A NEW CLASS OF HIGH ORDER FINITE VOLUME METHODS FOR SECOND ORDER ELLIPTIC EQUATIONS

LONG CHEN *

Abstract. In the numerical simulation of many practical problems in physics and engineering, finite volume methods are an important and popular class of discretization methods due to the local conservation and the capability of discretizing domains with complex geometry. However they are limited by low order approximation since most existing finite volume methods use piecewise constant or linear function space to approximate the solution. In this paper, a new class of high order finite volume methods for second order elliptic equations is developed by combining high order finite element methods and linear finite volume methods. Optimal convergence rate in H^1-norm for our new quadratic finite volume methods over two dimensional triangular or rectangular grids is obtained.

Key words. finite element, finite volume, discretization, error estimates, high order methods

AMS subject classifications. 65N10, 65N30

1. Introduction.

In this paper, we shall develop a new class of high order finite volume methods for solving the second order elliptic equation:

$$-\nabla\cdot(\boldsymbol{K}(\boldsymbol{x})\nabla u) = f \quad \text{for all } \boldsymbol{x}\in\Omega\subset\mathbb{R}^n, \tag{1.1}$$

with appropriate Dirichlet or Neumann boundary condition. The diffusion coefficient $\boldsymbol{K}(\boldsymbol{x})$ is a symmetric and positive definite $n\times n$ matrix function satisfying

$$0<a_0|\boldsymbol{\xi}|^2\leq\boldsymbol{\xi}^t\boldsymbol{K}(\boldsymbol{x})\boldsymbol{\xi}\leq a_1|\boldsymbol{\xi}|^2<\infty \quad \text{for all } \boldsymbol{x}\in\Omega \text{ and } \boldsymbol{\xi}\in\mathbb{R}^n. \tag{1.2}$$

It is well known that the smoothness requirement of the classic solution to (1.1), i.e., $u\in C^2(\Omega)$, excludes interesting solutions for many physical problems.

The weak solution of (1.1) is a function $u\in H_0^1(\Omega)$ such that

$$\int_\Omega(\boldsymbol{K}\nabla u)\cdot\nabla v\,\mathrm{d}\boldsymbol{x}=\int_\Omega fv\,\mathrm{d}\boldsymbol{x} \quad \text{for all } v\in H_0^1(\Omega). \tag{1.3}$$

Here, to fix ideas, we consider the homogenous Dirichlet boundary condition, i.e., $u|_{\partial\Omega}=0$ and $f\in L^2(\Omega)$. The existence and uniqueness of the weak solution can be easily established by the Lax-Milligram lemma. Restriction of the weak formulation (1.3) to finite element subspaces of $H_0^1(\Omega)$ leads to finite element methods (FEMs) which are flexible to deal with complex domains and various boundary conditions. Furthermore, the theory on the convergence of finite element methods is well established. The main drawback of FEM might be the loss of the local conservation property which can be fundamental for the simulation of many physical models, e.g., in computational fluid dynamics.

To derive discretization methods with local conservation property, we note that many physical models can be written as the following balance equation [18]

$$-\int_{\partial b}(\boldsymbol{K}\nabla u)\cdot\boldsymbol{n}\,\mathrm{d}S=\int_b f\,\mathrm{d}\boldsymbol{x} \quad \text{for all } b\subset\Omega. \tag{1.4}$$

The discretization of (1.4) by choosing an appropriate finite element space $\mathbb{V}$ to approximate u and a finite number of sub-domains b, the so-called control volume, will be called finite volume methods (FVMs).

*Department of Mathematics, University of California at Irvine, CA, 92697 (chenlong@math.uci.edu)
The author was supported in part by NSF Grant DMS-0811272, and in part by NIH Grant P50GM76516 and R01GM75309.

Since FVM discretizes the balance equation (1.4) directly, an obvious virtue is the local conservation property which is not the case for FEM. On the other hand, FVM inherits the intrinsic geometric flexibility of FEM and thus is more flexible than standard finite difference methods which mainly defined on structured grids of simple domains.

One of the main limitation of FVM is the low approximation order. For most existing finite volume methods, the space $\mathbb{V}$ is either a piecewise constant or a linear finite element spaces. Few work [23, 22, 8, 29, 25] is devoted to high order finite volume methods. Among them, a systematic way of deriving high order finite volume methods is presented in [25] for one dimensional elliptic problems and in [8] for cell-centered finite volume methods over rectangular grids.

We shall propose a new class of vertex-centered high order FVM by mixing the discretization of the balance equation (1.4) and the weak formulation (1.3). Our new method can be thought as a hybridization of high order finite element methods and a linear finite volume method. More precisely, we shall first formulate (1.4) into a Petrov-Galerkin formulation, by translating the left hand side of (1.4) into a bilinear form involving different trial and test function spaces. Then we design new high order finite volume methods by the following choices of trial and test spaces. Given a triangulation $\mathcal{T}$ of Ω, the trial space will be chosen as kth-order finite element space $\mathbb{V}_{k,\mathcal{T}}$ in which the function u is approximated. A novelty of our new method is on the choice of the test space. Using the hierarchical decomposition $\mathbb{V}_{k,\mathcal{T}} = \mathbb{V}_{1,\mathcal{T}} \oplus \mathbb{W}_{k,\mathcal{T}}$, where $\mathbb{V}_{1,\mathcal{T}}$ is the linear finite element space, the test space will be chosen by replacing $\mathbb{V}_{1,\mathcal{T}}$ by $\mathbb{V}_{0,\mathcal{B}}$, a piecewise constant function space on a dual mesh $\mathcal{B}$.

The error analysis is not easy for arbitrary orders since the stability (or in general the inf-sup condition) for the resulting algebraic system is difficult to establish. In this paper we only obtain inf-sup condition for quadratic finite volume methods on two dimensional triangular grids (assuming the geometry of the mesh is not too extreme) and rectangular grids. Optimal rate of convergence in H^1-norm is then obtained following the framework of Xu and Zou [29]. Due to the hierarchical structure of the trial and test spaces, our analysis is simplified to the verification of the positive semi-definiteness of the symmetrization of the local stiffness matrix in each element.

Note that existing quadratic finite volume methods [23, 22, 29] require control volumes for all basis of the trial space. While in our new method, we only need to choose control volumes for vertices of the triangulation. This will simplify the geometry of control volumes and in turn simplify the implementation and analysis. Indeed we shall show when $\boldsymbol{K}$ is piecewise constant, the resulting matrix equation is different from that of standard finite element methods only in one small block. Thus we can make use of vast existing finite element codes to easily implement our new method.

The rest of this paper is organized as follows. In Section 2, we present finite volume methods including our new class of high order finite volume methods. In Section 3, we give general error analysis of our methods. In Section 4, we study new quadratic finite volume methods in detail on triangular and rectangular grids in one and two dimensions. In Section 5, we present a numerical example to show the effectiveness of our methods. In the last section, we summarize our results and outline future work.

2. Finite volume methods. In this section, we shall present a general form of finite volume methods and give two examples: cell-centered and vertex-centered FVMs. We then formulate the vertex-centered FVM into a Petrov-Galerkin method and develop high order schemes using different choices of trial and test spaces.

2.1. General form of finite volume methods. Finite volume methods are discretizations of the balance equation (1.4) consisting of three approximations:

1. approximate the function u by u_h in an N-dimensional subspace $\mathbb{V}$;

2. approximate "arbitrary domain $b \subset \Omega$" by a finite subset $\mathcal{B} = \{b_i, i = 1 : M\}$;
3. approximate boundary flux $(\boldsymbol{K}\nabla u) \cdot \boldsymbol{n}$ on ∂b_i by a discrete one $(\boldsymbol{K}\nabla_h u_h) \cdot \boldsymbol{n}$.

We then end up with a method: find $u_h \in \mathbb{V}$ such that:

$$-\int_{\partial b_i} (\boldsymbol{K}(\boldsymbol{x})\nabla_h u_h) \cdot \boldsymbol{n}\, \mathrm{d}S = \int_{b_i} f\, \mathrm{d}\boldsymbol{x} \quad \text{for all } b_i \subset \Omega, i = 1 : M. \tag{2.1}$$

We call any method in the form (2.1) *finite volume methods* (FVMs).

Usually $\mathcal{B}$ forms a partition of Ω or an approximation Ω_h of Ω such that we have local conservation property on each b_i and thus the whole domain Ω or Ω_h by linear composition of control volumes. Furthermore $\mathbb{V}$ and $\mathcal{B}$ should be chosen so that the resulting matrix equation is solvable. We shall give two examples below.

Example 1: Cell-centered finite volume method. Let $\mathcal{T}$ be a triangular or rectangular grid of Ω. We choose the finite dimensional space $\mathbb{V} = \{v \in L^2(\Omega) : v|_\tau \text{ is constant } \}$. Then $\dim \mathbb{V} = N$, the number of elements of $\mathcal{T}$. We also choose $\mathcal{B} = \mathcal{T}$.

Since a control volume is an element (also called cell) of the mesh and the unknown is associated to each cell, it is often called cell-centered finite volume methods or cell-centered difference methods.

The boundary flux of each element can be approximated in a finite difference fashion. Theory and computation along this approach are summarized in the book [19]. Another approach to discretize the boundary flux is through mixed finite element methods. Optimal error estimate can be easily obtained by using that of mixed finite element methods [26].

Deriving high order finite volume methods from mixed finite element methods is a promising approach since theories on mixed methods are well established [5]. We refer to [8] for high order cell-centered finite volume methods over rectangle grids. However, the derivation on general unstructured triangulation is still open. Partially it is due to the loss of symmetry and good numerical quadrature for simplicial grids.

Example 2: Vertex-centered finite volume method. We now discuss another popular choice of $\mathbb{V}$ and $\mathcal{B}$. To fix ideas, we consider two dimensional triangular grids and homogenous Dirichlet boundary condition. We refer to [29] for a general treatment on simplicial grids in any dimensions.

Let $\Omega \subset \mathbb{R}^2$ be a polygon and let $\mathcal{T}$ be a triangulation of Ω. Denoted by $\mathbb{V}_{1,\mathcal{T}}$ the linear finite element spaces of $H_0^1(\Omega)$ based on $\mathcal{T}$:

$$\mathbb{V}_{1,\mathcal{T}} = \{v \in H_0^1(\Omega) : v|_\tau \in \mathcal{P}_1(\tau),\ \forall\, \tau \in \mathcal{T}\},$$

where $\mathcal{P}_k(\tau)$ is the kth order polynomial space on τ. We shall choose $\mathbb{V} = \mathbb{V}_{1,\mathcal{T}}$. The dimension $N = \dim \mathbb{V}$ is the number of interior vertices of $\mathcal{T}$.

The control volume will be given by another mesh $\bar{\mathcal{B}} = \{b_i, i = 1, \cdots, M\}$ satisfying

$$\bar{\Omega} = \cup_{i=1}^M b_i, \quad \text{and} \quad \overset{\circ}{b}_i \cap \overset{\circ}{b}_j = \varnothing \quad \text{for all } 1 \le i, j \le M \text{ and } i \ne j.$$

To reflect to the Dirichlet boundary condition, we set

$$\mathcal{B} = \{b_i \in \bar{\mathcal{B}}, b_i \subset \overset{\circ}{\Omega}\}.$$

Obviously for Neumann boundary condition, we should use $\bar{\mathcal{B}}$. The control volume b_i is not necessary to be polygons. But for practical reasons, each b_i is chosen as a polygon so that the boundary integral is easy to evaluate.

Note that for a function $u \in \mathbb{V}_{1,\mathcal{T}}$, the flux $(\boldsymbol{K}\nabla u)\cdot \boldsymbol{n}$ is not well defined on the edges of triangles. Therefore we further require that

$$(\partial b_i \cap \tau) \subset \overset{\circ}{\tau} \quad \text{for all } b_i \in \mathcal{B} \text{ and } \tau \in \mathcal{T}.$$

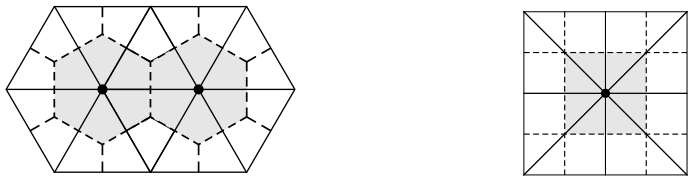
We then get a natural approximation $(\boldsymbol{K}\nabla u_h)\cdot \boldsymbol{n}$ of the flux $(\boldsymbol{K}\nabla u)\cdot \boldsymbol{n}$ on ∂b_i since $u|_\tau$ is a polynomial.

Given a triangulation $\mathcal{T}$, one popular construction of $\bar{\mathcal{B}}$ is given as follows: for each triangle $\tau \in \mathcal{T}$, select a point $c_\tau \in \tau$. The point c_τ can coincide with one of the middle points of edges, but not the vertices of triangles (to avoid the degeneracy of the control volume). In each triangle, we connect c_τ to three middle points on the edges of τ. This will divide each triangle in $\mathcal{T}$ into three regions. For each vertex x_i of $\mathcal{T}$, we collect all regions containing this vertex and define it as b_i. In Figure 2.1 we draw the control volume for interior vertices since the unknown is associated to interior vertices only.

The classical choices of the point c_τ include the circumcenter and the barycenter. When c_τ is chosen as the circumcenter of τ, the edges of control volumes will be orthogonal to the intersected edges of triangles, and if the mesh $\mathcal{T}$ is a Delaunay triangulation, $\mathcal{B}$ will be a Voronoi diagram. When c_τ is the barycenter of τ, then τ will be divided into three parts with equal areas. This symmetric property is important to get optimal L^2 convergence rate for the FVMs [20]. In this paper, we shall consider the choice of c_τ of the following two types:

- Type A: c_τ is the barycenter of τ.
- Type B: c_τ is the middle point of the longest edge

Type A is preferable for triangulations composed by equilateral triangles and type B is better for right triangles; See Figure 2.1. We shall call $\mathcal{B}$ a *dual mesh* of $\mathcal{T}$.



(a) Type A (b) Type B

FIG. 2.1. *Two types of meshes and dual meshes. The gray area is the control volume of interior nodes. Type A: the point c_τ is the barycenter of τ. Type B: the point c_τ is the middle point of the longest edge.*

Since we associate control volumes and unknowns to vertices, it is called vertex-centered finite volume method. It is also known as box method [28, 3, 20] (since the control volume is called box in these work), and finite volume element methods [9, 10, 7, 21] (to emphasis the approximation of u is in a finite element space).

2.2. Petrov-Galerkin formulation. We shall follow Bank and Rose [3] to formulate the vertex-centered linear finite volume method as a Petrov-Galerkin method.

We first introduce a function space defined on control volumes. Let $\mathcal{B}$ be the dual mesh of a triangulation $\mathcal{T}$ constructed in the previous subsection. We define a piecewise constant function space on $\mathcal{B}$ by:

$$\mathbb{V}_{0,\mathcal{B}} = \{v \in L^2(\Omega) : v|_{b_i} = \text{ constant }, \text{ for all } b_i \in \mathcal{B}\}. \tag{2.2}$$

The set of interior edges of the mesh $\mathcal{B}$ is denoted by $\mathcal{E}(\mathcal{B})$. For each $e \in \mathcal{E}(\mathcal{B})$, we shall fix a unit normal direction $\boldsymbol{n}_e$ of e. Suppose e is shared by two control volumes b_i and b_j. Without

loss of generality, we assume the outward normal direction of e in b_i coincides with $\boldsymbol{n}_e$. For any function $v \in \mathbb{V}_{0,\mathcal{B}}$, the jump of v across e is denoted by $[v] = v|_{b_i} - v|_{b_j}$.

We define a bilinear form on $\mathbb{V}_{1,\mathcal{T}} \times \mathbb{V}_{0,\mathcal{B}}$ as

$$\bar{a}(u,v) = -\sum_{e\in\mathcal{E}(\mathcal{B})}\int_e (\boldsymbol{K}\nabla u)\cdot\boldsymbol{n}_e[v]\,\mathrm{d}S \quad \text{for all } u \in \mathbb{V}_{1,\mathcal{T}}, v \in \mathbb{V}_{0,\mathcal{B}}, \tag{2.3}$$

and formulate the linear finite volume method as: find $\bar{u} \in \mathbb{V}_{1,\mathcal{T}}$ such that

$$\bar{a}(\bar{u},v) = (f,v) \quad \text{for all } v \in \mathbb{V}_{0,\mathcal{B}}. \tag{2.4}$$

REMARK 2.1. For Neumann boundary condition, we shall choose

$$\mathbb{V}_{1,\mathcal{T}} = \{v \in H^1(\Omega) : v|_\tau \in \mathcal{P}_1(\tau) \quad \text{for all } \tau \in \mathcal{T}\}, \text{ and}$$
$$\mathbb{V}_{0,\mathcal{B}} = \{v \in L^2(\Omega) : v|_{b_i} = \text{constant} \quad \text{for all } b_i \in \bar{\mathcal{B}}\}.$$

For $e \in \partial b_i \cap \partial\Omega$, the flux $(\boldsymbol{K}\nabla u)\cdot\boldsymbol{n}_e$ will be given by the boundary condition. Other type of boundary conditions can be built into the finite element space or the weak formulation. All algorithms and analysis in this paper can be applied to these boundary conditions in a straightforward way. □

We now show a close relation between the linear finite element method and the linear finite volume method. Let $a(u,v)$ be the bilinear form

$$a(u,v) = \int_\Omega (\boldsymbol{K}(\boldsymbol{x})\nabla u)\cdot\nabla v\,\mathrm{d}\boldsymbol{x}. \tag{2.5}$$

The linear finite element method is: find $u_L \in \mathbb{V}_{1,\mathcal{T}}$ such that

$$a(u_L,v) = (f,v) \quad \text{for all } v \in \mathbb{V}_{1,\mathcal{T}}. \tag{2.6}$$

To see the close relation, we formulate the corresponding matrix equations for (2.4) and (2.6). Let $\mathcal{N}(\mathcal{T})$ be the set of interior nodes of $\mathcal{T}$ and $N = \#\mathcal{N}(\mathcal{T})$. Then $\dim\mathbb{V}_{0,\mathcal{B}} = \dim\mathbb{V}_{1,\mathcal{T}} = N$. A basis of $\mathbb{V}_{0,\mathcal{B}}$ can be chosen as the characteristic function of each $b_i, i = 1,\cdots,N$:

$$\psi_i = \chi_{b_i}(x) = \begin{cases} 1 & x \in b_i, \\ 0 & \text{otherwise}. \end{cases}$$

The nodal basis of linear finite element space $\mathbb{V}_{1,\mathcal{T}}$ is the standard hat function:

$$\phi_i \in \mathbb{V}_{1,\mathcal{T}},\ \phi_i(x_j) = \delta_{ij} \quad \text{for all } x_j \in \mathcal{N}(\mathcal{T}), i = 1,\cdots,N.$$

Let $\bar{u} = \sum_{j=1}^N \bar{U}_j\phi_j$. Choosing $v = \psi_i, i = 1,\cdots,N$ in (2.4), we obtain a linear algebraic equation

$$\bar{A}\bar{U} = \bar{F}, \tag{2.7}$$

with $\bar{A}_{ij} = -\int_{\partial b_i}(\boldsymbol{K}\nabla\phi_j)\cdot\boldsymbol{n},\ \bar{F}_i = \int_{b_i} f dx$. Let $u_L = \sum_{j=1}^N U_j\phi_j$. Choosing $v = \phi_i, i = 1,\cdots,N$ in (2.6), we obtain another linear algebraic equation

$$AU = F, \tag{2.8}$$

with $A_{ij} = \int_\Omega (\boldsymbol{K}\nabla\phi_j)\cdot\nabla\phi_i,\ F_i = \int_\Omega f\phi_i dx$.

It is well known that when $\boldsymbol{K}(\boldsymbol{x})$ is piecewise constant on each triangle, then $A = \bar{A}$; See [3, 20, 29]. The solution vectors are point values for u_L and $\bar{u}$ at vertices. The only difference is the different way to compute the right hand side. For FEM, $F_i = \int_{\Omega_i} f\phi_i \,\mathrm{d}\boldsymbol{x}$, is a weighted average over the star Ω_i of a vertex, i.e., the support of ϕ_i. For FVM, $\bar{F}_i = \int_{b_i} f \,\mathrm{d}\boldsymbol{x}$ is the average over the control volume b_i. When we choose type A control volume, i.e. choosing c_τ to be the barycenter of τ, $\bar{F}_i$ can be thought as an approximation of F_i using mass lumping. In this sense, linear FVM approximation $\bar{u}$ can be thought as a perturbation of the linear FEM approximation u_L. First order optimal convergence rate in the energy norm can be obtained using this relation [3, 20].

Note that the right hand sides may be quite different for type B dual mesh. For example, let $f = 1$ and consider the control volume in Figure 2.1(b). Then $F_i = |\Omega|/3$ while $\bar{F}_i = |\Omega|/4$. Nerveless optimal first order convergence in H^1 norm can still be derived by comparing them in the discrete H^{-1} norm [20]. Optimal second order convergence in L^2-norm holds for type A dual mesh [20] but not type B dual mesh [21].

2.3. High order finite volume method. The Petrov-Galerkin formulation can be used to develop high order finite volume methods. Given a triangulation $\mathcal{T}$ of Ω and an integer $k \geq 1$, we shall choose the trial space as

$$\mathbb{V}_{k,\mathcal{T}} = \{v \in H_0^1(\Omega) : v|_\tau \in \mathcal{P}_k(\tau),\ \forall\ \tau \in \mathcal{T}\}. \tag{2.9}$$

To construct the test function space, the traditional way is to introduce a control volume for each basis of $\mathbb{V}_{k,\mathcal{T}}$ [22, 23, 29]. For example, for quadratic finite element space, in addition to the control volumes of vertices, control volumes for middle points of edges of $\mathcal{T}$ are needed. The geometry of the control volumes will complicate the analysis and implementation of high order FVMs especially on unstructured triangular grids.

We shall propose a new choice of the test function space based on the hierarchical decomposition of $\mathbb{V}_{k,\mathcal{T}}$:

$$\mathbb{V}_{k,\mathcal{T}} = \mathbb{V}_{1,\mathcal{T}} \oplus \mathbb{W}_{k,\mathcal{T}}, \tag{2.10}$$

where recall that $\mathbb{V}_{1,\mathcal{T}}$ is the linear finite element space, and $\mathbb{W}_{k,\mathcal{T}}$ is spanned by the hierarchical basis function up to order k by excluding linear basis. For example, for quadratic finite element space, $\mathbb{W}_{2,\mathcal{T}}$ consists of quadratic bubble functions on interior edges of $\mathcal{T}$.

Let $\mathcal{B}$ be the dual mesh of $\mathcal{T}$ used in the *linear* FVM. We shall choose the test function space as

$$\mathbb{V}_{k,\mathcal{B}} := \mathbb{V}_{0,\mathcal{B}} \oplus \mathbb{W}_{k,\mathcal{T}}, \tag{2.11}$$

where $\mathbb{V}_{0,B}$ is the piecewise constant function defined on $\mathcal{B}$; see (2.2). Obviously $\mathbb{V}_{k,B} \subset L^2(\Omega)$ and $\mathbb{V}_{0,B}$ is linearly independent with $\mathbb{W}_{k,\mathcal{T}}$.

Our kth-order order FVM is: given $f \in L^2(\Omega)$, find $u \in \mathbb{V}_{k,\mathcal{T}}$ such that

$$\bar{a}(u,v) = (f,v) \quad \text{for all } v \in \mathbb{V}_{0,\mathcal{B}}, \quad \text{and} \tag{2.12}$$

$$a(u,v) = (f,v) \quad \text{for all } v \in \mathbb{W}_{k,\mathcal{T}}, \tag{2.13}$$

where recall that $\bar{a}(u,v), a(u,v)$ are bilinear forms defined in (2.3) and (2.5), respectively.

Let us compare our kth-order finite volume method with standard kth-order finite element methods. Using hierarchical decomposition (2.10), we can rewrite finite element method in the following form: given $f \in L^2(\Omega)$, to find $u \in \mathbb{V}_{k,\mathcal{T}}$

$$a(u,v) = (f,v) \quad \text{for all } v \in \mathbb{V}_{1,\mathcal{T}}, \quad \text{and} \tag{2.14}$$

$$a(u,v) = (f,v) \quad \text{for all } v \in \mathbb{W}_{k,\mathcal{T}}. \tag{2.15}$$

Therefore our method can be thought as a hybridization of high order finite element methods (2.15) and a linear finite volume method (2.12). By choosing $v = \psi_i = \chi_{b_i}$ in (2.12), we get local conservation property on b_i, which also leads to a global conservation property by linear composition of control volumes. On the other hand, we are looking for the solution in the finite element space $\mathbb{V}_{k,\mathcal{T}}$ which could give high order approximation as that in finite element methods.

We now formulate (2.12)-(2.13) and (2.14)-(2.15) as operator equations. Let X' denote the dual of a space X and $\langle\cdot,\cdot\rangle$ the duality pair. We define the following operators introduced by the bilinear form $\bar{a}(\cdot,\cdot)$ or $a(\cdot,\cdot)$.

$$
\begin{aligned}
&\bar{A} : \mathbb{V}_{1,\mathcal{T}} \to \mathbb{V}'_{0,\mathcal{B}} && \text{for } u \in \mathbb{V}_{1,\mathcal{T}}, \quad \langle \bar{A}u, v\rangle = \bar{a}(u,v) \quad \text{for all } v \in \mathbb{V}_{0,\mathcal{B}},\\
&A : \mathbb{V}_{1,\mathcal{T}} \to \mathbb{V}'_{1,\mathcal{T}} && \text{for } u \in \mathbb{V}_{1,\mathcal{T}}, \quad \langle Au, v\rangle = a(u,v) \quad \text{for all } v \in \mathbb{V}_{1,\mathcal{T}},\\
&B : \mathbb{V}_{1,\mathcal{T}} \to \mathbb{W}'_{k,\mathcal{T}} && \text{for } u \in \mathbb{V}_{1,\mathcal{T}}, \quad \langle Bu, v\rangle = a(u,v) \quad \text{for all } v \in \mathbb{W}_{k,\mathcal{T}},\\
&B^t : \mathbb{W}_{k,\mathcal{T}} \to \mathbb{V}'_{1,\mathcal{T}} && \text{for } u \in \mathbb{W}_{k,\mathcal{T}}, \quad \langle B^t u, v\rangle = a(u,v) \quad \text{for all } v \in \mathbb{V}_{1,\mathcal{T}},\\
&C : \mathbb{W}_{k,\mathcal{T}} \to \mathbb{V}'_{0,\mathcal{B}} && \text{for } u \in \mathbb{W}_{k,\mathcal{T}}, \quad \langle Cu, v\rangle = \bar{a}(u,v) \quad \text{for all } v \in \mathbb{V}_{0,\mathcal{B}},\\
&D : \mathbb{W}_{k,\mathcal{T}} \to \mathbb{W}'_{k,\mathcal{T}} && \text{for } u \in \mathbb{W}_{k,\mathcal{T}}, \quad \langle Du, v\rangle = a(u,v) \quad \text{for all } v \in \mathbb{W}_{k,\mathcal{T}}.
\end{aligned}
$$

For any $v \in \mathbb{V}_{k,\mathcal{T}}$ or $\mathbb{V}_{k,\mathcal{B}}$, let us split it as $v = v_1 + v_2$ with $v_1 \in \mathbb{V}_{1,\mathcal{T}}$ or $\mathbb{V}_{0,\mathcal{B}}$, respectively, and $v_2 \in \mathbb{W}_{k,\mathcal{T}}$. Given an $f \in L^2(\Omega)$, we define $f_1, \bar{f}_1 \in \mathbb{V}_{1,\mathcal{T}}$ and $f_2 \in \mathbb{W}_{k,\mathcal{T}}$ as follows

$$
\begin{aligned}
(f_1, v) &= (f, v) \quad \text{for all } v \in \mathbb{V}_{1,\mathcal{T}},\\
(\bar{f}_1, v) &= (f, v) \quad \text{for all } v \in \mathbb{V}_{0,\mathcal{B}},\\
\text{and} \quad (f_2, v) &= (f, v) \quad \text{for all } v \in \mathbb{W}_{k,\mathcal{T}}.
\end{aligned}
$$

Then the kth-order FVM (2.12)-(2.13) can be written as: find $\bar{u} = \bar{u}_1 + \bar{u}_2 \in \mathbb{V}_{k,\mathcal{T}}$ such that

$$
\begin{bmatrix} \bar{A} & C \\ B & D \end{bmatrix} \begin{bmatrix} \bar{u}_1 \\ \bar{u}_2 \end{bmatrix} = \begin{bmatrix} \bar{f}_1 \\ f_2 \end{bmatrix} \tag{2.16}
$$

and kth-order FEM (2.14)-(2.15) is: find $u = u_1 + u_2 \in \mathbb{V}_{k,\mathcal{T}}$ such that

$$
\begin{bmatrix} A & B^t \\ B & D \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \tag{2.17}
$$

Let $\phi_i, \psi_i, i = 1, \cdots N_L$ still denote the basis of $\mathbb{V}_{1,\mathcal{T}}$ and $\mathbb{V}_{0,\mathcal{B}}$, respectively. We choose a basis of $\mathbb{W}_{k,\mathcal{T}}$ as $\{\omega_i, i = 1, \cdots, N_W\}$. Then there are isomorphisms

$$
P_{\mathcal{T}} : \mathbb{R}^{N_L+N_W} \to \mathbb{V}_{k,\mathcal{T}} \text{ with } P_{\mathcal{T}}(U^1, U^2) = \sum_{i=1}^{N_L} U_i^1 \phi_i + \sum_{i=1}^{N_W} U_i^2 \omega_i \tag{2.18}
$$

$$
P_{\mathcal{B}} : \mathbb{R}^{N_L+N_W} \to \mathbb{V}_{k,\mathcal{B}} \text{ with } P_{\mathcal{B}}(U^1, U^2) = \sum_{i=1}^{N_L} U_i^1 \psi_i + \sum_{i=1}^{N_W} U_i^2 \omega_i. \tag{2.19}
$$

With this identification, (2.16) and (2.17) can be also understood as linear algebraic equations. For the simplicity of notation, we shall still use the same letter of the operator for its corresponding matrix representation. This should not be a source of confusion.

When $\boldsymbol{K}(\boldsymbol{x})$ is piecewise constant, in matrix form, $A = \bar{A}$, the system (2.16) is simply replacing B^t in (2.17) by C which make the system non-symmetric.

The big system (2.17) for FEM is symmetric and positive definite and thus ensures the existence and uniqueness of the solution. The stability and accuracy of our new kth-order FVM will be studied in the next section.

3. Error Analysis of finite volume methods. We shall analyze our method using the framework of Petrov-Galerkin methods as pioneered in Bank and Rose [3] and developed independently by Chinese mathematicians Li et al.; see the book [22] and the reference therein. Here we shall mainly follow a recent work of Xu and Zou [29]. The special hierarchical structure of our methods will simplify the verification of the inf-sup condition.

In the sequel, we are considering a set of triangulations $\mathscr{T} = \{\mathcal{T}_h, h \in \mathcal{H}\}$ with the parameter $h \to 0$. We assume $\mathscr{T}$ is shape regular in the sense of [16]. The parameter h has meaning of the maximal diameter of triangle in $\mathcal{T}_h$. When $\mathscr{T}$ is quasi-uniform in the sense of [16], h is a good measurement of the convergence rate.

For simplicity, the analysis is restricted to the Poisson equation, i.e. $\boldsymbol{K}(\boldsymbol{x}) = 1$. Consequently $A = \bar{A}$. Our analysis can be easily generalized to the case when $\boldsymbol{K}(\boldsymbol{x})$ is piecewise constant in each element of $\mathcal{T}$. See Remark 3.4 and 4.3.

3.1. Mesh dependent norms and continuity. We first assign norms on $\mathbb{V}_{k,\mathcal{T}}$ and $\mathbb{V}_{k,\mathcal{B}}$, and prove the continuity of the bilinear form $\bar{a}(\cdot,\cdot)$ with respect to these norms.

Since the space $\mathbb{V}_{k,\mathcal{T}} \subset H_0^1(\mathcal{T})$, H^1 semi-norm is a natural choice. But the bilinear form $\bar{a}(\cdot,\cdot)$ involves line integrals of u, an additional smoothness on u is required. Given a triangulation $\mathcal{T}$, we shall consider the space

$$H_0^1(\Omega) \cap H_{\mathcal{T}}^2(\Omega) = \{u \in H_0^1(\Omega) : u|_\tau \in H^2(\tau) \quad \text{for all } \tau \in \mathcal{T}\},$$

endowed with a mesh dependent semi-norm

$$|u|_{1,\mathcal{T}} = \left[\sum_{\tau\in\mathcal{T}} \left(|u|_{1,\tau}^2 + h_\tau^2 |u|_{2,\tau}^2\right)\right]^{1/2},$$

where $h_\tau = \operatorname{diam}(\tau)$ is the size of the element τ. Obviously

$$|u|_1 \le |u|_{1,\mathcal{T}} \quad \text{for all } u \in H_0^1(\Omega) \cap H_{\mathcal{T}}^2(\Omega).$$

Consequently, by the Poincáre inequality, $|\cdot|_{1,\mathcal{T}}$ is a norm for the space $H_0^1(\Omega) \cap H_{\mathcal{T}}^2(\Omega)$ and its subspace $\mathbb{V}_{k,\mathcal{T}}$. By the inverse inequality for finite element functions, we also have

$$|u|_{1,\mathcal{T}} \le C|u|_1 \quad \text{for all } u \in \mathbb{V}_{k,\mathcal{T}}, \tag{3.1}$$

with a constant C depending only on the shape regularity of $\mathscr{T}$ and the polynomial degree k.

The piecewise constant space $\mathbb{V}_{0,\mathcal{B}} \not\subseteq H_0^1(\Omega)$, thus we need to define a discrete "H^1 norm". With an appropriate scaling, we use the following mesh dependent semi-norm

$$|u|_{1,\mathcal{B}} = \left(\sum_{e\in\mathcal{E}(\mathcal{B})} [u]^2\right)^{1/2} = \left(\sum_{e\in\mathcal{E}(\mathcal{B})} h_e^{-1} \int_e [u]^2\right)^{1/2}, \tag{3.2}$$

where $h_e = \operatorname{diam}(e)$ is the size of e. Note that the control volumes intersect the boundary $\partial\Omega$ is not included in $\mathcal{B}$ or equivalently for $v \in \mathbb{V}_{0,\mathcal{B}}, v|_{b_i} = 0$ if $b_i \cap \partial\Omega \neq \varnothing$. Based on this observation, it is easy to show $|\cdot|_{1,\mathcal{B}}$ defines a norm on $\mathbb{V}_{0,\mathcal{B}}$. See, for example, [4] for a Poincaré type inequality for discontinuous function space.

For $u = u_1 + u_2 \in \mathbb{V}_{k,\mathcal{B}}, u_1 \in \mathbb{V}_{0,\mathcal{B}}$ and $u_2 \in \mathbb{W}_{k,\mathcal{T}}$, we define a norm

$$|u|_{1,\mathcal{B}} = \left(|u_1|_{1,\mathcal{B}}^2 + |u_2|_1^2\right)^{1/2}.$$

We then define the bilinear from $\mathcal{A}: H_0^1(\Omega)\cap H_{\mathcal{T}}^2(\Omega)\times \mathbb{V}_{k,\mathcal{B}}\to\mathbb{R}$ as

$$\mathcal{A}(u,v)=\bar{a}(u,v_1)+a(u,v_2). \tag{3.3}$$

THEOREM 3.1. *The bilinear form $\mathcal{A}: H_0^1(\Omega)\cap H_{\mathcal{T}}^2(\Omega)\times \mathbb{V}_{k,\mathcal{B}}\to\mathbb{R}$ is uniformly continuous with respect to the norm $|\cdot|_{1,\mathcal{T}}$ and $|\cdot|_{1,\mathcal{B}}$. Namely there exists a constant C depending only on the shape regularity of the mesh such that*

$$\mathcal{A}(u,v)\le C|u|_{1,\mathcal{T}}|v|_{1,\mathcal{B}}. \tag{3.4}$$

Proof. For any $u\in H_0^1(\Omega)\cap H_{\mathcal{T}}^2(\Omega), v=v_1+v_2\in\mathbb{V}_{k,\mathcal{B}}, v_1\in\mathbb{V}_{0,\mathcal{B}}, v_2\in\mathbb{W}_{k,\mathcal{B}}$,

$$\begin{aligned}\bar{a}(u,v_1)&\le\sum_{e\in\mathcal{E}(\mathcal{B})}\|\nabla u\cdot\boldsymbol{n}\|_{0,e}\|[v_1]\|_{0,e}\le\left(\sum_{e\in\mathcal{E}(\mathcal{B})}h_e\|\nabla u\cdot\boldsymbol{n}\|_{0,e}^2\right)^{1/2}|v_1|_{1,\mathcal{B}}\\&\le C\left(\sum_{\tau\in\mathcal{T}}|u|_{1,\tau}^2+h_\tau^2|u|_{2,\tau}^2\right)^{1/2}|v_1|_{1,\mathcal{B}}=C|u|_{1,\mathcal{T}}|v_1|_{1,\mathcal{B}}.\end{aligned}$$

The last inequality is an application of the trace theorem and the scaling argument. The constant depends only the shape regularity of the mesh.

For any $u\in H^1(\Omega)$ and $v\in\mathbb{W}_{k,\mathcal{T}}$, by Cauchy-Schwarz inequality

$$a(u,v_2)\le|u|_1|v_2|_1\le|u|_{1,\mathcal{T}}|v_2|_1.$$

The desired result then follows from the definition of $|v|_{1,\mathcal{B}}$. ∎

3.2. Error analysis. To analyze the error, we need to firstly clarify what do we mean by the exact solution u of the Poisson equation and specify the right function space for u.

This is not a question for finite element methods. Given an $f\in L^2(\Omega)$, let $u\in H_0^1(\Omega)$ be the solution of Poisson equation, denoted by $u=-\Delta^{-1}f$, in the sense that

$$a(u,v)=(f,v)\quad\forall\, v\in H_0^1(\Omega). \tag{3.5}$$

The existence and uniqueness of such a weak solution $u\in H_0^1(\Omega)$ is a consequence of Riesz representation theorem.

The following theorem shows that the weak solution of Poisson equation is also variational exact for the finite volume formulation provided additional smoothness of u.

THEOREM 3.2. *For a given $f\in L^2(\Omega)$, let $u=-\Delta^{-1}f$ satisfy (3.5). If $u\in H_0^1(\Omega)\cap H_{\mathcal{T}}^2(\Omega)$, then*

$$\mathcal{A}(u,v)=(f,v)\quad \textit{for all}\ \ v\in\mathbb{V}_{\mathcal{B}}. \tag{3.6}$$

Proof. Obviously (3.6) holds for $v\in\mathbb{W}_{k,\mathcal{T}}$. We only need to prove (3.6) for $v\in\mathbb{V}_{0,B}$.

By choosing $v\in C_0^\infty(\Omega)\subset H_0^1(\Omega)$ in (3.5), we know $-\Delta u=f$ in the distribution sense. Since $f\in L^2(\Omega)$, we conclude $-\Delta u=f$ holds in L^2 sense. In particular, choosing $v=\sum_{i=1}^N v_i\psi_i\in\mathbb{V}_{0,\mathcal{B}}$, we get

$$-\sum_{i=1}^N\int_{b_i}\Delta u\, v_i\,\mathrm{d}\boldsymbol{x}=\int_\Omega fv\,\mathrm{d}\boldsymbol{x}. \tag{3.7}$$

For each control volume b_i, we consider a triangulation $\mathcal{T}(b_i)$ of b_i given by connecting the vertices of b_i to the node x_i which forms a refinement of the mesh $\mathcal{T}$ restricted to b_i. Then

$$\int_{b_i} \Delta u \, \mathrm{d}\boldsymbol{x} = \sum_{\tau \in \mathcal{T}(b_i)} \int_{\tau} \Delta u \, \mathrm{d}\boldsymbol{x} = \sum_{\tau \in \mathcal{T}(b_i)} \int_{\partial \tau} \nabla u \cdot \boldsymbol{n} \, dS = \sum_{e \in \partial b_i} \int_e \nabla u \cdot \boldsymbol{n}_e \, dS. \quad (3.8)$$

In the last step, we use the fact that $u \in H^2_{\mathcal{T}}(b_i)$ and thus the boundary flux is canceled for interior edges of $\mathcal{T}(b_i)$. The smoothness assumption $u \in H^2_{\mathcal{T}}(\Omega)$ ensures the trace $\nabla u \cdot \boldsymbol{n}$ is in $L^2(e)$. Applying (3.8) to (3.7), we obtain the desired result

$$-\sum_{e \in \partial b_i} \int_e \nabla u \cdot \boldsymbol{n}[v] \, dS = -\sum_{i=1}^{N} \sum_{e \in \partial b_i} \int_e \nabla u \cdot \boldsymbol{n} \, v_i \, dS = \int_{\Omega} f v \, \mathrm{d}\boldsymbol{x}. \quad (3.9)$$

□

To derive the optimal error estimates, besides the continuity and variational exactness, we need the following uniform inf-sup condition: there exists a constant α depending only on the shape regularity of $\mathscr{T}$ such that for all $\mathcal{T} \in \mathscr{T}$:

$$\inf_{u \in \mathbb{V}_{k,\mathcal{T}}} \sup_{v \in \mathbb{V}_{k,\mathcal{B}}} \frac{\mathcal{A}(u,v)}{|u|_{1,\mathcal{T}} |v|_{1,\mathcal{B}}} \geq \alpha. \quad (3.10)$$

THEOREM 3.3. *Suppose the inf-sup condition (3.10) holds. Given an $f \in L^2(\Omega)$, let $u = -\Delta^{-1} f$ satisfy (3.5) and $u_{\mathcal{T}} \in \mathbb{V}_{k,\mathcal{T}}$ satisfy (2.12)-(2.13). If $u \in H^1_0(\Omega) \cap H^2_{\mathcal{T}}(\Omega)$, we then have the quasi-optimal error estimate:*

$$|u - u_{\mathcal{T}}|_{1,\mathcal{T}} \leq C \inf_{v_{\mathcal{T}} \in \mathbb{V}_{\mathcal{T}}} |u - v_{\mathcal{T}}|_{1,\mathcal{T}}. \quad (3.11)$$

Furthermore if $u \in H^1_0(\Omega) \cap H^{k+1}(\Omega)$, we have optimal order convergence in H^1-norm

$$|u - u_{\mathcal{T}}|_1 \leq C h_{\mathcal{T}}^k \|u\|_{k+1}, \quad (3.12)$$

where $h_{\mathcal{T}} = \max_{\tau \in \mathcal{T}} \operatorname{diam}(\tau)$.

Proof. For any $v_{\mathcal{T}} \in \mathbb{V}_{\mathcal{T}}$, by the inf-sup condition (3.10), exactness of the solution (3.6), continuity of $\mathcal{A}$, we have

$$\begin{aligned} |v_{\mathcal{T}} - u_{\mathcal{T}}|_{1,\mathcal{T}} &\leq \alpha^{-1} \sup_{v \in \mathbb{V}_{k,\mathcal{B}}} \frac{\mathcal{A}(v_{\mathcal{T}} - u_{\mathcal{T}}, v)}{|v|_{1,\mathcal{B}}} = \alpha^{-1} \sup_{v \in \mathbb{V}_{k,\mathcal{B}}} \frac{\mathcal{A}(v_{\mathcal{T}}, v) - (f, v)}{|v|_{1,\mathcal{B}}} \\ &= \alpha^{-1} \sup_{v \in \mathbb{V}_{k,\mathcal{B}}} \frac{\mathcal{A}(v_{\mathcal{T}} - u, v)}{|v|_{1,\mathcal{B}}} \leq \alpha^{-1} C |u - v_{\mathcal{T}}|_{1,\mathcal{T}}. \end{aligned}$$

Therefore, for any $v_{\mathcal{T}} \in \mathbb{V}_{\mathcal{T}}$,

$$|u - u_{\mathcal{T}}|_{1,\mathcal{T}} \leq |u - v_{\mathcal{T}}|_{1,\mathcal{T}} + |v_{\mathcal{T}} - u_{\mathcal{T}}|_{1,\mathcal{T}} \leq (1 + \alpha^{-1} C) |u - v_{\mathcal{T}}|_{1,\mathcal{T}},$$

which leads to (3.11).

Taking $v_{\mathcal{T}}$ as the Lagrange interpolation of u in $\mathbb{V}_{k,\mathcal{T}}$ and applying the standard interpolation error estimate [16], we get (3.12). □

REMARK 3.4. With the uniform bound (1.2) of the coefficient $\boldsymbol{K}$, the analysis and the result in Theorem 3.3 can be easily extended to a general tensor $\boldsymbol{K} \in \boldsymbol{L}^{\infty}(\Omega)$. In this case, the constants in (3.11) and (3.12) will depend on the ratio of a_1/a_0. The difficulty is to verify the inf-sup condition (3.10); see Remark 4.3. □

3.3. Inf-sup condition. To verify the inf-sup condition, we shall make use of the hierarchical structure of our trial and test function spaces. The following strengthened Cauchy-Bunyakowski-Schwarz (CBS) inequality [17, 1] is well known in the multigrid community.

THEOREM 3.5. *Let M be a symmetric, positive semi-definite 2×2 block matrix*

$$M = \begin{bmatrix} A & B^t \\ B & D \end{bmatrix}.$$

Let $\mathbb{U}$ and $\mathbb{V}$ be the space of vectors with only non-zero first and second components, respectively, i.e. $\boldsymbol{u} \in \mathbb{U}$ is of the form $\boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_1 \\ 0 \end{bmatrix}$ and $\boldsymbol{v} \in \mathbb{V}$ is of the form $\boldsymbol{v} = \begin{bmatrix} 0 \\ \boldsymbol{u}_2 \end{bmatrix}$. If

$$\ker(M) \subset \mathbb{U},$$

then there exists a $\gamma \in [0, 1)$ satisfying

$$(\boldsymbol{u}^t M \boldsymbol{v})^2 \leq \gamma^2 (\boldsymbol{u}^t M \boldsymbol{u})(\boldsymbol{v}^t M \boldsymbol{v}) \quad \textit{for all } \boldsymbol{u} \in \mathbb{U}, \boldsymbol{v} \in \mathbb{V}. \tag{3.13}$$

In this section, we shall use matrix representation for $u \in \mathbb{V}_{k,\mathcal{T}}$ and $v \in \mathbb{V}_{k,\mathcal{B}}$. Due to the hierarchical structure, we can identify them with $\mathbb{R}^{N_L+N_W}$, where N_L is the dimension of the linear finite element space and N_W is the dimension of its complement.

To simplify the notation, we use boldface letters, instead of capital letters, to denote the vector representation of the solution using basis ϕ_i, ψ_i and ω_i. For example, for $u \in \mathbb{V}_{k,\mathcal{T}}$, then $\boldsymbol{u} \in \mathbb{R}^{N_L+N_W}$ such that $P_{\mathcal{T}} \boldsymbol{u} = u$, c.f. (2.18) for the definition of the isomorphism $P_{\mathcal{T}}$. Note that for $u_1 \in \mathbb{V}_{1,\mathcal{T}}$, $\boldsymbol{u}_1 \in \mathbb{R}^{N_L+N_W}$ with only possible nonzero entries in the first N_L components. With an abuse of notation, we also use $\boldsymbol{u}_1$ to represent a chopped vector in $\mathbb{R}^{N_L}$. Similar notation will be applied to the spaces $\mathbb{V}_{k,\mathcal{B}}$ and $\mathbb{V}_{0,\mathcal{B}}$.

We denote the stiffness matrix corresponding to the finite element method by

$$\mathcal{A}^{FE} = \begin{bmatrix} A & B^t \\ B & D \end{bmatrix}.$$

It is a symmetric and positive definite matrix. The direct application of Theorem 3.5 will give a constant γ which may depend on the triangulation $\mathcal{T}_h$ and could tend to 1 as the mesh parameter $h \to 0$. We shall use local stiffness matrix to show this is not the case.

To this end, we denote by $\mathbb{V}_{k,\tau}, \mathbb{V}_{1,\tau}$ and $\mathbb{W}_{k,\tau}$ the kth-order finite element space and its decomposition restricted to one triangle τ, respectively. The function and its vector representation will be denoted accordingly by a subscript τ. The bilinear form $a(\cdot, \cdot)$ restricted to these subspaces gives the local stiffness matrix for a triangle $\tau \in \mathcal{T}$

$$\mathcal{A}^{FE}_{\tau} = \begin{bmatrix} A_\tau & B^t_\tau \\ B_\tau & D_\tau \end{bmatrix}.$$

LEMMA 3.6. *Let $\mathscr{T} = \{\mathcal{T}_h, h \in \mathcal{H}\}$ be a sequence of shape regular triangulations. There exists a constant $\gamma \in [0, 1)$ depending only on the shape regularity and the polynomial order k such that for all $\mathcal{T} \in \mathscr{T}$ and all $u = u_1 + u_2, u_1 \in \mathbb{V}_{1,\mathcal{T}}, u_2 \in \mathbb{W}_{k,\mathcal{T}}$*

$$(1-\gamma)(\boldsymbol{u}_1^t A \boldsymbol{u}_1 + \boldsymbol{u}_2^t D \boldsymbol{u}_2) \leq \boldsymbol{u}^t \mathcal{A}^{FE} \boldsymbol{u} \leq (1+\gamma)(\boldsymbol{u}_1^t A \boldsymbol{u}_1 + \boldsymbol{u}_2^t D \boldsymbol{u}_2). \tag{3.14}$$

Proof. Although the global matrix $\mathcal{A}^{FE}$ is symmetric positive definite, the local stiffness matrix $\mathcal{A}^{FE}_{\tau}$ is only positive semi-definite. To apply Theorem 3.5, we need to show the kernel

of $\mathcal{A}_\tau^{FE}$ is contained in $\mathbb{V}_{1,\tau}$. This can be easily proved from the fact: $a(u,u)=0$ implies that u is constant in τ.

Then by Theorem 3.5, for any $\tau \in \mathcal{T} \subset \mathscr{T}$, there exists a constant $\gamma_\tau \in [0,1)$ such that

$$(\boldsymbol{u}_{1,\tau}^t B_\tau^t \boldsymbol{u}_{2,\tau})^2 \le \gamma_\tau^2 (\boldsymbol{u}_{1,\tau}^t A \boldsymbol{u}_{1,\tau}^t)(\boldsymbol{u}_{2,\tau}^t D \boldsymbol{u}_{2,\tau}) \quad \text{for all } u_1 \in \mathbb{V}_{1,\tau}, u_2 \in \mathbb{W}_{k,\tau}. \tag{3.15}$$

By transferring back to the reference triangle, we see the constant γ_τ depends continuously on the geometry of the triangle [17]. Let $\theta_1 \ge \theta_2 \ge \theta_3$ be three angles of the triangle τ. The ordering of angles restrict possible configuration of triangles to the domain Θ on the $\theta_1 - \theta_3$ (max angle – min angle) plane

$$\Theta = \{(\theta_1,\theta_3) : \theta_1 \ge 60^\circ, 0 < \theta_3 \le 60^\circ, \theta_3 \le \theta_1, 0 < \theta_2 \le \theta_1, \theta_3 \le \theta_2.\} \tag{3.16}$$

The shape regularity of triangulations implies all angles have a lower bound denoted by θ_0. Let $\Theta_0 = \{(\theta_1,\theta_3) \in \Theta, \theta_3 \ge \theta_0, \theta_1 \le \pi - 2\theta_0\}$ be a compact subset of Θ. The inequality (3.15) implies that $\gamma(\theta_1,\theta_3) \in [0,1)$ for all $(\theta_1,\theta_3) \in \Theta_0$. Then we take $\gamma = \max_{\Theta_0} \gamma(\theta_1,\theta_3)$, which also belongs to $[0,1)$, to get a uniform version of (3.15).

Using standard Cauchy inequality, we get

$$\begin{aligned} \boldsymbol{u}_\tau^t \mathcal{A}^{FE} \boldsymbol{u}_\tau &= \boldsymbol{u}_{1,\tau}^t A \boldsymbol{u}_{1,\tau} + \boldsymbol{u}_{2,\tau}^t D \boldsymbol{u}_{2,\tau} + 2\boldsymbol{u}_{1,\tau}^t B_\tau^t \boldsymbol{u}_{2,\tau} \\ &\le (1+\gamma)(\boldsymbol{u}_{1,\tau}^t A \boldsymbol{u}_{1,\tau} + \boldsymbol{u}_{2,\tau}^t D \boldsymbol{u}_{2,\tau}). \end{aligned}$$

Summing over all $\tau \in \mathcal{T}$, we obtain the second inequality in (3.17). The first inequality is proved similarly. □

We now turn to the matrix obtained from kth-order finite volume methods. Recall that

$$\mathcal{A} = \begin{bmatrix} A & C \\ B & D \end{bmatrix}.$$

We define its symmetrization of $\mathcal{A}$ as $\mathcal{A}^s = (\mathcal{A} + \mathcal{A}^t)/2$. In matrix form, it is

$$\mathcal{A}^s = \begin{bmatrix} A & \bar{B}^t \\ \bar{B} & D \end{bmatrix},$$

with $\bar{B} = (B + C^t)/2$. Similar notation will be applied to the local matrix in each triangle.

LEMMA 3.7. *Let $\mathscr{T} = \{\mathcal{T}_h, h \in \mathcal{H}\}$ be a sequence of shape regular triangulations. If $\mathcal{A}_\tau^s$ is positive semi-definite for all $\tau \in \mathcal{T} \in \mathscr{T}$, then there exists a constant $\bar{\gamma} \in [0,1)$ depending only on the shape regularity and the polynomial order k such that for all $\mathcal{T} \in \mathscr{T}$ and all $u = u_1 + u_2, u_1 \in \mathbb{V}_{1,\mathcal{T}}, u_2 \in \mathbb{W}_{k,\mathcal{T}}$*

$$(1-\bar{\gamma})(\boldsymbol{u}_1^t A \boldsymbol{u}_1 + \boldsymbol{u}_2^t D \boldsymbol{u}_2) \le \boldsymbol{u}^t \mathcal{A}^s \boldsymbol{u} \le (1+\bar{\gamma})(\boldsymbol{u}_1^t A \boldsymbol{u}_1 + \boldsymbol{u}_2^t D \boldsymbol{u}_2). \tag{3.17}$$

Proof. Since D is symmetric and positive definite and A has rank 2, we conclude the kernel of $\mathcal{A}^s$ has dimension one. For any $u \in \mathbb{V}_{k,\tau}$, by the definition of the bilinear form

$$\bar{a}_\tau(u,1) = -\sum_{e \in \mathcal{E}(\mathcal{B}) \cap \tau} \int_e \nabla u \cdot \boldsymbol{n}_e[1] \, \mathrm{d}S = 0. \tag{3.18}$$

Therefore the kernel of $\mathcal{A}^s$ is spanned by the constant vector $[1,1,1,0,0,0]^t$ which is contained in $\mathbb{V}_{1,\mathcal{T}}$. We can then apply Theorem 3.5 and the rest is identical to Lemma 3.6. □

Let us introduce an isomorphism

$$G = P_{\mathcal{T}} P_{\mathcal{B}}^{-1} : \mathbb{V}_{k,\mathcal{B}} \to \mathbb{V}_{k,\mathcal{T}}, \quad \psi_i \to \phi_i, 1 \le i \le N.$$

Then v and Gv share the same vector representation. Since $\mathbb{V}_{k,\mathcal{T}} \subset H_0^1(\Omega)$, we can use this map to define an H^1-norm on $\mathbb{V}_{k,\mathcal{B}}$. It turns out this norm is equivalent to the discrete H^1-norm defined by (3.2).

LEMMA 3.8. *There exist constants c_1 and c_2 depending only the shape regularity of the mesh such that for any $v \in \mathbb{V}_{k,\mathcal{B}}$*

$$c_1 |v|_{1,\mathcal{B}} \le |Gv|_1 \le c_2 |v|_{1,\mathcal{B}}. \tag{3.19}$$

Proof. By the duality of $\mathcal{B}$ and $\mathcal{T}$, in matrix form $|v|_{1,\mathcal{B}}^2 = \boldsymbol{v}^t A^G \boldsymbol{v}$, where A^G is a graph Laplacian based on the mesh $\mathcal{T}$. It is not difficult to verify that the graph Laplacian A^G and the stiffness matrix A are spectral equivalent [3] and thus (3.19) holds for $v \in \mathbb{V}_{0,\mathcal{B}}$.

For $v = v_1 + v_2, v_1 \in \mathbb{V}_{0,\mathcal{B}}, v_2 \in \mathbb{W}_{k,\mathcal{T}}$, $Gv = Gv_1 + v_2$ and thus by the definition of the norm and Lemma 3.6:

$$|v|_{1,\mathcal{B}}^2 = |v_1|_{1,\mathcal{B}}^2 + |v_2|_1^2 \le C(|Gv_1|_1^2 + |v_2|_1^2) \le \frac{C}{(1-\gamma)^2} |Gv|_1^2.$$

The right inequality in (3.19) is proved similarly. ∎

The following theorem reduces the verification of the inf-sup condition to the positive semi-definite of the local stiffness matrix.

THEOREM 3.9. *Let $\mathscr{T} = \{\mathcal{T}_h, h \in \mathcal{H}\}$ be a sequence of shape regular triangulations. If $\mathcal{A}_\tau^s$ is positive semi-definite for all $\tau \in \mathcal{T} \in \mathscr{T}$, then there exists a constant $\alpha > 0$ depending only on the shape regularity of $\mathscr{T}$ and the polynomial order k such that the inf-sup condition*

$$\inf_{u \in \mathbb{V}_{k,\mathcal{T}}} \sup_{v \in \mathbb{V}_{k,\mathcal{B}}} \frac{\mathcal{A}(u,v)}{|u|_{1,\mathcal{T}} |v|_{1,\mathcal{B}}} \ge \alpha \tag{3.20}$$

holds.

Proof. For $u = u_1 + u_2 \in \mathbb{V}_{k,\mathcal{T}}$ with $u_1 \in \mathbb{V}_{1,\mathcal{T}}$ and $u_2 \in \mathbb{W}_{k,\mathcal{T}}$, we shall choose $v = P_{\mathcal{B}} P_{\mathcal{T}}^{-1} u \in \mathbb{V}_{k,\mathcal{B}}$ and prove that

$$(\mathcal{A}u, v) \ge \alpha |u|_{1,\mathcal{T}} |v|_{1,\mathcal{B}}, \tag{3.21}$$

with a constant α depending only on the shape regularity and the polynomial order k. Then (3.20) follows.

Note that $\boldsymbol{u} = \boldsymbol{v} = \boldsymbol{u}_1 + \boldsymbol{u}_2$, i.e., $\boldsymbol{u}$ and $\boldsymbol{v}$ share the same vector representation. In the matrix notation, we have

$$\boldsymbol{u}^t \mathcal{A} \boldsymbol{u} = (\mathcal{A}^t \boldsymbol{u})^t \boldsymbol{u} = \boldsymbol{u}^t \mathcal{A}^t \boldsymbol{u} = \boldsymbol{u}^t \mathcal{A}^s \boldsymbol{u}.$$

We then apply Lemma 3.6 and 3.7 to conclude that

$$\boldsymbol{u}^t \mathcal{A} \boldsymbol{u} = \boldsymbol{u}^t \mathcal{A}^s \boldsymbol{u} \ge (1-\bar{\gamma})(\boldsymbol{u}_1^t A \boldsymbol{u}_1 + \boldsymbol{u}_2^t D \boldsymbol{u}_2) \ge \frac{1-\bar{\gamma}}{1+\gamma} \boldsymbol{u}^t \mathcal{A}^{FE} \boldsymbol{u}.$$

This is equivalent to

$$(\mathcal{A}u, v) \ge \frac{1-\bar{\gamma}}{1+\gamma} |u|_1 |v|_1 \ge C |u|_{1,\mathcal{T}} |v|_{1,\mathcal{B}}.$$

In the last step, we have used the inequalities (3.1) and (3.19). ∎

4. Quadratic finite volume method. In this section, we shall consider quadratic finite volume methods on triangular and rectangular grids in one and two dimensions. Recall that our quadratic finite volume methods is: find $u_{\mathcal{T}} \in \mathbb{V}_{2,\mathcal{T}}$ such that

$$\mathcal{A}(u_{\mathcal{T}}, v) = (f, v) \quad \text{for all } v \in \mathbb{V}_{2,\mathcal{B}}, \tag{4.1}$$

(cf. (3.3) for the bilinear form $\mathcal{A}(\cdot,\cdot)$). The trial space and the test space will be more precise in the context.

4.1. Quadratic finite volume method in one dimension. Without loss of generality, we assume $\Omega = (0,1)$. Let $\mathcal{T} = \{0 = x_0 < x_1 < \cdots < x_N < x_{N+1} = 1\}$ be a grid of Ω and let $\mathcal{B} = \{0 = x_0 < x_{1/2} < x_{1+1/2} < \cdots < x_{N+1/2} < x_{N+1} = 1\}$ with $x_{k+1/2} = (x_k + x_{k+1})/2$ be the dual mesh. The quadratic finite element space $\mathbb{V}_{2,\mathcal{T}} \subset H_0^1(\Omega)$ is spanned by piecewise linear nodal basis $\phi_i, i = 1, \cdots, N$ at all interior nodes, and quadratic bubble functions $q_i, i = 1, \cdots, N+1$:

$$q_i = \frac{4(x - x_{i-1})(x_i - x)}{(x_i - x_{i-1})^2}, \quad i = 1, \cdots, N+1. \tag{4.2}$$

The test space $\mathbb{V}_{2,\mathcal{B}}$ will be obtained by replacing $\mathbb{V}_{1,\mathcal{T}}$ by $\mathbb{V}_{0,\mathcal{B}}$.

Since the contribution of the boundary flux from the quadratic bubble function is zero, i.e., $Kq_i'(x_{i-1/2}) = 0$, we have the special structure $C = 0$. Using the notation in Section 2.3, the matrix equation is in the form

$$\begin{bmatrix} \bar{A} & 0 \\ B & D \end{bmatrix} \begin{bmatrix} \bar{u}_1 \\ \bar{u}_2 \end{bmatrix} = \begin{bmatrix} \bar{f}_1 \\ f_2 \end{bmatrix}. \tag{4.3}$$

Since $C = 0$, $\bar{u}_1$ and $\bar{u}_2$ is decoupled and can be solved as $\bar{u}_1 = \bar{A}^{-1}\bar{f}_1$ and $\bar{u}_2 = D^{-1}(f_2 - B\bar{A}^{-1}\bar{f}_1)$. The computation of $\bar{u}_2$ can be done efficiently since the matrix D is diagonal and the procedure can be thought as a post-processing of $\bar{u}_1$ by solving the residual equation in the quadratic bubble function spaces.

THEOREM 4.1. *Let u be the solution of the variable Poisson equation $-(Ku')' = f, u(0) = u(1) = 0$ in the weak sense and $K \in L^\infty(\Omega)$. For quasi-uniform grids $\mathscr{T} = \{\mathcal{T}_h, h \in \mathcal{H}, h \to 0\}$, when h is small enough, the inf-sup condition (3.10) holds for the quadratic finite volume method (4.1).*

Let $u_h^L = \bar{u}_1, u_h^Q$ be the linear or quadratic finite volume approximation, respectively, and let u_I^L denote the nodal linear interpolation of u. We have

- *the optimal convergence rate for quadratic finite volume approximation*

$$|u - u_h^Q|_1 \le Ch^2\|u\|_3, \quad \text{for all } u \in H^3(\Omega);$$

- *the superconvergence of linear finite volume approximation*

$$|u_I^L - u_h^L|_1 \le Ch^2\|u\|_3, \quad \text{for all } u \in H^3(\Omega);$$

- *the optimal convergence rate of linear finite volume approximation in L^∞ norm*

$$\|u - u_h^L\|_\infty \le Ch^2\|u\|_{3,\infty}, \quad \text{for all } u \in W^{3,\infty}(\Omega).$$

Proof. If K is piecewise constant, then $\bar{A} = A$ and $B = 0$. We then obtain the same stiffness matrix as that from quadratic finite element method. The inf-sup condition for piecewise constant K is then from that of FEM. For general variable coefficients $K \in L^\infty(\Omega)$, we can consider the system obtained using $\bar{K}_h$, the piecewise constant approximation of K. Since

$\lim_{h\to 0}\|K-\bar{K}_h\|\to 0$, when h is sufficiently small, the inf-sup condition will hold by the perturbation argument; see [29] for details. The optimal convergence rate of u_h^Q in H^1 norm then follows easily.

Since $C=0$, $\bar{u}_1$ is also the solution of linear finite volume method. We can write $u_h^L=\bar{u}_1$ and $u_h^Q=\bar{u}_1+\bar{u}_2$. Let u_I^Q, u_I^L denote the quadratic interpolation and linear interpolation of u respectively, and let $e_h^Q=u_I^Q-u_h^Q, e_h^L=u_I^L-u_h^L$. We have the following decomposition

$$e_h^Q=e_h^L+(e_h^Q-e_h^L). \tag{4.4}$$

Due to the special feature $u_h^Q-u_h^L\in\mathbb{W}_{2,\mathcal{T}}$, $(e_h^Q-e_h^L)\in\mathbb{W}_{2,\mathcal{T}}$ and (4.4) is a hierarchical decomposition. Since $\mathbb{V}_{1,\mathcal{T}}$ is orthogonal to $\mathbb{W}_{2,\mathcal{T}}$ in the H^1 semi-inner product, we obtain

$$|u_I^L-u_h^L|_1=|e_h^L|_1\le|e_h^Q|_1\le Ch^2\|u\|_3.$$

Note that this superconvergence result does not use the uniformity of the mesh $\mathcal{T}$.

The optimal L^∞ norm estimate for u_h^L follows from:

- the the embedding theorem $\|u_I^L-u_h^L\|_\infty\le C|u_I^L-u_h^L|_1$;
- the triangle inequality $\|u-u_h^L\|_\infty\le\|u_I^L-u_h^L\|_\infty+\|u-u_I^L\|_\infty$;
- and the interpolation error estimate $\|u-u_I^L\|_\infty\le Ch^3\|u\|_{3,\infty}$.

□

4.2. Quadratic finite volume method on triangular grids. In this subsection, we shall provide explicit formula for quadratic finite volume methods for Poisson equation on 2-D triangular grids and verify the positive semi-definiteness condition.

We first compute the local stiffness matrix in one triangle. Let θ_i denote the angle of the triangle at the vertex x_i for $i=1,2,3$, and index the edge opposite to vertex x_i by e_i. Let λ_i be the barycentric coordinates corresponding to x_i which is the basis of linear polynomial space. Then the quadratic bubble function on the edge x_ix_j is given by $4\lambda_i\lambda_j$. Following [24, 2], we introduce the notation

$$c_i=\cot\theta_i, i=1\text{ to }3,\ \text{and}\ c=\sum_{i=1}^3 c_i. \tag{4.5}$$

By direct computation, we obtain the corresponding matrices:

$$A=\frac{1}{2}\begin{bmatrix}c_2+c_3 & -c_3 & -c_2\\ -c_3 & c_3+c_1 & -c_1\\ -c_2 & -c_1 & c_1+c_2\end{bmatrix},\ B=-\frac{4}{3}A,\ \text{and}\ D=\frac{4}{3}\begin{bmatrix}c & -c_3 & -c_2\\ -c_3 & c & -c_1\\ -c_2 & -c_1 & c\end{bmatrix}.$$

We shall compute the matrix C for two typical choices of control volumes.

Type A control volumes. In this case, we connect the centroid to the middle points of edges. See Figure 1 (a). The area of each control volume is one third of the area of the triangle. We list the matrix C below:

$$C=-\begin{bmatrix}-\frac{1}{3}c_1+\frac{1}{2}c_2+\frac{1}{2}c_3 & \frac{1}{6}c_2-\frac{1}{2}c_3 & \frac{1}{6}c_3-\frac{1}{2}c_2\\ \frac{1}{6}c_1-\frac{1}{2}c_3 & \frac{1}{2}c_1-\frac{1}{3}c_2+\frac{1}{2}c_3 & \frac{1}{6}c_3-\frac{1}{2}c_1\\ \frac{1}{6}c_1-\frac{1}{2}c_2 & \frac{1}{6}c_2-\frac{1}{2}c_1 & \frac{1}{2}c_1+\frac{1}{2}c_2-\frac{1}{3}c_3\end{bmatrix}. \tag{4.6}$$

When $c_1 = c_2 = c_3 = \cot 60^\circ$, i.e., the triangle is equilateral, it is simplified as $C = B/2$. For this special case, we have

$$\mathcal{A}^s = \frac{3}{4}\mathcal{A}^{FE} + \frac{1}{4}\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix},$$

which is symmetric and positive definite and satisfy the inf-sup condition. Optimal error estimates then follows for this special case.

Type B control volumes. Without loss of generality, we assume θ_1 is the largest angle. For each triangle, we divide it into three parts by connecting middle points of e_2 and e_3 to the middle point of e_1. See Figure 1 (b). We list the matrix C below:

$$C = -\frac{1}{2}\begin{bmatrix} c_2 + c_3 - 2c_1 & -c_2 - c_3 & -c_2 - c_3 \\ c_1 - c_3 & c_1 + c_3 & c_3 - c_1 \\ c_1 - c_2 & c_2 - c_1 & c_1 + c_2 \end{bmatrix}. \tag{4.7}$$

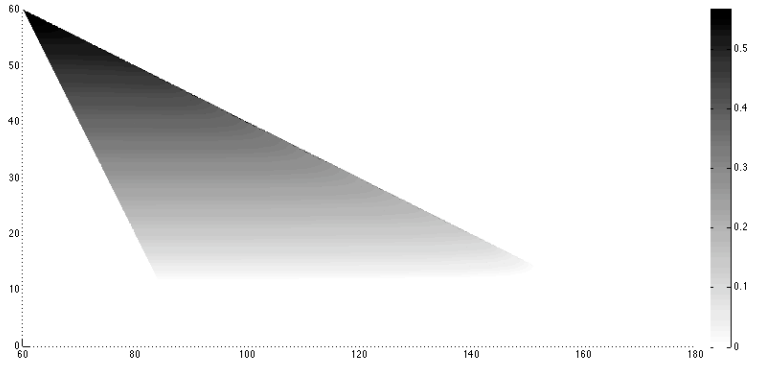


FIG. 4.1. *The second eigenvalue of symmetrized local stiffness matrix of quadratic FVM: type A dual mesh. x-axis: maximal angle θ_1; y-axis: minimal angle θ_3.*

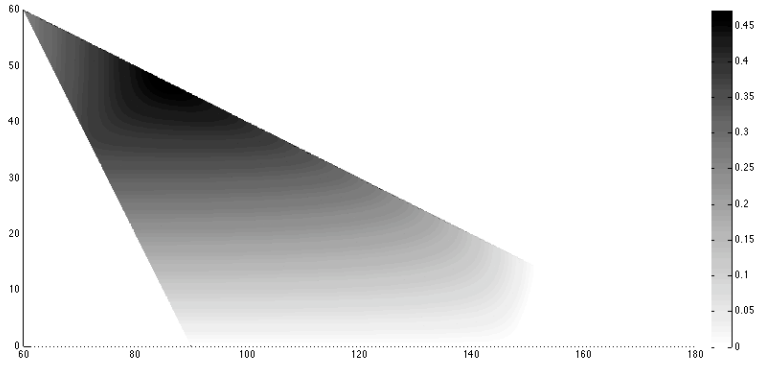


FIG. 4.2. *The second eigenvalue of symmetrized local stiffness matrix of quadratic FVM: type B dual. x-axis: maximal angle θ_1; y-axis: minimal angle θ_3*

For triangles of general shape, we use the following procedure to verify the positive semi-definite of $\mathcal{A}^s_\tau$. Suppose the eigenvalues of the symmetric matrix $\mathcal{A}^s_\tau$ are sorted by

$\mu_1 \le \mu_2 \le \cdots \le \mu_6$. From (3.18), we know zero is an eigenvalue of $\mathcal{A}_\tau^s$. We compute the second eigenvalue μ_2. If $\mu_2 > 0$, then $\mu_1 = 0$ and thus $\mathcal{A}_\tau^s$ is positive semi-definite.

Obviously from the formulation of the local stiffness matrix, μ_2 depends continuously on angles of the triangle τ. Without loss of generality, we assume $\theta_1 \ge \theta_2 \ge \theta_3$ and consider the domain Θ defined in (3.16). We discretize the rectangular domain $(0, 180^\circ) \times (0, 180^\circ)$ on the $\theta_1 - \theta_3$ plane by a uniform grid with mesh size 0.1 and compute $\mu_2(\theta_1, \theta_3)$ at grid points. By the ordering of angles, we restrict our computation to the domain Θ and set $\mu_2 = 0$ outside of Θ.

The contour of the computed μ_2 is plotted in Fig. 4.1 and 4.2. We say the triangle is admissible if $\mu_2(\tau) > 0$. From Fig. 4.1 and 4.2, it is evident that when the maximal angle is less than a threshold θ_1^* and the minimal angle is greater than θ_3^*, then the triangle is admissible. Numerical computation shows that for both Type A and B dual mesh, $\theta_1^* = 151.6^\circ$ which is consistent with the maximal angle condition 151.56° obtained in [23] (since our computation is one digit accurate after the decimal point). We also observe that type B dual mesh requires less restriction on the minimal angle.

We summarize the convergence of our new quadratic FVM in the following theorem.

THEOREM 4.2. *Suppose every triangle in the triangulation $\mathcal{T} \in \mathscr{T}$ is admissible. Then the inf-sup condition (3.10) of the quadratic FVM (4.1) holds with a constant depending only on the shape regularity of $\mathscr{T}$.*

Consequently, given an $f \in L^2(\Omega)$, let $u = -\Delta^{-1} f$ satisfy (3.5) and $u_{\mathcal{T}} \in \mathbb{V}_{k,\mathcal{T}}$ satisfy (2.12)-(2.13). If $u \in H_0^1(\Omega) \cap H_{\mathcal{T}}^2(\Omega)$, we have the quasi-optimal error estimate:

$$|u - u_{\mathcal{T}}|_{1,\mathcal{T}} \le C \inf_{v_{\mathcal{T}} \in \mathbb{V}_{\mathcal{T}}} |u - v_{\mathcal{T}}|_{1,\mathcal{T}}. \tag{4.8}$$

Furthermore if $u \in H_0^1(\Omega) \cap H^3(\Omega)$, we have optimal order convergence in H^1-norm

$$|u - u_{\mathcal{T}}|_1 \le C h_{\mathcal{T}}^2 \|u\|_3. \tag{4.9}$$

The convergence analysis on the quadratic finite volume method presented in the book [22] (Chapter 3, page 148) requires stronger geometrical conditions: the maximum angle of each triangle is not greater than $\pi/2$, and that the ratio of the lengths of the two sides of the maximum angle is in the range $[\sqrt{2/3}, \sqrt{3/2}]$. In [23], the maximal angle condition is relaxed to 151.56°. But the proof is complicated and not easy to verify since the key steps are skipped. In a recent work [29], the angle condition is improved.

Quadratic finite volume methods discussed in these works [22, 29, 23], however, requires a control volume for each quadratic bubble basis and the local stiffness matrix is more complicated. Instead in our new approach the local stiffness matrix can be easily modified from hierarchical basis finite element code.

The angle condition for each triangle is a sufficient condition to prove the inf-sup condition and is by no means a necessary condition. There could be cancelation when assembling local stiffness matrix to a big one. In this sense, if there are only few number of 'bad triangles' are not admissible, the scheme may still have optimal convergent rate. The angle condition for the error analysis may not be a constrain for practical computation.

REMARK 4.3. When $\boldsymbol{K}$ is piecewise constant, we can write

$$\int_\tau \boldsymbol{K} \nabla u \cdot \nabla v \, \mathrm{d}\boldsymbol{x} = \int_\tau (\boldsymbol{K}^{1/2} \nabla u) \cdot (\boldsymbol{K}^{1/2} \nabla v) \, \mathrm{d}\boldsymbol{x} = \frac{1}{\det(\boldsymbol{K}^{1/2})} \int_{\tilde{\tau}} \tilde{\nabla} \tilde{u} \cdot \tilde{\nabla} \tilde{v} \, d\tilde{\boldsymbol{x}},$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{K}^{1/2} \boldsymbol{x}$. Therefore similar results will hold when the transformed triangle $\tilde{\tau}$ is admissible. We refer to [13] for a method on generating quasi-uniform grids under general Riemannian metrics.

For variable coefficients, results hold by assuming h is small enough; see the argument in Theorem 4.1. □

4.3. Quadratic finite volume method on rectangular grids. In this subsection, we shall consider a biquadratic finite volume method for solving Poisson equation over rectangular grids. The convergence analysis for bilinear finite volume methods on rectangular grids can be found at [27, 6].

For the simplicity of exposition, we consider homogenous Dirichlet boundary condition and assume $\Omega = (0,1) \times (0,1)$. The domain is discretized by a non-uniform mesh $\mathcal{T} = \mathcal{T}_x \otimes \mathcal{T}_y$, which is the Cartesian product of the one-dimensional meshes

$$\mathcal{T}_x = \{x_i, i = 0, \cdots, M : x_0 = 0, x_i - x_{i-1} = h_i, x_M = 1\},$$
$$\mathcal{T}_y = \{y_j, j = 0, \cdots, N : y_0 = 0, y_j - y_{j-1} = k_j, y_N = 1\}.$$

We choose the trial space $\mathbb{V}_{2,\mathcal{T}}$ as 8-nodes biquadratic finite element space of $H_0^1(\Omega)$. Let $\mathbb{V}_{1,\mathcal{T}}$ be the bilinear finite element space. For a rectangle $\tau_{i,j} = (x_i, x_{i+1}) \times (y_j, y_{j+1})$ in $\mathcal{T}$, we label the four nodes $v_i, i = 1:4$ and four middle points on edges in Figure 4.3. For a point $(x,y) \in \tau$, it can be denoted by barycentric coordinates in x and y direction as $(x,y) = (\lambda_1^x, \lambda_2^x, \lambda_1^y, \lambda_4^y)$, where $\lambda_i^x(x)$ is a linear function of x such that $\lambda_i^x(v_j) = \delta_i^j$ for $i,j = 1,2$ and $\lambda_i^y(y)$ is a linear function of y such that $\lambda_i^y(v_j) = \delta_i^j$ for $i,j = 1,4$. Then the hierarchical basis in τ can be written as

$$\phi_1 = \lambda_1^x\lambda_1^y, \qquad \phi_2 = \lambda_2^x\lambda_1^y, \qquad \phi_3 = \lambda_2^x\lambda_4^y, \qquad \phi_4 = \lambda_1^x\lambda_4^y,$$
$$\omega_5 = 4\lambda_1^x\lambda_2^x\lambda_1^y, \qquad \omega_6 = 4\lambda_1^y\lambda_4^y\lambda_2^x, \qquad \omega_7 = 4\lambda_1^x\lambda_2^x\lambda_4^y, \qquad \omega_8 = 4\lambda_1^y\lambda_4^y\lambda_1^x.$$

Restricted to one element τ, the space is $\mathbb{V}_{2,\tau} = \mathbb{V}_{1,\tau} \oplus \mathbb{W}_{2,\tau}$, where

$$\mathbb{V}_{1,\tau} = \text{span}\{\phi_1, \phi_2, \phi_3, \phi_4\}, \quad \text{and } \mathbb{W}_{2,\tau} = \text{span}\{\omega_5, \omega_6, \omega_7, \omega_8\}.$$

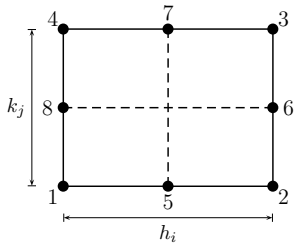


FIG. 4.3. *Q8 biquadratic element*

For each vertex $(x_i, y_j) \in \mathcal{T}$, the control volume is choose as

$$b_{ij} = (x_{i-1/2}, x_{i+1/2}) \times (y_{j-1/2}, y_{j+1/2}),$$

where $x_{i-1/2} = (x_i + x_{i-1})/2, x_{i+1/2} = (x_i + x_{i+1})/2, y_{i-1/2} = (y_i + y_{i-1})/2, y_{i+1/2} = (y_i + y_{i+1})/2$. The dual mesh $\mathcal{B} = \{b_{ij} : (x_i, y_j)$ is an interior node of $\mathcal{T}\}$ and the test space will be $\mathbb{V}_{2,\mathcal{B}} = \mathbb{V}_{0,\mathcal{B}} + \mathbb{W}_{2,\mathcal{B}}$.

The convergence could be analyzed similarly using the framework developed for triangular grids. For rectangular grids, however, we have a more direct way to establish the continuity and stability. We shall sketch the proof below and skip details.

The following lemma can be found in [15]. Here recall that $G : \mathbb{V}_{2,\mathcal{B}} \to \mathbb{V}_{2,\mathcal{T}}$ is the isomorphism between the trial and test spaces.

LEMMA 4.4. *For any $u_1 \in \mathbb{V}_{1,\mathcal{T}}, v_1 \in \mathbb{V}_{0,\mathcal{B}}$, we have*

$$-\sum_{e\in\mathcal{E}(\mathcal{B})}\int_e \nabla u_1 \cdot \boldsymbol{n}_e[v_1]dS = \int_\Omega \nabla u_1 \cdot \nabla(Gv_1)\,\mathrm{d}\boldsymbol{x} + Q(u_1, v_1), \tag{4.10}$$

where

$$Q(u_1, v_1) = \frac{1}{24}\sum_{\tau_{ij}\in\mathcal{T}}(h_i^3 k_j + h_i k_j^3)\frac{\partial^2 u_1}{\partial x\partial y}\frac{\partial^2 (Gv_1)}{\partial x\partial y}.$$

By direct computations we have the following identity.

LEMMA 4.5. *In one rectangle τ, we have*

$$-\int_{\partial b_i}\frac{\partial \omega_j}{\partial n}dS = \int_\tau \nabla\omega_j \cdot \nabla\phi_i\,dxdy, \quad i = 1, \cdots, 4,\ j = 5, \cdots, 8. \tag{4.11}$$

We then obtain a *symmetric* quadratic finite volume methods with the following matrix formulation for the stiffness matrix:

$$\mathcal{A} = \begin{bmatrix} \bar{A} & B^t \\ B & D \end{bmatrix}.$$

Comparing with the stiffness matrix of the quadratic finite element

$$\mathcal{A}^{FE} = \begin{bmatrix} A & B^t \\ B & D \end{bmatrix},$$

the implementation of our quadratic finite volume methods can be easily modified from quadratic finite element methods. The resulting matrix is symmetric and thus can borrow efficient iterative methods designed for finite element methods.

THEOREM 4.6. *The bilinear form $\mathcal{A}(\cdot,\cdot)$ is symmetric, positive definite, and continuous in the sense that for any $u \in \mathbb{V}_{2,\mathcal{T}}, v \in \mathbb{V}_{2,\mathcal{B}}$*

$$\mathcal{A}(u, v) = \mathcal{A}(Gv, G^{-1}u), \tag{4.12}$$

$$\mathcal{A}(u, G^{-1}u) \geq |u|_1^2, \tag{4.13}$$

$$\mathcal{A}(u, v) \leq C|u|_1|Gv|_1, \tag{4.14}$$

where the constant in (4.14) depending only the aspect ratio of rectangles.

Proof. By (4.10) and (4.11), for $u = u_1 + u_2 \in \mathbb{V}_{2,\mathcal{T}}, v = v_1 + v_2 \in \mathbb{V}_{2,\mathcal{B}}$, we have

$$\mathcal{A}(u, v) = \mathcal{A}^{FE}(u, Gv) + Q(u_1, v_1).$$

Then (4.12) is from the symmetric of $Q(\cdot,\cdot)$ and (4.13) is consequence of $Q(u_1, u_1) \geq 0$. Using the inverse inequality, it is easy to show (c.f [15])

$$Q(u_1, v_1) \leq C|u_1|_1|Gv_1|_1.$$

Then using the strengthened Cauchy inequality, we get $|u_1| \leq C|u|_1$ with a constant C depending only on the aspect ratio of rectangles. Using the continuity of $\mathcal{A}^{FE}(\cdot,\cdot)$, we obtain (4.14). ∎

The variational exactness and error estimate can be proved similarly. We summarize results in the following theorem.

THEOREM 4.7. *Let $\mathscr{T} = \{\mathcal{T}_h, h \in \mathcal{H}\}$ be a sequence of shape regular rectangular grids. Given an $f \in L^2(\Omega)$, let $u = -\Delta^{-1} f$ satisfy (3.5) and $u_{\mathcal{T}} \in \mathbb{V}_{2,\mathcal{T}}$ satisfy (4.1). If $u \in H_0^1(\Omega) \cap H_{\mathcal{T}}^2(\Omega)$, we have the quasi-optimal error estimate:*

$$|u - u_{\mathcal{T}}|_{1,\mathcal{T}} \leq C \inf_{v_{\mathcal{T}} \in \mathbb{V}_{\mathcal{T}}} |u - v_{\mathcal{T}}|_{1,\mathcal{T}}. \tag{4.15}$$

Furthermore if $u \in H_0^1(\Omega) \cap H^3(\Omega)$, we have optimal order convergence in H^1-norm

$$|u - u_{\mathcal{T}}|_1 \leq C h_{\mathcal{T}}^2 \|u\|_3. \tag{4.16}$$

5. Numerical Experiment. In this section, we shall present a numerical example to support our theoretical results. Let $\Omega = (-1,1) \times (-1,1) \backslash ([0,1] \times [-1,0])$ be a L-shape domain and consider the Poisson equation

$$-\Delta u = 0, \text{ in } \Omega \quad u = u_D \text{ on } \partial\Omega.$$

We choose u_D and f such that the exact solution u in polar coordinates is

$$u(r,\theta) = r^{\frac{2}{3}} \sin \frac{2}{3}\theta.$$

It is well known that the solution presents a singularity at the origin. Mesh adaptation based on the procedure for adaptive finite element methods (AFEMs) is applied to get a suitable locally refined mesh for quadratic elements. See Fig. 5.1 for an example of such a grid. We refer to [14] and references therein for the detailed description of AFEM.

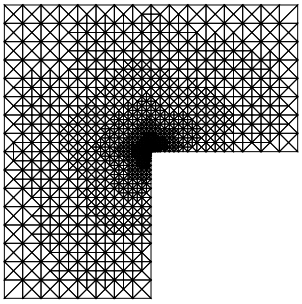


FIG. 5.1. *L-shape domain and a locally refined mesh*

We replace the quadratic finite element approximation in AFEM by quadratic finite volume approximation with type B dual mesh. We plot the error in H^1 norm. Since the mesh is not quasi-uniform, we use $N = \#\text{dof}$, the number of degree of freedom, to measure the convergent rate. In two dimensions, $h^2 = O(N^{-1})$ for quasi-uniform grids. From Fig. 5.2, it is evident that it achieves optimal order in H^1 norm. The simulation is implemented using *i*FEM [11].

6. Conclusion and future work. In this paper, we have developed a new class of high order finite volume methods using the hierarchical high order finite element methods. Our new method is easy to implement comparing with other quadratic finite volume methods. We

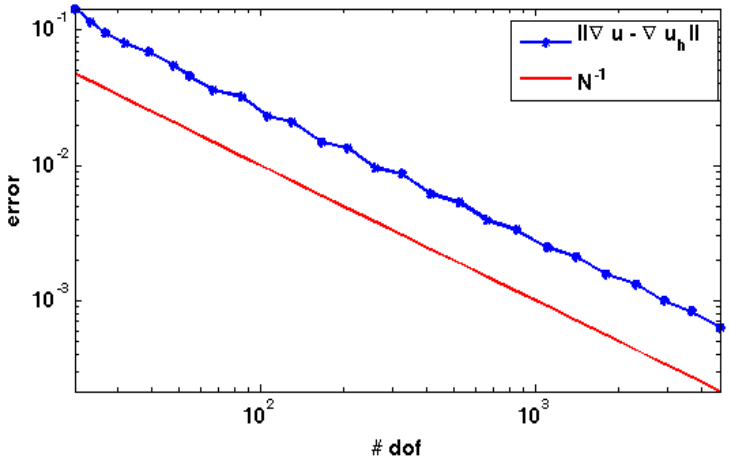


FIG. 5.2. *Error of a quadratic finite volume approximation in H^1 norm*

also verified the inf-sup condition for our quadratic finite volume methods on two dimensional triangular and rectangular grids and thus obtain optimal convergence rate in H^1 norm.

We showed that in two dimensional rectangular grids, our new quadratic finite volume method results in a symmetric matrix. We note that, however, the resulting matrix for triangular mesh is non-symmetric. In general for variable coefficients, the system for both triangular and rectangular grids are non-symmetric.

We have not discussed efficient iterative solvers for the resulting non-symmetric matrix. Since the matrix is not far away from that from finite element methods, we expect existing multilevel methods for solving linear algebraic equation developed in finite element methods will help.

Most existing finite volume methods of Stokes equations are restricted to lower order pairs. With our new quadratic finite volume discretization of the Laplacian operator, we will be able to examine the P_2–P_1 or Q_2–Q_1 Taylor-Hood type elements for the finite volume approximation of Stokes equations. We shall report our finding in another work [12].

Acknowledgement. The author would like to thank Professor Jinchao Xu for fruitful discussions, encouragement to write this paper, and the summer workshop he organized in Beijing University in 2007, and the referees and the editor for careful reading and helpful suggestions on the improvement of the presentation.

REFERENCES

[1] B. ACHCHAB, O. AXELSSON, L. LAAYOUNI, AND A. SOUISSI, *Strengthened Cauchy-Bunyakowski-Schwarz inequality for a three-dimensional elasticity system*, Numerical Linear Algebra with Applications, 8 (2001), pp. 191–205.

[2] R. E. BANK, *Hierarchical bases and the finite element method*, Acta Numerica, 5 (1996), pp. 1–43.

[3] R. E. BANK AND D. J. ROSE, *Some error estimates for the box scheme*, SIAM Journal on Numerical Analysis, 24 (1987), pp. 777–787.

[4] S. C. BRENNER, *Poincaré–Friedrichs inequalities for piecewise H^1 functions*, SIAM Journal on Numerical Analysis, 41 (2003), pp. 306–324.

[5] F. BREZZI AND M. FORTIN, *Mixed and hybrid finite element methods*, Springer-Verlag, 1991.

[6] Z. CAI, *A theoretical foundation for the finite volume element method*, PhD thesis, in partial fulfillment of the requirements for the degree, Doctor of Philosophy, Department of Mathematics.University of Colorado at Denver, 1990.
[7] Z. CAI, *On the finite volume element method*, Numerische Mathematik, 58 (1990), pp. 713–735.
[8] Z. CAI, J. DOUGLAS, JR., AND M. PARK, *Development and analysis of higher order finite volume methods over rectangles for elliptic equations*, Advances in Computational Mathematics, 19 (2003), pp. 3–33.
[9] Z. CAI, J. MANDEL, AND S. F. MCCORMICK, *The finite volume element method for diffusion equations on general triangulations*, SIAM Journal on Numerical Analysis, 28 (1991), pp. 392–402.
[10] Z. CAI AND S. F. MCCORMICK, *On the accuracy of the finite volume element method for diffusion equations on composite grids*, SIAM Journal on Numerical Analysis, 27 (1990), pp. 636–655.
[11] L. CHEN, *iFEM: an innovative finite element methods package in MATLAB*, Submitted, (2009).
[12] ———, *Some first and second order finite volume methods for the Stokes problem*, Submitted, (2009).
[13] L. CHEN, P. SUN, AND J. XU, *Optimal anisotropic simplicial meshes for minimizing interpolation errors in L^p-norm*, Mathematics of Computation, 76 (2007), pp. 179–204.
[14] L. CHEN AND J. XU, *Topics on adaptive finite element methods*, in Adaptive Computations: Theory and Algorithms, T. Tang and J. Xu, eds., Science Press, Beijing, 2007, pp. 1–31.
[15] S. H. CHOU AND D. Y. KWAK, *Analysis and Convergence of a MAC-like Scheme for the Generalized Stokes Problem*, Numerical Methods for Partial Differential Equations, 13 (1997), pp. 147–162.
[16] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, vol. 4 of Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978.
[17] V. EIJKHOUT AND P. S. VASSILEVSKI, *The role of the strengthened Cauchy–Buniakowskii–Schwarz inequality in multilevel methods*, SIAM Review, 33 (1991), pp. 405–419.
[18] A. ERINGEN, *Mechanics of continua*, New York: Krieger Pub Co, 2 nd ed., 1980.
[19] R. EYMARD, T. GALLOUËT, AND R. HERBIN, *Finite volume methods*, in Handbook of numerical analysis, Vol. VII, Handb. Numer. Anal., VII, North-Holland, Amsterdam, 2000, pp. 713–1020.
[20] W. HACKBUSCH, *On first and second order box schemes*, Computing, 41 (1989), pp. 277–296.
[21] J. HUANG AND S. XI, *On the finite volume element method for general self-adjoint elliptic problems*, SIAM Journal on Numerical Analysis, 35 (1998), pp. 1762–1774.
[22] R. LI, Z. CHEN, AND W. WU, *Generalized difference methods for differential equations*, vol. 226 of Monographs and Textbooks in Pure and Applied Mathematics, Marcel Dekker Inc., New York, 2000. Numerical analysis of finite volume methods.
[23] F. LIEBAU, *The finite volume element method with quadratic basis functions*, Computing, 57 (1996), pp. 281–299.
[24] J.-F. MAITRE AND F. MUSY, *The contraction number of a class of two-level methods; an exact evaluation for some finite element subspaces and model problems*, in Multigrid methods (Cologne, 1981), vol. 960 of Lecture Notes in Math., Springer, Berlin, 1982, pp. 535–544.
[25] M. PLEXOUSAKIS AND G. E. ZOURARIS, *On the construction and analysis of high order locally conservative finite volume-type methods for one-dimensional elliptic problems*, SIAM J. Numer. Anal., 42 (2004), pp. 1226–1260 (electronic).
[26] T. RUSSELL AND M. WHEELER, *Finite element and finite difference method for continuous flows in porous media*, Frontiers in Applied Mathematics, 1 (1983), p. 35.
[27] E. SÜLI, *Convergence of finite volume schemes for Poisson's equation on nonuniform meshes*, SIAM J. Numer. Anal., 28 (1991), pp. 1419–1430.
[28] R. S. VARGA, *Matrix iterative analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962.
[29] J. XU AND Q. ZOU, *Analysis of linear and quadratic simplicial finite volume methods for elliptic equations*, Numer. Math., 111 (2009), pp. 469–492.