Adv Comput Math (2020) 46: 71
https://doi.org/10.1007/s10444-020-09809-8

A unified analysis of a class of quadratic finite volume element schemes on triangular meshes

Yanhui Zhou[1] · Jiming Wu[1]

Received: 21 December 2019 / Accepted: 13 July 2020 /
Published online: 27 August 2020

Abstract

This paper presents a general framework for the coercivity analysis of a class of quadratic finite volume element (FVE) schemes on triangular meshes for solving elliptic boundary value problems. This class of schemes covers all the existing quadratic schemes of Lagrange type. With the help of a new mapping from the trial function space to the test function space, we find that each element matrix can be decomposed into three parts: the first part is the element stiffness matrix of the standard quadratic finite element method (FEM), the second part is the difference between the FVE and FEM on the element boundary, while the third part can be expressed as the tensor product of two vectors. Thanks to this decomposition, we obtain a sufficient condition to guarantee the existence, uniqueness, and coercivity result of the FVE solution on triangular meshes. Moreover, based on this sufficient condition, some minimum angle conditions with simple, analytic, and computable expressions can be derived and they depend only on the constructive parameters of the schemes. As a byproduct, some existing coercivity results are improved. Finally, an optimal H^1 error estimate is proved by the standard techniques.

Keywords Quadratic finite volume element schemes · Triangular meshes · Coercivity result · Minimum angle condition · Optimal H^1 error estimate

Mathematics Subject Classification (2010) 65N08 · 65N12 · 35J25

Communicated by: Aihui Zhou

✉ Jiming Wu
wu_jiming@iapcm.ac.cn

Yanhui Zhou
zhouyh9@mail2.sysu.edu.cn

[1] Institute of Applied Physics and Computational Mathematics, Beijing, 100088, People's Republic of China

1 Introduction

The finite volume method (FVM) is one of the major numerical methods for solving partial differential equations (c.f. [2, 22, 28, 29, 32]), since it preserves the local conservation law. The finite volume element method (FVEM) is one type of FVM, and its mathematical progress can be found in [23, 26, 44] and the references cited therein.

The coercivity result is one of the most challenging works for the analysis of FVEMs, especially for the high-order schemes. For the linear FVEM on triangular meshes, its element stiffness matrix can be regarded as a small perturbation of linear FEM for variable coefficient, then the coercivity result can be proved (c.f. [1, 4, 18, 19, 40]), and the error estimates were presented in [1, 4, 12, 13, 18, 19, 38, 40] for incomplete references. Recently, [16, 17, 47] studied the adaptive linear FVEM on triangular meshes, and [35] studied the conditioning of linear FVEM on arbitrary simplicial meshes.

Unlike the linear scheme, the existing quadratic scheme is constructed by two parameters α and β, where $\alpha \in (0, 1/2)$ on the element boundary and $\beta \in (0, 2/3)$ in the interior of element (c.f. [40]). For the coercivity result of quadratic FVE schemes, its mathematical progress still lags far behind compared with the linear FVEM. For example, for the first proposed quadratic scheme $(\alpha, \beta) = (1/3, 1/3)$, by assuming that the maximum angle of each triangular element is not greater than $90°$, and the ratio of the lengths of the two sides of the maximum angle belong to $[\sqrt{2/3}, \sqrt{3/2}]$, Tian and Chen [31] presented a coercivity result. In 1996, Liebau [25] studied the scheme $(\alpha, \beta) = (1/4, 1/3)$, and required that the geometry of the triangulation triangles is not too extreme. In 2009, Xu and Zou [40] improved some earlier coercivity results, the minimum angle should be greater than or equal to $7.11°$ for the scheme proposed in [15] $(\alpha, \beta) = (1/6, 1/4)$, $9.98°$ for the scheme proposed in [25] $(\alpha, \beta) = (1/4, 1/3)$, and $20.95°$ for the scheme proposed in [31] $(\alpha, \beta) = (1/3, 1/3)$. In 2012, a general framework for the construction and analysis of higher-order FVMs was established in [8] by Chen, Wu, and Xu. For a specific quadratic scheme, its minimum angle condition can be obtained by a computer program, and the coercivity result is the same as [40] for the schemes in [15, 25, 31]. Later, the relationship of the uniform ellipticity, inf-sup condition, and uniform local ellipticity of high-order FVMs was presented in [10]. In 2017, by introducing a novel mapping from the trial function space to test function space, Zou [49] proposed an unconditionally stable quadratic scheme with $(\alpha, \beta) = ((3 - \sqrt{3})/6, (3 - \sqrt{3})/6)$. Recently, Zhou and Wu [46] analyze a family of quadratic schemes with a parameter β and α is fixed as $(3 - \sqrt{3})/6$, and improved the minimum angle condition in [36] to $1.42°$.

Based on the coercivity result of quadratic scheme on triangular meshes, one can study its convergence properties and apply it to solve more complicated problems, e.g., [7, 14, 21, 33, 34, 36, 37, 39, 41, 42]. The relevant studies of hybrid FVMs and Hermite FVMs were presented in [5, 8, 9, 11] and the references cited therein. On the other hand, for the coercivity analysis on quadrilateral meshes, we refer the reader to a non-exhaustive literature [6, 20, 23, 24, 27, 30, 43, 45, 48]. From another viewpoint, one can postprocess the continuous Galerkin finite element solution, to obtain a finite-volume-like solution which satisfies the conservation law on each dual cell. For example, in 2013 Bush and Ginting [3] postprocess the linear FEM, while in 2017 Zou, Guo, and Deng [50] consider the high-order FEM.

In this work, we intend to generalize the coercivity analysis in [46] to any scheme parameter pair (α, β), which cover all the existing quadratic schemes of Lagrange type. In this case, some new difficulties arise. Firstly, there are two scheme parameters α and β need us to consider, while [46] only concentrate on one scheme parameter β and α is fixed as $(3-\sqrt{3})/6$. Therefore, the representation, computation, and analysis of the element stiffness matrix are more complicated than [46]. Secondly, the analysis technique in [46] heavily depends on the orthogonality on the boundary of triangle. However, here we consider a class of schemes with two scheme parameters α and β, and the orthogonality on the boundary of triangle does not hold when $\alpha \neq (3-\sqrt{3})/6$.

In order to overcome these difficulties, in this paper, we introduce a novel mapping from the trial function space to the test function space. Precisely, we first convert the FVE element bilinear form to a quadratic form with respect to a 6-by-6 singular element matrix. With the help of the mapping, a weak orthogonality on the boundary of triangle holds for any α, then this element matrix can be split as three parts: the first part is the element stiffness matrix of the standard quadratic FEM, the second part is the difference between the FVE and FEM on the boundary of triangle, while the third part is the difference in the interior of triangle and can be simplified to the tensor product of two vectors. Then, the analysis of this element matrix can be further transformed to that of a 5-by-5 symmetric matrix. Thanks to this finding, we obtain a sufficient condition to guarantee the existence, uniqueness, and coercivity result of the FVE solution on triangular meshes. Under the coercivity result, the optimal H^1 error estimate is proved.

Compared with the prior works, the present work has some contributions. Firstly, we present a general framework to study a class of quadratic FVE schemes with two parameters α and β, covering all the existing quadratic schemes of Lagrange type [15, 25, 31, 36, 46, 49]. Secondly, for any quadratic scheme which determined by α and β, we proved that the coercivity result is valid on equilateral triangular mesh. Thirdly, in order to ensure the coercivity result of these schemes on general triangular meshes, we obtain the corresponding minimum angle condition with a simple, analytic, and computable expression, and moreover the minimum angle only relies on the scheme parameters. As a direct consequence, some existing coercivity results are improved, see Table 1. Throughout the analysis, two factors play important roles, i.e., the proper choice of the mapping from the trial function space to the test function space, and the weak orthogonality properties.

Here we mention a closely related work [8] where a general framework was proposed to analyze high-order FVE schemes on triangular meshes. Specifically, for the quadratic FVE schemes, there exist some similarities and differences between the present work and [8]. Firstly, both works adopt the element analysis approach. Secondly, the mappings from the trial function space to the test function space are different. [8] uses a fixed mapping while this work uses a special one depending on α, leading to different bilinear forms. Finally, the coercivity condition obtained in

Table 1 The minimum angle conditions for some special quadratic FVE schemes

(α, β)	Existing results	Our results
$\left(\frac{1}{3}, \frac{1}{3}\right)$, in 1991, [31]	20.95°, in 2009, [40]	10.08°
	20.95°, in 2012, [8]	
$\left(\frac{1}{6}, \frac{1}{4}\right)$, in 1992, [15]	7.11°, in 2009, [40]	7.11°
	7.11°, in 2012, [8]	
$\left(\frac{1}{4}, \frac{1}{3}\right)$, in 1996, [25]	9.98°, in 2009, [40]	4.14°
	9.98°, in 2012, [8]	
$\left(\frac{3-\sqrt{3}}{6}, \frac{6+\sqrt{3}-\sqrt{21+6\sqrt{3}}}{9}\right)$, in 2016, [36]	5.24°, in 2016, [36]	1.42°
	1.42°, in 2020, [46]	

[8] does not have an analytic expression and it can only be verified by a computer program.

We organize the rest of the paper as follows. In Section 2, we present a class of quadratic finite volume element schemes. In order to prove the coercivity result, we first give some preliminaries in Section 3. The coercivity analysis of these schemes on equilateral triangular mesh and general triangular meshes is presented in Sections 4 and 5, respectively. In Section 6, we provide some analytic expressions to approximate the minimal angle condition. In Section 7, we discuss the minimum angle condition for some special quadratic schemes, and give a simple and analytic expression of this angle. The optimal H^1 error estimate is shown in Section 8 and the conclusions are given in Section 9.

In the sequential discussion, to avoid repetition, we sometimes write "$A \lesssim B$" to indicate that A can be bounded by B multiplied by a constant irrelative to the parameters which A and B may depend on. Analogously, "$A \gtrsim B$" means that B can be bounded by A, while "$A \sim B$" stands for the fact that we have both "$A \lesssim B$" and "$B \lesssim A$."

2 A class of quadratic FVE schemes

We consider the following elliptic equation

$$-\nabla \cdot (\kappa \nabla u) = f, \quad \text{in } \Omega, \tag{2.1}$$

$$u = 0, \quad \text{on } \partial\Omega, \tag{2.2}$$

where $\Omega \subset \mathbb{R}^2$ is a bounded polygonal domain, $f \in L^2(\Omega)$, and κ is a piecewise smooth function that can be bounded above and below, i.e., there exist positive constants $\kappa_{\min}$ and $\kappa_{\max}$, such that

$$\kappa_{\min} \le \kappa(x, y) \le \kappa_{\max}, \quad \text{for a.e. } (x, y) \in \Omega.$$

For the polygonal domain Ω, the *primary mesh* $\mathcal{T}_h := \{K\}$ is a *conforming* triangular partition of Ω, where $h = \max_{K\in\mathcal{T}_h} h_K$ and h_K is the diameter of K. We assume that κ is smooth inside each cell and $\mathcal{T}_h$ is *shape regular*, i.e., there exists a positive constant θ_0, independent of h and K, such that

$$\theta_K \geq \theta_0 > 0, \quad \forall K \in \mathcal{T}_h, \tag{2.3}$$

where θ_K is the minimum interior angle of K. With respect to the primary mesh $\mathcal{T}_h$, the standard k-th order finite element space of Lagrange type is

$$U_h^k = \{u_h \in C(\overline{\Omega}) : u_h|_K \in \mathbb{P}_k, \ \forall K \in \mathcal{T}_h; \ u_h|_{\partial\Omega} = 0\}, \tag{2.4}$$

where $\mathbb{P}_k$ is the set of all polynomials of degree not greater than k. Obviously, we have $U_h^k \subset H_0^1(\Omega)$. Throughout the paper, we choose the *trial function space* U_h as $U_h := U_h^2$.

Next, we introduce the construction of the dual cells. To this end, for any triangular element $K \in \mathcal{T}_h$, let $\mathcal{N}_K$ and $\mathcal{E}_K$ be the set of six nodes (three vertices and three edge midpoints) and the set of three edges of K, respectively. Moreover, let

$$\mathcal{N}_h = \bigcup_{K\in\mathcal{T}_h} \mathcal{N}_K, \quad \mathcal{N}_h^\circ = \mathcal{N}_h \backslash \partial\Omega, \quad \mathcal{E}_h = \bigcup_{K\in\mathcal{T}_h} \mathcal{E}_K.$$

For any $K = \triangle P_1P_2P_3 \in \mathcal{T}_h$, see Fig. 1, we denote by Q the barycenter of K and M_i $(i = 1, 2, 3)$ the midpoint of the line segment P_iP_{i+1}, here and hereafter i denotes, without special mention, a periodic index with period 3. For each $\alpha \in (0, 1/2)$, $P_{i,i+1}^\alpha$ and $P_{i+1,i}^\alpha$ are the two points on the line segment P_iP_{i+1}, subjected to

$$\frac{|P_iP_{i,i+1}^\alpha|}{|P_iP_{i+1}|} = \frac{|P_{i+1,i}^\alpha P_{i+1}|}{|P_iP_{i+1}|} = \alpha. \tag{2.5}$$

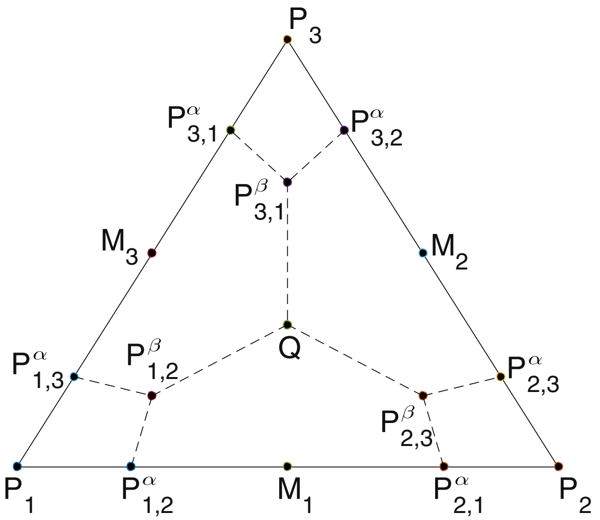

Fig. 1 Partition of the triangular element K

For each $\beta \in (0, 2/3)$, the point $P^{\beta}_{i,i+1}$ is located at the line segment $P_i M_{i+1}$, satisfying

$$\frac{|P_i P^{\beta}_{i,i+1}|}{|P_i M_{i+1}|} = \beta. \tag{2.6}$$

Using these notations, we obtain a partition of K, consisting of three quadrilaterals and three pentagons, see Fig. 1. For each node $P \in \mathcal{N}_h$, the dual cell associated with P is a polygonal domain surrounding P and denoted as $\mathcal{V}_P$. If $P = P_i$ is a vertex of K, then the contribution of K to $\mathcal{V}_P$ is the quadrilateral $P_i P^{\alpha}_{i,i+1} P^{\beta}_{i,i+1} P^{\alpha}_{i,i+2}$. If $P = M_i$ is an edge midpoint of K, then the contribution of K to $\mathcal{V}_P$ is the pentagon $P^{\alpha}_{i,i+1} P^{\alpha}_{i+1,i} P^{\beta}_{i+1,i+2} Q P^{\beta}_{i,i+1}$. The *dual mesh* $\mathcal{T}'_h$ consists of all dual cells, i.e.,

$$\mathcal{T}'_h = \{\mathcal{V}_P : P \in \mathcal{N}_h\},$$

see Fig. 2 for an example of $\mathcal{T}'_h$.

The corresponding *test function space* is defined as

$$V_h = \mathrm{Span}\{\psi_P : P \in \mathcal{N}^{\circ}_h\},$$

where ψ_P is the characteristic function associated with $\mathcal{V}_P$. Then, for any $v_h \in V_h$, we have $v_h = \sum_{P \in \mathcal{N}^{\circ}_h} v_P \psi_P$ with $v_P = v_h(P)$. Moreover, there holds $\dim U_h = \dim V_h = \#\mathcal{N}^{\circ}_h$.

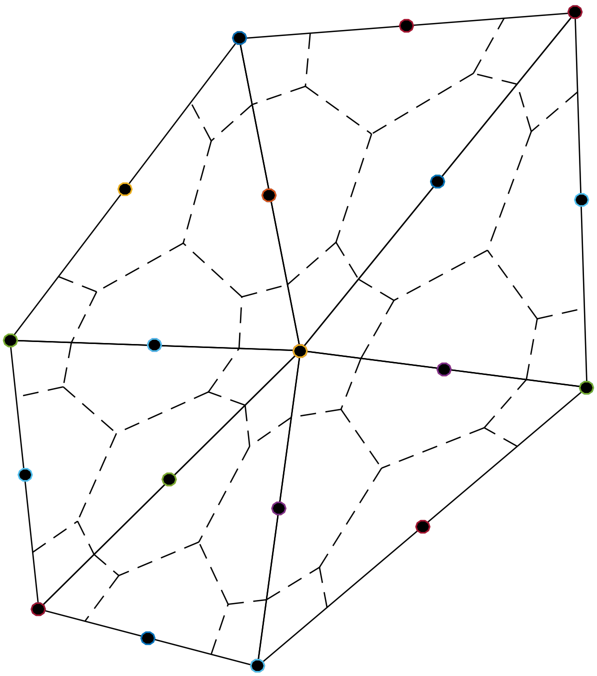

Fig. 2 The primary mesh $\mathcal{T}_h$ (solid lines) and its associated dual mesh $\mathcal{T}'_h$ (dotted lines)

The quadratic finite volume element solution of (2.1) and (2.2) is a function $u_h \in U_h$, satisfying the following local conservation law

$$-\int_{\partial \mathcal{V}_P} \kappa \frac{\partial u_h}{\partial \boldsymbol{n}} \,\mathrm{d}s = \int_{\mathcal{V}_P} f \,\mathrm{d}x\mathrm{d}y, \quad \forall\, P \in \mathcal{N}_h^\circ,$$

where $\boldsymbol{n}$ is the unit normal outward to $\partial \mathcal{V}_P$. Consequently, the above quadratic finite volume element method can be reformulated as: Find $u_h \in U_h$ such that

$$a_h(u_h, v_h) = (f, v_h), \quad \forall\, v_h \in V_h, \tag{2.7}$$

where

$$a_h(u, v_h) = -\sum_{P \in \mathcal{N}_h^\circ} v_P \int_{\partial \mathcal{V}_P} \kappa \frac{\partial u}{\partial \boldsymbol{n}} \,\mathrm{d}s, \quad u \in H_0^1(\Omega),\ v_h \in V_h$$

and (f, v_h) denotes the standard L^2 inner product of f and v_h.

Recalling the construction of the dual mesh, one can see that the above problem depends on two parameters $\alpha \in (0, 1/2)$ and $\beta \in (0, 2/3)$. Therefore, (2.7) actually leads to a class of quadratic finite volume element schemes. When (α, β) is endowed with specific values, we recover all the existing schemes listed below.

- If $\alpha = \beta = 1/3$, (2.7) reduces to the FVE scheme in [31].
- If $\alpha = 1/6$ and $\beta = 1/4$, (2.7) leads to the FVE scheme in [15].
- If $\alpha = 1/4$ and $\beta = 1/3$, (2.7) reduces to the FVE scheme in [25].
- If
$$\alpha = \frac{3-\sqrt{3}}{6}, \quad \beta = \frac{6+\sqrt{3}-\sqrt{21+6\sqrt{3}}}{9},$$
(2.7) is identical to the quadratic FVE scheme in [36].
- If $\alpha = \beta = (3-\sqrt{3})/6$, (2.7) reduces to the FVE scheme in [49].
- If $\alpha = (3-\sqrt{3})/6$ and $\beta \in (0, 2/3)$, (2.7) reduces to the family of FVE schemes studied in [46].

In the subsequential discussion, we shall study the quadratic finite volume element scheme (2.7) for any $\alpha \in (0, 1/2)$ and $\beta \in (0, 2/3)$.

3 Preliminaries

In this section, we first introduce a novel mapping from the trial function space to the test function space, then present a sketch of the standard element analysis to prove the coercivity result.

3.1 A novel mapping from the trial function space to the test function space

For a given $\omega \neq 0$, let Π_ω be an interpolation operator that maps $u_h \in U_h$ to $\Pi_\omega u_h \in V_h$, given by

$$\Pi_\omega u_h(P_i) = u_h(P_i)$$

and

$$\Pi_\omega u_h(M_i) = \frac{1-\omega}{2}\left(u_h(P_i) + u_h(P_{i+1})\right) + \omega u_h(M_i), \tag{3.1}$$

where $P_i(i = 1, 2, 3)$ and $M_i(i = 1, 2, 3)$ denote the vertices and midpoints of the triangle K, respectively. Obviously, Π_ω is a bijection from the trial function space U_h to the test function space V_h. We remark that if $\omega = 1$ (resp. $\omega = 2/\sqrt{3}$), then Π_ω reduces to the mapping in [23, 31, 36] (resp. [46, 49]).

In the following coercivity analysis, ω plays an important role and it is determined by the scheme parameter α. Here, we give two choices below.

– In Section 4, for the discussion of the equilateral triangular mesh, we choose

$$\omega \in \left(\omega^-,\ \omega^+\right), \tag{3.2}$$

where

$$\omega^\pm = \frac{4\left(1 \pm \sqrt{3\alpha(1-\alpha)}\right)^2}{3\,(1-2\alpha)}.$$

– In Section 5, for the discussion of the general triangular meshes, we choose

$$\omega = \frac{2}{3\,(1-2\alpha)}. \tag{3.3}$$

Suppose that λ_i, $i = 1, 2, 3$ are the three linear nodal basis functions corresponding to P_i, $i = 1, 2, 3$. Then, for any $u_h \in U_h$

$$u_h|_K = \sum_{i=1}^{3} u_h(P_i)\phi_{P_i} + \sum_{i=1}^{3} u_h(M_i)\phi_{M_i}, \quad \forall K \in \mathcal{T}_h,$$

where

$$\phi_{P_i} = \lambda_i(2\lambda_i - 1), \quad \phi_{M_i} = 4\lambda_i\lambda_{i+1} \tag{3.4}$$

are the quadratic nodal basis functions corresponding to P_i and M_i, $i = 1, 2, 3$. Some properties of ϕ_{P_i} and ϕ_{M_i} are listed in the following Lemma whose proof is trivial and we omit it here.

Lemma 3.1 *For each ϕ_{P_i} and ϕ_{M_i} defined in (3.4), we have*

$$\Pi_\omega\phi_{P_i}(P_j) = \delta_{ij}, \quad j = 1, 2, 3,$$

$$\Pi_\omega\phi_{P_i}(M_i) = \Pi_\omega\phi_{P_i}(M_{i+2}) = \frac{1-\omega}{2}, \quad \Pi_\omega\phi_{P_i}(M_{i+1}) = 0,$$

and

$$\Pi_\omega\phi_{M_i}(P_j) = \Pi_\omega\phi_{M_i}(M_{i+1}) = \Pi_\omega\phi_{M_i}(M_{i+2}) = 0, \quad j = 1, 2, 3,$$

$$\Pi_\omega\phi_{M_i}(M_i) = \omega.$$

Moreover, we have

$$\sum_{i=1}^{3}\phi_{P_i} + \sum_{i=1}^{3}\phi_{M_i} = \sum_{i=1}^{3}\Pi_\omega\phi_{P_i} + \sum_{i=1}^{3}\Pi_\omega\phi_{M_i} = 1.$$

3.2 The sketch of the element analysis for the coercivity analysis

To prove the global coercivity result

$$a_h(u_h, \Pi_\omega u_h) \gtrsim |u_h|_1^2, \quad \forall\, u_h \in U_h, \tag{3.5}$$

it suffices to prove

$$a_h^K(u_h, \Pi_\omega u_h) \gtrsim |u_h|_{1,K}^2, \quad \forall\, u_h \in U_h, \quad \forall\, K \in \mathcal{T}_h,$$

where

$$a_h^K(u_h, \Pi_\omega u_h) = -\sum_{P \in \mathcal{N}_h} \Pi_\omega u_h(P) \int_{\partial \mathcal{V}_P \cap K} \kappa \frac{\partial u_h}{\partial \boldsymbol{n}}\, \mathrm{d}s. \tag{3.6}$$

For any $u_h \in U_h$, in each K, we define the vector

$$\boldsymbol{u} = (u_h(P_1), \cdots, u_h(P_6))^T, \tag{3.7}$$

where $P_{i+3} := M_i$, $i = 1, 2, 3$. Hence, there holds

$$a_h^K(u_h, \Pi_\omega u_h) = a_h^K\left(\sum_{j=1}^{6} u_h(P_j)\phi_{P_j}, \sum_{i=1}^{6} u_h(P_i)\Pi_\omega \phi_{P_i}\right) = \boldsymbol{u}^T \mathbb{A}_K \boldsymbol{u}, \tag{3.8}$$

where $\mathbb{A}_K = (a_{ij})_{6\times 6}$ with

$$a_{ij} = a_h^K(\phi_{P_j}, \Pi_\omega \phi_{P_i}). \tag{3.9}$$

Thus, the proof of (3.5) reduces to the spectral analysis of the element matrix $\mathbb{A}_K$.

3.3 Preliminaries for the spectral analysis of $\mathbb{A}_K$

Lemma 3.2 *Assume that $\mathcal{T}_h$ is shape regular, then for each $K \in \mathcal{T}_h$,*

$$|u_h|_{1,K} \sim \|\mathbb{G}\boldsymbol{u}\|, \quad \forall\, u_h \in U_h, \tag{3.10}$$

where

$$\mathbb{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \tag{3.11}$$

$\boldsymbol{u}$ is defined in (3.7) and $\|\cdot\|$ denotes the Euclidean norm.

Proof The proof of (3.10) can be found in Lemma 1 of [31] or Lemma 3.4.1 in [23]. □

Lemma 3.3 *The element matrix $\mathbb{A}_K$ defined in (3.8) is singular and*

$$\sum_{k=1}^{6} a_{ik} = \sum_{k=1}^{6} a_{kj} = 0, \quad i, j = 1, \cdots, 6. \tag{3.12}$$

Proof For any $i = 1, \cdots, 6$, it follows from (3.9) and Lemma 3.1 that

$$\sum_{k=1}^{6} a_{ik} = \sum_{k=1}^{6} a_h^K(\phi_{P_k}, \Pi_\omega \phi_{P_i}) = a_h^K\left(\sum_{k=1}^{6} \phi_{P_k},\ \Pi_\omega \phi_{P_i}\right) = a_h^K(1, \Pi_\omega \phi_{P_i}) = 0.$$

On the other hand, for any $j = 1, \cdots, 6$,

$$\sum_{k=1}^{6} a_{kj} = \sum_{k=1}^{6} a_h^K(\phi_{P_j}, \Pi_\omega \phi_{P_k}) = a_h^K\left(\phi_{P_j},\ \sum_{k=1}^{6} \Pi_\omega \phi_{P_k}\right) = a_h^K(\phi_{P_j}, 1) = 0,$$

where we have used the fact that $\kappa \nabla \phi_{P_j} \cdot \boldsymbol{n}$ is continuous inside K in the last equality. Thus, (3.12) is proved and $\mathbb{A}_K$ is singular. □

Lemma 3.4 *Let*

$$\mathbb{T} = \frac{1}{6}\begin{pmatrix} 5 & -1 & -1 & -2 & 2 \\ -1 & 5 & -1 & -2 & -4 \\ -1 & -1 & 5 & 4 & 2 \\ -1 & -1 & -1 & 4 & 2 \\ -1 & -1 & -1 & -2 & 2 \\ -1 & -1 & -1 & -2 & -4 \end{pmatrix} \tag{3.13}$$

and define

$$\mathbb{B}_K = \frac{1}{2}\mathbb{T}^T\left(\mathbb{A}_K + \mathbb{A}_K^T\right)\mathbb{T}, \tag{3.14}$$

where $\mathbb{A}_K$ is given by (3.9). Then, we have

$$\mathbb{G}^T \mathbb{B}_K \mathbb{G} = \frac{1}{2}\left(\mathbb{A}_K + \mathbb{A}_K^T\right). \tag{3.15}$$

Proof By (3.13) and (3.11), we have

$$\mathbb{T}\mathbb{G} = \mathbb{I} - \frac{1}{6}\mathbb{1},$$

where $\mathbb{I}$ is the identity matrix and $\mathbb{1}$ is a 6×6 matrix with all entries equal to 1. Then, using this identity, (3.12) and (3.14), we reach (3.15) by direct calculations. □

Lemma 3.5 *If $\kappa = 1$ on K, then, for a_{ij} defined in (3.9), we have*

$$a_{ij} = \int_{\partial K} \nabla \phi_{P_j} \cdot \boldsymbol{n} \Pi_\omega \phi_{P_i}\, ds - \int_K (\Delta \phi_{P_j}) \Pi_\omega \phi_{P_i}\, dxdy. \tag{3.16}$$

Proof Recalling (3.9) and (3.6), we find that (3.16) is a direct consequence of Green's formula. □

Lemma 3.6 *For any $K = \triangle P_1P_2P_3 \in \mathcal{T}_h$, we have*

$$\nabla\lambda_i \cdot \nabla\lambda_{i+1} = -\frac{1}{2|K|}\cot\theta_{i+2}, \quad \nabla\lambda_i \cdot \nabla\lambda_i = \frac{1}{2|K|}\left(\cot\theta_{i+1} + \cot\theta_{i+2}\right),$$
$$\Delta\phi_{P_i} = \frac{2}{|K|}\left(\cot\theta_{i+1} + \cot\theta_{i+2}\right), \quad \Delta\phi_{M_i} = -\frac{4}{|K|}\cot\theta_{i+2}$$

and

$$\boldsymbol{n}_i = -\frac{2|K|}{|P_iP_{i+1}|}\nabla\lambda_{i+2},$$

where $\boldsymbol{n}_i$ denotes the unit normal outward to the edge P_iP_{i+1} of K, and $\theta_i = \angle P_{i+2}P_iP_{i+1}$ is the interior angle of K corresponding to vertex P_i.

Proof The proof can be found, e.g., in Lemma 3 of [46]. □

From (3.15), one can see that its right-hand side is the symmetric part of $\mathbb{A}_K$. In order to investigate the spectral property of $\mathbb{A}_K$, we shall study $\mathbb{B}_K$ for the equilateral triangular element and general triangular element in Sections 4 and 5, respectively. For simplicity of exposition, we introduce the following notations,

$$a_1 = \frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \lambda_i \Pi_\omega\phi_{P_i}\,\mathrm{d}s, \tag{3.17}$$
$$a_2 = \frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \lambda_{i+1} \Pi_\omega\phi_{P_i}\,\mathrm{d}s, \tag{3.18}$$
$$a_3 = \frac{1}{|K|}\int_K \Pi_\omega\phi_{P_i}\,\mathrm{d}x\mathrm{d}y. \tag{3.19}$$

By Lemma 3.1 and through some straightforward calculations, we have

$$a_1 = \frac{1}{4}\left(2\alpha\omega - \omega + 1 - 2\alpha^2 + 2\alpha\right), \tag{3.20}$$
$$a_2 = \frac{1}{4}\left(2\alpha\omega - \omega + 1 + 2\alpha^2 - 2\alpha\right), \tag{3.21}$$
$$a_3 = \alpha\beta\omega + \frac{1}{3}\left(1 - \omega\right), \tag{3.22}$$

where α, β, and ω are defined by (2.5), (2.6), and (3.1), respectively. The above results indicate that a_1, a_2, and a_3 are independent of index i. Moreover, it follows that

$$\begin{aligned}
&\frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \Pi_\omega\phi_{P_i}\,\mathrm{d}s = \frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \Pi_\omega\phi_{P_{i+1}}\,\mathrm{d}s = a_1 + a_2,\\
&\frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \lambda_i\Pi_\omega\phi_{M_i}\,\mathrm{d}s = \frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \lambda_{i+1}\Pi_\omega\phi_{M_i}\,\mathrm{d}s = \frac{1}{2} - a_1 - a_2,\\
&\frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \Pi_\omega\phi_{M_i}\,\mathrm{d}s = 1 - 2a_1 - 2a_2,\\
&\frac{1}{|K|}\int_K \Pi_\omega\phi_{M_i}\,\mathrm{d}x\mathrm{d}y = \frac{1}{3} - a_3.
\end{aligned} \tag{3.23}$$

4 The coercivity result for the equilateral triangular mesh

In this section, for any $\alpha \in (0, 1/2)$ and $\beta \in (0, 2/3)$, we assume that ω satisfies (3.2) and prove that the coercivity result of these schemes is valid on the equilateral triangular mesh.

Lemma 4.1 *If $\kappa = 1$ on K and K is an equilateral triangle, then for any given $\alpha \in (0, 1/2)$ and $\beta \in (0, 2/3)$, $\mathbb{B}_K$ is a positive definite matrix if and only if ω satisfies* (3.2).

Proof From (3.16), (3.4), Lemma 3.6, (3.17), (3.18), and (3.19), a direct calculation yields that

$$\mathbb{A}_K = \frac{1}{\sqrt{3}} \begin{pmatrix} b_1 & b_2 & b_2 & b_3 & b_4 & b_3 \\ b_2 & b_1 & b_2 & b_3 & b_3 & b_4 \\ b_2 & b_2 & b_1 & b_4 & b_3 & b_3 \\ b_5 & b_5 & b_6 & b_7 & b_8 & b_8 \\ b_6 & b_5 & b_5 & b_8 & b_7 & b_8 \\ b_5 & b_6 & b_5 & b_8 & b_8 & b_7 \end{pmatrix},$$

where

$$\begin{aligned} b_1 &= 6a_1 - 2a_2 - 4a_3, & b_2 &= a_1 + 5a_2 - 4a_3, \\ b_3 &= -4a_1 + 4a_2 + 4a_3, & b_4 &= -16a_2 + 4a_3, \\ b_5 &= -\frac{1}{3} - 2a_1 - 2a_2 + 4a_3, & b_6 &= \frac{2}{3} - 4a_1 - 4a_2 + 4a_3, \\ b_7 &= \frac{16}{3} - 8a_1 - 8a_2 - 4a_3, & b_8 &= -\frac{8}{3} + 8a_1 + 8a_2 - 4a_3. \end{aligned}$$

From the equality (3.14), still by straightforward calculations, we obtain that

$$\mathbb{C}^T \mathbb{B}_K \mathbb{C} = \frac{1}{\sqrt{3}} \mathbb{B}'_K,$$

where

$$\mathbb{C} = \begin{pmatrix} 0 & -1 & 0 & -1 & -1 \\ -1 & 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

and

$$\mathbb{B}'_K = \begin{pmatrix} 5-8a_1-8a_2 & 3-8a_1-8a_2 & 5a_1-7a_2-2 & \frac{5}{2}-4a_1-4a_2 & \frac{17}{6}-5a_1-9a_2 \\ 3-8a_1-8a_2 & 5-8a_1-8a_2 & \frac{5}{2}-4a_1-4a_2 & 5-13a_1-a_2 & \frac{31}{6}-11a_1-7a_2 \\ 5a_1-7a_2-2 & \frac{5}{2}-4a_1-4a_2 & 5-8a_1-8a_2 & 3-8a_1-8a_2 & \frac{17}{6}-5a_1-9a_2 \\ \frac{5}{2}-4a_1-4a_2 & 5-13a_1-a_2 & 3-8a_1-8a_2 & 5-8a_1-8a_2 & \frac{31}{6}-11a_1-7a_2 \\ \frac{17}{6}-5a_1-9a_2 & \frac{31}{6}-11a_1-7a_2 & \frac{17}{6}-5a_1-9a_2 & \frac{31}{6}-11a_1-7a_2 & \frac{16}{3}-8a_1-8a_2-4a_3 \end{pmatrix}.$$

Using *Matlab*, we find that

$$\det\left(\mathbb{B}'_K(1:1,1:1)\right) = 5-8(a_1+a_2) = 1+4\omega(1-2\alpha),$$
$$\det\left(\mathbb{B}'_K(1:2,1:2)\right) = 16-32(a_1+a_2) = 16\omega(1-2\alpha),$$
$$\det\left(\mathbb{B}'_K(1:3,1:3)\right) = 3\,(5-8a_1-8a_2)\,p_1(a_1,a_2) = 3\,(1+4\omega(1-2\alpha))\,p_1(a_1,a_2),$$
$$\det\left(\mathbb{B}'_K(1:4,1:4)\right) = 9\,(p_1(a_1,a_2))^2$$

and

$$\det\left(\mathbb{B}'_K\right) = 12(2a_1+2a_2-3a_3)\,(p_1(a_1,a_2))^2 = 12\omega\alpha(2-3\beta)\,(p_1(a_1,a_2))^2\,,$$

where

$$p_1(a_1,a_2) = -27a_1^2+18a_1a_2-3a_2^2+13a_1-15a_2-\frac{1}{12}.$$

From (3.20) and (3.21), we find that

$$p_1(a_1,a_2) = -3\left(\frac{1}{2}-\alpha\right)^2(\omega-\omega^-)(\omega-\omega^+),$$

Therefore, for any given $\alpha \in (0,1/2)$ and $\beta \in (0,2/3)$, $\mathbb{B}'_K$ is a positive definite matrix if and only if (3.2) holds. Since $\mathbb{C}$ is a nonsingular matrix, thus the proof is complete. □

From (3.2), we see that for any given $\alpha \in (0,1/2)$ and $\beta \in (0,2/3)$, there exists at least one ω such that $\mathbb{B}_K$ is a positive definite matrix. Thus, we have the following Theorem 4.1.

Theorem 4.1 *Assume that $\mathcal{T}_h$ consists of equilateral triangles, κ is piecewise constant with respect to $\mathcal{T}_h$ or alternatively, κ is piecewise $W^{1,\infty}$ with respect to $\mathcal{T}_h$, and the mesh size h is small enough. Then, for the scheme (2.7) with $\alpha \in (0,1/2)$ and $\beta \in (0,2/3)$, we have the coercivity result (3.5) with ω subjected to (3.2).*

Proof The proof is similar to that of Theorem 1 in [46] and we omit it here. □

5 The coercivity result for general triangular meshes

Throughout this section, for any $\alpha \in (0,1/2)$ and $\beta \in (0,2/3)$, we shall assume that ω satisfies (3.3). Thanks to (3.3), a certain weak orthogonality holds for any α.

Consequently, we are able to obtain a sufficient condition to ensure the coercivity result of the schemes on general triangular meshes.

5.1 The properties of $\mathbb{A}_K$ and $\mathbb{B}_K$

Lemma 5.1 *Assume that* $\mathbb{A}_K^{(1)} = (a_{ij}^{(1)})_{6\times 6}$ *is the element stiffness matrix of the standard quadratic finite element method, given by*

$$a_{ij}^{(1)} = \int_K \nabla\phi_{P_i} \cdot \nabla\phi_{P_j}\, dxdy. \tag{5.1}$$

Then, we have

$$\mathbb{A}_K^{(1)} = \frac{1}{6}\begin{pmatrix} 3(r_2+r_3) & r_3 & r_2 & -4r_3 & 0 & -4r_2 \\ r_3 & 3(r_1+r_3) & r_1 & -4r_3 & -4r_1 & 0 \\ r_2 & r_1 & 3(r_1+r_2) & 0 & -4r_1 & -4r_2 \\ -4r_3 & -4r_3 & 0 & 8(r_1+r_2+r_3) & -8r_2 & -8r_1 \\ 0 & -4r_1 & -4r_1 & -8r_2 & 8(r_1+r_2+r_3) & -8r_3 \\ -4r_2 & 0 & -4r_2 & -8r_1 & -8r_3 & 8(r_1+r_2+r_3) \end{pmatrix}, \tag{5.2}$$

where

$$r_i = \cot\theta_i, \quad i = 1, 2, 3 \tag{5.3}$$

and $\theta_i = \angle P_{i+2}P_iP_{i+1}$ $(i = 1, 2, 3)$ *are the three interior angles of* K.

Proof By (3.4), Lemma 3.6, and straightforward calculations, we can verify (5.2). □

Lemma 5.2 *For any* $\alpha \in (0, 1/2)$ *and* $\beta \in (0, 2/3)$*, if only if* ω *satisfies* (3.3)*, we have*

$$a_1 + a_2 = \frac{1}{6}, \tag{5.4}$$

$$\int_e (u_h - \Pi_\omega u_h)\, ds = 0, \quad \forall\, u_h \in U_h^2, \quad \forall\, e \in \mathcal{E}_h \tag{5.5}$$

and

$$\int_e v_h\, (\phi_{M'} - \Pi_\omega \phi_{M'})\, ds = 0, \quad \forall\, v_h \in U_h^1, \quad \forall\, e \in \mathcal{E}_h, \tag{5.6}$$

where U_h^1 *and* U_h^2 *are defined by* (2.4)*,* M' *is the midpoint of* $e' \in \mathcal{E}_h$*. Moreover,* (3.3) *implies* (3.2) *if and only if*

$$\alpha \in I_0 := \left(\frac{1}{2}\left(1 - \sqrt{\frac{4}{3}\sqrt{2} - 1}\right),\ \frac{1}{2}\right). \tag{5.7}$$

Proof From (3.20), (3.21), and (3.3), there holds

$$a_1 + a_2 = \frac{1}{2}(2\alpha\omega - \omega + 1) = \frac{1}{6},$$

which verifies (5.4). For each edge $e \in \mathcal{E}_h$, let its two vertices be P_1 and P_2, and the midpoint M. Thus, to verify (5.5), it suffices to verify the cases $u_h = \lambda_1$ and $\lambda_1\lambda_2$. From the definition of Π_ω, we see that

$$\Pi_\omega\lambda_1(P_1) = 1, \quad \Pi_\omega\lambda_1(P_2) = 0, \quad \Pi_\omega\lambda_1(M) = \frac{1}{2}.$$

Then,

$$\int_e \Pi_\omega\lambda_1 \, \mathrm{d}s = \alpha|e| + \frac{1}{2}(1-2\alpha)|e| = \frac{1}{2}|e| = \int_e \lambda_1 \, \mathrm{d}s.$$

As for $u_h = \lambda_1\lambda_2$, we have

$$\Pi_\omega u_h(P_1) = 0, \quad \Pi_\omega u_h(P_2) = 0, \quad \Pi_\omega u_h(M) = \frac{1}{4}\omega.$$

Hence, it follows from (3.3) that

$$\int_e \Pi_\omega u_h \, \mathrm{d}s = \frac{1}{4}\omega(1-2\alpha)|e| = \frac{1}{6}|e| = \int_e u_h \, \mathrm{d}s.$$

Next, to prove (5.6), it suffices to verify (5.6) for $v_h = \lambda_1$ and $M' = M$ (namely $\phi_{M'} = \phi_M$). Note that in this case

$$\int_e v_h\phi_M \, \mathrm{d}s = 4\int_e \lambda_1^2\lambda_2 \, \mathrm{d}s = \frac{1}{3}|e|.$$

From (3.23) and (5.4), we get

$$\int_e v_h \Pi_\omega\phi_M \, \mathrm{d}s = \left(\frac{1}{2} - a_1 - a_2\right)|e| = \frac{1}{3}|e|,$$

which verifies (5.6).

Finally, if ω given by (3.3) satisfies (3.2), we have

$$2\left(1 - \sqrt{3\alpha(1-\alpha)}\right)^2 < 1 < 2\left(1 + \sqrt{3\alpha(1-\alpha)}\right)^2,$$

which implies (5.7). The proof is complete. □

Lemma 5.3 *For the basis functions ϕ_{P_i} and ϕ_{M_i} defined in (3.4), we have*

$$\int_K \left(\phi_{P_i} - \Pi_\omega\phi_{P_i}\right) dxdy = -a_3\,|K|\,, \quad \int_K \left(\phi_{M_i} - \Pi_\omega\phi_{M_i}\right) dxdy = a_3\,|K|\,. \tag{5.8}$$

Proof A direct calculation yields that

$$\int_K \phi_{P_i} \, \mathrm{d}x\mathrm{d}y = 0, \quad \int_K \phi_{M_i} \, \mathrm{d}x\mathrm{d}y = \frac{1}{3}\,|K|\,, \quad i = 1, 2, 3.$$

It follows from (3.19) and (3.23) that

$$\int_K \Pi_\omega \phi_{P_i}\,\mathrm{d}x\mathrm{d}y = a_3\,|K|\,, \quad \int_K \Pi_\omega \phi_{M_i}\,\mathrm{d}x\mathrm{d}y = \left(\frac{1}{3} - a_3\right)|K|\,, \quad i = 1, 2, 3.$$

The desired equalities in (5.8) follow immediately. □

Lemma 5.4 *For the r_i defined in (5.3), we have*

$$r_1 r_2 + r_1 r_3 + r_2 r_3 = 1, \tag{5.9}$$

$$r_1^2 r_2^2 + r_1^2 r_3^2 + r_2^2 r_3^2 = 1 - 2 r_1 r_2 r_3 (r_1 + r_2 + r_3), \tag{5.10}$$

$$r_1^3 r_2 + r_1^3 r_3 + r_2^3 r_1 + r_2^3 r_3 + r_3^3 r_1 + r_3^3 r_2 = \left(r_1^2 + r_2^2 + r_3^2\right) - r_1 r_2 r_3 (r_1 + r_2 + r_3) \tag{5.11}$$

and

$$r_1 + r_2 + r_3 \geq \sqrt{3}. \tag{5.12}$$

Proof Note that

$$\tan\theta_3 = -\tan(\theta_1 + \theta_2) = -\frac{1 - \tan\theta_1 \tan\theta_2}{\tan\theta_1 + \tan\theta_2},$$

which leads to (5.9). (5.10) and (5.11) follow from the relations

$$r_1^2 r_2^2 + r_1^2 r_3^2 + r_2^2 r_3^2 = (r_1 r_2 + r_1 r_3 + r_2 r_3)^2 - 2 r_1 r_2 r_3 (r_1 + r_2 + r_3)$$

and

$$r_1^3 r_2 + r_1^3 r_3 + r_2^3 r_1 + r_2^3 r_3 + r_3^3 r_1 + r_3^3 r_2 = (r_1 r_2 + r_1 r_3 + r_2 r_3)\left(r_1^2 + r_2^2 + r_3^2\right) - r_1 r_2 r_3 (r_1 + r_2 + r_3),$$

respectively. The proof of (5.12) can be found in Lemma 6 of [46]. □

Lemma 5.5 *Assume that (3.3) holds and $\kappa = 1$ on K. Then, for $\mathbb{A}_K$ defined in (3.8) and (3.9), we have*

$$\mathbb{A}_K = \mathbb{A}_K^{(1)} + a_2 \mathbb{A}_K^{(2)} + a_3 \mathbb{A}_K^{(3)}, \tag{5.13}$$

where $\mathbb{A}_K^{(1)}$ is given by (5.2),

$$\mathbb{A}_K^{(2)} = 4 \begin{pmatrix} -r_2 - r_3 & r_1 & r_1 & r_2 + r_3 & -2r_1 - r_2 - r_3 & r_2 + r_3 \\ r_2 & -r_1 - r_3 & r_2 & r_1 + r_3 & r_1 + r_3 & -r_1 - 2r_2 - r_3 \\ r_3 & r_3 & -r_1 - r_2 & -r_1 - r_2 - 2r_3 & r_1 + r_2 & r_1 + r_2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{5.14}$$

$\mathbb{A}_K^{(3)} = \boldsymbol{\xi} \otimes \boldsymbol{\eta}$ *(i.e., $\boldsymbol{\xi}\boldsymbol{\eta}^T$) with*

$$\boldsymbol{\xi} = (-1, \quad -1, \quad -1, \quad 1, \quad 1, \quad 1)^T,$$
$$\boldsymbol{\eta} = 2\,(r_2 + r_3, \quad r_1 + r_3, \quad r_1 + r_2, \quad -2r_3, \quad -2r_1, \quad -2r_2)^T.$$

Consequently, for $\mathbb{B}_K$ defined in (3.14), we have

$$\det(\mathbb{B}_K) = \frac{16}{243}(r_1+r_2+r_3)p(r_1,r_2,r_3), \tag{5.15}$$

where

$$p(r_1,r_2,r_3) = A + Br_1r_2r_3(r_1+r_2+r_3) + C\left(r_1^2+r_2^2+r_3^2\right) \tag{5.16}$$

and

$$\begin{aligned} A &= 2C - (12a_2-1)(12a_2+9a_3-2)^2, \\ B &= -486a_2^2(4a_2-3a_3)^2, \\ C &= -9\left(18a_2^2+12a_2-1\right)\left(96a_2^2a_3-16a_2^2+8a_2a_3-3a_3^2\right). \end{aligned} \tag{5.17}$$

Proof It follows from (3.16) that

$$\begin{aligned} a_{ij} &= \int_{\partial K} \nabla\phi_{P_j}\cdot \boldsymbol{n}\Pi_\omega\phi_{P_i}\,\mathrm{d}s - \int_K \left(\Delta\phi_{P_j}\right)\Pi_\omega\phi_{P_i}\,\mathrm{d}x\mathrm{d}y \\ &= \int_{\partial K} \nabla\phi_{P_j}\cdot \boldsymbol{n}\phi_{P_i}\,\mathrm{d}s - \int_K \left(\Delta\phi_{P_j}\right)\phi_{P_i}\,\mathrm{d}x\mathrm{d}y + \int_{\partial K}\nabla\phi_{P_j}\cdot\boldsymbol{n}\left(\Pi_\omega\phi_{P_i}-\phi_{P_i}\right)\mathrm{d}s \\ &\quad - \int_K \left(\Delta\phi_{P_j}\right)\left(\Pi_\omega\phi_{P_i}-\phi_{P_i}\right)\mathrm{d}x\mathrm{d}y \\ &= \int_K \nabla\phi_{P_j}\cdot\nabla\phi_{P_i}\,\mathrm{d}x\mathrm{d}y + \int_{\partial K}\nabla\phi_{P_j}\cdot\boldsymbol{n}\left(\Pi_\omega\phi_{P_i}-\phi_{P_i}\right)\mathrm{d}s \\ &\quad - \left(\Delta\phi_{P_j}\right)\int_K \left(\Pi_\omega\phi_{P_i}-\phi_{P_i}\right)\mathrm{d}x\mathrm{d}y, \end{aligned}$$

where we have used Green's formula in the last equality. Consequently,

$$\mathbb{A}_K = \mathbb{A}_K^{(1)} + \widetilde{\mathbb{A}}_K^{(2)} + \widetilde{\boldsymbol{\xi}}\otimes\boldsymbol{\eta}, \quad \widetilde{\boldsymbol{\xi}} = \left(\widetilde{\xi}_i\right), \quad \boldsymbol{\eta} = (\eta_j),$$

where $\mathbb{A}_K^{(1)} = (a_{ij}^{(1)})_{6\times6}$ is defined in (5.1) and given by (5.2), $\widetilde{\mathbb{A}}_K^{(2)} = (\widetilde{a}_{ij}^{(2)})_{6\times6}$ is given by

$$\widetilde{a}_{ij}^{(2)} = \int_{\partial K}\nabla\phi_{P_j}\cdot\boldsymbol{n}\left(\Pi_\omega\phi_{P_i}-\phi_{P_i}\right)\mathrm{d}s,$$

and the last part is the tensor product of $\widetilde{\boldsymbol{\xi}}$ and $\boldsymbol{\eta}$, given by

$$\widetilde{\xi}_i = \frac{1}{|K|}\int_K\left(\phi_{P_i}-\Pi_\omega\phi_{P_i}\right)\mathrm{d}x\mathrm{d}y, \quad \eta_j = |K|\left(\Delta\phi_{P_j}\right), \quad i,j=1,\cdots,6.$$

From (5.8), (3.4), and Lemma 3.6, we obtain

$$\begin{aligned} \widetilde{\boldsymbol{\xi}} &= a_3(-1,\quad -1,\quad -1,\quad 1,\quad 1,\quad 1)^T = a_3\boldsymbol{\xi}, \\ \boldsymbol{\eta} &= 2(r_2+r_3,\quad r_1+r_3,\quad r_1+r_2,\quad -2r_3,\quad -2r_1,\quad -2r_2)^T. \end{aligned}$$

By direct calculations, we deduce from (3.4) that

$$\frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}}\lambda_i\phi_{P_i}\mathrm{d}s = \frac{1}{6}, \quad \frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}}\lambda_{i+1}\phi_{P_i}\mathrm{d}s = 0, \quad i=1,2,3.$$

Consequently, for $i = 1, 2, 3$, it follows from (5.4) that

$$\begin{aligned}\frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \lambda_i\left(\Pi_\omega\phi_{P_i} - \phi_{P_i}\right)\mathrm{d}s = -a_2,\\ \frac{1}{|P_iP_{i+1}|}\int_{P_iP_{i+1}} \lambda_{i+1}\left(\Pi_\omega\phi_{P_i} - \phi_{P_i}\right)\mathrm{d}s = a_2.\end{aligned} \tag{5.18}$$

By (3.4), Lemma 3.6, (5.5), and (5.18)

$$\begin{aligned}&\int_{P_1P_2} \nabla\phi_{P_1}\cdot\boldsymbol{n}_1\left(\Pi_\omega\phi_{P_1} - \phi_{P_1}\right)\,\mathrm{d}s\\ &= \int_{P_1P_2}(4\lambda_1 - 1)\nabla\lambda_1\cdot\left(-\frac{2|K|}{|P_1P_2|}\nabla\lambda_3\right)\left(\Pi_\omega\phi_{P_1} - \phi_{P_1}\right)\,\mathrm{d}s\\ &= \frac{4r_2}{|P_1P_2|}\int_{P_1P_2}\lambda_1\left(\Pi_\omega\phi_{P_1} - \phi_{P_1}\right)\,\mathrm{d}s\\ &= -4a_2r_2.\end{aligned}$$

Similarly,

$$\int_{P_1P_3} \nabla\phi_{P_1}\cdot\boldsymbol{n}_3\left(\Pi_\omega\phi_{P_1} - \phi_{P_1}\right)\,\mathrm{d}s = -4a_2r_3.$$

Therefore, we have

$$\widetilde{a}_{11}^{(2)} = \int_{\partial K}\nabla\phi_{P_1}\cdot\boldsymbol{n}\left(\Pi_\omega\phi_{P_1} - \phi_{P_1}\right)\,\mathrm{d}s = 4a_2(-r_2 - r_3).$$

By the same arguments,

$$\widetilde{a}_{12}^{(2)} = 4a_2r_1, \quad \widetilde{a}_{14}^{(2)} = 4a_2(r_2 + r_3), \quad \widetilde{a}_{15}^{(2)} = 4a_2(-2r_1 - r_2 - r_3).$$

By the symmetric property of the index r_i,

$$\widetilde{a}_{13}^{(2)} = \mathcal{R}\left(\widetilde{a}_{12}^{(2)},\ \{r_1, r_3, r_2\}\right), \quad \widetilde{a}_{16}^{(2)} = \mathcal{R}\left(\widetilde{a}_{14}^{(2)},\ \{r_1, r_3, r_2\}\right),$$

where $\mathcal{R}(a, \{r_i, r_j, r_k\})$ is an index replace function such that r_1, r_2, and r_3 in a are replaced by r_i, r_j, and r_k, respectively. Moreover, we have

$$\widetilde{\mathbb{A}}_K^{(2)}(2,:) = \mathcal{R}\left(\mathcal{P}\left(\widetilde{\mathbb{A}}_K^{(2)}(1,:),\ \{3,1,2,6,4,5\}\right),\ \{r_2, r_3, r_1\}\right)$$

and

$$\widetilde{\mathbb{A}}_K^{(2)}(3,:) = \mathcal{R}\left(\mathcal{P}\left(\widetilde{\mathbb{A}}_K^{(2)}(1,:),\ \{2,3,1,5,6,4\}\right),\ \{r_3, r_1, r_2\}\right),$$

where for a 1×6 vector $\mathbb{A}$, $\widehat{\mathbb{A}} = \mathcal{P}(\mathbb{A}, \{i_1, i_2, \cdots, i_6\})$ is a permutate function such that $\widehat{\mathbb{A}}(j) = \mathbb{A}(i_j)$, $j = 1, \cdots, 6$. It follows from (5.6) that

$$\widetilde{a}_{ij}^{(2)} = 0, \quad i \in \{4,5,6\},\ j \in \{1, \cdots, 6\}.$$

In other words, we obtain that $\widetilde{\mathbb{A}}_K^{(2)} = a_3\mathbb{A}_K^{(2)}$ with $\mathbb{A}_K^{(2)}$ defined in (5.14), and (5.13) is verified.

Finally, using *Matlab*, we get from (3.14) and (5.13) that

$$\det\left(\mathbb{B}_K\right) = \frac{16}{243}(r_1 + r_2 + r_3)\widehat{p}(r_1, r_2, r_3),$$

where

$$\widehat{p}(r_1,r_2,r_3) = \widehat{A}r_1r_2r_3(r_1+r_2+r_3)+\widehat{B}\left(r_1^2r_2^2+r_1^2r_3^2+r_2^2r_3^2\right)$$
$$+\widehat{C}\left(r_1^3r_2+r_1^3r_3+r_2^3r_1+r_2^3r_3+r_3^3r_1+r_3^3r_2\right) \tag{5.19}$$

and

$$\widehat{A}=2A+B+C, \quad \widehat{B}=A, \quad \widehat{C}=C, \tag{5.20}$$

and A, B, and C are defined in (5.17). Then from (5.19), (5.10), (5.11), and (5.20),

$$\widehat{p}(r_1,r_2,r_3)=A+Br_1r_2r_3(r_1+r_2+r_3)+C\left(r_1^2+r_2^2+r_3^2\right)=p(r_1,r_2,r_3),$$

where the polynomial p is defined by (5.16), thus (5.15) is proved and the proof is complete. □

5.2 The coercivity result

For any triangular element K, without loss of generality, we assume that $\theta_1 \le \theta_2 \le \theta_3$, then $\theta_1 \le \pi/3$, $r_1 \ge 1/\sqrt{3}$ and

$$r_1 \ge r_2 = \cot\theta_2 \ge \cot\left(\frac{\pi-\theta_1}{2}\right) = \tan\frac{\theta_1}{2} = \sqrt{1+r_1^2}-r_1. \tag{5.21}$$

In other words, (r_1,r_2,r_3) is defined in

$$\mathcal{D}=\left\{(r_1,r_2,r_3):\ r_1\ge\frac{1}{\sqrt{3}},\ \sqrt{1+r_1^2}-r_1\le r_2\le r_1,\ r_3=\frac{1-r_1r_2}{r_1+r_2}\right\}.$$

Then, we denote

$$\theta_{\min}=\operatorname{arccot}\sup\left\{r_1:\ p(r_1,r_2,r_3)>0,\ \forall\,(r_1,r_2,r_3)\in\mathcal{D}\right\}, \tag{5.22}$$

where the polynomial p is defined in (5.16), and suppose that θ_K is the minimum interior angle of K. In order to present the coercivity result, we introduce the following two geometric assumptions.

(**A1**) For any $K\in\mathcal{T}_h$, there holds

$$\theta_K>\theta_{\min}.$$

(**A2**) There exists a positive constant ε_0, independent of h, such that

$$\theta_K\ge\theta_{\min}+\varepsilon_0,\quad \forall\,K\in\mathcal{T}_h.$$

Thus, we have the following Lemma 5.6.

Lemma 5.6 *Assume that the diffusion coefficient κ is piecewise constant with respect to $\mathcal{T}_h$. For any $\alpha\in I_0$ (defined by* (5.7)*) and $\beta\in(0,2/3)$, let ω satisfy* (3.3)*. Then, for each $\mathbb{B}_K$ defined by* (3.14)*, we have the following results.*

(1) *If* (**A1**) *holds, then $\mathbb{B}_K$ is a positive definite matrix.*
(2) *Under the assumptions* (2.3) *and* (**A2**)*,*

$$\boldsymbol{u}^T\mathbb{B}_K\boldsymbol{u}\gtrsim\|\boldsymbol{u}\|^2,\quad\forall\,\boldsymbol{u}\in\mathbb{R}^5.$$

Proof Note that κ is piecewise constant with respect to $\mathcal{T}_h$, we can assume further that, without losing generality, $\kappa = 1$ on K. If (**A1**) holds, then it follows from (5.15), (5.12), and (5.22) that

$$\det(\mathbb{B}_K) > 0. \tag{5.23}$$

Moreover, we suppose that μ_i, $i = 1, \cdots, 5$ are the five eigenvalues of $\mathbb{B}_K$, then they satisfy the equation $F(r_1, r_2, r_3, \mu) := \det(\mu\mathbb{I} - \mathbb{B}_K) = 0$, where $\mathbb{I}$ is the unit matrix. Note that $F(r_1, r_2, r_3, \mu)$ is a polynomial about the four variables r_1, r_2, r_3, and μ, then $F(r_1, r_2, r_3, \mu)$ is smooth enough. Therefore, μ_i relies on r_1, r_2, and r_3 continuously. For any $\alpha \in I_0$ and $\beta \in (0, 2/3)$, from Lemma 5.2 and Lemma 4.1, we get that μ_i, $i = 1, \cdots, 5$ are all positive provided $r_1 = r_2 = r_3 = 1/\sqrt{3}$. On the other hand, we have $\det(\mathbb{B}_K) = \prod_{i=1}^{5} \mu_i$ for any K. Thus, from (5.23) we obtain that μ_i, $i = 1, \cdots, 5$ are all positive by view of continuity argument. That is, $\mathbb{B}_K$ is a positive definite matrix.

Finally, by (3.14), (5.13), Gershgorin disk theorem, and (2.3), we have, for the spectral radius of $\mathbb{B}_K$,

$$\rho\left(\mathbb{B}_K\right) \lesssim |r_1| + |r_2| + |r_3| \lesssim \cot\theta_0.$$

Under the assumption (**A2**), we obtain that there exists a positive constant C_{ε_0} such that

$$p(r_1, r_2, r_3) \geq C_{\varepsilon_0}.$$

Consequently, we deduce from (5.15) and (5.12) that

$$\boldsymbol{u}^T \mathbb{B}_K \boldsymbol{u} \geq \frac{16\sqrt{3}C_{\varepsilon_0}}{243\left[\rho\left(\mathbb{B}_K\right)\right]^4} \|\boldsymbol{u}\|^2 \gtrsim \|\boldsymbol{u}\|^2, \quad \forall \boldsymbol{u} \in \mathbb{R}^5,$$

and complete the proof. □

Theorem 5.1 *Assume that κ is piecewise constant with respect to $\mathcal{T}_h$ or alternatively, κ is piecewise $W^{1,\infty}$ with respect to $\mathcal{T}_h$ and the mesh size h is small enough. For any $\alpha \in I_0$ (defined by (5.7)) and $\beta \in (0, 2/3)$, let ω satisfy (3.3). Then, we have the following results.*

(1) *(Existence and uniqueness) Under the assumption* (**A1**), *(2.7) has a unique solution.*
(2) *(Coercivity) Under the assumptions (2.3) and* (**A2**), *the coercivity result (3.5) holds.*

Proof The proof is similar to that of Theorem 1 in [46] and we omit it here. □

6 Some analytic expressions for $\theta_{\min}$

In this section, for any $\alpha \in I_0$ (defined by (5.7)) and $\beta \in (0, 2/3)$, we are going to find some analytic expressions to approximate the minimum angle condition $\theta_{\min}$ which defined in (5.22). From (5.17), one can see that $B \leq 0$, and here we consider two cases: $A > 0$, $B = 0$, $C < 0$ and $A > 0$, $B < 0$, $C < 0$.

6.1 Case 1: $A > 0, B = 0, C < 0$

Lemma 6.1 *Let θ_i $(i = 1, 2, 3)$ be the three interior angles of a triangle, satisfying $\theta_1 \le \theta_2 \le \theta_3$. Then for a given r_1,*

$$g(r_2) = r_1 + r_2 + r_3$$

is a strictly increasing function with respect to r_2, where r_i is defined by (5.3).

Proof From (5.9), we have

$$g(r_2) = r_1 + r_2 + \frac{1 - r_1 r_2}{r_1 + r_2},$$

which yields that

$$\frac{\mathrm{d}g}{\mathrm{d}r_2} = \frac{r_2^2 + 2r_1 r_2 - 1}{(r_1 + r_2)^2}.$$

By (5.21), we deduce that

$$r_2^2 + 2r_1 r_2 - 1 \ge \tan^2\frac{\theta_1}{2} + 2\frac{\tan\frac{\theta_1}{2}}{\tan\theta_1} - 1 = \tan^2\frac{\theta_1}{2} + \left(1 - \tan^2\frac{\theta_1}{2}\right) - 1 = 0,$$

and the equality holds if and only if $r_2 = \tan(\theta_1/2)$, namely $\theta_2 = \theta_3$, which implies $g(r_2)$ is a strictly increasing function. □

Theorem 6.1 *Assume that $\theta_1 \le \theta_2 \le \theta_3$ and A, B, and C are defined in (5.17), subjected to $A > 0$, $B = 0$ and $C < 0$. If r_1 satisfies*

$$r_1 \in (D_1^-, D_1^+),$$

where

$$D_1^{\pm} = \frac{1}{3}\left(\sqrt{2 - \frac{A}{C}} \pm \sqrt{-1 - \frac{A}{C}}\right),$$

then we have

$$p(r_1, r_2, r_3) = A + C\left(r_1^2 + r_2^2 + r_3^2\right) > 0. \tag{6.1}$$

Consequently, if $D_1^- < 1/\sqrt{3}$, there holds

$$\theta_{\min} = \operatorname{arccot} D_1^+. \tag{6.2}$$

Proof If $A > 0$, $B = 0$, and $C < 0$, then from (5.9)

$$p(r_1, r_2, r_3) = A + C\left(r_1^2 + r_2^2 + r_3^2\right) = A - 2C + C(r_1 + r_2 + r_3)^2.$$

Thus, from Lemma 6.1, for any given r_1, the polynomial p attains its minimum at the point $(r_1, r_2, r_3) = (r_1, r_1, (1 - r_1^2)/(2r_1))$. Consequently, we obtain (6.1) provided

$$p(r_1, r_1, (1 - r_1^2)/(2r_1)) > 0,$$

which is equivalent to $r_1 \in (D_1^-, D_1^+)$. Finally, we get the minimum angle (6.2) provided $D_1^- < 1/\sqrt{3}$. □

6.2 Case 2: A> 0, B< 0, C< 0

Theorem 6.2 *Assume that $\theta_1 \leq \theta_2 \leq \theta_3$ and A, B, and C are defined in (5.17), subjected to $A > -C > -B/5 > 0$. If r_1 satisfies*

$$\max\left\{D_2^+, D_3^+, D_4\right\} < r_1^2 < \min\left\{-\frac{A+C}{B+C}, \frac{2C}{B}, D_2^-, D_3^-, D_5\right\}, \tag{6.3}$$

where

$$D_2^{\pm} = \frac{B+5C \pm \sqrt{(B+5C)^2 - 7B(2C-B)}}{7B}, \tag{6.4}$$

$$D_3^{\pm} = \frac{B+C-2A \pm \sqrt{(B+C-2A)^2 - (5C-B)(3B+5C)}}{5C-B}, \tag{6.5}$$

D_4 and D_5 are the smallest two roots of

$$-3Bx^3 + (2B+9C)x^2 + (4A+B-2C)x + C > 0, \tag{6.6}$$

and satisfies $D_4 < D_5$, then we have

$$p(r_1, r_2, r_3) = A + Br_1r_2r_3(r_1+r_2+r_3) + C\left(r_1^2 + r_2^2 + r_3^2\right) > 0. \tag{6.7}$$

Consequently, if

$$\max\left\{D_2^+, D_3^+, D_4\right\} < 1/3, \tag{6.8}$$

then

$$\theta_{\min} = arccot\left(\min\left\{-\frac{A+C}{B+C}, \frac{2C}{B}, D_2^-, D_3^-, D_5\right\}\right)^{1/2}. \tag{6.9}$$

Proof We first claim that

$$r_1 \geq r_2 \geq |r_3|. \tag{6.10}$$

For the case where $\theta_3 \leq \pi/2$, namely $r_3 \geq 0$, by (6.10) and (5.9), we have

$$p(r_1, r_2, r_3) \geq A + Br_1^2(r_1r_3 + r_2r_3 + r_1r_2) + C(r_1^2 + r_1r_2 + r_2r_3 + r_1r_3) = A + C + (B+C)r_1^2,$$

which verifies (6.7) by recalling (6.3).

Consider the case where $\theta_3 > \pi/2$, namely $r_3 < 0$, then we have $\theta_1 = \pi - \theta_2 - \theta_3 < \pi/2 - \theta_1$, i.e., $r_1 > 1$. Let

$$\sigma = \frac{r_2}{r_1}, \quad \tau = -\frac{r_3}{r_1}. \tag{6.11}$$

It follows that

$$\begin{aligned} p(r_1, r_2, r_3) &= A - Br_1^4\sigma\tau(1+\sigma-\tau) + Cr_1^2\left(1+\sigma^2+\tau^2\right) \\ &= r_1^2\left(C - Br_1^2\tau\right)\sigma^2 - Br_1^4\tau(1-\tau)\sigma + A + Cr_1^2\left(1+\tau^2\right) \\ &\triangleq q(\sigma, \tau). \end{aligned} \tag{6.12}$$

Note that

$$r_2 = \cot\theta_2 = \cot(\pi - \theta_1 - \theta_3) > \cot(\pi/2 - \theta_1) = \tan\theta_1 = \frac{1}{r_1},$$

and from (6.11) and (5.9)

$$\tau = \frac{r_1 r_2 - 1}{r_1 r_2 + r_1^2} = \frac{\sigma - \frac{1}{r_1^2}}{\sigma + 1} = 1 - \frac{1 + \frac{1}{r_1^2}}{1 + \sigma}, \tag{6.13}$$

which implies that for any given r_1, $q(\sigma, \tau)$ is defined in the curve

$$\mathcal{C} = \left\{ (\sigma, \tau) : \frac{1}{r_1^2} < \sigma \leq 1,\ \tau = 1 - \frac{1 + \frac{1}{r_1^2}}{1 + \sigma} \right\}.$$

From (6.13), it is easy to verify that τ is a strictly increasing and upper convex function with respect to σ. Thus, for any given r_1, the curve $\mathcal{C}$ is contained in the right triangle $\widetilde{K} = \triangle \widetilde{P}_1 \widetilde{P}_2 \widetilde{P}_3$ (see Fig. 3a) with

$$\widetilde{P}_1 = \left(\frac{1}{r_1^2},\ 0 \right), \quad \widetilde{P}_2 = \left(1,\ \frac{1}{2} - \frac{1}{2r_1^2} \right), \quad \widetilde{P}_3 = \left(\frac{1}{r_1^2},\ \frac{1}{2} - \frac{1}{2r_1^2} \right).$$

Moreover, we have

$$0 < \tau \leq \frac{1}{2} - \frac{1}{2r_1^2} < \frac{1}{2}.$$

Recalling $B < 0$ and (6.3), we deduce that

$$C - Br_1^2 \tau < C - \frac{B}{2} r_1^2 < 0,$$

which implies that, for a fixed τ, $q(\sigma, \tau)$ is a quadratic function with respect to σ, opening downward. As a result,

$$\min_{(\sigma,\tau)\in\mathcal{C}} q(\sigma, \tau) \geq \min_{(\sigma,\tau)\in\widetilde{K}} q(\sigma, \tau) = \min_{(\sigma,\tau)\in\widetilde{P}_1\widetilde{P}_3 \cup \widetilde{P}_1\widetilde{P}_2} q(\sigma, \tau). \tag{6.14}$$

On the line segment $\widetilde{P}_1 \widetilde{P}_3$, namely $\sigma = 1/r_1^2$, there holds

$$q(1/r_1^2, \tau) = r_1^2 (B + C)\tau^2 - B\left(1 + r_1^2\right)\tau + A + Cr_1^2 + \frac{C}{r_1^2}.$$

Obviously, it is a quadratic function of τ and opens downward, which implies that

$$\min_{(\sigma,\tau)\in\widetilde{P}_1\widetilde{P}_3} q(\sigma, \tau) = \min_{(\sigma,\tau)\in\{\widetilde{P}_1, \widetilde{P}_3\}} q(\sigma, \tau). \tag{6.15}$$

On the line segment $\widetilde{P}_1 \widetilde{P}_2$, we have the relationship

$$\tau = \frac{1}{2}\left(\sigma - \frac{1}{r_1^2}\right)$$

and

$$q(\sigma, \tau) = -\frac{1}{4}Br_1^4\sigma^3 + \frac{1}{4}r_1^2\left(5C - 2Br_1^2\right)\sigma^2 + \frac{1}{4}\left(B - 2C + 2Br_1^2\right)\sigma + A + Cr_1^2 + \frac{C}{4r_1^2},$$

which implies that

$$\frac{\mathrm{d}q}{\mathrm{d}\sigma} = -\frac{3}{4}Br_1^4\sigma^2 + \frac{1}{2}r_1^2\left(5C - 2Br_1^2\right)\sigma + \frac{1}{4}\left(B - 2C + 2Br_1^2\right).$$

Note that $B < 0$, $r_1 > 1$ and using (6.3) once again,

$$\frac{\mathrm{d}q}{\mathrm{d}\sigma}\Big|_{\sigma=1/r_1^2} = \frac{1}{2}\left(4C - B - Br_1^2\right) < \frac{1}{2}\left(4C - Br_1^2 - Br_1^2\right) = 2C - Br_1^2 < 0$$

and

$$\frac{\mathrm{d}q}{\mathrm{d}\sigma}\Big|_{\sigma=1} = -\frac{1}{4}\left(7Br_1^4 - 2(B+5C)r_1^2 + (2C-B)\right) = -\frac{7}{4}B\left(r_1^2 - D_2^+\right)\left(r_1^2 - D_2^-\right) < 0,$$

where $D_2^\pm$ is defined in (6.4). It follows that

$$\frac{\mathrm{d}q}{\mathrm{d}\sigma}\Big|_{(\sigma,\tau)\in\widetilde{P}_1\widetilde{P}_2} < 0,$$

which implies that $p(\sigma, (\sigma - 1/r_1^2)/2)$ is a strictly decreasing function of σ when $1/r_1^2 < \sigma \le 1$. Consequently,

$$\min_{(\sigma,\tau)\in\widetilde{P}_1\widetilde{P}_2} q(\sigma,\tau) = \min_{(\sigma,\tau)=\widetilde{P}_2} q(\sigma,\tau)$$

and combining the facts (6.14) and (6.15)

$$\min_{(\sigma,\tau)\in\widetilde{K}} q(\sigma,\tau) = \min_{(\sigma,\tau)=\{\widetilde{P}_2,\widetilde{P}_3\}} q(\sigma,\tau).$$

At the point $\widetilde{P}_3$, note that $5C - B < 0$ and (6.3), we find that

$$\begin{aligned} q\left(\frac{1}{r_1^2}, \frac{1}{2} - \frac{1}{2r_1^2}\right) &= \frac{1}{4r_1^2}\left((5C - B)r_1^4 + 2(2A - B - C)r_1^2 + (3B + 5C)\right) \\ &= \frac{5C - B}{4r_1^2}\left(r_1^2 - D_3^+\right)\left(r_1^2 - D_3^-\right) > 0, \end{aligned}$$

where $D_3^\pm$ is defined in (6.5). At the point $\widetilde{P}_2$,

$$q\left(1, \frac{1}{2} - \frac{1}{2r_1^2}\right) = \frac{1}{4r_1^2}\left(-3Br_1^6 + (2B + 9C)r_1^4 + (4A + B - 2C)r_1^2 + C\right) > 0,$$

where the last inequality is obtained by (6.3) and (6.6). Recalling (6.14), we have

$$\min_{(\sigma,\tau)\in\mathcal{C}} q(\sigma,\tau) > 0.$$

Note that (6.12), then (6.7) is verified. The minimum angle (6.9) follows from (6.3) and (6.8) immediately, and completes the proof. □

Remark 6.1 In Theorem 6.2, for any given r_1, the polynomial p is defined in the curve $\mathcal{C}$. In order to analyze the property of p more easily, here we study its property in a right triangle $\widetilde{K} = \triangle\widetilde{P}_1\widetilde{P}_2\widetilde{P}_3$ which contained $\mathcal{C}$, see Fig. 3a. In fact, if the

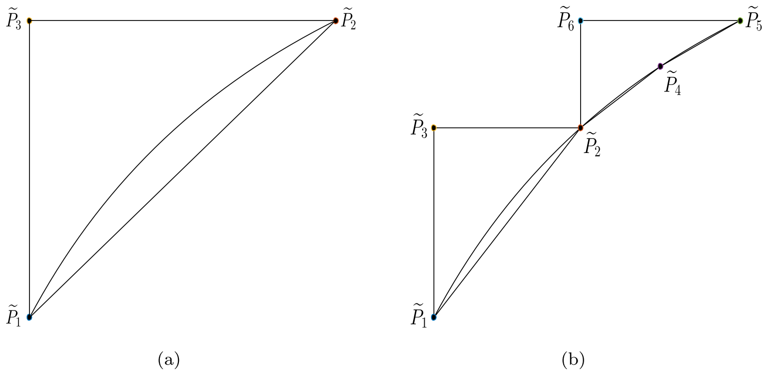

Fig. 3 The curve $\mathcal{C}$ and its associated region **(a)** a right triangle to approximate $\mathcal{C}$ **(b)** a smaller region to approximate $\mathcal{C}$

minimum angle (6.9) is not the optimal and $\cot\theta_{\min} > \sqrt{2}$, then one can choose other appropriate smaller region instead of $\widetilde{K}$, e.g., see Fig. 3b, where

$$\widetilde{P}_1 = \left(\tfrac{1}{r_1^2},\, 0\right), \quad \widetilde{P}_2 = \left(\tfrac{1}{2},\, \tfrac{1}{3} - \tfrac{2}{3r_1^2}\right), \quad \widetilde{P}_3 = \left(\tfrac{1}{r_1^2},\, \tfrac{1}{3} - \tfrac{2}{3r_1^2}\right),$$
$$\widetilde{P}_4 = \left(\tfrac{3}{4},\, \tfrac{3}{7} - \tfrac{4}{7r_1^2}\right), \quad \widetilde{P}_5 = \left(1,\, \tfrac{1}{2} - \tfrac{1}{2r_1^2}\right), \quad \widetilde{P}_6 = \left(\tfrac{1}{2},\, \tfrac{1}{2} - \tfrac{1}{2r_1^2}\right).$$

7 Discussions of minimum angle for some existing schemes

In this section, we will discuss the minimum angle condition for some special schemes. Precisely, we have given the analytic expressions of these minimum angle conditions, and improved some existing minimum angle conditions.

7.1 The family of schemes with $\alpha = (1 - 1/\sqrt{3})/2$ and $\beta \in (0, 2/3)$

Theorem 7.1 *Let $\alpha = \beta = (1 - 1/\sqrt{3})/2$, then the minimum angle $\theta_{\min}$ in (**A1**) can be expressed as*

$$\theta_{\min} = 0.$$

Proof Note that $\alpha \in I_0$ (defined by (5.7)), then from (3.3), (3.21), and (3.22), we get

$$\omega = \frac{2}{\sqrt{3}}, \quad a_2 = 0, \quad a_3 = 0.$$

By (5.17)

$$A = 4, \quad B = 0, \quad C = 0,$$

which implies that $p(r_1, r_2, r_3) = 4 > 0$ and $\theta_{\min} = 0$. □

Remark 7.1 Theorem 7.1 indicates that when $\alpha = \beta = (1 - 1/\sqrt{3})/2$, (2.7) leads to an unconditionally stable quadratic scheme on shape regular mesh $\mathcal{T}_h$, which is consistent with the result in [49].

Theorem 7.2 *Let $\alpha = (1 - 1/\sqrt{3})/2$, then for any $\beta \neq \alpha$, the minimum angle $\theta_{\min}$ in (**A1**) can be expressed as*

$$\theta_{\min} = arccot\left(\frac{1}{3}\left(\sqrt{2 - \frac{A}{C}} + \sqrt{-1 - \frac{A}{C}}\right)\right), \tag{7.1}$$

where

$$A = 2C + \left(9\widetilde{\beta} - 2\right)^2, \quad C = -27\widetilde{\beta}^2, \quad \widetilde{\beta} = \frac{2}{3}(1 - 3\alpha)(\beta - \alpha). \tag{7.2}$$

Proof Since $\alpha \in I_0$, then from (3.3), (3.21), and (3.22), we find that

$$\omega = \frac{2}{\sqrt{3}}, \quad a_2 = 0, \quad a_3 = \frac{1}{\sqrt{3}}\left(1 - \frac{1}{\sqrt{3}}\right)\beta + \frac{1}{3}\left(1 - \frac{2}{\sqrt{3}}\right) = \widetilde{\beta}.$$

By (5.17)

$$A = 2C + \left(9\widetilde{\beta} - 2\right)^2, \quad B = 0, \quad C = -27\widetilde{\beta}^2.$$

It follows from (7.2) that

$$\widetilde{\beta} = \frac{\sqrt{3}-1}{3}\left(\beta - \frac{3-\sqrt{3}}{6}\right) \in \left(-\frac{2\sqrt{3}-3}{9},\ 0\right) \cup \left(0,\ \frac{1}{9}\right), \quad \forall \beta \in (0, \alpha) \cup \left(\alpha,\ \frac{2}{3}\right),$$

which implies that $C < 0$ and $A = 27(\widetilde{\beta} - 2/3)^2 - 8 > 0$. Moreover, we have

$$\frac{A}{C} = 2 - \frac{1}{27}\left(\frac{2}{\widetilde{\beta}} - 9\right)^2 < -1,$$

which yields that

$$D_1^- = \frac{1}{3}\left(\sqrt{2 - \frac{A}{C}} - \sqrt{-1 - \frac{A}{C}}\right) = \frac{1}{\sqrt{2 - \frac{A}{C}} + \sqrt{-1 - \frac{A}{C}}} < \frac{1}{\sqrt{3}}.$$

Recalling the Theorem 6.1, then we complete the proof. □

Remark 7.2 By a simple calculation, the minimum angle (7.1) is the same as (18) in [46]. Consequently, for the quadratic scheme proposed in [36], here the minimum angle is the same as [46], namely 1.42°.

7.2 The scheme $\alpha = 1/4$ and $\beta = 1/3$

This scheme was proposed in [25], we improved the minimum angle 9.98° in [8, 40] to 4.14°.

Theorem 7.3 *Let $\alpha = 1/4$ and $\beta = 1/3$, then the minimum angle $\theta_{\min}$ in (**A1**) can be expressed as*

$$\theta_{\min} = arccot \left(\sqrt{\frac{575}{3}} \right) \approx 4.14°. \tag{7.3}$$

Proof Note that $\alpha \in I_0$, then from (3.3), (3.21), and (3.22), we obtain

$$\omega = \frac{4}{3}, \quad a_2 = -\frac{1}{96}, \quad a_3 = 0.$$

By (5.17)

$$A = \frac{82657}{16384}, \quad B = -\frac{3}{32768}, \quad C = -\frac{575}{32768}.$$

Then, the polynomial in (6.6) can be expressed as

$$\frac{1}{32768}(3x - 575)\left(3x^2 - 1152x + 1\right).$$

Thus, we have

$$D_4 = \frac{576 - \sqrt{331773}}{3}, \quad D_5 = \frac{575}{3}$$

and

$$\min \left\{ -\frac{A+C}{B+C}, \frac{2C}{B}, D_2^-, D_3^-, D_5 \right\} = D_5.$$

Note that

$$\max \left\{ D_2^+, D_3^+, D_4 \right\} = D_2^+ < 1/3$$

and Theorem 6.2, we get the desired result (7.3). □

7.3 The scheme $\alpha = 1/6$ and $\beta = 1/4$

For the scheme $\alpha = 1/6$ and $\beta = 1/4$ proposed in [15], we find that the minimum angle is 7.11°, which is the same as [8, 40].

Theorem 7.4 *Let $\alpha = 1/6$ and $\beta = 1/4$, then the minimum angle $\theta_{\min}$ in (**A1**) can be expressed as*

$$\theta_{\min} = arccot \left(\sqrt{\frac{161 + 2\sqrt{6479}}{5}} \right) \approx 7.11°. \tag{7.4}$$

Proof Note that $\alpha \in I_0$, then from (3.3), (3.21), and (3.22), we have

$$\omega = 1, \quad a_2 = \frac{1}{72}, \quad a_3 = \frac{1}{24}.$$

From (5.17)

$$A = \frac{15935}{9216}, \quad B = -\frac{25}{55296}, \quad C = -\frac{1195}{55296}.$$

Then, the polynomial in (6.6) can be expressed as

$$\frac{5}{55296}(3x-239)\left(5x^2-322x+1\right),$$

which implies that

$$D_4=\frac{161-2\sqrt{6479}}{5}, \quad D_5=\frac{161+2\sqrt{6479}}{5}$$

and

$$\min\left\{-\frac{A+C}{B+C}, \frac{2C}{B}, D_2^-, D_3^-, D_5\right\}=D_5.$$

Note that

$$\max\left\{D_2^+, D_3^+, D_4\right\}=D_2^+<1/3$$

and Theorem 6.2, the minimum angle (7.4) is obtained. □

7.4 The scheme $\alpha=\beta=1/3$

This scheme was proposed in [31], we improved the minimum angle 20.95° in [8, 40] to 10.08°. The result is given below.

Theorem 7.5 *Let $\alpha=\beta=1/3$, then the minimum angle $\theta_{\min}$ in (**A1**) can be expressed as*

$$\theta_{\min}=arccot\left(\sqrt{\frac{95}{3}}\right)\approx 10.08°. \tag{7.5}$$

Proof Since $\alpha\in I_0$, then from (3.3), (3.21), and (3.22), we deduce that

$$\omega=2, \quad a_2=-\frac{1}{36}, \quad a_3=-\frac{1}{9}.$$

By (5.17)

$$A=\frac{3410}{243}, \quad B=-\frac{1}{54}, \quad C=-\frac{95}{243}.$$

Then, the polynomial in (6.6) can be expressed as

$$\frac{1}{486}(3x-95)\left(9x^2-291x+2\right).$$

It follows that

$$D_4=\frac{97-\sqrt{9401}}{6}, \quad D_5=\frac{95}{3}$$

and

$$\min\left\{-\frac{A+C}{B+C}, \frac{2C}{B}, D_2^-, D_3^-, D_5\right\}=D_3^-.$$

Note that

$$\max\left\{D_2^+, D_3^+, D_4\right\}=D_2^+<1/3$$

and Theorem 6.2, we have

$$\theta_{\min}=\operatorname{arccot}\sqrt{D_3^-}\approx 10.46°.$$

Finally, we get the desired result (7.5) by the analysis in Fig. 3b of Remark 6.1. □

8 H^1 error estimates

Theorem 8.1 *Assume that $\mathcal{T}_h$ is shape regular and κ is piecewise $W^{1,\infty}$ with respect to $\mathcal{T}_h$. Assume also that the exact solution $u \in H_0^1(\Omega) \cap H^3(\Omega)$. Then, for any $\alpha \in I_0$ (defined by (5.7)) and $\beta \in (0, 2/3)$, under the assumption* (**A2**)*, we have*

$$|u - u_h|_1 \lesssim h^2 \|u\|_3. \tag{8.1}$$

Proof The proof is similar to that of Theorem 2 in [46] and we omit it here. □

Remark 8.1 If $\mathcal{T}_h$ consists of equilateral triangles, then it follows from Theorem 4.1 that the optimal H^1 error estimates (8.1) hold for any $\alpha \in (0, 1/2)$ and $\beta \in (0, 2/3)$.

9 Conclusions

This paper provides a general framework for the coercivity analysis of a class of quadratic FVE schemes on triangular meshes. This class of schemes have two parameters $\alpha \in (0, 1/2)$ and $\beta \in (0, 2/3)$, which cover all the existing quadratic schemes of Lagrange type. By the element analysis and a novel mapping from the trial function space to the test function space, we obtain the geometry assumption (**A1**) (resp. (**A2**)) that ensures the existence and uniqueness (resp. the coercivity result) of these schemes. Moreover, we give some minimum angle conditions with simple, analytic, and computable expressions. By these results, the minimum angle conditions for some existing schemes are improved, which is summarized in Table 1.

Acknowledgments The authors would like to thank the reviewers for their careful readings and valuable suggestions.

Funding information This work was partially supported by the National Natural Science Foundation of China (No. 11871009), CAEP Foundation (No. CX2019028), and Guangdong Natural Science Foundation (No. 2017B030311001).

References

1. Bank, R.E., Rose, D.J.: Some error estimates for the box method. SIAM J. Numer. Anal. **24**, 777–787 (1987)
2. Barth, T., Ohlberger, M.: Finite volume methods: foundation and analysis. In: Encyclopedia of Computational Mechanics, vol. 1, chapter 15. Wiley (2004)
3. Bush, L., Ginting, V.: On the application of the continuous Galerkin finite element method for conservation problems. SIAM J. Sci. Comput. **35**, A2953–A2975 (2013)
4. Cai, Z.: On the finite volume element method. Numer. Math. **58**, 713–735 (1991)
5. Chen, L.: A new class of high order finite volume methods for second order elliptic equations. SIAM J. Numer. Anal. **47**, 4021–4043 (2010)

6. Chen, Y., Li, Y.: Optimal bicubic finite volume methods on quadrilateral meshes. Adv. Appl. Math. Mech. **7**, 454–471 (2015)
7. Chen, Z.: A generalized difference method for the equations of heat conduction. Acta Sci. Natur. Univ. Sunyatseni **29**, 6–13 (1990)
8. Chen, Z., Wu, J., Xu, Y.: Higher-order finite volume methods for elliptic boundary value problems. Adv. Comput. Math. **37**, 191–253 (2012)
9. Chen, Z., Xu, Y., Zhang, J.: A second-order hybrid finite volume method for solving the Stokes equation. Appl. Numer. Math. **119**, 213–224 (2017)
10. Chen, Z., Xu, Y., Zhang, Y.: Higher-order finite volume methods II: Inf-sup condition and uniform local ellipticity. J. Comput. Appl. Math. **265**, 96–109 (2014)
11. Chen, Z., Xu, Y., Zhang, Y.: A construction of higher-order finite volume methods. Math. Comp. **84**, 599–628 (2015)
12. Chou, S., Li, Q.: Error estimates in L^2, H^1 and L^∞ in covolume methods for elliptic and parabolic problems: a unified approach. Math. Comp. **69**, 103–120 (2000)
13. Chou, S., Ye, X.: Unified analysis of finite volume methods for second order elliptic problems. SIAM J. Numer. Anal. **45**, 1639–1653 (2007)
14. Du, Y., Li, Y., Sheng, Z.: Quadratic finite volume method for a nonlinear elliptic problem. Adv. Appl. Math. Mech. **11**, 838–869 (2019)
15. Emonot, P.h.: Methodes de volumes elements finis: applications aux equations de Navier-Stokes et resultats de convergence. Dissertation, (1992)
16. Erath, C., Praetorius, D.: Adaptive vertex-centered finite volume methods with convergence rates. SIAM J. Numer. Anal. **54**, 2228–2255 (2016)
17. Erath, C., Praetorius, D.: Adaptive vertex-centered finite volume methods for general second-order linear elliptic partial differential equations. IMA J. Numer. Anal. **39**, 983–1008 (2019)
18. Ewing, R.E., Lin, T., Lin, Y.: On the accuracy of the finite volume element method based on piecewise linear polynomials. SIAM J. Numer. Anal. **39**, 1865–1888 (2002)
19. Hackbusch, W.: On first and second order box schemes. Computing **41**, 277–296 (1989)
20. Hong, Q., Wu, J.: Coercivity results of a modified Q_1-finite volume element scheme for anisotropic diffusion problems. Adv. Comput. Math. **44**, 897–922 (2018)
21. Jin, G., Li, H., Zhang, Q., Zou, Q.: Linear and quadratic finite volume methods on triangular meshes for elliptic equations with singular solutions. Int. J. Numer. Anal. Mod. **13**, 244–264 (2016)
22. LeVeque, R.J.: Finite Volume Methods for Hyperbolic Problems. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge (2002)
23. Li, R., Chen, Z., Wu, W.: The Generalized Difference Methods for Partial Differential Equations: Numerical Analysis of Finite Volume Methods. Marcel Dikker, New York (2000)
24. Li, Y., Li, R.: Generalized difference methods on arbitrary quadrilateral networks. J. Comput. Math. **17**, 653–672 (1999)
25. Liebau, F.: The finite volume element method with quadratic basis functions. Computing **57**, 281–299 (1996)
26. Lin, Y., Liu, J., Yang, M.: Finite volume element methods: an overview on recent developments. Int. J. Numer. Anal. Mod. **4**, 14–34 (2013)
27. Lv, J., Li, Y.: Optimal biquadratic finite volume element methods on quadrilateral meshes. SIAM J. Numer. Anal. **50**, 2379–2399 (2012)
28. Moukalled, F., Mangani, L., Darwish, M.: The Finite Volume Method in Computational Fluid Dynamics: an Advanced Introduction with OpenFoam and Matlab. Springer, Switherland (2016)
29. Petrila, T., Trif, D.: Basics of Fluid Mechanics and Introduction to Computational Fluid Dynamics. Springer, Berlin (2005)
30. Schmidt, T.: Box schemes on quadrilateral meshes. Computing **51**, 271–292 (1993)
31. Tian, M., Chen, Z.: Quadratic element generalized differential methods for elliptic equations. Numer. Math. J. Chin. Univ. **13**, 99–113 (1991)
32. Versteeg, H.K., Malalasekra, W. An Introduction to Computational Fluid Dynamics: the Finite Volume Method, 2nd edn. Pearson Education, England (2007)
33. Vogel, A., Xu, J., Wittum, G.: A generalization of the vertex-centered finite volume scheme to arbitrary high order. Comput. Visual. Sci. **13**, 221–228 (2010)
34. Wang, P., Zhang, Z.: Quadratic finite volume element method for the air pollution model. Int. J. Comput. Math. **87**, 2925–2944 (2010)
35. Wang, X., Huang, W., Li, Y.: Conditioning of the finite volume element method for diffusion problems with general simplicial meshes. Math. Comp. **88**, 2665–2696 (2019)

36. Wang, X., Li, Y.: L^2 error estimates for high order finite volume methods on triangular meshes. SIAM J. Numer. Anal. **54**, 2729–2749 (2016)
37. Wang, X., Li, Y.: Superconvergence of quadratic finite volume method on triangular meshes. J. Comput. Appl. Math. **348**, 181–199 (2019)
38. Wu, H., Li, R.: Error estimates for finite volume element methods for general second-order elliptic problems. Numer. Meth. PDEs **19**, 693–708 (2003)
39. Xiong, Z., Deng, K.: A quadratic triangular finite volume element method for a semilinear elliptic equation. Adv. Appl. Math. Mech. **9**, 186–204 (2017)
40. Xu, J., Zou, Q.: Analysis of linear and quadratic simplicial finite volume methods for elliptic equations. Numer. Math. **111**, 469–492 (2009)
41. Yang, M.: Quadratic finite volume element methods for nonlinear parabolic equations. Numer. Math. J. Chin. Univ. **26**, 257–266 (2004)
42. Yang, M.: Error estimates of quadratic finite volume element methods for nonlinear parabolic systems. Acta Math., Appl. Sin. **29**, 29–38 (2006)
43. Yang, M.: A second-order finite volume element method on quadrilateral meshes for elliptic equations. ESAIM: M2AN **40**, 1053–1067 (2006)
44. Zhang, Z., Zou, Q.: Some recent advances on vertex centered finite volume element methods for elliptic equations. Sci. China Math. **56**, 2507–2522 (2013)
45. Zhang, Z., Zou, Q.: Vertex-centered finite volume schemes of any order over quadrilateral meshes for elliptic boundary value problems. Numer. Math. **130**, 363–393 (2015)
46. Zhou, Y., Wu, J.: A family of quadratic finite volume element schemes over triangular meshes for elliptic equations. Comput. Math. Appl. **79**, 2473–2491 (2020)
47. Zhou, Y., Zou, Q.: A novel adaptive finite volume method for elliptic equations. Int. J. Numer. Anal. Mod. **14**, 879–892 (2017)
48. Zhu, P., Li, R.: Generalized difference methods for second order elliptic partial differential equations. II. Quadrilateral subdivision. Numer. Math. J. Chin. Univ. **4**, 360–375 (1982)
49. Zou, Q.: An unconditionally stable quadratic finite volume scheme over triangular meshes for elliptic equations. J. Sci. Comput. **70**, 112–124 (2017)
50. Zou, Q., Guo, L., Deng, Q.: High order continuous local-conserving fluxes and finite-volume-like finite element solutions for elliptic equations. SIAM J. Numer. Anal. **55**, 2666–2686 (2017)