)=\frac{1}{2}+\frac{1}{2}(1+x)^{-\frac{1}{4}}. \tag{7.2}$$

We applied the uniform hexahedral mesh here. The numerical results are displaced in Tables 6 and 7, where h here denote the face length of the cube before subdivision (readers can refer to Fig. 2a).

8 Conclusion

In this paper, we propose a staggered DG method for quasi-Newtonian Stokes flow problems. By introducing the flux and tensor gradient as additional variables and eliminating the pressure variable by the incompressibility condition, this problem becomes a non-linear two-fold saddle point problem, where the well-posedness of our scheme can be proved by the related abstract theorem. Even though only the normal continuity over dual faces is adopted on the finite element space for velocity, we prove that the obtained velocity is $H(\text{div},\Omega)$-conforming with the divergence-free condition. We provide a prior error analysis involving

Table 7 convergence history of Example 7.4 in Carreau-Law

k	mesh h^{-1}	$\|\boldsymbol{u}-\boldsymbol{u}_h\|_0$ Error	Rate	$\|G-G_h\|_0$ Error	Rate	$\|L-L_h\|_0$ Error	Rate
0	1	5.6045e−01	NA	1.0487e+00	NA	9.5133e−01	NA
	2	2.7820e−01	1.0105	5.6857e−01	0.8832	5.4546e−01	0.8024
	4	1.3876e−01	1.0036	2.7594e−01	1.0430	2.9247e−01	0.8991
	8	6.9330e−02	1.0010	1.2968e−01	1.0894	1.5071e−01	0.9565
	16	3.4658e−02	1.0003	6.2298e−02	1.0577	7.6239e−02	0.9831
1	1	2.5003e−01	NA	3.6851e−01	NA	4.7047e−01	NA
	2	6.0600e−02	2.0447	8.3611e−02	2.1399	1.1288e−01	2.0593
	4	1.4974e−02	2.0168	2.0324e−02	2.0405	2.7915e−02	2.0157
	8	3.7319e−03	2.0045	5.0478e−03	2.0095	6.9738e−03	2.0010

all the unknowns involved in the scheme. Additionally, we suggest the hybridization of the proposed scheme, where the resulting global system only involves the trace variables, rendering the method computationally efficient. Finally, we carry out several numerical experiments to show the performance of our scheme.

Funding Open access publishing enabled by City University of Hong Kong Library's agreement with Springer Nature The research of Lina Zhao is partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, China. [Project No. 21309522].

Data Availibility Statement The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Agouzal, A.: A posteriori error estimator for finite element discretizations of quasi-Newtonian Stokes flows. Int. J. Numer. Anal. Model. **2**(2), 221–239 (2005)
2. Ambartsumyan, I., Ervin, V.J., Nguyen, T., Yotov, I.: A nonlinear stokes-biot model for the interaction of a non-Newtonian fluid with poroelastic media. Esaim. Math. Model. Numer. Anal. **53**(6), 1915–1955 (2019)
3. Andreianov, B., Bendahmane, M., Ruiz-Baier, R.: Analysis of a finite volume method for a cross-diffusion model in population dynamics. Math. Models Meth. Appl. Sci. **21**(02), 307–344 (2011)

4. Babuska, I., Aziz, A.K.: Survey lectures on the mathematical foundations of the finite element method. In: The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations. Academic Press, Boca Raton (1972)
5. Barrett, J.W., Liu, W.B.: Quasi-norm error bounds for the finite element approximation of a non-Newtonian flow. Numer. Math. **68**, 437–456 (1994)
6. Berrone, S., Süli, E.: Two-sided a posteriori error bounds for incompressible quasi-Newtonian flows. IMA J. Numer. Anal. **28**(2), 382–421 (2008)
7. Botti, M., Di Pietro, D.A., Sochala, P.: A Hybrid High-Order method for nonlinear elasticity. SIAM J. Numer. Anal. **55**(6), 2687–2717 (2017)
8. Bustinza, R., Gatica, G.N.: A local discontinuous Galerkin method for nonlinear diffusion problems with mixed boundary conditions. SIAM J. Sci. Comput. **26**(1), 152–177 (2004)
9. Bustinza, R., Gatica, G.N.: A mixed local discontinuous Galerkin method for a class of nonlinear problems in fluid mechanics. J. Comput. Phys. **207**(2), 427–456 (2005)
10. Cáceres, E., Gatica, G.N., Sequeira, F.A.: A mixed virtual element method for quasi-Newtonian Stokes flows. SIAM J. Numer. Anal. **56**(1), 317–343 (2018)
11. Chung, E.T., Engquist, B.: Optimal discontinuous Galerkin methods for wave propagation. SIAM J. Numer. Anal. **44**(5), 2131–2158 (2006)
12. Chung, E.T., Engquist, B.: Optimal discontinuous Galerkin methods for the acoustic wave equation in higher dimensions. SIAM J. Numer. Anal. **47**(5), 3820–3848 (2009)
13. Ciarlet, P.G.: The Finite Element Method for Elliptic Problems. SIAM, Philadelphia (2002)
14. Ciarlet, P.G.: Linear and Nonlinear Functional Analysis with Applications, vol. 130. SIAM, Philadelphia (2013)
15. Congreve, S., Houston, P., Süli, E., Wihlet, T.P.: Discontinuous Galerkin finite element approximation of quasilinear elliptic boundary value problems II: strongly monotone quasi-Newtonian flows. IMA J. Numer. Anal. **33**(4), 1386–1415 (2013)
16. Diening, L., Kreuzer, C., Süli, E.: Finite element approximation of steady flows of incompressible fluids with implicit power-law-like rheology. SIAM J. Numer. Anal. **51**(2), 984–1015 (2013)
17. Gatica, G.N., González, M., Meddahi, S.: A low-order mixed finite element method for a class of quasi-Newtonian Stokes flows. Part I: a priori error analysis. Comput. Meth. Appl. Mech. Eng. **193**(9–11), 881–892 (2004)
18. Gatica, G.N., Heuer, N., Meddahi, S.: On the numerical analysis of nonlinear twofold saddle point problems. IMA J. Numer. Anal. **23**(2), 301–330 (2003)
19. Gatica, G.N., Márquez, A., Sánchez, M.A.: A priori and a posteriori error analyses of a velocity-pseudostress formulation for a class of quasi-Newtonian Stokes flows. Comput. Meth. Appl. Mech. Eng. **200**(17–20), 1619–1636 (2011)
20. Gatica, G.N., Sequeira, F.A.: Analysis of an augmented HDG method for a class of quasi-Newtonian Stokes flows. J. Sci. Comput. **65**(3), 1270–1308 (2015)
21. Gatica, G.N., Sequeira, F.A.: A priori and a posteriori error analyses of an augmented HDG method for a class of quasi-Newtonian Stokes flows. J. Sci. Comput. **69**, 1192–1250 (2016)
22. Gudi, T., Nataraj, N., Pani, A.K.: hp-discontinuous Galerkin methods for strongly nonlinear elliptic boundary value problems. Numer. Math. **109**(2), 233–268 (2008)
23. Gudi, T., Nataraj, N., Pani, A.K.: An hp-local discontinuous Galerkin method for some quasilinear elliptic boundary value problems of non-monotone type. Math. Comput. **77**(262), 731–756 (2008)
24. Gudi, T., Pani, A.K.: Discontinuous Galerkin methods for quasi-linear elliptic problems of non-monotone type. SIAM J. Numer. Anal. **45**(1), 163–192 (2007)
25. Houston, P., Robson, J., Süli, E.: Discontinuous Galerkin finite element approximation of quasilinear elliptic boundary value problems I: The scalar case. IMA J. Numer. Anal. **25**(4), 726–749 (2005)
26. Howell, J.S.: Dual-mixed finite element approximation of Stokes and nonlinear Stokes problems using trace-free velocity gradients. J. Comput. Appl. Math. **231**(2), 780–792 (2009)
27. Kim, D., Zhao, L., Park, E.-J.: Staggered DG methods for the pseudostress-velocity formulation of the Stokes equations on general meshes. SIAM J. Sci. Comput. **42**(4), A2537–A2560 (2020)
28. Kim, H., Chung, E.T., Lee, C.: A staggered discontinuous Galerkin method for the Stokes system. SIAM J. Numer. Anal. **51**(6), 3327–3350 (2013)
29. Oyarzúa, R., Solano, M., Zúniga, P.: Analysis of an unfitted mixed finite element method for a class of quasi-Newtonian Stokes flow. Comput. Math. with Appl. **114**, 225–243 (2022)
30. Yadav, S., Pani, A.K., Park, E.-J.: Superconvergent discontinuous Galerkin methods for nonlinear elliptic equations. Math. Comput. **82**(283), 1297–1335 (2013)
31. Zhao, L., Chung, E.T., Lam, M.: A new staggered DG method for the Brinkman problem robust in the Darcy and Stokes limits. Comput. Meth. Appl. Mech. Eng. **364**, 112986 (2020)

32. Zhao, L., Park, E.-J.: A staggered discontinuous Galerkin method of minimal dimension on quadrilateral and polygonal meshes. SIAM J. Sci. Comput. **40**(4), A2543–A2567 (2018)
33. Zhao, L., Park, E.-J.: A lowest-order staggered DG method for the coupled Stokes-Darcy problem. IMA J. Numer. Anal. **40**(4), 2871–2897 (2020)
34. Zhao, L., Park, E.-J., Chung, E.T.: A pressure robust staggered discontinuous Galerkin method for the Stokes equations. Comput. Math. Appl. **128**, 163–179 (2022)
35. Zhao, L., Park, E.-J., Shin, D.-W.: A staggered DG method of minimal dimension for the Stokes equations on general meshes. Comput. Meth. Appl. Mech. Eng. **345**, 854–875 (2019)
36. Zhao, L., Sun, S.: A strongly mass conservative method for the coupled Brinkman-Darcy flow and transport. SIAM J. Sci. Comput. **45**(2), B166–B199 (2023)