# A CONSTRUCTION OF HIGHER-ORDER FINITE VOLUME METHODS

## ZHONGYING CHEN, YUESHENG XU, AND YUANYUAN ZHANG

ABSTRACT. We provide a method for the construction of higher-order finite volume methods (FVMs) for solving boundary value problems of the two dimensional elliptic equations. Specifically, when the trial space of the FVM is chosen to be a conforming triangle mesh finite element space, we describe a construction of the associated test space that guarantees the uniform localellipticity of the family of the resulting discrete bilinear forms. We show that the uniform local-ellipticity ensures that the resulting FVM has a unique solution which enjoys an optimal error estimate. We characterize the uniform localellipticity in terms of the uniform boundedness (below by a positive constant) of the smallest eigenvalues of the matrices associated with the FVMs. We then translate the characterization to equivalent requirements on the shapes of the triangle meshes for the trial spaces. Four convenient sufficient conditions for the family of the discrete bilinear forms to be uniformly local-elliptic are derived from the characterization. Following the general procedure, we construct four specific FVMs which satisfy the uniform local-ellipticity. Numerical results are presented to verify the theoretical results on the convergence order of the FVMs.

# 1. INTRODUCTION

This paper continues the general theme of the recent paper [12] in studying higher-order FVMs for solving the elliptic boundary value problems, with the focus on the general construction of the test spaces. In [12], a theoretical framework was developed for convergence analysis of FVMs, establishing the uniform boundedness and the uniform ellipticity of the discrete bilinear forms for the methods and showing that they lead to the optimal error estimate of the methods in the  $H^1$ -norm, and several specific constructions of the methods were provided. Specifically, the uniform boundedness estimate of the discrete bilinear forms of FVMs was given in the form different from finite element methods (FEMs), and a systematic study of mesh geometric requirements was provided, which guarantee the uniform ellipticity of the discrete bilinear forms. Although paper [12] provided a theoretical framework for analysis of FVMs, it is desirable to supply a construction of the test

Received by the editor October 4, 2012 and, in revised form, June 19, 2013.

<sup>2010</sup> Mathematics Subject Classification. Primary 65N30, 65N12.

Key words and phrases. Finite volume methods.

This work was supported in part by Guangdong provincial government of China through the "Computational Science Innovative Research Team" program.

The first author was also supported in part by the Natural Science Foundation of China under grants 10771224 and 11071264.

The second author was supported in part by US Air Force Office of Scientific Research under grant FA9550-09-1-0511, by the US National Science Foundation under grants DMS-0712827, DMS-1115523, and by the Natural Science Foundation of China under grants 11071286 and 91130009. All correspondence should be sent to this author.

spaces which match the trial spaces for FVMs. The main purpose of this paper is to present a general setting for the construction of test spaces in the case that the trial spaces for FVMs are assumed to be standard conforming finite element (FE) spaces and to provide the unified analysis for the FVMs constructed here. Even though the construction given in this paper does not cover *all* FVMs known in the literature (for example, the mixed FVMs (cf. [16,17,30]) and the cell-centered FVMs (cf. [3,23,25]) are not included in this construction framework), it covers all specific FVMs presented in [12] and most triangulation based FVMs in other literature which employ conforming FE spaces as trial spaces. More importantly, it provides many new interesting FVM schemes.

The FVM has a long history. Historically, it appeared under many different names. Due to different starting points on algorithm constructions, FVMs were also called finite difference methods on irregular networks (cf. [27, 37, 40]), generalized difference methods (cf. [33]), box methods (cf. [2, 26, 38]), finite volume element methods (cf. [5]-[7]) and control-volume methods or covolume methods (cf. [15]). The numerical solution of the linear system resulted from FVMs is an important issue, which was addressed in [14, 35]. FVMs can be regarded essentially as special types of Petrov-Galerkin FEMs, in which the trial spaces are chosen as finite element spaces so as to deal with complex geometry domains and enjoy higher-order accuracy, while the choice of the test spaces needs a special care. We divide the domain of the problem into a finite number of control volumes, and the test spaces include characteristic functions of the control volume. Such local conservativeness can be fundamentally important for simulations of many physical models.

A key issue in the construction of FVMs is how the test space is chosen to match the FE trial space. According to different choices of the test spaces in the algorithm construction, the FVMs can be classified into three types: the Lagrange FVMs, the Hermite FVMs and the hybrid FVMs. In the Lagrange FVMs, the trial spaces are the Lagrange type FE spaces and the test spaces are spanned by the characteristic functions of the control volumes in the associated volume partitions (cf. [2,5–7,21,31,34,39,42]). In the Hermite FVMs, the trial spaces are the Hermite type FE spaces and the test spaces are spanned by the *generalized characteristic* functions of the control volumes in the associated volume partitions (cf. [10,32,33]). The construction of test spaces in the two types of FVMs mentioned above is based only on the volume partitions, and as a result, the volume partitions are complex when a higher-order scheme is constructed. In the hybrid FVMs, the trial spaces are the Lagrange or Hermite type FE spaces and the test spaces are spanned by the lower-order generalized characteristic functions of the control volumes of the volume partitions combined with certain linearly independent functions of the trial spaces. FVMs of this type were initially constructed for a quadratic FVM in [9]. The hybrid FVMs have not only the advantages of dealing with complex geometric domains and preserving local conservation laws, but also simple volume partitions for higher-order schemes and flexible algorithm construction which allows us to construct many more new schemes. Moreover, we show that the theoretical framework established in [12], which is applicable for the Lagrange and Hermite type of FVMs with affine invariant volume partitions, is also applicable to FVMs of this type, and the optimal error estimates can be derived when the primary triangulation satisfies certain geometric requirements. In other words, we may establish a unified way to compute the geometric requirements for the Lagrange, Hermite and Hybrid FVMs with affine invariant volume partitions, so as to obtain their optimal error estimate.

A typical FVM includes characteristic functions of the control volumes as a part of the basis of the test spaces to ensure that its resulting discrete equation preserves the local conservation laws. This results in the corresponding discrete bilinear form depending on the grids and being non-symmetric and non-conforming. These features of the FVMs introduce a major obstacle for their numerical analysis. Unlike the FEMs, the uniform boundedness and ellipticity of the bilinear form of the FVMs cannot be inherited directly from the original variational form of the boundary value problem. To overcome this challenge, we shall develop a general convergence theorem for FVMs which shows that if two sequences of trial and test spaces are connected by proper invertible linear mappings such that the uniform boundedness and ellipticity of the discrete bilinear forms can be defined, then the FVM equation has a unique solution and the convergence estimate can also be derived. We then study the construction of test spaces for FVMs. By ensuring that the test spaces have the same set of degrees of freedom as the associated trial spaces, we can obtain the bases for the trial and test spaces by using the set of degrees of freedom. By using these two bases, a natural invertible linear mapping between the trial and test spaces can be obtained. There are hypotheses that the volume partition, the set of degrees of freedom and the test space must be satisfied. Under these hypotheses, we can establish the uniform boundedness of the resulting discrete bilinear forms with the help of the equivalent discrete norms. In addition, with the help of the equivalence, we reduce the uniform ellipticity of the discrete bilinear forms to their uniform local-ellipticity. The geometric requirements of the primary triangle mesh to ensure the uniform local-ellipticity will be established under the hypotheses.

This paper is organized in eight sections. In section 2, we describe the framework of FVMs for elliptic boundary value problems and provide a way of constructing the test spaces. In section 3, we introduce the equivalent norms, which are important tools for the numerical analysis of FVMs. Section 4 is devoted to the convergence analysis of FVMs. With the help of the equivalent norms, the optimal error estimate of FVMs is derived under the uniform local-ellipticity of the discrete bilinear forms. Moreover, superconvergence properties of the FVM solutions are discussed. In section 5, algebraic and geometric sufficient and necessary conditions for the uniform local-ellipticity are established. In section 6, we characterize the admissible region which plays an important role in developing equivalent geometric conditions for the uniform local-ellipticity. In section 7, we present four new FVM schemes derived from the general construction and study their uniform local-ellipticity. In the last section, we present two numerical examples to verify the convergence results of the FVMs.

# 2. General FVMs for elliptic boundary problems

In this section, we present a framework of FVMs for solving elliptic equations and provide a convergence theorem which allows us to establish error analysis of FVMs. The general construction of the test spaces that match the conforming FE trial spaces is also described here.

Let  $\Omega$  be a polygonal domain in  $\mathbb{R}^2$  with boundary  $\partial\Omega$ . Suppose that  $\mathbf{a} := [a_{ij}(x)]$ is a 2 × 2 symmetric matrix of functions  $a_{ij} \in W^{1,\infty}(\Omega)$  and that  $b \in L^{\infty}(\Omega)$  and  $f \in L^2(\Omega)$ . We consider the Dirichlet problem of the second order partial differential equation

(2.1) 
$$\begin{cases} -\nabla \cdot (\mathbf{a}\nabla u) + bu = f, & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega, \end{cases}$$

where u is the unknown to be determined. We assume that the coefficients in equation (2.1) satisfy the elliptic condition  $\sum_{i,j=1}^{2} a_{ij}(x)\xi_i\xi_j \ge r \sum_{j=1}^{2} \xi_j^2$ , for some r > 0, for all  $(\xi_i, \xi_j) \in \mathbb{R}^2$  and  $b(x) \ge 0$ ,  $x \in \Omega$ .

We shall use the standard Sobolev notation. For a non-negative integer k and a subdomain  $\mathcal{D} \subseteq \mathbb{R}^2$ , let  $\mathbb{H}^k(\mathcal{D})$  denote the Sobolev space with the norm  $\|\cdot\|_{k,\mathcal{D}}$  and the corresponding semi-norm  $\|\cdot\|_{k,\mathcal{D}}$ . When  $\mathcal{D} = \Omega$ , we use  $\|\cdot\|_k$  and  $|\cdot|_k$  instead of  $\|\cdot\|_{k,\mathcal{D}}$  for simplicity. By  $\mathbb{H}^1_0(\Omega)$  we denote the subspace of  $\mathbb{H}^1(\Omega)$  whose functions have the vanishing trace on  $\partial\Omega$ .

We now derive a variational form of equation (2.1) suitable for establishing the FVMs. To this end, we need two partitions of the domain  $\Omega$ . Let  $\mathcal{T} := \{K\}$  be a triangulation of  $\Omega$ . We denote by  $\mathscr{T} := \{\mathcal{T}\}$  a family of triangulations  $\mathcal{T}$  of  $\Omega$ . Let  $\mathcal{T}^* := \{K^*\}$  be another partition of  $\Omega$  associated with  $\mathcal{T}$ . The partition  $\mathcal{T}^*$  will be called the volume partition or dual partition of  $\mathcal{T}$ , and the elements in  $\mathcal{T}^*$  will be called control volumes or dual elements. On  $\Omega$ , associated with each  $\mathcal{T}$  and  $\mathcal{T}^*$ , we define respectively the space  $\mathbb{H}^2_{\mathcal{T}}(\Omega) := \{v : v \in L^2(\Omega), v|_K \in \mathbb{H}^2(K), \text{ for all } K \in \mathcal{T}\}$  and the space

$$\mathbb{H}^{1}_{\mathcal{T}^{*}}(\Omega) := \{ v : v \in L^{2}(\Omega), \ v|_{K^{*}} \in \mathbb{H}^{1}(K^{*}), \text{ for all } K^{*} \in \mathcal{T}^{*}, \text{ and } v|_{\partial\Omega} = 0 \}.$$

For a domain E in  $\mathbb{R}^2$  the Lebesgue integral of g on E is denoted by  $\int_E g$ , and for a curve  $\ell$  in  $\mathbb{R}^2$  the line integral of h on  $\ell$  is denoted by  $\int_{\ell} h$ . With  $\mathcal{T}$  and the associated  $\mathcal{T}^*$ , we define the discrete bilinear form for  $w \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$  and  $v \in \mathbb{H}^1_{\mathcal{T}^*}(\Omega)$  by setting

$$a_{\mathcal{T}}(w,v) := \sum_{K^* \in \mathcal{T}^*} \left\{ \int_{K^*} \left( \nabla w^T \mathbf{a} \nabla v + b w v \right) - \int_{\partial K^* \setminus \partial \Omega} v(\mathbf{a} \nabla w) \cdot \mathbf{n} \right\},\$$

where **n** is the outward unit normal vector on  $\partial K^*$ . Employing the Green formula on the dual elements, we can show for  $w \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2(\Omega)$  and  $v \in \mathbb{H}^1_{\mathcal{T}^*}(\Omega)$  that

$$a_{\mathcal{T}}(w,v) = \int_{\Omega} \left( -\nabla \cdot (\mathbf{a} \nabla w) + bw \right) v.$$

Let  $(f, v) := \int_{\Omega} f v$ . The variational form for (2.1) is written as finding  $u \in \mathbb{H}_0^1(\Omega) \cap \mathbb{H}_{\mathcal{T}}^2(\Omega)$  such that

(2.2) 
$$a_{\mathcal{T}}(u,v) = (f,v), \text{ for all } v \in \mathbb{H}^1_{\mathcal{T}^*}(\Omega)$$

Unlike the standard variational form in the FEM, variational form (2.2) depends on the discretization. As in the FEM theory, we can easily prove that  $u \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2(\Omega)$  satisfies equation (2.1) if and only if it satisfies (2.2).

The FVM is developed based on the variational form (2.2) by projecting the spaces  $\mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$  and  $\mathbb{H}^1_{\mathcal{T}^*}(\Omega)$  respectively onto finite dimensional trial and test spaces. We choose the trial space  $\mathbb{U}_{\mathcal{T}} \subset \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$  as a standard conforming

finite element space with respect to  $\mathcal{T}$  and demand that the test space  $\mathbb{V}_{\mathcal{T}^*} \subset \mathbb{H}^1_{\mathcal{T}^*}(\Omega)$  and dim  $\mathbb{V}_{\mathcal{T}^*} = \dim \mathbb{U}_{\mathcal{T}}$ . The construction of the test space will be discussed in detail later in this section. In this way, the FVM for solving (2.1) is a finite dimensional approximation scheme which finds  $u_{\mathcal{T}} \in \mathbb{U}_{\mathcal{T}}$  such that

(2.3) 
$$a_{\mathcal{T}}(u_{\mathcal{T}}, v) = (f, v), \text{ for all } v \in \mathbb{V}_{\mathcal{T}^*}.$$

Since  $\mathbb{V}_{\mathcal{T}^*} \subset \mathbb{H}^1_{\mathcal{T}^*}(\Omega)$ ,  $v \in \mathbb{V}_{\mathcal{T}^*}$  may have a jump between the adjoining control volumes in  $\mathcal{T}^*$ , the integral along the boundary of  $K^* \in \mathcal{T}^*$  cannot be ignored in the discrete bilinear form  $a_{\mathcal{T}}(u_{\mathcal{T}}, v)$ . This is one of the major differences between FVMs and the conforming FEMs.

Unlike the bilinear form in the FEMs, the discrete bilinear form  $a_{\mathcal{T}}(\cdot, \cdot)$  in the FVMs does not necessarily inherit the boundedness and ellipticity directly from the original problem, because  $a_{\mathcal{T}}(\cdot, \cdot)$  depends on the grids and is non-symmetric and non-conforming. To tackle this difficulty, we next establish a convergence theorem which serves as a guide for the numerical analysis of the FVMs. To this end, we assume that there exist positive constants c and  $\sigma$  such that for all  $\mathcal{T} \in \mathscr{T}$  and the associated  $\mathcal{T}^*$  there are linear mappings  $\Pi_{\mathcal{T}^*} : \mathbb{U}_{\mathcal{T}} \to \mathbb{V}_{\mathcal{T}^*}$  with  $\Pi_{\mathcal{T}^*}\mathbb{U}_{\mathcal{T}} = \mathbb{V}_{\mathcal{T}^*}$  satisfying the conditions that

(C-1)  $|a_{\mathcal{T}}(w, \Pi_{\mathcal{T}^*}v)| \leq c \cdot e_{\mathcal{T}}(w) ||v||_1$ , for all  $w \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$ ,  $v \in \mathbb{U}_{\mathcal{T}}$ , (C-2)  $a_{\mathcal{T}}(w, \Pi_{\mathcal{T}^*}w) \geq \sigma ||w||_1^2$ , for all  $w \in \mathbb{U}_{\mathcal{T}}$ ,

where  $e_{\mathcal{T}}(\cdot)$  is a norm on  $\mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$ . If there exists a constant c > 0 such that condition (C-1) holds for all  $\mathcal{T} \in \mathscr{T}$  and the associated  $\mathcal{T}^*$ , we say that the family  $\mathscr{A}_{\mathscr{T}} := \{a_{\mathcal{T}}(\cdot, \Pi_{\mathcal{T}^*} \cdot) : \mathcal{T} \in \mathscr{T}\}$  of the discrete bilinear forms is uniformly bounded. If there exists a constant  $\sigma > 0$  such that condition (C-2) holds for all  $\mathcal{T} \in \mathscr{T}$  and the associated  $\mathcal{T}^*$ , we say that the family  $\mathscr{A}_{\mathscr{T}}$  is uniformly elliptic.

We present the convergence of the FVMs. Its proof is similar to that of Theorem 3.18 in [12].

**Theorem 2.1.** Let  $u \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2(\Omega)$  be the solution of (2.1). If conditions (C-1) and (C-2) hold, then for each  $\mathcal{T} \in \mathscr{T}$  the FVM equation (2.3) has a unique solution  $u_{\mathcal{T}} \in \mathbb{U}_{\mathcal{T}}$ , and there exists a positive constant c such that for all  $\mathcal{T} \in \mathscr{T}$ ,

$$||u - u_{\mathcal{T}}||_1 \le c \inf_{w \in \mathbb{U}_{\mathcal{T}}} (||u - w||_1 + e_{\mathcal{T}}(u - w)).$$

In the remaining part of this section we discuss a general construction of the test spaces that match the trial spaces. Such a construction may be reduced to the construction of a triple element on a reference triangle.

We choose the triangle  $\hat{K}$  with vertices  $\hat{P}_1 := (0,0), \hat{P}_2 := (1,0)$  and  $\hat{P}_3 := (0,1)$  as the reference triangle. For a triangle K, there is a unique invertible affine mapping  $\mathcal{F}_K$  from  $\hat{K}$  to K (cf. [32]). Since the trial space  $\mathbb{U}_{\mathcal{T}}$  is a FE space with respect to the triangulation  $\mathcal{T}$ , from the FE theory (cf. [1,4,19]), it is determined by the global continuity condition  $\mathbb{U}_{\mathcal{T}} \subset \mathbb{H}_0^1(\Omega) \cap \mathbb{H}_{\mathcal{T}}^2(\Omega)$  and the FE triple element  $(\hat{K}, \hat{\Sigma}, \mathbb{U}_{\hat{K}})$ . For a positive integer m, we let  $\mathbb{N}_m := \{1, 2, \ldots, m\}$ . The set of degrees of freedom  $\hat{\Sigma} := \{\hat{\zeta}_i : i \in \mathbb{N}_{\hat{n}}\}$  is a set of linearly independent linear functionals. The trial space  $\mathbb{U}_{\hat{K}}$  on  $\hat{K}$  is a space of polynomials defined on  $\hat{K}$  with dim  $\mathbb{U}_{\hat{K}} = \hat{n}$ , and  $\hat{\Sigma}$  is  $\mathbb{U}_{\hat{K}}$ -unisolvent. The test space  $\mathbb{V}_{\mathcal{T}^*}$  is likewise determined by the global continuity condition  $\mathbb{V}_{\mathcal{T}^*} \subset \mathbb{H}_{\mathcal{T}^*}^1(\Omega)$  and the triple element  $(\hat{\mathcal{T}}^*, \hat{\Sigma}^*, \mathbb{V}_{\hat{\mathcal{T}}^*})$ . We describe below the triple element in detail.

(i) The volume partition or dual partition  $\hat{\mathcal{T}}^* := \{\hat{K}^*\}$  of  $\hat{K}$  satisfies the following hypothesis:

Hypothesis 1.  $\hat{K} = \bigcup_{\hat{K}^* \in \hat{\mathcal{T}}^*} \hat{K}^*$  and  $\operatorname{meas}(\hat{K}^* \cap \hat{K}'^*) = 0$ , for all  $\hat{K}^*, \hat{K}'^* \in \hat{\mathcal{T}}^*$  and  $\hat{K}^* \neq \hat{K}'^*$ , where for a subdomain E of  $\mathbb{R}^2$ ,  $\operatorname{meas}(E)$  denotes the Lebesgue measure of E.

(ii) The set  $\hat{\Sigma}^* := {\hat{\eta}_j : j \in \mathbb{N}_{\hat{m}}}$  of degrees of freedom for the triple is not necessarily the same as  $\hat{\Sigma}$  but satisfies the following two hypotheses.

Hypothesis 2.  $\hat{\Sigma}^*$  is equivalent to  $\hat{\Sigma}$  in the sense that  $\hat{m} = \hat{n}$  and the FE triple elements  $(\hat{K}, \hat{\Sigma}^*, \mathbb{U}_{\hat{K}})$  and  $(\hat{K}, \hat{\Sigma}, \mathbb{U}_{\hat{K}})$  are equal (cf. [19]).

From this hypothesis we learn that  $\hat{\Sigma}^*$  is also  $\mathbb{U}_{\hat{K}}$ -unisolvent; that is, there is a basis  $\Phi_{\hat{K}} := \{\hat{\phi}_i : i \in \mathbb{N}_{\hat{K}}\}$  for  $\mathbb{U}_{\hat{K}}$  such that

(2.4) 
$$\hat{\eta}_j(\hat{\phi}_i) = \delta_{i,j}, \quad i, j \in \mathbb{N}_{\hat{n}}.$$

The basis  $\Phi_{\hat{K}}$  may be different from the classical basis of the space  $\mathbb{U}_{\hat{K}}$ .

Hypothesis 3. There is a positive integer  $\hat{s} \leq \hat{n}$  such that for each  $j \in \mathbb{N}_{\hat{s}}$ ,  $\hat{\eta}_j$  is a point evaluation functional at point  $\hat{x}_i \in \hat{K}$ , and the remaining functionals satisfy

(2.5) 
$$\hat{\eta}_j(\chi_{\hat{K}}) = 0, \quad j \in \mathbb{N}_{\hat{n}} \setminus \mathbb{N}_{\hat{s}},$$

where  $\chi_E$  denotes the characteristic function of  $E \subset \mathbb{R}^2$ .

Hypothesis 3 describes the forms of the functionals in  $\hat{\Sigma}^*$ . The first  $\hat{s}$  functionals in  $\hat{\Sigma}^*$  are the point evaluation functionals, and the remaining functionals in  $\hat{\Sigma}^*$  may be chosen in various ways as long as condition (2.5) is satisfied. In the Lagrange FVMs, all the functionals in  $\hat{\Sigma}^*$  are point evaluation functionals. In the Hermite FVMs,  $\hat{s} < \hat{n}$  and the functionals  $\hat{\eta}_j, j \in \mathbb{N}_{\hat{n}} \setminus \mathbb{N}_{\hat{s}}$  are directional derivatives or even higher-order directional derivatives at some points in  $\hat{K}$ . In the hybrid FVMs,  $\hat{s} < \hat{n}$  and the functionals  $\hat{\eta}_j, j \in \mathbb{N}_{\hat{n}} \setminus \mathbb{N}_{\hat{s}}$  may be chosen as differences, directional derivatives, higher-order directional derivatives or other forms which satisfy condition (2.5).

(iii) The test space  $\mathbb{V}_{\hat{\mathcal{T}}^*}$  on  $\hat{K}$  satisfies the following two hypotheses:

Hypothesis 4. The functions in  $\mathbb{V}_{\hat{\mathcal{T}}^*}$  are piecewise polynomials with respect to the partition  $\hat{\mathcal{T}}^*$ , with the characteristic functions of the control volumes in  $\hat{\mathcal{T}}^*$  being included in  $\mathbb{V}_{\hat{\mathcal{T}}^*}$ .

Hypothesis 5. The set  $\hat{\Sigma}^*$  is  $\mathbb{V}_{\hat{\mathcal{T}}^*}$ -unisolvent; that is, there is a basis  $\Psi_{\hat{\mathcal{T}}^*} := \{\hat{\psi}_i : i \in \mathbb{N}_{\hat{n}}\}$  for  $\mathbb{V}_{\hat{\mathcal{T}}^*}$  such that

(2.6) 
$$\hat{\eta}_j(\hat{\psi}_i) = \delta_{i,j}, \quad i, j \in \mathbb{N}_{\hat{n}}.$$

We comment that Hypotheses 1 and 4 are trivial. The key point is to properly choose the set  $\hat{\Sigma}^*$  of degrees of freedom which has the form described in Hypothesis 3 such that there exist a basis  $\Phi_{\hat{K}}$  for the trial space  $\mathbb{U}_{\hat{K}}$  and a basis  $\Psi_{\hat{\mathcal{T}}^*}$  for the test space  $\mathbb{V}_{\hat{\mathcal{T}}^*}$  satisfying (2.4) and (2.6).

For an affine mapping  $\mathcal{F}_{\hat{K}}$  from  $\hat{K}$  to itself, we let  $\Phi_{\hat{K}} \circ \mathcal{F}_{\hat{K}} := \{\hat{\phi}_i \circ \mathcal{F}_{\hat{K}} : i \in \mathbb{N}_{\hat{n}}\}$ and  $\Psi_{\hat{K}} \circ \mathcal{F}_{\hat{K}} := \{\hat{\psi}_i \circ \mathcal{F}_{\hat{K}} : i \in \mathbb{N}_{\hat{n}}\}$ . We further require that the bases  $\Phi_{\hat{K}}$  and  $\Psi_{\hat{\tau}^*}$  satisfy the following hypothesis.

Hypothesis 6. For any affine mapping  $\mathcal{F}_{\hat{K}}$  from  $\hat{K}$  to itself,  $\Phi_{\hat{K}} \circ \mathcal{F}_{\hat{K}} = \Phi_{\hat{K}}$  and  $\Psi_{\hat{K}} \circ \mathcal{F}_{\hat{K}} = \Psi_{\hat{K}}$ .

It can be verified that all the test spaces of the FVM presented in [12] satisfy all the hypotheses listed above. We shall also show later that other new and interesting FVM schemes may be constructed with test spaces that satisfy these hypotheses. Noting that  $\hat{K}^*$  is a subset of  $\hat{K}$ , the set  $\mathcal{F}_K(\hat{K}^*)$  is well-defined. We let  $\mathcal{F}_K(\hat{\mathcal{T}}^*) := \{\mathcal{F}_K(\hat{K}^*) : \hat{K}^* \in \hat{\mathcal{T}}^*\}.$ 

**Definition 2.2.** The volume partition  $\hat{\mathcal{T}}^*$  is said to be self-affine invariant if for any affine mapping  $\mathcal{F}_{\hat{K}}$  from  $\hat{K}$  to itself, there holds  $\mathcal{F}_{\hat{K}}(\hat{\mathcal{T}}^*) = \hat{\mathcal{T}}^*$ .

From Hypothesis 4, we get that the functions in  $\Phi_{\hat{K}}$  are piecewise polynomials with respect to  $\hat{\mathcal{T}}^*$ . It is clear that the functions in  $\Psi_{\hat{K}} \circ \mathcal{F}_{\hat{K}}$  are piecewise polynomials with respect to  $\mathcal{F}_{\hat{K}}(\hat{\mathcal{T}}^*)$ . Hypothesis 6 ensures that  $\Psi_{\hat{K}} \circ \mathcal{F}_{\hat{K}} = \Psi_{\hat{K}}$ . Thus,  $\mathcal{F}_{\hat{K}}(\hat{\mathcal{T}}^*) = \hat{\mathcal{T}}^*$ ; that is,  $\hat{\mathcal{T}}^*$  is self-affine invariant. Therefore, the volume partition  $\mathcal{T}^*$  can be formed automatically by  $F_K(\hat{\mathcal{T}}^*), K \in \mathcal{T}$ . Then, by Hypothesis 4, the global test space  $\mathbb{V}_{\mathcal{T}^*}$  contains the characteristic functions of the control volumes in the global dual partition  $\mathcal{T}^*$ . This is an essential requirement of FVMs.

Let  $n := \dim \mathbb{U}_{\mathcal{T}}$ . From (2.4) and (2.6), there exist a set of degrees of freedom  $\Sigma := \{\eta_i : i \in \mathbb{N}_n\}$ , a basis  $\Phi_{\mathcal{T}} := \{\phi_i : i \in \mathbb{N}_n\}$  for  $\mathbb{U}_{\mathcal{T}}$  and a basis  $\Psi_{\mathcal{T}^*} = \{\psi_i : i \in \mathbb{N}_n\}$  for  $\mathbb{V}_{\mathcal{T}^*}$  such that

(2.7) 
$$\eta_j(\phi_i) = \eta_j(\psi_i) = \delta_{i,j}, \quad i, j \in \mathbb{N}_n$$

Using  $\Phi_{\mathcal{T}}$  and  $\Psi_{\mathcal{T}^*}$ , we define a natural invertible linear mapping  $\Pi_{\mathcal{T}^*} : \mathbb{U}_{\mathcal{T}} \to \mathbb{V}_{\mathcal{T}^*}$ for any  $w = \sum_{i \in \mathbb{N}_n} w_i \phi_i \in \mathbb{U}_{\mathcal{T}}$  by  $\Pi_{\mathcal{T}^*} w := \sum_{i \in \mathbb{N}_n} w_i \psi_i$ . With the mapping  $\Pi_{\mathcal{T}^*}$ , the FVM equation (2.3) is reformulated as finding  $u_{\mathcal{T}} \in \mathbb{U}_{\mathcal{T}}$  such that

$$a_{\mathcal{T}}(u_{\mathcal{T}}, \Pi_{\mathcal{T}^*} w) = (f, \Pi_{\mathcal{T}^*} w), \text{ for all } w \in \mathbb{U}_{\mathcal{T}}.$$

The stiffness matrix  $\mathbf{A}_{\mathcal{T}} := [a_{\mathcal{T}}(\phi_j, \psi_i) : i, j \in \mathbb{N}_n]$  of the linear system of the FVM may be formed by using the bilinear form  $a_{\mathcal{T}}$  with the bases  $\Phi_{\mathcal{T}}$  and  $\Psi_{\mathcal{T}^*}$ . However, in practice, we first form the element matrices  $\mathbf{A}_K$ , for each  $K \in \mathcal{T}$ . Specifically, by making use of the bases  $\Phi_{\hat{K}}$  and  $\Psi_{\hat{\mathcal{T}}^*}$  on the reference triangle  $\hat{K}$  satisfying (2.4) and (2.6) respectively, and the affine mapping  $\mathcal{F}_K$  from  $\hat{K}$  to the triangle  $K \in \mathcal{T}$ , we derive the corresponding bases on K which are used to form the element stiffness matrix  $\mathbf{A}_K$  on K. We then assemble all element matrices  $\mathbf{A}_K$  to form the stiffness matrix  $\mathbf{A}_{\mathcal{T}}$ .

# 3. Discrete norm equivalence

We investigate in this section discrete norm equivalences. These norm equivalence results will play a vital role in establishing the error estimate of FVMs.

We first present a preliminary lemma, which describes a property of the bases of the trial and the test spaces on  $\hat{K}$ .

**Lemma 3.1.** (i) If Hypotheses 2 and 3 hold, then the basis  $\Phi_{\hat{K}} = \{\hat{\phi}_i : i \in \mathbb{N}_{\hat{n}}\}$ for  $\mathbb{U}_{\hat{K}}$  determined by (2.4) satisfies  $\sum_{i \in \mathbb{N}_{\hat{s}}} \hat{\phi}_i = \chi_{\hat{K}}$ .

(ii) If Hypotheses 1, 3, 4 and 5 hold, then the basis  $\Psi_{\hat{\mathcal{T}}^*} = \{\hat{\psi}_i : i \in \mathbb{N}_{\hat{n}}\}$  for  $\mathbb{V}_{\hat{\mathcal{T}}^*}$  determined by (2.6) satisfies  $\sum_{i \in \mathbb{N}_{\hat{n}}} \hat{\psi}_i = \chi_{\hat{K}}$ .

*Proof.* It follows from Hypothesis 3 and (2.4) that (3.1)

$$\hat{\eta}_j(\chi_{\hat{K}}) = \hat{\eta}_j\left(\sum_{i \in \mathbb{N}_{\hat{s}}} \hat{\phi}_i\right) = 1, \ j \in \mathbb{N}_{\hat{s}} \quad \text{and} \quad \hat{\eta}_j(\chi_{\hat{K}}) = \hat{\eta}_j\left(\sum_{i \in \mathbb{N}_{\hat{s}}} \hat{\phi}_i\right) = 0, \ j \in \mathbb{N}_{\hat{n}} \setminus \mathbb{N}_{\hat{s}}.$$

Since by Hypothesis 2,  $\{\hat{\eta}_j : j \in \mathbb{N}_{\hat{n}}\}$  is  $\mathbb{U}_{\hat{K}}$ -unisolvent, and since  $\chi_{\hat{K}} \in \mathbb{U}_{\hat{K}}$ , equation (3.1) ensures that  $\sum_{i \in \mathbb{N}_{\hat{s}}} \hat{\phi}_i = \chi_{\hat{K}}$ . 

The proof for (ii) can be similarly carried out.

If  $\hat{\Sigma}^* = {\hat{\eta}_j : j \in \mathbb{N}_{\hat{n}}}$  satisfies Hypothesis 3, we define a discrete norm on the trial space on K. For  $w \in \mathbb{U}_{\hat{K}}$ , we let (3.2)

$$w_{i} := \hat{\eta}_{i}(w), \ i \in \mathbb{N}_{\hat{n}}, \quad \bar{w} := \frac{1}{\hat{s}} \sum_{i \in \mathbb{N}_{\hat{s}}} w_{i}, \quad |w|_{1,\mathbb{U}_{\hat{K}}} := \left( \sum_{i \in \mathbb{N}_{\hat{s}}} (w_{i} - \bar{w})^{2} + \sum_{i \in \mathbb{N}_{\hat{n}} \setminus \mathbb{N}_{\hat{s}}} w_{i}^{2} \right)^{1/2}.$$

From Hypothesis 3,  $\bar{w}$  as defined in (3.2) is in fact the mean value of w at the points  $\hat{x}_i, i \in \mathbb{N}_{\hat{s}}$  and the first term in  $|w|_{1,\mathbb{U}_{\hat{\kappa}}}^2$  is the quadratic sum of the difference between the point evaluations of w at the points  $\hat{x}_i, i \in \mathbb{N}_{\hat{s}}$ , and the mean value.

We define a discrete norm on the test space on  $\hat{K}$ . Let  $L^*_{\hat{K}}$  denote the dual gridlines in  $\hat{K}$ . For any  $\ell^*$  in  $L^*_{\hat{K}}$ , it is actually a common edge of two control volumes in  $\hat{\mathcal{T}}^*$ . For  $x \in \ell^*$ , let

$$[v](x) := \lim_{\delta \to 0^+} v(x - \delta \mathbf{n}) - \lim_{\delta \to 0^+} v(x + \delta \mathbf{n})$$

denote the jump of v from a volume  $\hat{K}_1^*$  to its neighboring volume  $\hat{K}_2^*$ , where **n** is the unit normal vector on  $\ell^*$  pointing from  $K_1^*$  to  $K_2^*$ . For  $v \in \mathbb{V}_{\hat{\tau}^*}$ , let

$$\begin{split} |v|_{1,\hat{\mathcal{T}}^*} &:= \left(\sum_{\hat{K}^* \in \hat{\mathcal{T}}^*} |v|_{1,\hat{K}^*}^2\right)^{1/2}, \ [v]_{L_{\hat{K}}^*} &:= \left(\sum_{\ell^* \in L_{\hat{K}}^*} |\ell^*|^{-1} \int_{\ell^*} [v]^2\right)^{1/2}, \ |v|_{1,\mathbb{V}_{\hat{\mathcal{T}}^*}} \\ &:= \left(|v|_{1,\hat{\mathcal{T}}^*}^2 + [v]_{L_{\hat{K}}^*}^2\right)^{1/2}. \end{split}$$

We introduce  $\Pi_{\hat{K}^*} : \mathbb{U}_{\hat{K}} \to \mathbb{V}_{\hat{\mathcal{T}}^*}$  for any  $w := \sum_{i \in \mathbb{N}_{\hat{n}}} x_i \hat{\phi}_i \in \mathbb{U}_{\hat{K}}$  by  $\Pi_{\hat{K}^*} w :=$  $\sum_{i\in\mathbb{N}_{\hat{n}}} x_i \hat{\psi}_i.$ 

**Lemma 3.2.** If Hypotheses 2 and 3 hold, then there exist positive constants  $c_1$  and  $c_2$  such that for all  $w \in \mathbb{U}_{\hat{K}}$ ,

(3.3) 
$$c_1 |w|_{1,\mathbb{U}_{\hat{K}}} \le |w|_{1,\hat{K}} \le c_2 |w|_{1,\mathbb{U}_{\hat{K}}}$$

If furthermore Hypotheses 1, 4 and 5 hold, then there exist positive constants  $c_3$ and  $c_4$  such that for all  $w \in \mathbb{U}_{\hat{K}}$ ,

$$(3.4) c_3|w|_{1,\mathbb{U}_{\hat{K}}} \le |\Pi_{\hat{K}^*}w|_{1,\mathbb{V}_{\hat{\mathcal{T}}^*}} \le c_4|w|_{1,\mathbb{U}_{\hat{K}}}.$$

*Proof.* Hypothesis 2 ensures (2.4), from which we get for any  $w \in \mathbb{U}_{\hat{K}}$  that w = $\sum_{i \in \mathbb{N}_{\hat{n}}} w_i \hat{\phi}_i$ , where  $w_i$  is defined as in (3.2). Let  $\mathbf{w} := [w_i : i \in \mathbb{N}_{\hat{n}}]^T$ . Note that both  $|w|_{1,\hat{K}}^2$  and  $|w|_{1,\mathbb{U}_{\hat{K}}}^2$  are non-negative quadratic forms of **w**. Thus, there exist positive semi-definite symmetric matrices A and B such that

(3.5) 
$$|w|_{1,\hat{K}}^2 = \mathbf{w}^T \mathbf{A} \mathbf{w} \text{ and } |w|_{1,\mathbb{U}_{\hat{K}}}^2 = \mathbf{w}^T \mathbf{B} \mathbf{w}.$$

We next verify that the null spaces of  $\mathbf{A}$  and  $\mathbf{B}$  are the same. It is sufficient to verify that  $|w|_{1,\hat{K}} = 0$  if and only if  $|w|_{1,\mathbb{U}_{\hat{K}}} = 0$ . Hypotheses 2 and 3 ensure the validity of (i) in Lemma 3.1. It follows from Hypothesis 3 and (i) of Lemma 3.1 that  $|w|_{1,\hat{K}} = 0$  if and only if there holds

(3.6) 
$$w_i = \bar{w}, \ i \in \mathbb{N}_{\hat{s}} \text{ and } w_i = 0, \ i \in \mathbb{N}_{\hat{n}} \setminus \mathbb{N}_{\hat{s}},$$

where  $\bar{w}$  is defined as in (3.2). From the definition of  $|\cdot|_{1,\mathbb{U}_{\hat{K}}}$ , it is clear that (3.6) is sufficient and necessary for  $|w|_{1,\mathbb{U}_{\hat{K}}} = 0$ .

It follows from [20] that if two positive semi-definite matrices have the same null space, then there exist positive constants  $c_1$  and  $c_2$  such that for all  $\mathbf{w} \in \mathbb{R}^{\hat{n}}$ ,  $c_1 \mathbf{w}^T \mathbf{B} \mathbf{w} \leq \mathbf{w}^T \mathbf{A} \mathbf{w} \leq c_2 \mathbf{w}^T \mathbf{B} \mathbf{w}$ . Combining the above inequality with (3.5) yields (3.3).

The proof for (3.4) can be similarly carried out.

We now define a discrete norm on the trial space  $\mathbb{U}_{\mathcal{T}}$ . According to the FE theory (cf. [19]), for each  $K \in \mathcal{T}$ , corresponding to the FE triple element  $(\hat{K}, \hat{\Sigma}^*, \mathbb{U}_{\hat{K}})$  on the reference triangle  $\hat{K}$ , there is a FE triple element  $(K, \Sigma_K, \mathbb{U}_K)$  on K. Note that the set of degrees of freedom  $\Sigma_K := \{\eta_{i,K} : i \in \mathbb{N}_{\hat{n}}\}$  are the functionals corresponding to  $\hat{\eta}_i$  in the sense that for all  $\hat{w} \in \mathbb{U}_{\hat{K}}$  with  $w := \hat{w} \circ \mathcal{F}_K^{-1}, \eta_{i,K}(w) = \hat{\eta}_i(\hat{w})$ . For  $K \in \mathcal{T}$ , we let

(3.7) 
$$\phi_{i,K} := \hat{\phi}_i \circ \mathcal{F}_K^{-1}, \quad i \in \mathbb{N}_{\hat{n}}.$$

By Hypothesis 2, we get that  $\{\phi_{i,K} : i \in \mathbb{N}_{\hat{n}}\}$  is a basis for  $\mathbb{U}_K$  and there holds the relationship

(3.8) 
$$\eta_{j,K}(\phi_{i,K}) = \delta_{i,j}, \quad i, j \in \mathbb{N}_{\hat{n}}$$

For each  $w \in \mathbb{U}_{\mathcal{T}}$  and  $K \in \mathcal{T}$ , we let (3.9)

$$w_{i,K} := \eta_{i,K}(w), \ i \in \mathbb{N}_{\hat{n}}, \quad \bar{w}_K := \frac{1}{\hat{s}} \sum_{i \in \mathbb{N}_{\hat{s}}} w_{i,K} \text{ and } \bar{w}_{i,K} := w_{i,K} - \bar{w}_K, \ i \in \mathbb{N}_{\hat{s}}.$$

Define

$$|w|_{1,\mathbb{U}_{\mathcal{T}},K} := \left(\sum_{i\in\mathbb{N}_{\hat{s}}} \bar{w}_{i,K}^2 + \sum_{i\in\mathbb{N}_{\hat{n}}\setminus\mathbb{N}_{\hat{s}}} w_{i,K}^2\right)^{1/2}$$

and

$$|w|_{1,\mathbb{U}_{\mathcal{T}}} := \left(\sum_{K\in\mathcal{T}} |w|^2_{1,\mathbb{U}_{\mathcal{T}},K}\right)^{1/2}$$

We introduce the discrete norm  $|\cdot|_{1,\mathbb{V}_{\mathcal{T}^*}}$  on the test space  $\mathbb{V}_{\mathcal{T}^*}$ . For any  $v \in \mathbb{V}_{\mathcal{T}^*}$ , define

$$|v|_{1,\mathbb{V}_{\mathcal{T}^*},K} := \left(\sum_{K^* \in \mathcal{T}^*} |v|_{1,K^* \cap K}^2 + \sum_{\ell^* \in L_K^*} |\ell^*|^{-1} \int_{\ell^*} [v]^2 \right)^{1/2} \text{ and}$$
$$|v|_{1,\mathbb{V}_{\mathcal{T}^*}} := \left(\sum_{K \in \mathcal{T}} |v|_{1,\mathbb{V}_{\mathcal{T}^*},K}^2 \right)^{1/2},$$

where for  $K \in \mathcal{T}$ ,  $L_K^*$  denotes the dual gridlines in K.

The next lemma shows the equivalence of the discrete norms  $|\cdot|_{1,\mathbb{U}_{\mathcal{T}}}$  and  $|\Pi_{\mathcal{T}^*} \cdot |_{1,\mathbb{V}_{\mathcal{T}^*}}$  and the Sobolev norm  $|\cdot|_1$ . We say a family  $\mathscr{T}$  of triangulations of  $\Omega$  is *regular* if there exists a positive constant  $\theta_{inf}$  such that

(3.10) 
$$\theta_{\min,K} \ge \theta_{\inf}, \text{ for all } K \in \bigcup_{\mathcal{T} \in \mathscr{T}} \mathcal{T},$$

where  $\theta_{\min,K}$  denotes the minimum angle of the triangle K. For a triangle K, we use  $\nabla \mathcal{F}_K$  to denote the Jacobian matrix of the affine mapping  $\mathcal{F}_K$  and let  $\mathbf{M}_K := |\det(\nabla \mathcal{F}_K)| \cdot \nabla \mathcal{F}_K^{-1} (\nabla \mathcal{F}_K^{-1})^T$ . For all  $w \in \mathbb{U}_T$ ,  $v \in \mathbb{V}_{T^*}$  and all  $K \in \mathcal{T}$ , we let

(3.11) 
$$\hat{w}_K := w|_K \circ \mathcal{F}_K \text{ and } \hat{v}_K := v|_K \circ \mathcal{F}_K.$$

**Lemma 3.3.** If  $\mathscr{T}$  is regular and Hypotheses 2 and 3 hold, then there exist positive constants  $c_1$  and  $c_2$  such that for all  $\mathcal{T} \in \mathscr{T}$  and for all  $w \in \mathbb{U}_{\mathcal{T}}$ ,

(3.12) 
$$c_1 |w|_{1, \mathbb{U}_{\mathcal{T}}} \le |w|_1 \le c_2 |w|_{1, \mathbb{U}_{\mathcal{T}}}$$

If furthermore Hypotheses 1, 4 and 5 hold, then there exist positive constants  $c_3$ and  $c_4$  such that for all  $\mathcal{T} \in \mathscr{T}$  and for all  $w \in \mathbb{U}_{\mathcal{T}}$ ,

(3.13) 
$$c_3|w|_1 \le |\Pi_{\mathcal{T}^*}w|_{1,\mathbb{V}_{\mathcal{T}^*}} \le c_4|w|_1.$$

*Proof.* If we can prove that there are positive constants  $c_1$  and  $c_2$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,

(3.14) 
$$c_1^2 |w|_{1,\mathbb{U}_{\mathcal{T}},K}^2 \le |w|_{1,K}^2 \le c_2^2 |w|_{1,\mathbb{U}_{\mathcal{T}},K}^2,$$

then summing up the above inequality over all  $K \in \mathcal{T}$ , we obtain (3.12). We now prove (3.14). Since  $\eta_{i,K}(w) = \hat{\eta}_i(\hat{w}_K), i \in \mathbb{N}_{\hat{n}}$ , for all  $w \in \mathbb{U}_{\mathcal{T}}$  and all  $K \in \mathcal{T}$ ,

(3.15) 
$$|w|_{1,\mathbb{U}_{\mathcal{T}},K} = |\hat{w}_K|_{1,\mathbb{U}_{\hat{K}}}$$

By changing variables, we have that

$$|w|_{1,K}^2 = \int_K \nabla w \cdot \nabla w = \int_{\hat{K}} (\nabla \hat{w}_K)^T \mathbf{M}_K \nabla \hat{w}_K.$$

It was proved in Lemma 3.3 of [12] that the eigenvalues  $\lambda(\mathbf{M}_K) \in [(1/4) \tan \theta_{\min,K}, 4 \cot \theta_{\min,K}]$ . This combined with the regularity of  $\mathscr{T}$  yields  $\lambda(\mathbf{M}_K) \in [(1/4) \tan \theta_{\inf}, 4 \cot \theta_{\inf}]$  for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ . Thus

(3.16) 
$$(1/4) \tan \theta_{\inf} |\hat{w}_K|^2_{1,\hat{K}} \le |w|^2_{1,K} \le 4 \cot \theta_{\inf} |\hat{w}_K|^2_{1,\hat{K}}.$$

Since Hypotheses 2 and 3 hold and  $\hat{w}_K \in \mathbb{U}_{\hat{K}}$ , by inequality (3.3) in Lemma 3.2, we get that there are positive constants  $d_1$  and  $d_2$  independent of the grids such that

(3.17) 
$$d_1 |\hat{w}_K|_{1,\mathbb{U}_{\hat{K}}} \le |w|_{1,\hat{K}} \le d_2 |\hat{w}_K|_{1,\mathbb{U}_{\hat{K}}}.$$

Combining equations (3.15), (3.16) and (3.17) yields inequality (3.14) with  $c_1 := (1/2) d_1 (\tan \theta_{\inf})^{1/2}$  and  $c_2 := 2 d_2 (\cot \theta_{\inf})^{1/2}$ . The proof for (3.13) can be similarly carried out.

In the next lemma, we show that the  $L^2$ -norms  $\|\Pi_{\mathcal{T}^*} \cdot \|_0$  and  $\|\cdot\|_0$  are equivalent.

**Lemma 3.4.** If Hypotheses 2 and 5 hold, then there exist positive constants  $c_1$  and  $c_2$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $w \in \mathbb{U}_{\mathcal{T}}$ ,

(3.18) 
$$c_1 \|w\|_0^2 \le \|\Pi_{\mathcal{T}^*} w\|_0^2 \le c_2 \|w\|_0^2$$

*Proof.* For each  $\mathcal{T} \in \mathscr{T}$  and  $w \in \mathbb{U}_{\mathcal{T}}$ , let  $v := \prod_{\mathcal{T}^*} w$ . It suffices to prove that there exist positive constants  $c_1$  and  $c_2$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,

$$(3.19) c_1 \|w\|_{0,K}^2 \le \|v\|_{0,K}^2 \le c_2 \|w\|_{0,K}^2$$

By changing variables, we find that

(3.20) 
$$||v||_{0,K}^2 = |\det(\nabla \mathcal{F}_K)|||\hat{v}_K||_{0,\hat{K}}^2$$
 and  $||w||_{0,K}^2 = |\det(\nabla \mathcal{F}_K)|||\hat{w}_K||_{0,\hat{K}}^2$ .

Since  $v = \prod_{\mathcal{T}^*} w$ , we know that  $\hat{v}_K = \prod_{\hat{K}^*} \hat{w}_K$ . Noting that both  $\|\prod_{\hat{K}^*} \cdot\|_{0,\hat{K}}$  and  $\|\cdot\|_{0,\hat{K}}$  are norms for the finite dimensional space  $\mathbb{U}_{\hat{K}}$ , there exist positive constants  $c_1$  and  $c_2$  such that for all  $\hat{w} \in \mathbb{U}_{\hat{K}}$ ,

(3.21) 
$$c_1 \|\hat{w}\|_{0,\hat{K}}^2 \le \|\Pi_{\hat{K}^*} \hat{w}\|_{0,\hat{K}}^2 \le c_2 \|\hat{w}\|_{0,\hat{K}}^2.$$

Note that  $c_1$  and  $c_2$  depend only on the reference triangle  $\hat{K}$ . Thus, from (3.20) and (3.21) we obtain (3.19).

## 4. Convergence and superconvergence

In this section, we establish that the uniform local-ellipticity of the family of the discrete bilinear forms leads to the optimal error estimate of FVMs. We also present a brief discussion of superconvergence properties of the FVM solution.

For convenience of analysis, we rewrite the discrete bilinear form as the sum of its continuous and discontinuous components. Notice that  $v \in \mathbb{H}^1_{\mathcal{T}^*}(\Omega)$  is not necessarily continuous on  $K \in \mathcal{T}$ . Hence, for each  $\mathcal{T}$  and each  $K \in \mathcal{T}$ , we introduce the continuous component  $a_{c,K}$  and the discontinuous component  $a_{d,K}$  respectively by

$$\begin{aligned} a_{c,K}(w,v) &:= \sum_{K^* \in \mathcal{T}^*} \int_{K^* \cap K} \left( \nabla w^T \mathbf{a} \nabla v + b w v \right), \quad a_{d,K}(w,v) \\ &:= -\sum_{K^* \in \mathcal{T}^*} \int_{\partial K^* \cap \operatorname{int} K} v(\mathbf{a} \nabla w) \cdot \mathbf{n}, \end{aligned}$$

for each  $w \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$  and  $v \in \mathbb{H}^1_{\mathcal{T}^*}(\Omega)$ . For each  $K \in \mathcal{T}$ , we define  $a_K(w,v) := a_{c,K}(w,v) + a_{d,K}(w,v)$ . The discrete bilinear form is then rewritten as

(4.1) 
$$a_{\mathcal{T}}(w,v) = \sum_{K \in \mathcal{T}} a_K(w,v) = a_{c,\mathcal{T}}(w,v) + a_{d,\mathcal{T}}(w,v),$$

where

$$a_{c,\mathcal{T}}(w,v) := \sum_{K \in \mathcal{T}} a_{c,K}(w,v), \qquad a_{d,\mathcal{T}}(w,v) := \sum_{K \in \mathcal{T}} a_{d,K}(w,v).$$

By making use of (3.13) and (3.18), we can establish the uniform boundedness of the family of the discrete bilinear forms of FVMs. For each  $w \in \mathbb{H}^2_{\mathcal{T}}(\Omega)$ , we define the semi-norm

$$|w|_{2,\mathcal{T}} := \left(\sum_{K\in\mathcal{T}} |w|_{2,K}^2\right)^{1/2}$$

**Lemma 4.1.** If  $\mathscr{T}$  is regular and Hypotheses 1-5 hold, then there exists a positive constant c such that for all  $\mathcal{T} \in \mathscr{T}$ , and for all  $w \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$  and  $v \in \mathbb{U}_{\mathcal{T}}$ ,

$$|a_{\mathcal{T}}(w, \Pi_{\mathcal{T}^*} v)| \le c(|w|_1 + h|w|_{2,\mathcal{T}}) ||v||_1,$$

where h is the largest diameter of  $K \in \mathcal{T}$  for all  $\mathcal{T} \in \mathscr{T}$ .

Lemma 4.1 shows that the uniform boundedness condition (C-1) holds with  $e(\cdot) := |\cdot|_1 + h| \cdot |_{2,\mathcal{T}}$ . It is clear that this function e is a norm on  $\mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2_{\mathcal{T}}(\Omega)$ .

We next investigate the uniform ellipticity of the family of the discrete bilinear forms. As in [12], we say that the family  $\mathscr{A}_{\mathscr{T}}$  of the bilinear forms is *uniformly local-elliptic* if there exists a positive constant  $\sigma$  such that for all  $\mathcal{T} \in \mathscr{T}$  and its associated  $\mathcal{T}^*$ , all  $K \in \mathcal{T}$  and all  $w \in \mathbb{U}_{\mathcal{T}}$ ,

(4.2) 
$$a_K(w, \Pi_{\mathcal{T}^*} w) \ge \sigma |w|_{1, \mathbb{U}_{\mathcal{T}^*} K}^2.$$

By summing up (4.2) over all  $K \in \mathcal{T}$  and using the discrete norm equivalence (3.12), we see that the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$  leads to its uniform ellipticity. Thus, by Theorem 2.1 and Lemma 4.1, we get the following optimal estimate of the FVMs.

**Theorem 4.2.** Let  $u \in \mathbb{H}^1_0(\Omega) \cap \mathbb{H}^2(\Omega)$  be the solution of (2.1). Suppose that Hypotheses 1-5 hold and  $\mathscr{T}$  is regular. If  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic, then for each  $\mathcal{T} \in \mathscr{T}$  the FVM equation (2.3) has a unique solution  $u_{\mathcal{T}} \in \mathbb{U}_{\mathcal{T}}$ , and there exists a positive constant c such that for all  $\mathcal{T} \in \mathscr{T}$ ,

$$||u - u_{\mathcal{T}}||_1 \le c \inf_{w \in \mathbb{U}_{\mathcal{T}}} (||u - w||_1 + h|u - w|_{2,\mathcal{T}}).$$

When the trial space  $\mathbb{U}_{\mathcal{T}}$  is chosen as the space of the conforming piecewise polynomial of degree k with respect to the triangulation  $\mathcal{T}$  and  $u \in \mathbb{H}_0^1(\Omega) \cap \mathbb{H}^{k+1}(\Omega)$ , then by Theorem 4.2 we obtain the optimal convergence estimate  $||u - u_{\mathcal{T}}||_1 = O(h^k)$ . This estimate cannot be improved even if u has a higher order of smoothness.

In the remaining part of this section, we briefly discuss superconvergence properties of the FVM solution. Solutions of certain FVMs were shown to have a higher-order convergence property at certain points in the domain. This property, called the superconvergence, has drawn the attention of a number of researchers. In the literature, mainly there are two approaches in studying this superconvergence property of FVMs. One approach considered the FVM as a *small* perturbation of a FEM, and superconvergence results of the FVM were derived from those of the FEM solutions (cf. [41, 42]). Another approach was to establish the superinterpolation weak estimate of the discrete bilinear form of the FVM (cf. [11, 32, 36, 43]). Most of these results are for the FVMs whose trial spaces are linear elements. Now we discuss superconvergence of the FVMs of higher orders. Since an FVM of a higher order may not be considered as a small perturbation of its FEM counterpart (cf. [12]), the first approach mentioned above may not be suitable for an FVM of a higher order. We thus take the second approach.

The following discussion is not restricted to the FVMs of the triangle meshes. For a non-negative integer k, a real number  $p \in [1, \infty]$  and a subdomain  $\mathcal{D} \subseteq \mathbb{R}^2$ , let  $\mathbb{W}^{k,p}(\mathcal{D})$  denote the Sobolev space with the norm  $\|\cdot\|_{k,p,\mathcal{D}}$ . When  $\mathcal{D} = \Omega$ , we use  $\|\cdot\|_{k,p}$  instead of  $\|\cdot\|_{k,p,\mathcal{D}}$  for simplicity. We choose the trial space  $\mathbb{U}_{\mathcal{T}} := \{v \in C(\bar{\Omega}) : v|_K \in \mathbf{P}_k, \text{ for all } K \in \mathcal{T}, v|_{\partial\Omega} = 0\}$ , where  $\mathbf{P}_k$  is the set of all polynomials of degree k. We use C to denote a positive constant (independent of meshes and functions), which may be different in different places. Let S denote the set of the optimal interpolation stress points; that is, there exists  $q \in (2, \infty]$  such that for all  $x \in S$ ,

(4.3) 
$$|(\nabla u - \bar{\nabla} u_I)(x)| \le Ch^{k+1-\frac{2}{q}} ||u||_{k+2,q,E},$$

where  $u_I$  is the interpolation of u in  $\mathbb{U}_{\mathcal{T}}$ ,  $\overline{\nabla}u_I(x)$  denotes the average of the gradients of  $u_I$  on the elements containing the point x and E is the union of all such elements. We next assume that the discrete bilinear form of the FVM satisfies the superinterpolation weak estimate: there exists a constant  $p \in [1, \infty]$  such that

(4.4) 
$$|a_{\mathcal{T}}(u-u_I,\Pi_{\mathcal{T}^*}w)| \le Ch^{k+1}(||u||_{k+2,p}+F(u))||w||_1, \text{ for all } w \in \mathbb{U}_{\mathcal{T}},$$

where F(u) is a non-negative function of u independent of meshes. This hypothesis has been proved for several special cases. For the linear FVM with p = 2 and  $F(\cdot) = 0$  on the uniform triangulation, p = 2 and  $F(\cdot) = \|\cdot\|_{2,\infty}$  on the C-uniform triangulation and  $p = \infty$  and  $F(\cdot) = 0$  on the piecewise C-uniform triangulation, property (4.4) was established in [11]. Condition (4.4) for the bilinear FVM on the regular quadrilateral partition with p = 2 and  $F(\cdot) = 0$  was derived in [36]. Condition (4.4) was established in [43] for the vertex-centered FVMs of any order on rectangular meshes with p = 2 and  $F(\cdot) = 0$ .

**Proposition 4.3.** Let  $u \in \mathbb{H}_0^1(\Omega) \cap \mathbb{W}^{k+2,\infty}(\Omega)$  be the solution of (2.1) and  $u_{\mathcal{T}} \in \mathbb{U}_{\mathcal{T}}$  be the solution of (2.3). If  $\mathscr{A}_{\mathscr{T}}$  is uniformly elliptic and (4.4) holds, then

(4.5) 
$$\left(\frac{1}{N}\sum_{x\in S} |(\nabla u - \bar{\nabla} u_{\mathcal{T}})(x)|^2\right)^{1/2} \le Ch^{k+1} \left(\|u\|_{k+2,p} + \|u\|_{k+2,q} + F(u)\right),$$

where N is the cardinality of S.

*Proof.* We first prove that  $u_I$  is super close to  $u_{\mathcal{T}}$ . Since  $\mathscr{A}_{\mathscr{T}}$  is uniformly elliptic, we obtain that there exists a positive constant  $\sigma$  such that

(4.6) 
$$||u_I - u_{\mathcal{T}}||_1^2 \le \sigma a_{\mathcal{T}}(u_I - u_{\mathcal{T}}, \Pi_{\mathcal{T}^*}(u_I - u_{\mathcal{T}})) = \sigma a_{\mathcal{T}}(u_I - u, \Pi_{\mathcal{T}^*}(u_I - u_{\mathcal{T}})).$$

Combining (4.4) and (4.6) yields

$$||u_I - u_{\mathcal{T}}||_1^2 \le Ch^{k+1}(||u||_{k+2,p} + F(u))||u_I - u_{\mathcal{T}}||_1,$$

which leads to the fact that

(4.7) 
$$\|u_I - u_{\mathcal{T}}\|_1 \le Ch^{k+1}(\|u\|_{k+2,p} + F(u)).$$

We next prove the inequality (4.5). Using (4.3) and noting  $N = O(h^{-2})$  and q > 2, we obtain that

(4.8) 
$$\left(\frac{1}{N}\sum_{x\in S} |(\nabla u - \bar{\nabla} u_I)(x)|^2\right)^{1/2} \le Ch^{k+2-\frac{2}{q}} ||u||_{k+2,q} \le Ch^{k+1} ||u||_{k+2,q}.$$

By the inverse inequality of the FEMs, we get that

(4.9) 
$$|(\nabla u_I - \bar{\nabla} u_T)(x)| \le Ch^{-1} ||u_I - u_T||_{1,E}, \quad x \in S.$$

From (4.9), the fact that  $N = O(h^{-2})$  and (4.7), we obtain that (4.10)

$$\left(\frac{1}{N}\sum_{x\in S} |(\bar{\nabla}u_I - \bar{\nabla}u_{\mathcal{T}})(x)|^2\right)^{1/2} \le C ||u_I - u_{\mathcal{T}}||_1 \le Ch^{k+1}(||u||_{k+2,p} + F(u)).$$

Combining (4.8) and (4.10) yields the desired inequality (4.5).

Proposition 4.3 provides a guidance for the study of the superconvergence property of the FVM solutions. In this regard, the superinterpolation weak estimate (4.4) is crucial. For the linear FVM for solving 2D elliptic equations on triangle meshes, (4.4) was derived (cf. [11,32]). Condition (4.4) for higher-order FVMs for solving 2D elliptic equations on triangle meshes requires further investigation.

#### 5. Equivalent conditions for the uniform local-ellipticity

We have seen in Theorem 4.2 that the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$  is crucial to ensure the optimal error estimate of FVMs. This section is devoted to the study of the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$ . Sufficient and necessary conditions for the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$  will be established when the matrix **a** in (2.1) is chosen as the identity matrix, and b = 0.

We first reexpress the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$  in an equivalent matrix form. To this end, for each  $K \in \mathcal{T}$ , we define  $\psi_{i,K} := \hat{\psi}_i \circ \mathcal{F}_K^{-1}$ ,  $i \in \mathbb{N}_{\hat{n}}$ ,  $\mathbf{A}_K := [a_K(\phi_{i,K}, \psi_{j,K}) : i, j \in \mathbb{N}_{\hat{n}}]$  and  $\tilde{\mathbf{A}}_K := (\mathbf{A}_K + \mathbf{A}_K^T)/2$ , where  $\phi_{i,K}$ ,  $i \in \mathbb{N}_{\hat{n}}$  are defined by (3.7). The matrix  $\tilde{\mathbf{A}}_K$  is the symmetrization of the element stiffness matrix  $\mathbf{A}_K$ . It is well-known that  $\tilde{\mathbf{A}}_K$  and  $\mathbf{A}_K$  have the same quadratic form; that is, for all  $\mathbf{w} \in \mathbb{R}^{\hat{n}}$ , there holds  $\mathbf{w}^T \tilde{\mathbf{A}}_K \mathbf{w} = \mathbf{w}^T \mathbf{A}_K \mathbf{w}$ . However, the eigenvalues of matrix  $\tilde{\mathbf{A}}_K$  are all real and it has complete orthogonal eigenvectors. Recall that Hypothesis 2 ensures that  $\{\phi_{i,K} : i \in \mathbb{N}_{\hat{n}}\}$  is a basis for  $\mathbb{U}_K$  and (3.8) holds. Thus, for each  $w \in \mathbb{U}_{\mathcal{T}}$  and each  $K \in \mathcal{T}$ ,

(5.1) 
$$w(x) = \sum_{i \in \mathbb{N}_{\hat{n}}} w_{i,K} \phi_{i,K}(x), \quad x \in K,$$

where  $w_{i,K}, i \in \mathbb{N}_{\hat{n}}$  are as defined in (3.9). For each  $w \in \mathbb{U}_{\mathcal{T}}$  and each  $K \in \mathcal{T}$ , we let  $\mathbf{w}_{K} := [w_{i,K} : i \in \mathbb{N}_{\hat{n}}]^{T}$ . We define a matrix of rank 1 by setting  $e_{i} := \hat{\eta}_{i}(\chi_{\hat{K}}), i \in \mathbb{N}_{\hat{n}}, \mathbf{e} := [e_{i} : i \in \mathbb{N}_{\hat{n}}]^{T}$  and  $\mathbf{E} := \frac{1}{\mathbf{e}^{T}\mathbf{e}}\mathbf{e}\mathbf{e}^{T}$ .

**Lemma 5.1.** If Hypotheses 2 and 3 hold, then the uniform local-ellipticity of the family  $\mathscr{A}_{\mathscr{T}}$  of the discrete bilinear forms is equivalent to the existence of a positive constant  $\sigma$  such that for all  $\mathcal{T} \in \mathscr{T}$ , all  $K \in \mathcal{T}$  and all  $w \in \mathbb{U}_{\mathcal{T}}$ ,

(5.2) 
$$\mathbf{w}_K^T \tilde{\mathbf{A}}_K \mathbf{w}_K \ge \sigma (\mathbf{w}_K - \mathbf{E} \mathbf{w}_K)^T (\mathbf{w}_K - \mathbf{E} \mathbf{w}_K).$$

*Proof.* It suffices to verify that for any positive constant  $\sigma$  inequality (4.2) is equivalent to (5.2). We prove this by identifying the bilinear form and the norm appearing in (4.2) with the quantities appearing in (5.2). For all  $w \in \mathbb{U}_{\mathcal{T}}$ , Hypothesis 2 ensures that (5.1) holds. From the definition of  $\Pi_{\mathcal{T}^*}$ , we get that

(5.3) 
$$(\Pi_{\mathcal{T}^*} w)(x) = \sum_{i \in \mathbb{N}_{\hat{n}}} w_{i,K} \psi_{i,K}(x), \quad x \in K.$$

Substituting (5.1) and (5.3) into the definition of  $a_K(\cdot, \cdot)$ , we have that

(5.4) 
$$a_K(w, \Pi_{\mathcal{T}^*} w) = \mathbf{w}_K^T \mathbf{A}_K \mathbf{w}_K = \mathbf{w}_K^T \tilde{\mathbf{A}}_K \mathbf{w}_K.$$

It follows from Hypothesis 3 that  $e_i = 1$  for  $i \in \mathbb{N}_{\hat{s}}$  and  $e_i = 0$  for  $i \in \mathbb{N}_{\hat{n}} \setminus \mathbb{N}_{\hat{s}}$ . By the definition of  $|\cdot|_{1,\mathbb{U}_{\tau},K}$ , we find that

(5.5) 
$$|w|_{1,\mathbb{U}_{\mathcal{T},K}}^2 = (\mathbf{w}_K - \mathbf{E}\mathbf{w}_K)^T (\mathbf{w}_K - \mathbf{E}\mathbf{w}_K)^T$$

The desired result follows from (5.4) and (5.5).



FIGURE 1. The region G and its lower boundary

Inequality (5.2) in Lemma 5.1 involves all  $w \in \mathbb{U}_{\mathcal{T}}$ , and as a result it is still not convenient for practical use. However, it provides a basis for us to derive convenient sufficient and necessary conditions on the uniform local-ellipticity of the family  $\mathscr{A}_{\mathscr{T}}$ . We next introduce matrices constructed from  $\tilde{\mathbf{A}}_K$  and  $\mathbf{E}$  for pairs  $(r_1, r_2)$  of real numbers. The equation  $\mathbf{E}^T \mathbf{E} = \mathbf{E} = \mathbf{E}^T$  together with (5.2) implies that

(5.6) 
$$\mathbf{w}_{K}^{T}(\tilde{\mathbf{A}}_{K}+\mathbf{E})\mathbf{w}_{K} \geq \frac{\sigma}{1+\sigma}\mathbf{w}_{K}^{T}\mathbf{w}_{K}.$$

For a triangle  $K \in \mathcal{T}$  with vertices  $P_i$ , i = 1, 2, 3, we denote by |K| its area and by  $\ell_i$  its edge opposite to  $P_i$  and by  $|\ell_i|$  the length of the edge  $\ell_i$ . Without loss of generality, we assume that  $|\ell_1| \ge |\ell_2| \ge |\ell_3|$ . For each triangle  $K \in \mathcal{T}$ , we introduce three geometric parameters:

(5.7) 
$$r_{1,K} := |\ell_2|^2 / |\ell_1|^2, \quad r_{2,K} := |\ell_3|^2 / |\ell_1|^2, \quad \alpha_K := |\ell_1|^2 / (4|K|).$$

Then, we have that  $0 < r_{2,K} \leq r_{1,K} \leq 1$ . Each non-degenerate triangle K is associated with a pair of numbers  $(r_{1,K}, r_{2,K})$  with  $0 < r_{2,K} \leq r_{1,K} \leq 1$  and  $(r_{1,K})^{1/2} + (r_{2,K})^{1/2} > 1$ . Conversely, a point in the set

$$\mathbb{G} := \left\{ (r_1, r_2) \in \mathbb{R}^2 : \ (r_1)^{1/2} + (r_2)^{1/2} > 1, \ 0 < r_2 \le r_1 \le 1 \right\}$$

determines a class of non-degenerate similar triangles. In fact, the set  $\mathbb{G}$  consists of all possible non-degenerate triangles in  $\mathbb{R}^2$ . We call the curve  $\Gamma := \{(r_1, r_2) \in \mathbb{R}^2 : (r_1)^{1/2} + (r_2)^{1/2} = 1, 1/4 \leq r_1 \leq 1\}$  the lower boundary of  $\mathbb{G}$ . We illustrate in Figure 1 the region  $\mathbb{G}$  in  $\mathbb{R}^2$  and its low boundary  $\Gamma$ . By changing variables from K to  $\hat{K}$ , we get that there are three symmetric matrices  $\tilde{\mathbf{A}}_i, i = 0, 1, 2$ , associated with the reference triangle  $\hat{K}$  only such that  $\tilde{\mathbf{A}}_K = \alpha_K(\tilde{\mathbf{A}}_0 + r_{1,K}\tilde{\mathbf{A}}_1 + r_{2,K}\tilde{\mathbf{A}}_2)$  (cf. [12]). If  $\mathscr{T}$  is regular, there is a positive constant  $\theta_{\text{inf}}$  such that (3.10) holds. By the area formula of a triangle, we get that

(5.8) 
$$1/\sqrt{3} \le 1/(2\sin\theta_{\min,K}) \le \alpha_K \le 1/\tan\theta_{\min,K} \le 1/\tan\theta_{\inf}.$$

Motivated from inequality (5.6) and (5.8), for each  $(r_1, r_2) \in \mathbb{R}^2$ , we introduce the matrix  $\mathbf{H}(r_1, r_2) := \tilde{\mathbf{A}}_0 + r_1 \tilde{\mathbf{A}}_1 + r_2 \tilde{\mathbf{A}}_2 + \mathbf{E}$ . It is obvious that the elements of  $\mathbf{H}(r_1, r_2)$  are continuous functions of  $(r_1, r_2) \in \mathbb{R}^2$  and  $\mathbf{H}(r_{1,K}, r_{2,K}) = \alpha_K^{-1} \tilde{\mathbf{A}}_K + \mathbf{E}$ .

We next characterize the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$  in terms of the uniform boundedness below from zero of the minimum eigenvalues of the matrices  $\mathbf{H}(\cdot, \cdot)$ . We denote by  $\mathcal{T}_b$  the set of triangles in  $\mathcal{T}$  which have non-empty intersection with the boundary  $\partial\Omega$ . That is, for  $K \in \mathcal{T}_b$ , it may have edges or vertices on  $\partial\Omega$ . Then,  $\mathcal{T} \setminus \mathcal{T}_b$  is the set of triangles in  $\mathcal{T}$ , contained in the interior of  $\Omega$ . Recall that for each  $K \in \mathcal{T}$ ,  $\phi_{i,K}, i \in \mathbb{N}_{\hat{n}}$ , are defined by (3.7). For  $K \in \mathcal{T}_{b}$ , there are  $1 \leq i_{1} \leq i_{2} \leq \cdots \leq i_{n_{K}} \leq \hat{n}$  with  $1 \leq n_{K} < \hat{n}$  dependent on K such that (5.9)

 $\phi_{j,K}|_{\partial K \cap \partial \Omega} = 0, \ j \in \{i_m : m \in \mathbb{N}_{n_K}\}, \quad \phi_{j,K}|_{\partial K \cap \partial \Omega} \neq 0, \ j \in \mathbb{N}_{\hat{n}} \setminus \{i_m : m \in \mathbb{N}_{n_K}\}.$ For all  $w \in \mathbb{U}_{\mathcal{T}}$  and all  $K \in \mathcal{T}_b$ , since w vanishes on the boundary  $\partial \Omega$  and since the intersection of K and  $\partial \Omega$  is non-empty, we have that

(5.10) 
$$w(x) = \sum_{m \in \mathbb{N}_{n_K}} w_{i_m,K} \phi_{i_m,K}(x), \quad x \in K.$$

For a vector  $\mathbf{v}$ , we use  $(\mathbf{v})_i$  to denote its *i*th component, and for a matrix  $\mathbf{M}$ , we use  $(\mathbf{M})_{i,j}$  to denote its (i,j) entry. For a  $K \in \mathcal{T}_b$  and all  $w \in \mathbb{U}_{\mathcal{T}}$ , by (5.10) and the definition of  $\mathbf{w}_K$ , we have that

(5.11) 
$$(\mathbf{w}_K)_j = 0, \quad j \in \mathbb{N}_{\hat{n}} \setminus \{i_m : m \in \mathbb{N}_{n_K}\}.$$

For  $K \in \mathcal{T}_b$ , we define a subvector  $\mathbf{w}_K^- := [(\mathbf{w}_K)_{i_l} : l \in \mathbb{N}_{n_K}]^T$  of  $\mathbf{w}_K$  and a submatrix of  $\mathbf{H}(r_{1,K}, r_{2,K})$  by  $(\mathbf{H}^-(r_{1,K}, r_{2,K}))_{l,m} := (\mathbf{H}(r_{1,K}, r_{2,K}))_{i_l,i_m}$ .

The next theorem gives a necessary condition and a sufficient condition for the uniform local-ellipticity of the family of the discrete bilinear forms.

**Theorem 5.2.** Suppose that  $\mathscr{T}$  is regular and Hypotheses 1-5 hold. If  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic, then there is a positive constant c such that for all  $\mathcal{T} \in \mathscr{T}$ ,

(5.12) 
$$\begin{cases} \lambda_{\min}(\mathbf{H}(r_{1,K}, r_{2,K})) \ge c, & K \in \mathcal{T} \setminus \mathcal{T}_b, \\ \lambda_{\min}(\mathbf{H}^-(r_{1,K}, r_{2,K})) \ge c, & K \in \mathcal{T}_b. \end{cases}$$

Conversely, if there exists a positive constant c such that

(5.13) 
$$\lambda_{\min}(\mathbf{H}(r_{1,K}, r_{2,K})) \ge c, \text{ for all } \mathcal{T} \in \mathscr{T} \text{ and all } K \in \mathcal{T},$$

then  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic.

*Proof.* In this proof we write **H** for  $\mathbf{H}(r_{1,K}, r_{2,K})$  for simple presentation. We suppose that  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic and prove that there is a positive constant c such that (5.12) holds. By Lemma 5.1, we conclude that there is a positive constant  $\sigma$  such that for all  $\mathcal{T} \in \mathscr{T}$ , all  $K \in \mathcal{T}$  and all  $w \in \mathbb{U}_{\mathcal{T}}$  inequality (5.2) holds, which leads to (5.6). The regularity of  $\mathscr{T}$  ensures the validity of (5.8). Combining (5.6) and (5.8) yields

(5.14) 
$$\mathbf{w}_{K}^{T}\mathbf{H}\mathbf{w}_{K} \geq \frac{\sigma \tan \theta_{\inf}}{1+\sigma}\mathbf{w}_{K}^{T}\mathbf{w}_{K}.$$

For  $K \in \mathcal{T} \setminus \mathcal{T}_b$ , since w is an arbitrary element in  $\mathbb{U}_{\mathcal{T}}$ ,  $\mathbf{w}_K$  can be any element in  $\mathbb{R}^{\hat{n}}$ . In inequality (5.14), choosing  $\mathbf{w}_K$  as the eigenvector associated with the minimum eigenvalue of **H** we get the first inequality of (5.12). For  $K \in \mathcal{T}_b$ , from (5.11) we derive that

$$(\mathbf{w}_K^-)^T \mathbf{H}^- \mathbf{w}_K^- = \mathbf{w}_K^T \mathbf{H} \mathbf{w}_K$$
 and  $(\mathbf{w}_K^-)^T \mathbf{w}_K^- = \mathbf{w}_K^T \mathbf{w}_K$ .

This combined with (5.14) yields that for all  $\mathcal{T} \in \mathscr{T}$ , all  $K \in \mathcal{T}_b$  and all  $w \in \mathbb{U}_{\mathcal{T}}$ ,

(5.15) 
$$(\mathbf{w}_{K}^{-})^{T}\mathbf{H}^{-}\mathbf{w}_{K}^{-} \geq \frac{\sigma \tan \theta_{\inf}}{1+\sigma} (\mathbf{w}_{K}^{-})^{T}\mathbf{w}_{K}^{-}.$$

For  $K \in \mathcal{T}_b$ , as w goes throng  $\mathbb{U}_{\mathcal{T}}$ ,  $\mathbf{w}_K^-$  can be any element in  $\mathbb{R}^{n_K}$ . In inequality (5.15), choosing  $\mathbf{w}_K^-$  as the eigenvector associated with the minimum eigenvalue of  $\mathbf{H}^-$ , we conclude that the second inequality of (5.12) holds.

Conversely, we prove that  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic by contradiction. If  $\mathscr{A}_{\mathscr{T}}$  is not local-elliptic, from (5.2) we learn that for any  $\sigma > 0$ , there exist a  $\mathcal{T} \in \mathscr{T}$ , a  $K \in \mathcal{T}$  and a  $w \in \mathbb{U}_{\mathcal{T}}$  such that

(5.16) 
$$\mathbf{w}_K^T \tilde{\mathbf{A}}_K \mathbf{w}_K < \sigma (\mathbf{w}_K - \mathbf{E} \mathbf{w}_K)^T (\mathbf{w}_K - \mathbf{E} \mathbf{w}_K).$$

Let  $V_1 := \operatorname{span}\{\mathbf{e}\}$  and  $V_2 := \operatorname{span}\{\mathbf{v}_i, i \in \mathbb{N}_{\hat{n}-1}\}$ , where  $\mathbf{v}_i, i \in \mathbb{N}_{\hat{n}-1}$ , are the orthogonal eigenvectors of  $\mathbf{E}$  associated with the eigenvalue 0. Then,  $V_1$  is the eigenspace of  $\mathbf{E}$  associated with the eigenvalue 1 and  $V_2$  is the eigenspace of  $\mathbf{E}$  associated with the eigenvalue 0. Hypotheses 1-5 ensure the validity of Lemma 3.1. By making use of Lemma 3.1, we can easily prove that  $V_1$  is contained in the null space of  $\tilde{\mathbf{A}}_K$ . Note that  $\mathbb{R}^2 = V_1 + V_2$ . There is  $\mathbf{w}_{1,K} \in V_1$  and  $\mathbf{w}_{2,K} \in V_2$  such that  $\mathbf{w}_K = \mathbf{w}_{1,K} + \mathbf{w}_{2,K}$ . Then, we get that

(5.17) 
$$\mathbf{w}_K^T \tilde{\mathbf{A}}_K \mathbf{w}_K = \mathbf{w}_{2,K}^T \tilde{\mathbf{A}}_K \mathbf{w}_{2,K}, \quad (\mathbf{w}_K - \mathbf{E}\mathbf{w}_K)^T (\mathbf{w}_K - \mathbf{E}\mathbf{w}_K) = \mathbf{w}_{2,K}^T \mathbf{w}_{2,K}.$$

Substituting (5.17) into (5.16), we get that for each  $\sigma > 0$ , there exist a  $\mathcal{T} \in \mathscr{T}$ , a  $K \in \mathcal{T}$  and a  $\mathbf{w}_{2,K} \in V_2$  such that

(5.18) 
$$\mathbf{w}_{2,K}^T \tilde{\mathbf{A}}_K \mathbf{w}_{2,K} < \sigma \mathbf{w}_{2,K}^T \mathbf{w}_{2,K}.$$

Since  $\sigma$  can be sufficiently small, from (5.18) we derive that there exists a  $\mathbf{w}_{2,K_0} \in V_2$  such that  $\mathbf{w}_{2,K_0}^T \tilde{\mathbf{A}}_K \mathbf{w}_{2,K_0} \leq 0$ . Hence,  $\mathbf{w}_{2,K_0}^T \mathbf{H}(r_{1,K_0}, r_{2,K_0}) \mathbf{w}_{2,K_0} = \alpha_{K_0}^{-1} \mathbf{w}_{2,K_0}^T \tilde{\mathbf{A}}_{K_0} \mathbf{w}_{2,K_0} \leq 0$ . This contradicts (5.13). Therefore, we conclude that  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic.

From Theorem 5.2, we see that (5.12) is slightly weaker than (5.13). If we ignore the possibly weaker requirement for the boundary triangles in  $\mathcal{T}_b$ , inequality (5.13) is an equivalent condition for the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$ .

The discussion on the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$ , presented in Theorem 5.2, is from the algebraic point of view and can be used to derive geometric conditions for the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$ . For this purpose, we introduce the *admissible region*  $\mathbb{G}_{\mathbf{H}}$  (of the parameters  $r_1$  and  $r_2$ ) for the FVM primary grid

$$\mathbb{G}_{\mathbf{H}} := \left\{ (r_1, r_2) \in \mathbb{R}^2 : \ \lambda_{\min}(\mathbf{H}(r_1, r_2)) > 0, \ 0 \le r_2 \le r_1 \le 1 \right\}.$$

**Corollary 5.3.** Suppose that  $\mathscr{T}$  is regular and Hypotheses 1-5 hold. If  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic, then there exists a compact subset  $\mathbb{G}_0$  of  $\mathbb{G}_{\mathbf{H}}$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T} \setminus \mathcal{T}_b$ ,  $(r_{1,K}, r_{2,K}) \in \mathbb{G}_0$ . Conversely, if there exists a compact subset  $\mathbb{G}_0$  of  $\mathbb{G}_{\mathbf{H}}$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,  $(r_{1,K}, r_{2,K}) \in \mathbb{G}_0$ , then  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic.

*Proof.* We prove the first part of this corollary. Since  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic, there is a positive constant c such that (5.12) in Theorem 5.2 holds. Associated with this constant c, we define the set

$$\mathbb{G}_0 := \{ (r_1, r_2) \in \mathbb{R}^2 : \lambda_{\min}(\mathbf{H}(r_1, r_2)) \ge c, 0 \le r_2 \le r_1 \le 1 \}$$

Then, for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T} \setminus \mathcal{T}_b$ , there holds  $(r_{1,K}, r_{2,K}) \in \mathbb{G}_0$ , and clearly,  $\mathbb{G}_0 \subset \mathbb{G}_{\mathbf{H}}$ . It remains to prove that  $\mathbb{G}_0$  is compact. To this end, we introduce a compact set  $\mathbb{G}_1 := \{(r_1, r_2) \in \mathbb{R}^2 : 0 \leq r_2 \leq r_1 \leq 1\}$ . Obviously,  $\mathbb{G}_0 \subset \mathbb{G}_1$ . Since eigenvalues of a matrix depend continuously upon its entries (cf. [28]), there exists a point  $(r_1^*, r_2^*) \in \mathbb{G}_1$  such that  $\lambda^* := \lambda_{\min}(\mathbf{H}(r_1^*, r_2^*)) =$  $\max_{(r_1, r_2) \in \mathbb{G}_1} \lambda_{\min}(\mathbf{H}(r_1, r_2))$ . Since  $\mathbb{G}_0 \subset \mathbb{G}_1$ , the set  $\mathbb{G}_0$  has the following form:  $\mathbb{G}_0 = \{(r_1, r_2) \in \mathbb{R}^2 : c \leq \lambda_{\min}(\mathbf{H}(r_1, r_2)) \leq \lambda^*, 0 \leq r_2 \leq r_1 \leq 1\}.$  This means that  $\mathbb{G}_0$  is a compact set.

We prove the second part of this corollary. Since for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,  $(r_{1,K}, r_{2,K}) \in \mathbb{G}_0$  and  $\lambda_{\min}(\mathbf{H}(r_1, r_2))$  is a continuous function of  $(r_1, r_2)$  on the compact set  $\mathbb{G}_0$ , there is a positive constant c such that (5.13) holds. Thus, by Theorem 5.2, we conclude that  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic.

From Corollary 5.3, we see that if we ignore the possibly weaker requirement for the triangles in  $\mathcal{T}_b$ , the family  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic if and only if there exists a compact subset  $\mathbb{G}_0$  of  $\mathbb{G}_{\mathbf{H}}$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,  $(r_{1,K}, r_{2,K}) \in$  $\mathbb{G}_0$ . A similar result was presented in [12] for specific FVMs. Corollary 5.3 extends that geometric result to more general FVMs in the setting under Hypotheses 1-5. Moreover, we establish in this paper an algebraic characterization of the uniform local-ellipticity in Theorem 5.2, which easily leads to the geometric result presented in Corollary 5.3.

# 6. The admissible region and sufficient conditions for the uniform local-ellipticity

The admissible region defined in the last section is crucial in determining the uniform local-ellipticity of the family of the discrete bilinear forms. It is difficult to use the definition of the region directly in determining the uniform local-ellipticity. In this section, we provide a characterization of the admissible region in a way convenient for us to derive its compact subsets so as to obtain various sufficient conditions for the uniform local-ellipticity.

We use  $\lambda_{\min}(\mathbf{A}, \mathbf{B})$  and  $\lambda_{\max}(\mathbf{A}, \mathbf{B})$ , respectively, to denote the minimum generalized eigenvalue and the maximum generalized eigenvalue of  $\mathbf{B}$  with respect to  $\mathbf{A}$ . Let  $\lambda_{1,\mathbf{H}} := \lambda_{\max}\left(\mathbf{H}(1,1), \tilde{\mathbf{A}}_1 + \tilde{\mathbf{A}}_2\right)$  and  $\lambda_{2,\mathbf{H}}(r_1) := \lambda_{\max}\left(\mathbf{H}(r_1, r_1), \tilde{\mathbf{A}}_2\right)$ , and define  $r_{1,\mathbf{H}} := 1 - 1/\lambda_{1,\mathbf{H}}$  and  $r_{2,\mathbf{H}}(r_1) := r_1 - 1/\lambda_{2,\mathbf{H}}(r_1)$ . The next proposition extends the description of the admissible region for specific FVMs in [12] to more general constructions of FVMs. Since the proof is similar, we omit it.

**Proposition 6.1.** If Hypotheses 1-6 hold and the admissible region  $\mathbb{G}_{\mathbf{H}}$  is nonempty, then  $\mathbb{G}_{\mathbf{H}}$  is a convex subset of  $\mathbb{G}$  and there holds

$$\mathbb{G}_{\mathbf{H}} = \{ (r_1, r_2) \in \mathbb{R}^2 : r_{1,\mathbf{H}} < r_1 \le 1 \text{ and } r_{2,\mathbf{H}}(r_1) < r_2 \le r_1 \}.$$



FIGURE 2. The possible shapes of  $\mathbb{G}_{\mathbf{H}}$  described in Proposition 6.1

The description of the set  $\mathbb{G}_{\mathbf{H}}$  in Proposition 6.1 does not provide a precise shape of its lower boundary near the left end-point of the interval  $(r_{1,\mathbf{H}}, 1]$ . We next provide such a precise description. To this end, we introduce the low boundary of  $\mathbb{G}_{\mathbf{H}}$ :

(6.1) 
$$\Gamma_{\mathbf{H}} := \left\{ (r_1, r_{2,\mathbf{H}}(r_1)) \in \mathbb{R}^2 : r_{1,\mathbf{H}} < r_1 \le 1 \right\}.$$

From the definition of  $\mathbb{G}_{\mathbf{H}}$ , the curve  $\Gamma_{\mathbf{H}}$  in fact has the form

(6.2) 
$$\Gamma_{\mathbf{H}} = \{ (r_1, r_2) \in \mathbb{R}^2 : \lambda_{\min}(\mathbf{H}(r_1, r_2)) = 0, r_{1,\mathbf{H}} < r_1 \le 1 \}.$$

Figure 2 illustrates possible shapes of the admissible region. The next lemma excludes the possibility of the shape of  $\mathbb{G}_{\mathbf{H}}$  being the case illustrated in the left of Figure 2.

**Lemma 6.2.** If Hypotheses 1-6 hold and  $\mathbb{G}_{\mathbf{H}}$  is non-empty, then  $\Gamma_{\mathbf{H}}$  is a continuous curve and  $r' := \lim_{r_1 \to r_{1+\mathbf{H}}^+} r_{2,\mathbf{H}}(r_1)$  exists and is equal to  $r_{1,\mathbf{H}}$ .

Proof. Since  $\lambda_{\min}(\mathbf{H}(r_1, r_2))$  is a continuous function of  $(r_1, r_2) \in \mathbb{R}^2$ , by (6.2) we find that the curve  $\Gamma_{\mathbf{H}}$  is continuous. That is,  $r_{2,\mathbf{H}}(r_1)$  is a continuous function for  $r_1 \in (r_{1,\mathbf{H}}, 1]$ . It follows from Proposition 6.1 that  $0 \leq r_{2,\mathbf{H}}(r_1) < r_1 \leq 1$  for all  $r_1 \in (r_{1,\mathbf{H}}, 1]$ . This means that the continuous function  $r_{2,\mathbf{H}}(\cdot)$  is bounded on the domain  $(r_{1,\mathbf{H}}, 1]$ . We thus conclude that r' exists and  $r' \leq r_{1,\mathbf{H}}$ .

We now prove that  $r' = r_{1,\mathbf{H}}$  by contradiction. To this end, we introduce an auxiliary set  $\tilde{\mathbb{G}}_{\mathbf{H}} := \{(r_1, r_2) : \lambda_{\min}(\mathbf{H}(r_1, r_2) > 0, 0 \le r_1, r_2 \le 1\}$ . It is in fact the union of  $\mathbb{G}_{\mathbf{H}}$  and its reflection set with respect to the line  $r_1 = r_2$ . It can be easily verified that  $\tilde{\mathbb{G}}_{\mathbf{H}}$  is convex. Assume to the contrary that  $r' < r_{1,\mathbf{H}}$ . Then, there exists  $r_2$  such that  $r' < r_2 < r_{1,\mathbf{H}}$ . Since  $r_{2,\mathbf{H}}$  is a continuous function on  $(r_{1,\mathbf{H}}, 1]$ , from the definition of r' there exists

(6.3) 
$$0 < \varepsilon \le \min\{1 - r_{1,\mathbf{H}}, r_{1,\mathbf{H}} - r_2\}$$

such that

(6.4) 
$$r_{2,\mathbf{H}}(r_{1,\mathbf{H}} + \varepsilon) < r_2 < r_{1,\mathbf{H}} + \varepsilon.$$

It follows from (6.3), (6.4) and Proposition 6.1 that  $(r_{1,\mathbf{H}} + \varepsilon, r_2) \in \mathbb{G}_{\mathbf{H}}$ . From the relationship between  $\mathbb{G}_{\mathbf{H}}$  and  $\mathbb{\tilde{G}}_{\mathbf{H}}$ , we get that both  $(r_{1,\mathbf{H}} + \varepsilon, r_2)$  and  $(r_2, r_{1,\mathbf{H}} + \varepsilon)$  are in  $\mathbb{\tilde{G}}_{\mathbf{H}}$ . By the convexity of  $\mathbb{\tilde{G}}_{\mathbf{H}}$ , we observe that the point  $(r_{1,0}, r_{1,0}) := ((r_{1,\mathbf{H}} + \varepsilon + r_2)/2, (r_{1,\mathbf{H}} + \varepsilon + r_2)/2) \in \mathbb{\tilde{G}}_{\mathbf{H}}$ . By Proposition 6.1,  $(1,1) \in \mathbb{G}_{\mathbf{H}}$ . This leads to that  $(1,1) \in \mathbb{\tilde{G}}_{\mathbf{H}}$ . From (6.3), we find that  $(r_{1,\mathbf{H}}, r_{1,\mathbf{H}})$  is on the line segment between  $(r_{1,0}, r_{1,0})$  and (1,1). The convexity of  $\mathbb{\tilde{G}}_{\mathbf{H}}$  leads to  $(r_{1,\mathbf{H}}, r_{1,\mathbf{H}}) \in \mathbb{\tilde{G}}_{\mathbf{H}}$ , which in turn leads to  $(r_{1,\mathbf{H}}, r_{1,\mathbf{H}}) \in \mathbb{G}_{\mathbf{H}}$ . This contradicts Proposition 6.1 and we conclude that  $r' = r_{1,\mathbf{H}}$ .

Combining Proposition 6.1 and Lemma 6.2, we obtain the following theorem.

**Theorem 6.3.** If Hypotheses 1-6 hold and  $\mathbb{G}_{\mathbf{H}}$  is non-empty, then  $\mathbb{G}_{\mathbf{H}}$  is a convex subset of  $\mathbb{G}$  surrounded by lines  $r_1 = r_2$ ,  $r_1 = 1$  and curve  $\Gamma_{\mathbf{H}}$ , being close on  $r_1 = r_2$ ,  $r_1 = 1$ , and open on  $\Gamma_{\mathbf{H}}$ .

Based on the characterization of the admissible region  $\mathbb{G}_{\mathbf{H}}$  presented above, we may easily get its various compact subsets and in turn according to Corollary 5.3 we derive many types of sufficient conditions for the uniform local-ellipticity of the family of the discrete bilinear forms. Three types of sufficient conditions for the uniform local-ellipticity of the family of the discrete bilinear forms were presented in [12] for specific constructions of FVMs. The same types of sufficient conditions can be derived for the general construction in the setting under Hypotheses 1-6. Since the proofs for these three types of sufficient conditions are similar to those presented in [12], we shall provide the main results without proofs. The interested readers are referred to [12] for more details. Moreover, we shall provide a new sufficient condition which uses the largest angle of the triangles in the primary meshes.

Recalling that for a triangle K,  $\theta_{\min,K}$  denotes its smallest angle, we use  $\theta_{\max,K}$  to denote the largest angle of K. For a triangle K with  $r_{1,K}$  and  $r_{2,K}$  defined as in (5.7), we find that  $\cos \theta_{\min,K} = \frac{1+r_{1,K}-r_{2,K}}{2r_{1,K}^{1/2}}$  and  $\cos \theta_{\max,K} = \frac{r_{1,K}+r_{2,K}-1}{2(r_{1,K}r_{2,K})^{1/2}}$ . Motivated from these formulas, for any  $(r_1, r_2) \in \mathbb{R}^2$  with  $r_1 \neq 0$  and  $r_2 \neq 0$ , we define  $G(r_1, r_2) := \frac{1+r_1-r_2}{2(r_1)^{1/2}}$ ,  $\theta_{\min}(r_1, r_2) := \cos^{-1}(G(r_1, r_2))$ ,  $F(r_1, r_2) := \frac{r_1+r_2-1}{2(r_1r_2)^{1/2}}$  and  $\theta_{\max}(r_1, r_2) := \cos^{-1}(F(r_1, r_2))$ . For a given integer N, we select N+1 points  $R_l := (r_{1,l}, r_{2,l})$ ,  $l = 0, 1, \ldots, N$  in  $\Gamma_{\mathbf{H}}$ , where (6.5)

$$r_{1,0} := r_{1,\mathbf{H}}, \ r_{2,0} := r_{1,\mathbf{H}} \text{ and } r_{1,l} := r_{1,\mathbf{H}} + \frac{l(1-r_{1,\mathbf{H}})}{N}, \ r_{2,l} := r_{2,\mathbf{H}}(r_{1,l}), \ l \in \mathbb{N}_N.$$

**Type I sufficient condition:** For any  $S := (s_1, s_2) \in \mathbb{G}$  and  $T := (t_1, t_2) \in \mathbb{G}$ such that  $t_1 > s_1$ , we introduce  $d_1 := 1 - \frac{t_2 - s_2}{t_1 - s_1}$ ,  $d_2 := 1 + s_1 - s_2 - d_1 s_1$ , and  $h(t) := \frac{d_1 t + d_2}{2t^{1/2}}$ . Define  $\eta_1 := h(t_1)$ ,  $\eta_2 := \min \{\eta_1, h(s_1)\}$  and  $\eta_3 := \min \{\eta_2, h(d_2/d_1)\}$ , and compute

$$\eta(S,T) := \begin{cases} \cos^{-1} \eta_1, & d_1 = 0, \\ \cos^{-1} \eta_2, & d_1 \neq 0 \text{ and } \frac{d_2}{d_1} \notin [s_1, t_1], \\ \cos^{-1} \eta_3, & \text{otherwise.} \end{cases}$$

We have the type I sufficient condition on the uniform local-ellipticity of the family of the discrete bilinear forms of FVMs.

**Theorem 6.4.** If Hypotheses 1-6 hold and there exists  $\epsilon_0 > 0$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,

$$\theta_{\min}(r_{1,K}, r_{2,K}) \ge \epsilon_0 + \max_{0 \le l \le N-1} \eta(R_l, R_{l+1}),$$

then the family  $\mathscr{A}_{\mathscr{T}}$  of the discrete bilinear forms is uniformly local-elliptic.

**Type II sufficient condition:** We consider the curve on which the largest angle is a constant:  $\Gamma_{\gamma} := \{(r_1, r_2) \in \mathbb{G} : \theta_{\max}(r_1, r_2) = \cos^{-1}\gamma\}$ , for some  $\gamma \in (-1, 0]$ . We construct circumscribed polygons that approximate  $\Gamma_{\gamma}$ . For  $l = 0, 1, \ldots, N$ , let  $\rho_{\gamma}(N, l) := \gamma \cdot \cos\left(\frac{l\cos^{-1}|\gamma|}{N}\right)$ ,  $\sigma_{\gamma}(N, l) := (1 - \gamma^2)^{1/2} \cdot \sin\left(\frac{l\cos^{-1}|\gamma|}{N}\right)$ , and

$$\tau_c(N,l) := \begin{cases} 1, & l \text{ is even,} \\ \cos\left(\frac{\cos^{-1}|\gamma|}{N}\right), & l \text{ is odd.} \end{cases}$$

For j = 1, 2, we define

$$r_{j,\gamma,l} := \frac{1}{2(1-\gamma^2)} \left( 1 + \frac{\rho_{\gamma}(N,l) + (-1)^{j+1}\sigma_{\gamma}(N,l)}{\tau_{\gamma}(N,l)} \right)$$

for l = 0, 1, ..., N. We further restrict  $\gamma \in (1 - 1/(2r_{1,\mathbf{H}}), 0]$  so that  $r_{1,\gamma}(\alpha) \in (r_{1,\mathbf{H}}, 1]$  for all  $\alpha \in [0, \cos^{-1} |\gamma|)$ . Using the construction above, we can find the index set

$$\mathbb{K}_{\gamma} := \{ l \in \{0, 1, \dots, N-1\} : r_{2,\gamma,l} \le r_{2,\mathbf{H}}(r_{1,\gamma,l}) \text{ or } r_{2,\gamma,l+1} \le r_{2,\mathbf{H}}(r_{1,\gamma,l+1}) \}.$$

Let  $S_{\gamma,l} := (r_{1,\gamma,l}, r_{2,\mathbf{H}}(r_{1,\gamma,l})), \ l \in \mathbb{K}_{\gamma}$ . We have the type II sufficient condition on the uniform local-ellipticity of the family of the discrete bilinear forms of FVMs.

**Theorem 6.5.** If Hypotheses 1-6 hold and there exist  $\gamma \in (1 - 1/(2r_{1,\mathbf{H}}), 0]$  and  $\epsilon_0 > 0$ , such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,  $\theta_{\max}(r_{1,K}, r_{2,K}) \leq \cos^{-1} \gamma$  and  $\theta_{\min}(r_{1,K}, r_{2,K}) \geq \epsilon_0 + \max_{l \in \mathbb{K}_{\gamma}} \eta(S_{\gamma,l}, S_{\gamma,l+1}))$ , then the family  $\mathscr{A}_{\mathscr{T}}$  of the discrete bilinear forms is uniformly local-elliptic.

**Type III sufficient condition:** We have the type III sufficient condition on the uniform local-ellipticity of the family of the discrete bilinear forms of FVMs.

**Theorem 6.6.** If Hypotheses 1-6 hold and there exist  $\gamma \in (1 - 1/(2r_{1,\mathbf{H}}), 0]$  and  $\epsilon_0 > 0$ , such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,

$$\theta_{\max}(r_{1,K}, r_{2,K}) \le \cos^{-1}\gamma \quad and \quad \frac{r_{2,K}}{r_{1,K}} \ge \epsilon_0 + \max_{l \in \mathbb{K}_{\gamma}} \left\{ \frac{r_{2,\mathbf{H}}(r_{1,\gamma,l})}{r_{1,\gamma,l}}, \frac{r_{2,\mathbf{H}}(r_{1,\gamma,l+1})}{r_{1,\gamma,l+1}} \right\},$$

then the family  $\mathscr{A}_{\mathscr{T}}$  of the discrete bilinear forms is uniformly local-elliptic.

**Type IV sufficient condition**: We now derive a new sufficient condition for the uniform local-ellipticity based on an upper bound on the largest angle of the triangle meshes.

It has been verified in Lemma 4.9 of [12] for  $(r_1, r_2) \in \mathbb{R}^2$  with  $r_1 \neq 0$  and  $r_2 \neq 0$  that

(6.6) 
$$\frac{d\theta_{\max}(r_1, r_1)}{dr_1} < 0 \quad \text{and} \quad \frac{\partial\theta_{\max}(r_1, r_2)}{\partial r_2} < 0,$$

which implies certain monotonicity of  $\theta_{\max}(\cdot, \cdot)$ . Define

$$\varrho_0 := \inf_{(r_1, r_2) \in \Gamma_{\mathbf{H}}} \theta_{\max}(r_1, r_2).$$

The next proposition presents a sufficient condition for the uniform localellipticity of  $\mathscr{A}_{\mathscr{T}}$ .

**Proposition 6.7.** If Hypotheses 1-6 hold and there exists an  $\epsilon_0 > 0$  such that for all  $\mathcal{T} \in \mathcal{T}$  and all  $K \in \mathcal{T}$ ,

(6.7) 
$$\theta_{\max}(r_{1,K}, r_{2,K}) \le \varrho_0 - \epsilon_0$$

then the family  $\mathscr{A}_{\mathscr{T}}$  of the discrete bilinear forms is uniformly local-elliptic.

*Proof.* We prove this result by employing Corollary 5.3. Introducing the set

$$\mathbb{G}_{0,\epsilon_0} := \{ (r_1, r_2) \in \mathbb{G} : \pi/3 \le \theta_{\max}(r_1, r_2) \le \varrho_0 - \epsilon_0 \},\$$

by hypothesis, for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,  $(r_{1,K}, r_{2,K}) \in \mathbb{G}_{0,\epsilon_0}$ . Clearly,  $\mathbb{G}_{0,\epsilon_0}$  is compact.

It remains to prove that  $\mathbb{G}_{0,\epsilon_0}$  is contained in  $\mathbb{G}_{\mathbf{H}}$ . For any  $(r_1, r_2) \in \mathbb{G}_{0,\epsilon_0}$ , we show that  $r_1 > r_{1,\mathbf{H}}$  by contradiction. Assume to the contrary that  $r_1 \leq r_{1,\mathbf{H}}$ . Then by (6.6) and Lemma 6.2, we get that

(6.8) 
$$\theta_{\max}(r_1, r_2) \ge \theta_{\max}(r_1, r_1) \ge \theta_{\max}(r_{1,\mathbf{H}}, r_{1,\mathbf{H}}) \ge \varrho_0.$$

On the other hand, since  $(r_1, r_2) \in \mathbb{G}_{0,\epsilon_0}$ , we have that  $\theta_{\max}(r_1, r_2) < \varrho_0$ . This contradicts (6.8). Thus, we conclude that  $r_1 > r_{1,\mathbf{H}}$ . We next verify that  $r_2 > r_{2,\mathbf{H}}(r_1)$ . From the definitions of  $\mathbb{G}_{0,\epsilon_0}$  and  $\varrho_0$ , we get that  $\theta_{\max}(r_1, r_2) \leq \varrho_0 - \epsilon_0 < \varrho_0 \leq \theta_{\max}(r_1, r_{2,\mathbf{H}}(r_1))$ . This together with the second inequality of (6.6) yields  $r_2 > r_{2,\mathbf{H}}(r_1)$ . Thus, from Proposition 6.1, we know that  $(r_1, r_2) \in \mathbb{G}_{\mathbf{H}}$ . That is,  $\mathbb{G}_{0,\epsilon_0} \subset \mathbb{G}_{\mathbf{H}}$ .

The computation of  $\rho_0$  is not easy. From (6.5), we get a polygonal curve that approximates  $\Gamma_{\mathbf{H}}$ . We next show that as N to infinity, the polygonal curves produce a sequence of the numbers that are smaller than  $\rho_0$  and converge to  $\rho_0$ . For points  $P_1, P_2 \in \mathbb{R}^2$ , we let  $\overline{P_1P_2}$  denote the closed line segment between  $P_1$  and  $P_2$ . Let  $\Gamma_{\mathbf{H},N} := \bigcup_{l=0}^{N-1} \overline{R_l R_{l+1}} \text{ and } \varrho_{0,N} := \inf_{(r_1,r_2) \in \Gamma_{\mathbf{H},N}} \theta_{\max}(r_1,r_2).$  Since the polygonal curves  $\Gamma_{\mathbf{H},N}$  can approximate  $\Gamma_{\mathbf{H}}$  to an arbitrary accuracy as N increases and the curve  $\Gamma_{\mathbf{H},N}$  is continuous, we have that  $\lim_{N\to\infty} \varrho_{0,N} = \varrho_0$ . The next lemma shows that  $\rho_{0,N}$  is smaller than  $\rho_0$ .

**Lemma 6.8.** If Hypotheses 1-6 hold, then for any positive integer N, there holds  $\varrho_{0,N} \leq \varrho_0.$ 

*Proof.* Hypotheses 1-6 ensure the validity of Theorem 6.3, from which we know that  $r_{2,\mathbf{H}}(r)$  is a convex function for  $r \in (r_{1,\mathbf{H}}, 1]$ . Hence, for each  $(r_1, r_2) \in \Gamma_{\mathbf{H},N}$ , we have that  $r_2 \ge r_{2,\mathbf{H}}(r_1)$ . This together with the second inequality of (6.6) leads to that  $\theta_{\max}(r_1, r_2) \leq \theta_{\max}(r_1, r_{2,\mathbf{H}}(r_1))$ , which yields the desired result. 

We next consider computing  $\rho_{0,N}$ . For  $S := (s_1, s_2) \in \mathbb{G}$  and  $T := (t_1, t_2) \in \mathbb{G}$ such that  $t_1 > s_1$ , we let  $d_1 := 1 + \frac{t_2 - s_2}{t_1 - s_1}$ ,  $d_2 := -1 + s_1 + s_2 - d_1 s_1$ ,  $d_3 := -1 + s_1 + s_2 - d_1 s_1$  $d_1 + 2d_2 - d_1d_2, d_4 := \frac{d_2(d_2+1)}{d_3}.$  We define  $g(t) := \frac{d_1t + d_2}{2[(d_1-1)t^2 + (d_2+1)t]^{1/2}},$  for t > 0, $\vartheta_1 := \max \{ g(s_1), g(t_1) \}, \ \vartheta_2 := \max \{ \vartheta_1, g(d_4) \}, \ \text{and compute}$ 

$$\vartheta\left(S,T\right) := \begin{cases} \cos^{-1}\vartheta_1, & d_3 = 0, \text{ or } d_3 \neq 0 \text{ and } d_4 \notin [s_1, t_1], \\ \cos^{-1}\vartheta_2, & \text{otherwise.} \end{cases}$$

**Lemma 6.9.** If  $\Gamma_{\mathbf{H}} \subset \mathbb{G}$ , then  $\varrho_{0,N} = \min_{l \in \mathbb{N}_N} \vartheta(R_{l-1}, R_l)$ .

*Proof.* From the definition of  $\rho_{0,N}$ , we find that

$$\varrho_{0,N} = \min_{l \in \mathbb{N}_N} \inf_{(r_1, r_2) \in \overline{R_{l-1}R_l}} \theta_{\max}\left(r_1, r_2\right).$$

Since  $\Gamma_{\mathbf{H}} \subset \mathbb{G}$ , it is clear that  $R_l \in \mathbb{G}$ ,  $l = 0, 1, \ldots, N$ . Thus, it suffices to prove that for any  $S := (s_1, s_2) \in \mathbb{G}$  and  $T := (t_1, t_2) \in \mathbb{G}$  with  $t_1 > s_1$ , there holds

(6.9) 
$$\inf_{(r_1,r_2)\in\overline{ST}}\theta_{\max}\left(r_1,r_2\right)=\vartheta\left(S,T\right).$$

Since  $\overline{ST}$  is a line segment between S and T, we have the relation

(6.10) 
$$r_2 = (d_1 - 1)r_1 + d_2 + 1, \quad (r_1, r_2) \in \overline{ST}.$$

Substituting (6.10) into the definition of  $\theta_{\max}$ , we get  $\theta_{\max}(r_1, r_2) = \cos^{-1}(g(r_1))$ ,  $(r_1, r_2) \in \overline{ST}$ . Noticing that the function  $\cos^{-1}(\cdot)$  is decreasing, we derive that

(6.11) 
$$\inf_{(r_1, r_2) \in \overline{ST}} \theta_{\max}(r_1, r_2) = \inf_{r_1 \in [s_1, t_1]} \cos^{-1}(g(r_1)) = \cos^{-1}\left(\sup_{r_1 \in [s_1, t_1]} g(r_1)\right).$$
  
By employing the monotonicity of  $g(\cdot)$ , we get (6.9).

By employing the monotonicity of  $g(\cdot)$ , we get (6.9).

Recall that the curve  $\Gamma$  is the lower boundary of the set  $\mathbb{G}$ . The next theorem gives the type IV convenient sufficient condition on the uniform local-ellipticity of Ag.

**Theorem 6.10.** Suppose that Hypotheses 1-6 hold. If  $\Gamma_{\mathbf{H}} = \Gamma$  and there exists  $\epsilon_1 > 0$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,

(6.12) 
$$\theta_{\max}(r_{1,K}, r_{2,K}) \le \pi - \epsilon_1,$$

then  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic. If  $\Gamma_{\mathbf{H}} \subset \mathbb{G}$  and there exists  $\epsilon_2 > 0$  such that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$ ,

(6.13) 
$$\theta_{\max}(r_{1,K}, r_{2,K}) \le \min_{l \in \mathbb{N}_N} \vartheta(R_{l-1}, R_l) - \epsilon_2,$$

then  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic.

*Proof.* Since Hypotheses 1-6 holds and  $\Gamma_{\mathbf{H}} = \Gamma$ , it follows from Theorem 6.3 that  $\mathbb{G}_{\mathbf{H}} = \mathbb{G}$ . We find that  $\mathbb{G}_{\epsilon_1} := \{(r_1, r_2) \in \mathbb{G} : \pi/3 \leq \theta_{\max}(r_1, r_2) \leq \pi - \epsilon_1\}$  is a compact subset of  $\mathbb{G}$ . By Corollary 5.3,  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic.

Since Hypotheses 1-6 hold and  $\Gamma_{\mathbf{H}} \subset \mathbb{G}$ , from inequality (6.13) and Lemmas 6.9 and 6.8, we get that for all  $\mathcal{T} \in \mathscr{T}$  and all  $K \in \mathcal{T}$  that  $\theta_{\max}(r_{1,K}, r_{2,K}) \leq \varrho_{0,N} - \epsilon_2 \leq \varrho_0 - \epsilon_2$ . By Proposition 6.7, we conclude that  $\mathscr{A}_{\mathscr{T}}$  is uniformly local-elliptic.  $\Box$ 

From Theorem 6.10, we get that when  $\Gamma_{\mathbf{H}} = \Gamma$  (that is,  $\mathbb{G}_{\mathbf{H}} = \mathbb{G}$ ), the uniform local-ellipticity of the family of the discrete bilinear forms of the FVM is always satisfied for FVMs with the trial spaces constructed from any non-degenerate triangulations.



FIGURE 3. Two types of sufficient conditions for the Lagrange 5-0 FVM

The type IV sufficient condition can provide information for determining the uniform local-ellipticity of  $\mathscr{A}_{\mathscr{T}}$  complementary to that provided by the first three type sufficient conditions. We provide an example to demonstrate this point. The Lagrange 5-0 FVM was constructed in [12]. Proposition 5.3 of [12] shows that if the smallest angle of the triangles in the triangulation is greater than 35.69°, then the family of the corresponding discrete bilinear forms is uniformly local-elliptic. We shall use the type IV sufficient condition for this FVM to give an alternative condition. Specifically, using Theorem 6.10 of this paper with N = 100 we calculate that if the greatest angle of the triangles in the triangulation is smaller than 78.76°, then the family of the discrete bilinear forms for the Lagrange 5-0 FVM is uniformly local-elliptic. We illustrate in Figure 3 these two types of sufficient conditions for the Lagrange 5-0 FVM. The red shadow region  $\mathbb{G}_r$  illustrates the set of triangles with the largest angles smaller than 78.76°. We see from the figure that these two regions are not identical.

# 7. Specific FVM constructions

We present in this section specific FVM constructions and examine their uniform local-ellipticity. According to the construction of the test spaces, there are three kinds of FVM schemes: Lagrange FVMs, Hermite FVMs and hybrid FVMs. Since

621

the first two types of FVMs have been studied extensively, our discussion will focus on the third type of FVMs.

Hybrid quadratic FVMs was introduced in [9]. In the following, we present more higher-order hybrid FVMs and derive their geometric conditions which ensure that the corresponding FVMs achieve optimal error estimate. From the construction described in Section 2, for a FVM scheme it suffices to present the FE triple element  $(\hat{K}, \hat{\Sigma}, \mathbb{U}_{\hat{K}})$  and the triple element  $(\hat{\mathcal{T}}^*, \hat{\Sigma}^*, \mathbb{V}_{\hat{\mathcal{T}}^*})$ . For m > 2 and m points  $Q_j \in \mathbb{R}^2$ ,  $j \in \mathbb{N}_m$ , we use  $\Theta\{Q_1, Q_2, \ldots, Q_m\}$  for the polygon with the vertices  $Q_i, i \in \mathbb{N}_m$ , being connected consecutively. For a non-negative integer k and a domain  $\mathcal{D} \subset \mathbb{R}^2$ , by  $\wp_k(\mathcal{D})$  we denote the space of functions that are polynomials of degree k on  $\mathcal{D}$ and zero elsewhere. For  $\hat{K}_i^* \in \hat{\mathcal{T}}^*$ , we write  $\chi_i$  for  $\chi_{\hat{K}^*}$  for a simple presentation.

**Example 7.1** (Hybrid Lagrange cubic FVM). For the hybrid Lagrange cubic FVM, we choose the following functional nodes:

$$\begin{array}{ll} P_1 := (0,0), & P_2 := (1,0), & P_3 := (0,1), & P_4 := (1/3,1/3), & P_5 := (2/3,1/3), \\ P_6 := (1/3,2/3), & P_7 := (0,2/3), & P_8 := (0,1/3), & P_9 := (1/3,0), & P_{10} := (2/3,0). \end{array}$$

For  $i \in \mathbb{N}_{10}$ , we let  $\hat{\zeta}_i$  be the point evaluation functional at the node  $P_i$  and we have the set of degrees of freedom  $\hat{\Sigma} := \left\{\hat{\zeta}_i : i \in \mathbb{N}_{10}\right\}$ . The trial space  $\mathbb{U}_{\hat{K}}$  is the cubic element space  $\wp_3(\hat{K})$ .

We choose the dual nodes  $Q_0 := (1/3, 1/3), Q_1 := (1/2, 1/2), Q_2 := (0, 1/2), Q_3$ := (1/2, 0), and define the dual partition  $\hat{\mathcal{T}}^* := \{\hat{K}_l^* : l \in \mathbb{N}_3\}$  of  $\hat{K}$  by setting  $\hat{K}_1^* := \Theta \{P_1, Q_3, Q_0, Q_2\}, \hat{K}_2^* := \Theta \{P_2, Q_1, Q_0, Q_3\}, \hat{K}_3^* := \Theta \{P_3, Q_2, Q_0, Q_1\}.$  Let  $\xi_1 := (-1, -1), \xi_2 := (2, -1)$  and  $\xi_3 := (-1, 2)$  and define  $\bar{w}_4 := \lim_{\epsilon \to 0+} \frac{1}{3} [w(P_4 + \epsilon \xi_1) + w(P_4 + \epsilon \xi_2) + w(P_4 + \epsilon \xi_3)]$ , which denotes the average of the function w at the node  $P_4$  in the directions  $\xi_i, i \in \mathbb{N}_3$ . Clearly,  $\bar{w}_4$  is well-defined for  $w \in \mathbb{U}_{\hat{K}}$  and when  $w \in \mathbb{U}_{\hat{K}}, \ \bar{w}_4 = w(P_4)$ . The set of degrees of freedom  $\hat{\Sigma}^* := \{\hat{\eta}_l, l \in \mathbb{N}_{\hat{n}}\}$ , where  $\hat{n} = 10$ , and

(7.1) 
$$\hat{\eta}_i := \hat{\zeta}_i, \ i \in \mathbb{N}_3, \quad \hat{\eta}_4(w) := \bar{w}_4 - \frac{1}{3}[w(P_1) + w(P_2) + w(P_3)],$$

$$\hat{\eta}_5(w) := w(P_5) - w(P_2), \quad \hat{\eta}_6(w) := w(P_6) - w(P_3), \quad \hat{\eta}_7(w) := w(P_7) - w(P_3), \\ \hat{\eta}_8(w) := w(P_8) - w(P_1), \quad \hat{\eta}_9(w) := w(P_9) - w(P_1), \quad \hat{\eta}_{10}(w) := w(P_{10}) - w(P_2).$$

The basis  $\Phi_{\hat{K}} := {\hat{\varphi}_i \in \varphi_3(\hat{K}), i \in \mathbb{N}_{10}}$  for  $\mathbb{U}_{\hat{K}}$  is determined by  $\hat{\Sigma}^*$  in the form of (2.4). The test space is chosen as  $\mathbb{V}_{\hat{\mathcal{T}}^*} := \operatorname{span}\Psi_{\hat{\mathcal{T}}^*}$ , where  $\Psi_{\hat{\mathcal{T}}^*}$  consists of  $\hat{\psi}_i := \chi_i, i \in \mathbb{N}_3$  and  $\hat{\psi}_i := \hat{\varphi}_i, i \in \mathbb{N}_{10} \setminus \mathbb{N}_3$ .

We remark that  $\bar{w}_4$  is well-defined for  $w \in \mathbb{V}_{\hat{\mathcal{T}}^*}$ . In Figure 4, we illustrate the control volumes  $\hat{K}_1^*$  (the red region),  $\hat{K}_2^*$  (the yellow region),  $\hat{K}_3^*$  (the green region), and the directions  $\xi_i, i \in \mathbb{N}_3$ . For  $w \in \mathbb{V}_{\hat{\mathcal{T}}^*}$ , since w may have a jump in the intersections of two adjacent control volumes,  $w(P_4)$  has no meanings. However, for each  $i \in \mathbb{N}_3$ , for any  $\epsilon > 0$ ,  $P_4 + \epsilon \xi_i$  is in the control volume  $\hat{K}_i^*$ , which means that  $w(P_4 + \epsilon \xi_i)$  is well-defined. Thus,  $\bar{w}_4$  is well-defined for  $w \in \mathbb{V}_{\hat{\mathcal{T}}^*}$ .

**Example 7.2** (Hybrid Zienkiewicz FVM). For the hybrid Zienkiewicz FVM, we choose the functional nodes  $P_1 := (0,0)$ ,  $P_2 := (1,0)$  and  $P_3 := (0,1)$ , and the



FIGURE 4. The dual partition and  $\bar{w}_4$  for the hybrid Lagrange cubic FVM

set of degrees of freedom  $\hat{\Sigma} := \left\{ \hat{\zeta}_i, i \in \mathbb{N}_9 \right\}$ , where  $\hat{\zeta}_1(w) := w(P_1), \hat{\zeta}_2 := w(P_2),$  $\hat{\zeta}_3 := w(P_3),$ 

$$\begin{array}{rclrcl} \hat{\zeta}_4(w) & := & \nabla w(P_1) \cdot (1,0), & \hat{\zeta}_5(w) & := & \nabla w(P_1) \cdot (0,1), & \hat{\zeta}_6(w) & := & \nabla w(P_2) \cdot (-1,1), \\ \hat{\zeta}_7(w) & := & \nabla w(P_2) \cdot (-1,0), & \hat{\zeta}_8(w) & := & \nabla w(P_3) \cdot (0,-1), & \hat{\zeta}_9(w) & := & \nabla w(P_3) \cdot (1,-1). \end{array}$$

Let  $P_0 := (1/3, 1/3)$ . The trial space  $\mathbb{U}_{\hat{K}}$  is the cubic element space with element w satisfying

(7.2) 
$$6w(P_0) - 2\sum_{i \in \mathbb{N}_3} w(P_i) + \sum_{i \in \mathbb{N}_3} \nabla w(P_i) \cdot (P_i - P_0) = 0.$$

The dual partition  $\hat{\mathcal{T}}^* := \{\hat{K}_l^* : l \in \mathbb{N}_3\}$  of this FVM is the same as that in Example 7.1. We choose the set of degrees of freedom  $\hat{\Sigma}^* := \hat{\Sigma}$ . The basis  $\Phi_{\hat{K}} := \{\hat{\varphi}_i \in \mathbb{U}_{\hat{K}}, i \in \mathbb{N}_9\}$  for  $\mathbb{U}_{\hat{K}}$  is determined by (2.4) and (7.2). The test space is chosen as  $\mathbb{V}_{\hat{\mathcal{T}}^*} := \operatorname{span}\Psi_{\hat{\mathcal{T}}^*}$ , where its basis  $\Psi_{\hat{\mathcal{T}}^*}$  consists of  $\hat{\psi}_i = \chi_i, i \in \mathbb{N}_3$ , and  $\hat{\psi}_i = \hat{\varphi}_i, i \in \mathbb{N}_9 \setminus \mathbb{N}_3$ .

**Example 7.3** (Type one hybrid hermite cubic FVM). We choose the functional nodes  $P_1 := (0,0), P_2 := (1,0), P_3 := (0,1), P_4 := (1/3,1/3)$  and the set of degrees of freedom  $\hat{\Sigma} := \left\{ \hat{\zeta}_i : i \in \mathbb{N}_{10} \right\}$ , where  $\hat{\zeta}_i(w) := w(P_i), i \in \mathbb{N}_4$ , and

$$\begin{aligned} \hat{\zeta}_5(w) &:= \nabla w(P_1) \cdot (1,0), \quad \hat{\zeta}_6(w) &:= \nabla w(P_1) \cdot (0,1), \quad \hat{\zeta}_7(w) &:= \nabla w(P_2) \cdot (-1,1), \\ \hat{\zeta}_8(w) &:= \nabla w(P_2) \cdot (-1,0), \quad \hat{\zeta}_9(w) &:= \nabla w(P_3) \cdot (0,-1), \quad \hat{\zeta}_{10}(w) &:= \nabla w(P_3) \cdot (1,-1). \end{aligned}$$

The trial space  $\mathbb{U}_{\hat{K}}$  is chosen as the complete cubic element space  $\wp_3(\hat{K})$ .

We choose the dual nodes  $Q_1 = (1/2, 1/2)$ ,  $Q_2 = (0, 1/2)$  and  $Q_3 = (1/2, 0)$  and define the dual partition  $\hat{\mathcal{T}}^* := {\hat{K}_l^* : l \in \mathbb{N}_4}$  of  $\hat{K}$  by setting

$$\begin{split} \hat{K}_1^* &:= \Theta \left\{ P_1, Q_3, Q_2 \right\}, \ \hat{K}_2^* &:= \Theta \left\{ P_2, Q_1, Q_3 \right\}, \\ \hat{K}_3^* &:= \Theta \left\{ P_3, Q_2, Q_1 \right\}, \ \hat{K}_4^* &:= \Theta \left\{ Q_1, Q_2, Q_3 \right\}. \end{split}$$

We choose the set of degrees of freedom  $\hat{\Sigma}^* := \hat{\Sigma}$ , which uniquely determines the basis  $\Phi_{\hat{K}} := \{\hat{\varphi}_i \in \varphi_3(\hat{K}) : i \in \mathbb{N}_{10}\}$  for  $\mathbb{U}_{\hat{K}}$  by (2.4). The test space is chosen as  $\mathbb{V}_{\hat{\mathcal{T}}^*} := \operatorname{span} \Psi_{\hat{\mathcal{T}}^*}$ , where its basis  $\Psi_{\hat{\mathcal{T}}^*}$  consists of  $\hat{\psi}_i := \chi_i, i \in \mathbb{N}_4$ , and  $\hat{\psi}_i := \hat{\varphi}_i, i \in \mathbb{N}_{10} \setminus \mathbb{N}_4$ .

**Example 7.4** (Type two hybrid hermite cubic FVM). The FE triple element  $(\hat{K}, \hat{\Sigma}, \mathbb{U}_{\hat{K}})$  for this example is the same as that for Example 7.3, with a different choice of the test space. Specifically, the dual partition  $\hat{\mathcal{T}}^* := {\hat{K}_l^* : l \in \mathbb{N}_3}$  of



FIGURE 5. The admissible regions  $\mathbb{G}_{\mathbf{H}}$  for the hybrid FVMs

this example is the same as that in Example 7.1. The set of degrees of freedom  $\hat{\Sigma}^* := {\hat{\eta}_l, l \in \mathbb{N}_{10}}$ , where  $\hat{\eta}_l := \hat{\zeta}_l, \ l \in \mathbb{N}_{10}$  and  $l \neq 4$ , and  $\hat{\eta}_4$  is defined as (7.1). By (2.4),  $\hat{\Sigma}^*$  uniquely determines a basis  $\Phi_{\hat{K}} := {\hat{\varphi}_i \in \varphi_3(\hat{K}), i \in \mathbb{N}_{10}}$  for  $\mathbb{U}_{\hat{K}}$ . The test space is chosen as  $\mathbb{V}_{\hat{\mathcal{T}}^*} := \operatorname{span}\Psi_{\hat{\mathcal{T}}^*}$ , where its basis  $\Psi_{\hat{\mathcal{T}}^*}$  consists of  $\hat{\psi}_i := \chi_i, i \in \mathbb{N}_3$ , and  $\hat{\psi}_i := \hat{\varphi}_i, i \in \mathbb{N}_{10} \setminus \mathbb{N}_3$ .

We now examine the uniform local-ellipticity conditions for these FVMs. It is straightforward to verify that the hybrid FVMs described above all satisfy Hypotheses 1-6. For convenience of representation, we shall use HLC, HZ, HHC1 and HHC2 to denote the FVMs described in Examples 7.1-7.4, respectively. The admissible regions  $\mathbb{G}_{\mathbf{H}}$  for the four hybrid FVMs are shown in Figure 5, where the curves labeled as "HLC", "HZ", "HHC1" and "HHC2" mark the lower boundary  $\Gamma_{\mathbf{H}}$  of  $\mathbb{G}_{\mathbf{H}}$  for the corresponding hybrid FVMs. In this figure, additional curves are included for better understanding. The curve labeled with "Valid Triangle" is the lower boundary  $\Gamma$  of  $\mathbb{G}$ . We compute the four types of sufficient conditions for these FVMs by using Theorems 6.4-6.6 and 6.10 with N = 100. We report the results in the following proposition.

**Proposition 7.5.** For the hybrid FVMs constructed in Examples 7.1-7.4, if for each triangular element in the triangulation its smallest angle is greater than the value listed in the fist line of Table 1, or its largest angle is not greater than  $90^{\circ}$  and its smallest angle is greater than the value listed in the second line of Table 1, or its largest angle is not greater than  $90^{\circ}$  and the ratio of its shortest edge length versus its second shortest edge length is greater than the value listed in the last line of Table 1, then the family of the corresponding bilinear forms is uniformly local-elliptic.

	HLC	ΗZ	HHC1	HHC2	
Type I	$20.96^{\circ}$	$37.70^{\circ}$	$42.93^{\circ}$	$37.70^{\circ}$	
	$(\approx 0.1164\pi)$	$(\approx 0.20994\pi)$	$(\approx 0.2385\pi)$	$(\approx 0.20944\pi)$	
Type II	$15.39^{\circ}$	$30.73^{\circ}$	$37.75^{\circ}$	$30.97^{\circ}$	
	$(\approx 0.0855\pi)$	$(\approx 0.1707\pi)$	$(\approx 0.2097\pi)$	$(\approx 0.1721\pi)$	
Type III	0.2756	0.5964	0.7749	0.6005	
Type IV	$82.31^{\circ}$	$75.02^{\circ}$	$72.29^{\circ}$	$74.88^{\circ}$	
	$(\approx 0.4573\pi)$	$(\approx 0.4168\pi)$	$(\approx 0.4016\pi)$	$(\approx 0.4160\pi)$	

TABLE 1. Types I-IV sufficient conditions for the hybrid FVMs



FIGURE 6. A triangulation of the region  $\overline{\Omega}$ 

# 8. Numerical examples

In this section, we present numerical examples to verify the theoretical convergence order of two new FVM schemes HLC and HHC2 constructed in this paper. All the experiments presented here are performed on a personal computer with 2.66 GHz CPU and 4 Gb RAM. Moreover, Matlab 7.7 is used as the testing platform and the direct algorithm is used to solve the resulting linear systems.

We consider solving the Poisson equation (2.1) with  $f(x, y) := 2(x^2 + y^2 - x - y)$ and  $\Omega := (0, 1) \times (0, 1)$ . The exact solution of the boundary value problem is given by u(x, y) = -x(x - 1)y(y - 1),  $(x, y) \in \overline{\Omega}$ . We subdivide the region  $\overline{\Omega}$  to  $M \times N$  equal rectangles. The triangle mesh of  $\overline{\Omega}$  is then obtained by connecting the diagonal lines of the resulting rectangles, as illustrated by Figure 6 with M = 2and N = 4. Without loss of generality, we assume that  $M \leq N$ . By  $\theta_{\min}$  we denote the minimum angle of all the triangles in the triangulation. Clearly, we have that  $\tan \theta_{\min} = M/N$ . We form the dual partition by using the barycenter dual partition which is obtained by connecting the barycenter and the midpoints of each edge of the triangles in the triangulation.

In the following examples, we use n to denote the number of the unknowns in the corresponding linear system of FVMs and  $\mathscr{E}_{\mathcal{T}} := |u - u_{\mathcal{T}}|_1$  to denote the  $H^1$ -error between the exact solution u of the boundary value problem and the approximate solution  $u_{\mathcal{T}}$ .

 $\theta_{\min}$ 45 $\theta_{\min}$ 18.43 $\theta_{\rm min}$ 1.43(M, N)C.O (M, N)C.O. (M, N) $\mathscr{E}_{\mathcal{T}}$ C.O. 81  $\mathscr{E}_{T}$ 4.90e-3 1.00e-2(2, 2)49(1, 3)40(4, 4)169 5.94e-4 3.04(2, 6)133 1.21e-3 3.05(1, 40)7.15e-4 484(8, 8)625 7.29e-5 3.02(4, 12)481 1.46e-4 3.04(2, 80)1687 1.60e-42.16(4, 160)(8, 320)1.55e-5 1.73e-6 (16, 16)24019.01e-6 3.02(8, 24)18251.80e-5 3.0262533.379409 (16.48)7105 2.23e-6 3.01240253.16(32, 32)1.12e-6 3.01(64.64)372491.50e-7 2.90(32.96)28033 2.84e-72.97(16, 640)94129 2.23e-62.01

TABLE 2. The numerical results of HLC

**Example 8.1** (Numerical results for HLC). In this example, we use the HLC described in Example 7.1 to solve the boundary value problem. In Table 2, we report the computed  $H^1$ -errors  $\mathscr{E}_{\mathcal{T}}$  and the convergence orders (C.O.) of the HLC with different triangulations having different minimum angles. When  $\theta_{\min} = 45^{\circ}$  or  $\theta_{\min} \approx 18.43^{\circ}$ , which satisfy the Type II sufficient condition listed in Proposition 7.5 and Theorem 4.2, the theoretical convergence order of  $\mathscr{E}_{\mathcal{T}}$  is 3. This is confirmed by the numerical results. We also list the numerical results for M : N = 1 : 40 ( $\theta_{\min} \approx 1.43^{\circ}$ ), in which case the parameter ( $r_{1,K}, r_{2,K}$ ) = (1600/1601, 1/1601) of

the triangle K in the triangulation is not in the admissible region  $\mathbb{G}_{\mathbf{H}}$  of the HLC. Therefore, the uniform local-ellipticity of the family of the discrete bilinear forms is not guaranteed to be satisfied. The numerical results show that the computed convergence order of  $\mathscr{E}_{\mathcal{T}}$  of the HLC oscillates around 3.

**Example 8.2** (Numerical results for HHC2). In this example, we use the HHC2 described in Example 7.4 to solve the boundary value problem. Numerical results of this example are reported in Table 3. It follows from the Type II sufficient condition listed in Proposition 7.5 and Theorem 4.2 that when  $\theta_{\min} = 45^{\circ}$  or  $\theta_{\min} \approx 33.69^{\circ}$ , the theoretical convergence order of  $\mathscr{E}_{\tau}$  is 3. This is confirmed by the numerical results. We also find that when M : N = 1 : 20 ( $\theta_{\min} \approx 2.86^{\circ}$ ), even though in this case the uniform local-ellipticity of the family of the discrete bilinear forms is not guaranteed to be satisfied, the computed convergence order for this case is still the optimal order 3. This result demonstrates that the uniform local-ellipticity of the family of the discrete bilinear forms is only sufficient but not necessary to ensure the optimal error estimate of the FVMs.

TABLE 3. The numerical results of HHC2

$\theta_{\min}$	=	$45^{\circ}$		$\theta_{\min}$	≈	$33.69^{\circ}$		$\theta_{\min}$	≈	$2.86^{\circ}$	
(M, N)	n	$\mathscr{E}_{\mathcal{T}}$	С.О.	(M, N)	n	$\mathscr{E}_{\mathcal{T}}$	С.О.	(M, N)	n	$\mathscr{E}_{\mathcal{T}}$	С.О.
(2, 2)	35	1.41e-2		(2, 3)	48	8.02e-3					
(4, 4)	107	1.79e-3	2.98	(4, 6)	153	1.02e-3	2.97	(1, 20)	166	4.33e-3	
(8, 8)	371	2.30e-4	2.96	(8, 12)	543	1.31e-4	2.96	(2, 40)	529	5.36e-4	3.01
(16, 16)	1379	2.93e-5	2.97	(16, 24)	2043	1.67e-5	2.98	(4, 80)	1855	6.72e-5	2.99
(32, 32)	5315	3.70e-6	2.99	(32, 48)	7923	2.10e-6	2.99	(8, 160)	6907	8.43e-6	3.00
(64, 64)	20867	4.67e-7	2.98	(64, 96)	31203	2.71e-7	2.95	(16, 320)	26611	1.07e-6	2.98

#### Acknowledgments

The authors of this paper are grateful to the referees for their constructive comments which leads to an improvement of the presentation of this paper. The third author would like to thank Dr. Junfeng Wu for providing computer codes for verifying the uniform local-ellipticity of the Lagrange and Hermite FVMs.

## References

- Ivov Babuška and A. K. Aziz, Lectures on the Mathematical Foundations of the Finite Element Method, University of Maryland, College Park, Washington DC, 1972, Technical Note BN-748.
- [2] Randolph E. Bank and Donald J. Rose, Some error estimates for the box method, SIAM J. Numer. Anal. 24 (1987), no. 4, 777–787, DOI 10.1137/0724050. MR899703 (88j:65235)
- [3] Martin Berzins and Justin M. Ware, Positive cell-centered finite volume discretization methods for hyperbolic equations on irregular meshes, Appl. Numer. Math. 16 (1995), no. 4, 417–438, DOI 10.1016/0168-9274(95)00007-H. MR1325257 (96a:65122)
- [4] Susanne C. Brenner and L. Ridgway Scott, The Mathematical Theory of Finite Element Methods, Texts in Applied Mathematics, vol. 15, Springer-Verlag, New York, 1994. MR1278258 (95f:65001)
- [5] Zhi Qiang Cai, On the finite volume element method, Numer. Math. 58 (1991), no. 7, 713–735, DOI 10.1007/BF01385651. MR1090257 (92d:65188)
- [6] Zhi Qiang Cai, Jan Mandel, and Steve McCormick, The finite volume element method for diffusion equations on general triangulations, SIAM J. Numer. Anal. 28 (1991), no. 2, 392– 402, DOI 10.1137/0728022. MR1087511 (92j:65165)
- [7] Zhi Qiang Cai and Steve McCormick, On the accuracy of the finite volume element method for diffusion equations on composite grids, SIAM J. Numer. Anal. 27 (1990), no. 3, 636–655, DOI 10.1137/0727039. MR1041256 (91d:65182)
- [8] Panagiotis Chatzipantelidis and Raytco D. Lazarov, Error estimates for a finite volume element method for elliptic PDEs in nonconvex polygonal domains, SIAM J. Numer. Anal. 42 (2005), no. 5, 1932–1958, DOI 10.1137/S0036142903427639. MR2139231 (2006f:65108)

- [9] Long Chen, A new class of high order finite volume methods for second order elliptic equations, SIAM J. Numer. Anal. 47 (2010), no. 6, 4021–4043, DOI 10.1137/080720164. MR2585177 (2011j:65251)
- [10] Zhong Ying Chen, The error estimate of generalized difference method of 3rd-order Hermite type for elliptic partial differential equations, Northeast. Math. J. 8 (1992), no. 2, 127–135. MR1182874 (93e:65125)
- [11] Zhong Ying Chen, Superconvergence of generalized difference method for elliptic boundary value problem, Numer. Math. J. Chinese Univ. (English Ser.) 3 (1994), no. 2, 163–171. MR1325662 (96b:65106)
- [12] Zhongying Chen, Junfeng Wu, and Yuesheng Xu, Higher-order finite volume methods for elliptic boundary value problems, Adv. Comput. Math. 37 (2012), no. 2, 191–253, DOI 10.1007/s10444-011-9201-8. MR2944051
- [13] Zhongying Chen and Yuesheng Xu, The Petrov-Galerkin and iterated Petrov-Galerkin methods for second-kind integral equations, SIAM J. Numer. Anal. 35 (1998), no. 1, 406–434 (electronic), DOI 10.1137/S0036142996297217. MR1618413 (99h:65214)
- [14] So-Hsiang Chou and Do Y. Kwak, Multigrid algorithms for a vertex-centered covolume method for elliptic problems, Numer. Math. 90 (2002), no. 3, 441–458, DOI 10.1007/s002110100288. MR1884225 (2002k:65209)
- [15] So-Hsiang Chou and Qian Li, Error estimates in L<sup>2</sup>, H<sup>1</sup> and L<sup>∞</sup> in covolume methods for elliptic and parabolic problems: a unified approach, Math. Comp. 69 (2000), no. 229, 103–120, DOI 10.1090/S0025-5718-99-01192-8. MR1680859 (2000i:65134)
- [16] So-Hsiang Chou, Do Y. Kwak, and Panayot S. Vassilevski, Mixed covolume methods for elliptic problems on triangular grids, SIAM J. Numer. Anal. 35 (1998), no. 5, 1850–1861 (electronic), DOI 10.1137/S0036142997321285. MR1639954 (99k:65105)
- [17] So-Hsiang Chou and Panayot S. Vassilevski, A general mixed covolume framework for constructing conservative schemes for elliptic problems, Math. Comp. 68 (1999), no. 227, 991– 1011, DOI 10.1090/S0025-5718-99-01090-X. MR1648371 (99j:65203)
- [18] So-Hsiang Chou and Xiu Ye, Unified analysis of finite volume methods for second order elliptic problems, SIAM J. Numer. Anal. 45 (2007), no. 4, 1639–1653 (electronic), DOI 10.1137/050643994. MR2338403 (2008i:65230)
- [19] Philippe G. Ciarlet, The Finite Element Method for Elliptic Problems, Studies in Mathematics and its Applications, Vol. 4, North-Holland Publishing Co., Amsterdam, 1978. MR0520174 (58 #25001)
- [20] Victor Eijkhout and Panayot Vassilevski, The role of the strengthened Cauchy-Buniakowskii-Schwarz inequality in multilevel methods, SIAM Rev. 33 (1991), no. 3, 405–419, DOI 10.1137/1033098. MR1124360 (92h:65051)
- [21] Philippe Emonot, Methodes de volumes elements finis: applications aux equations de Navier-Stokes et resultats de convergence, Dissertation, Lyon (1992)
- [22] Richard E. Ewing, Tao Lin, and Yanping Lin, On the accuracy of the finite volume element method based on piecewise linear polynomials, SIAM J. Numer. Anal. 39 (2002), no. 6, 1865– 1888, DOI 10.1137/S0036142900368873. MR1897941 (2003d:65105)
- [23] Robert Eymard, Tierry Gallouët, and Raphaèle Herbin, A cell-centered finite-volume approximation for anisotropic diffusion operators on unstructured meshes in any space dimension, IMA J. Numer. Anal. 26 (2006), no. 2, 326–353, DOI 10.1093/imanum/dri036. MR2218636 (2007a:65170)
- [24] Isabelle Faille, A control volume method to solve an elliptic equation on a two-dimensional irregular mesh, Comput. Methods Appl. Mech. Engrg. 100 (1992), no. 2, 275–290, DOI 10.1016/0045-7825(92)90186-N. MR1187634 (93j:65167)
- [25] Helmer André Friis, Michael G. Edwards, and Johannes Mykkeltveit, Symmetric positive definite flux-continuous full-tensor finite-volume schemes on unstructured cell-centered triangular grids, SIAM J. Sci. Comput. **31** (2008/09), no. 2, 1192–1220, DOI 10.1137/070692182. MR2466154 (2009i:65193)
- [26] Wolfgang Hackbusch, On first and second order box schemes (English, with German summary), Computing 41 (1989), no. 4, 277–296, DOI 10.1007/BF02241218. MR993825 (90g:65136)
- [27] Bernd Heinrich, Finite Difference Methods on Irregular Networks, ISNM 82, 1987, Birkhauser Verlag. MR1015930

- [28] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985. MR832183 (87e:15001)
- [29] Jianguo Huang and Shitong Xi, On the finite volume element method for general selfadjoint elliptic problems, SIAM J. Numer. Anal. 35 (1998), no. 5, 1762–1774, DOI 10.1137/S0036142994264699. MR1640017 (99g:65104)
- [30] Kwang Y. Kim, Error estimates for a mixed finite volume method for the p-Laplacian problem, Numer. Math. 101 (2005), no. 1, 121–142, DOI 10.1007/s00211-005-0610-9. MR2194721 (2006j:65315)
- [31] Rong Hua Li, Generalized difference methods for a nonlinear Dirichlet problem, SIAM J. Numer. Anal. 24 (1987), no. 1, 77–88, DOI 10.1137/0724007. MR874736 (88c:65091)
- [32] Ronghua Li, Zhongying Chen, and Wei Wu, Generalized Difference Methods for Differential Equations, Monographs and Textbooks in Pure and Applied Mathematics, vol. 226, Marcel Dekker Inc., New York, 2000. Numerical analysis of finite volume methods. MR1731376 (2000j:65003)
- [33] Ronghua Li and P. Zhu, Generalized difference methods for second order elliptic partial differential equations (I) - triangle grids, Numer. Math. J. Chinese Universities, 2 (1982), 140-152.
- [34] Frank Liebau, The finite volume element method with quadratic basis functions (English, with English and German summaries), Computing 57 (1996), no. 4, 281–299, DOI 10.1007/BF02252250. MR1422087 (97m:65206)
- [35] Yonghai Li, Shi Shu, Yuesheng Xu, and Qingsong Zou, Multilevel preconditioning for the finite volume method, Math. Comp. 81 (2012), no. 279, 1399–1428, DOI 10.1090/S0025-5718-2012-02582-8. MR2904584
- [36] Junliang Lv and Yonghai Li, L<sup>2</sup> error estimates and superconvergence of the finite volume element methods on quadrilateral meshes, Adv. Comput. Math. 37 (2012), no. 3, 393–416, DOI 10.1007/s10444-011-9215-2. MR2970858
- [37] Richard H. Macneal, An asymmetrical finite difference network, Quart. Math. Appl. 11 (1953), 295–310. MR0057631 (15,257d)
- [38] Thorsten Schmidt, Box schemes on quadrilateral meshes (English, with English and German summaries), Computing 51 (1993), no. 3-4, 271–292, DOI 10.1007/BF02238536. MR1253406 (94k:65151)
- [39] Ming Zhong Tian and Zhong Ying Chen, A generalized difference method with quadratic elements for elliptic equations (Chinese, with English summary), Numer. Math. J. Chinese Univ. 13 (1991), no. 2, 99–113. MR1142295 (92j:65156)
- [40] Alan M. Winslow, Numerical solution of the quasilinear Poisson equation in a nonuniform triangle mesh, J. Computational Phys. 1 (1967), 149–172. MR0241008 (39 #2353)
- [41] Haijun Wu and Ronghua Li, Error estimates for finite volume element methods for general second-order elliptic problems, Numer. Methods Partial Differential Equations 19 (2003), no. 6, 693–708, DOI 10.1002/num.10068. MR2009589 (2004i:65108)
- [42] Jinchao Xu and Qingsong Zou, Analysis of linear and quadratic simplicial finite volume methods for elliptic equations, Numer. Math. 111 (2009), no. 3, 469–492, DOI 10.1007/s00211-008-0189-z. MR2470148 (2009i:65199)
- [43] Zhimin Zhang and Qingsong Zou, A family of finite volume schemes of arbitrary order on rectangular meshes, J. Scientific Computing 58 (2014), no. 2, 308–330. MR3150261

GUANGDONG PROVINCE KEY LABORATORY OF COMPUTATIONAL SCIENCE, SCHOOL OF MATHE-MATICS AND COMPUTATIONAL SCIENCES, SUN YAT-SEN UNIVERSITY, GUANGZHOU 510275, PEOPLE'S REPUBLIC OF CHINA

E-mail address: lnsczy@mail.sysu.edu.cn

GUANGDONG PROVINCE KEY LABORATORY OF COMPUTATIONAL SCIENCE, SCHOOL OF MATHE-MATICS AND COMPUTATIONAL SCIENCES, SUN YAT-SEN UNIVERSITY, GUANGZHOU 510275, PEOPLE'S REPUBLIC OF CHINA – AND – DEPARTMENT OF MATHEMATICS, SYRACUSE UNIVERSITY, SYRACUSE, NEW YORK 13244

E-mail address: yxu06@syr.edu

GUANGDONG PROVINCE KEY LABORATORY OF COMPUTATIONAL SCIENCE, SCHOOL OF MATHE-MATICS AND COMPUTATIONAL SCIENCES, SUN YAT-SEN UNIVERSITY, GUANGZHOU 510275, PEOPLE'S REPUBLIC OF CHINA

E-mail address: yy0dd@126.com

628