MATHEMATICS OF COMPUTATION
Volume 88, Number 320, November 2019, Pages 2665–2696
https://doi.org/10.1090/mcom/3423
Article electronically published on March 26, 2019

CONDITIONING OF THE FINITE VOLUME ELEMENT METHOD FOR DIFFUSION PROBLEMS WITH GENERAL SIMPLICIAL MESHES

XIANG WANG, WEIZHANG HUANG, AND YONGHAI LI

Abstract. The conditioning of the linear finite volume element discretization for general diffusion equations is studied on arbitrary simplicial meshes. The condition number is defined as the ratio of the maximal singular value of the stiffness matrix to the minimal eigenvalue of its symmetric part. This definition is motivated by the fact that the convergence rate of the generalized minimal residual method for the corresponding linear systems is determined by the ratio. An upper bound for the ratio is established by developing an upper bound for the maximal singular value and a lower bound for the minimal eigenvalue of the symmetric part. It is shown that the bound depends on three factors: the number of the elements in the mesh, the mesh nonuniformity measured in the Euclidean metric, and the mesh nonuniformity measured in the metric specified by the inverse diffusion matrix. It is also shown that the diagonal scaling can effectively eliminate the effects from the mesh nonuniformity measured in the Euclidean metric. Numerical results for a selection of examples in one, two, and three dimensions are presented.

1. Introduction

The finite volume element method (FVEM) is a type of finite volume method that approximates the solution of partial differential equations (PDEs) in a finite element space. It inherits many advantages of finite volume methods such as the local conservation property while avoiding the complexity other types of finite volume methods have in defining the gradient of the approximate solution needed in the discretization of diffusion equations. FVEM has been successfully applied to a broad range of problems and studied extensively in theory; e.g., see [3, 5, 7–9, 11, 12, 14, 18, 22, 30, 32, 35, 41]. To date, significant progress has been made in understanding FVEM's stability and superconvergence, establishing error bounds, and developing high-order FVEMs. For example, stability analysis and error estimates in the L^2 or H^1 norm are developed for triangular and quadrilateral meshes in [13, 33, 37, 40, 42] while superconvergence results are established recently in [10, 34, 38, 42].

Received by the editor February 4, 2018, and, in revised form, August 2, 2018, and January 11, 2019.

2010 *Mathematics Subject Classification.* Primary 65N08, 65F35.

This work was supported in part by the National Natural Science Foundation of China through grants 11701211 and 11371170, the China Postdoctoral Science Foundation through grant 2017M620106, the Joint Fund of the National Natural Science Foundation of China and the China Academy of Engineering Physics (NASF) through grant U1630249, and the Science Challenge Program (China) through grant JCKY2016212A502.

The first-named author was supported by China Scholarship Council (CSC) under grant 201506170088 for his research visit to the University of Kansas from September of 2015 to September of 2016.

On the other hand, little progress has been made in understanding the conditioning of FVEM discretization on general meshes. There are two major barriers toward this. The first one is that FVEM does not preserve the symmetry of the underlying differential operator and has a nonsymmetric stiffness matrix in general. It is well known that standard condition numbers provide little useful information for the solution of nonsymmetric algebraic systems. A common alternative for measuring the conditioning of a nonsymmetric matrix is the ratio of its largest singular value to the minimal eigenvalue of its symmetric part. This is largely motivated by the work of Eisenstat et al. [16] (or see (16) below) stating that the ratio determines the convergence rate of the generalized minimal residual method (GMRES) for the corresponding linear systems. Establishing an upper bound for the ratio requires the development of an upper bound for the maximal singular value and a lower bound for the minimal eigenvalue of the symmetric part. This process is more difficult and complicated in general than that used to establish bounds for the extremal eigenvalues for symmetric and positive definite matrices.

The second barrier comes from mesh nonuniformity. A main advantage of FVEM is its flexibility to work with (nonuniform) adaptive meshes needed in many applications. It thus makes sense that the analysis is carried out for general nonuniform meshes without prior requirements on their uniformity and regularity. However, this is not a trivial task in general since it will need to have a mathematical characterization for nonuniform meshes and take the interplay between the mesh geometry and the underlying differential operator (or the diffusion matrix in the case of diffusion equations) into full consideration. For example, Li et al. [31] study a multilevel preconditioning technique for FVEM and establish a uniform bound on the ratio of the largest singular value to the minimal eigenvalue of the symmetric part of the preconditioned stiffness matrix but their analysis and results are valid only for quasi-uniform meshes. Moreover, FVEM analysis (such as L^2 error estimation) typically obtains relevant properties from the finite element (FE) discretization of the underlying problem by estimating the difference between the corresponding bilinear forms. This type of estimation has so far been carried out only for quasi-uniform or regular meshes too; e.g., see [30, 31, 33, 37].

It is interesting to point out that much more effort and progress have been made to understand the conditioning of FE discretization on general meshes. Noticeably, Fried [19] obtains a bound on the condition number of the stiffness matrix for the linear FE approximation of the Laplace operator for a general mesh. Bank and Scott [4] show that the condition number of the diagonally scaled stiffness matrix for the Laplace operator on an isotropic adaptive mesh is essentially the same as for a quasi-uniform mesh. Ainsworth, McLean, and Tran [2] and Graham and McLean [21] extend this result to the boundary element equations for locally quasi-uniform meshes. Du et al. [15] obtain a bound on the condition number of the stiffness matrix for a general diffusion operator on a general mesh which reveals the relation between the condition number and some mesh quality measures. The result is extended by Zhu and Du [43, 44] to parabolic problems. Shewchuk [36] provides a bound on the largest eigenvalue of the stiffness matrix scaled by the lumped mass matrix in terms of the maximum eigenvalues of local element matrices. More recently, bounds for the condition number of the stiffness matrix for the linear FE equations of a general diffusion operator (and implicit Runge-Kutta schemes of the corresponding parabolic problem) on an arbitrary mesh are developed in [27–29]

while the largest permissible time steps for explicit Runge-Kutta schemes for both linear and high-order FE approximations of parabolic problems are established in [24, 25]. These bounds take into full consideration the interplay between the mesh geometry and the diffusion matrix. Indeed, they show that the condition number of the stiffness matrix depends on three factors: the factor depending on the number of mesh elements and corresponding to the condition number of the linear FE equations for the Laplace operator on a uniform mesh, the mesh nonuniformity measured in the metric specified by the inverse diffusion matrix, and the mesh nonuniformity measured in the Euclidean metric. Moreover, the Jacobi preconditioning, or called the diagonal scaling, can effectively eliminate the effects of mesh nonuniformity and reduce those of the mesh nonuniformity with respect to the inverse diffusion matrix.

The objective of this paper is to study the conditioning for linear FVEM applied to anisotropic diffusion problems on general simplex meshes in any dimension. We shall use the ratio of the maximal singular value to the minimal eigenvalue of the symmetric part of the stiffness matrix to measure its conditioning (cf. (17) below). The task of estimating the condition number is then to develop an upper bound for the maximal singular value and a lower bound for the minimal eigenvalue of the symmetric part of the stiffness matrix. To this end, we use the FE bilinear form and show that the difference between the FE and FVE bilinear forms is small when the mesh is sufficiently fine. We also use a strategy similar to that in [29] for establishing a lower bound for the minimal eigenvalue of the symmetric part of the FVEM stiffness matrix. The results of this work are similar to those in [29]. In particular, the bound for the above-mentioned ratio depends on three factors too, i.e., the number of mesh elements and the mesh nonuniformity measured in the Euclidean metric and in the metric specified by the inverse diffusion matrix. Moreover, the analysis shows that the diagonal scaling can effectively eliminate the effects of mesh nonuniformity in the Euclidean metric. To a large extent, the current work can be viewed as an extension of [29] from FEM to FVEM. However, this extension is by no means trivial. As mentioned earlier, we have to deal with the nonsymmetric nature of the stiffness matrix in the current situation. Moreover, the current analysis is more technical and difficult since FVEM depends heavily on the specific geometry of the dual mesh elements which are formed by partitioning primary mesh elements in a certain manner. It should be emphasized that this work is also substantially different from that of [31] where FVEM has been studied for quasi-uniform meshes. Here, we consider not only the conditioning of FVEM on general simplicial meshes but also the effects of mesh nonuniformity and alignment with the diffusion matrix on the conditioning. Understanding these effects is crucial to the development of algorithms for efficient solution of FVEM algebraic systems.

The outline of this paper is as follows. The linear FVEM is described in §2 for the boundary value problem of an anisotropic diffusion equation. The definition of the condition number of the stiffness matrix and its estimates are given in §3. A similar analysis is carried out for the mass matrix in §4, followed by a selection of numerical examples in one, two, and three dimensions in §5. Conclusions are made in §6. Finally, the derivation for the expressions of two parameters in the mass matrix is given in Appendix A.

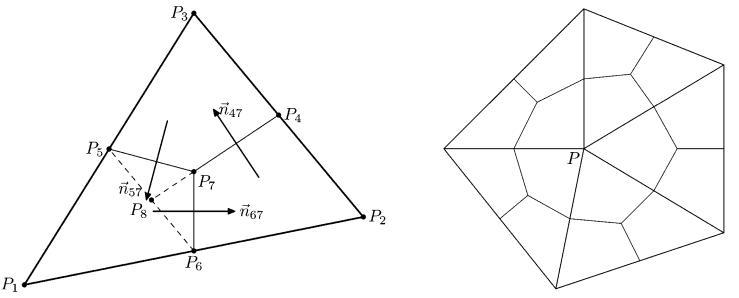
2. Linear finite volume element formulation

We consider the boundary value problem (BVP) of an anisotropic diffusion equation as

$$\left\{ \begin{array}{rcll} Lu \equiv -\nabla \cdot (\mathbb{D}\nabla u) & = & f, & \text{in } \Omega, \\ u & = & 0, & \text{on } \partial\Omega, \end{array} \right. \tag{1}$$

where $\Omega \subset \mathbb{R}^d$ $(d = 1, 2, 3, \dots)$ is a bounded polygonal/polyhedral domain, f is a given function, and $\mathbb{D} = \mathbb{D}(\boldsymbol{x}) = (d_{ij})_{d\times d}$ is the diffusion matrix. We assume that $\mathbb{D}$ is sufficiently smooth, symmetric, and strictly positive definite on Ω in the sense that there exist positive constants $0 < \underline{d} \le \overline{d}$ such that

$$\underline{d}\,|\xi|^2 \le \xi^T \mathbb{D}(\boldsymbol{x})\xi \le \overline{d}\,|\xi|^2 \quad \forall \xi \in \mathbb{R}^d, \quad \forall \boldsymbol{x} \in \Omega. \tag{2}$$



(a) (b)

Figure 1. Illustration of the dual element of the linear FVEM on element K or at vertex P (2-dimension).

Let $\mathcal{T}_h = \{K\}$ be a given simplicial mesh for Ω, let $\mathcal{N}_K$ be the set of all computing nodes in K, and let $\mathcal{N} = \bigcup_{K\in\mathcal{T}_h} \mathcal{N}_K$. For linear FVEM discretization, $\mathcal{N}_K$ is the set of the $(d+1)$ vertices of K. Divide each simplex into $(d+1)$ subregions by plane or line segments connecting the centroids of the simplex and its faces and edges. The dual element K_P^* associated with the vertex P is formed by the subregions surrounding P. The dual mesh is then defined as $\mathcal{T}_h^* = \{K_P^* \mid \forall P \in \mathcal{N}\}$ for the linear FVEM. The structures of the dual mesh on element $K \in \mathcal{T}_h$ (left) and at vertex P (right) in two dimensions are illustrated in Figure 1, where P_i $(i = 4, 5, 6)$ are the edge midpoints and P_7 is the centroid (i.e., the barycenter) of K. The trial and test function spaces are chosen as

$$U^h = \left\{u^h \mid u^h \in C(\Omega),\, u^h|_K \in P^1(K)\ \forall K \in \mathcal{T}_h \text{ and } u^h(P) = 0\ \forall P \in \mathcal{N} \cap (\partial\Omega)\right\},$$
$$V^h = \left\{v^h \mid v^h \in L^2(\Omega),\, v^h|_{K_P^*} = \text{constant}\ \forall K_P^* \in \mathcal{T}_h^* \text{ and } v^h|_{K_P^*} = 0\ \forall P \in \partial\Omega\right\}.$$

Denote the mapping from U^h to V^h by Π_h^*, i.e.,

$$\Pi_h^* u^h = u^h(P) \quad \forall \boldsymbol{x} \in K_P^*, \quad P \in \mathcal{N}.$$

We also denote the diameter of element K by h_K and define

$$|\mathbb{D}|_{1,\infty,K} = \max_{i,j=1,\dots,d} |d_{ij}|_{1,\infty,K}, \quad |\mathbb{D}|_{2,\infty,K} = \max_{i,j=1,\dots,d} |d_{ij}|_{2,\infty,K},$$
$$\mathbb{D}_K = \frac{1}{|K|}\int_K \mathbb{D}\,\mathrm{d}\boldsymbol{x},$$

where $|\cdot|_{m,p,K}$ is the seminorm of Sobolev space $W^{m,p}(K)$, $|K|$ is the volume (or the d-dimensional measure) of K, and $\mathbb{D}_K$ is the average of $\mathbb{D}$ over K.

The linear FVEM approximation of (1) is to find $u^h \in U^h$ such that

$$a_h(u^h, v^h) = (f, v^h) \qquad \forall v^h \in V^h, \tag{3}$$

where

$$a_h(u^h, v^h) = -\sum_{K_P^* \in \mathcal{T}_h^*} \int_{\partial K_P^*} (\mathbb{D}\nabla u^h) \cdot \mathbf{n}\, v^h \mathrm{d}s, \tag{4}$$

$$(f, v^h) = \sum_{K_P^* \in \mathcal{T}_h^*} \int_{K_P^*} f\, v^h \mathrm{d}\boldsymbol{x},$$

and $\mathbf{n}$ is the unit outward normal of ∂K_P^*. Denote the number of the elements and interior vertices (computing nodes) of $\mathcal{T}_h$ by N and N_{vi}, respectively. Assume that the vertices are ordered in such a way that the first N_{vi} vertices are the interior vertices. Then, U^h and u^h can be expressed as

$$U^h = \mathrm{span}\{\phi_1, \ldots, \phi_{N_{vi}}\}, \tag{5}$$

$$u^h = \sum_{j=1}^{N_{vi}} u_j \phi_j, \tag{6}$$

where ϕ_j is the linear basis function associated with the jth vertex P_j. Substituting (6) into (3) and taking v^h as the characteristic function of $K_{P_i}^*$ $(i = 1, \ldots, N_{vi})$ successively, we can rewrite (3) into matrix form as

$$A_{FV}\, \boldsymbol{u} = \boldsymbol{f}, \tag{7}$$

where $\boldsymbol{u} = (u_1, \ldots, u_{N_{vi}})^T$, $\boldsymbol{f} = (f_1, \ldots, f_{N_{vi}})^T$, and the entries of the stiffness matrix A and the right-hand-side vector $\boldsymbol{f}$ are given by

$$a_{ij}^{FV} = -\int_{\partial K_{P_i}^*} (\mathbb{D}\nabla\phi_j) \cdot \mathbf{n}\, \mathrm{d}s, \quad i, j = 1, \ldots, N_{vi}, \tag{8}$$

$$f_i = \int_{K_{P_i}^*} f\, \mathrm{d}\boldsymbol{x}, \quad i = 1, \ldots, N_{vi}. \tag{9}$$

Let ω_i be the element patch associated with P_i and let $\omega_{ij} = \omega_i \cap \omega_j$. Then we can rewrite a_{ij}^{FV} as

$$a_{ij}^{FV} = \sum_{K \in \omega_{ij}} a_{ij,K}^{FV}, \qquad a_{ij,K}^{FV} = -\int_{\partial K_{P_i}^* \cap K} (\mathbb{D}\nabla\phi_j) \cdot \mathbf{n}\, \mathrm{d}s. \tag{10}$$

Lemma 2.1. *The stiffness matrix A_{FV} is symmetric when $\mathbb{D}$ is piecewise constant on $\mathcal{T}_h$.*

Proof. Denote the face of K opposite to P_i by $l_{i,K}$ and the distance from P_i to $l_{i,K}$ by $\delta(P_i, l_{i,K})$. It is easy to see that

$$|K| = \frac{1}{d}\, \delta(P_i, l_{i,K})\, |l_{i,K}|, \quad \nabla\phi_i = -\frac{\mathbf{n}_{l_{i,K}}}{\delta(P_i, l_{i,K})},$$

where $|l_{i,K}|$ is the area (for 3D) or the $(d-1)$-dimensional measure of $l_{i,K}$, and $\mathbf{n}_{l_{i,K}}$ is the unit outward normal of $l_{i,K}$. Let $S_{i,k}$ be the kth face of $(\partial K_{P_i}^*) \cap K$ $(k = 1, \ldots, d)$. It is noted that the $S_{i,k}$'s are different from the $l_{i,K}$'s: $S_{i,k}$ is in the

interior of K while $l_{i,K}$ is a part of ∂K. Moreover, $\bigcup_{k=1}^{d} S_{i,k}$ separates the dual element corresponding to P_i from other dual elements restricted in K. We have

$$\begin{aligned}
\int_{\partial K^*_{P_i}\cap K} \mathbf{n}\,\mathrm{d}s &= \sum_{k=1}^{d}(\mathbf{n}_{S_{i,k}} \int_{S_{i,k}} 1\,\mathrm{d}s) = \sum_{k=1}^{d} \mathbf{n}_{S_{i,k}}\,|S_{i,k}| \\
&= \frac{2}{d(d+1)} \sum_{k=1}^{d} \mathbf{n}_{S_{i,k}}\,\frac{d(d+1)}{2}|S_{i,k}| \\
&= \frac{2}{d(d+1)} \sum_{k=1}^{d}(\mathbf{n}_{l_{i,K}}\,|l_{i,K}| + \frac{1}{2} \sum_{\substack{t\,=\,1,\,\ldots\,,\,d \\ t\,\neq\,k}} \mathbf{n}_{l_{t,K}}\,|l_{t,K}|) \\
&= \frac{2}{d(d+1)}\,(d\mathbf{n}_{l_{i,K}}\,|l_{i,K}| + \frac{d-1}{2} \sum_{k=1}^{d} \mathbf{n}_{l_{k,K}}\,|l_{k,K}|) \\
&= \frac{2}{d(d+1)}\,(d\mathbf{n}_{l_{i,K}}\,|l_{i,K}| - \frac{d-1}{2}\mathbf{n}_{l_{i,K}}\,|l_{i,K}|) \\
&= \frac{1}{d} \quad \mathbf{n}_{l_{i,K}}\,|l_{i,K}| = \frac{1}{d}\,(-\delta(P_i, l_{i,K})\,\nabla\phi_i)\,|l_{i,K}| = -|K|\,\nabla\phi_i,
\end{aligned}$$

where we have used the equalities

$$\mathbf{n}_{S_{i,k}}\,\frac{d(d+1)}{2}|S_{i,k}| = \mathbf{n}_{l_{i,K}}\,|l_{i,K}| + \frac{1}{2} \sum_{\substack{t\,=\,1,\,\ldots\,,\,d, \\ t\,\neq\,k}} \mathbf{n}_{l_{t,K}}\,|l_{t,K}|,$$

$$\mathbf{n}_{l_{i,K}}\,|l_{i,K}| + \sum_{k=1}^{d} \mathbf{n}_{l_{k,K}}\,|l_{k,K}| = 0.$$

(The second equality states the fact that the sum of the unit outward normal vectors of all faces multiplied by their $(d-1)$-dimensional measures vanishes for any polyhedron.)

When $\mathbb{D}$ is piecewise constant on $\mathcal{T}_h$, we get, for $i \neq j$,

$$\begin{aligned}
a^{FV}_{ij,K} &= -\int_{\partial K^*_{P_i}\cap K} (\mathbb{D}\nabla\phi_j)\cdot\mathbf{n}\,\mathrm{d}s = -(\mathbb{D}\nabla\phi_j)|_K \cdot \int_{\partial K^*_{P_i}\cap K} \mathbf{n}\,\mathrm{d}s \\
&= |K|(\mathbb{D}\nabla\phi_j)|_K \cdot \nabla\phi_i|_K = |K|(\mathbb{D}\nabla\phi_i)|_K \cdot \nabla\phi_j|_K = a^{FV}_{ji,K}. \qquad (11)
\end{aligned}$$

From this and (10), we get $a^{FV}_{ij} = a^{FV}_{ji}$, which implies that A_{FV} is symmetric. □

Remark 1. When $\mathbb{D}$ is piecewise constant on $\mathcal{T}_h$, we actually have $A_{FV} = A_{FE}$, where A_{FE} is the stiffness matrix for a linear finite element approximation of (1). This can be seen readily from (11) and (18). A similar result has been proven in [40] for the case with scalar diffusion coefficients in any dimension.

The above proof also shows that A_{FV} is not symmetric in general when $\mathbb{D}$ is not piecewise constant.

To conclude this section, we prove two orthogonality properties which are needed in the later analysis.

Lemma 2.2. *For any $K \in \mathcal{T}_h$, there hold*

$$\int_K g(v - \Pi_h^* v)\mathrm{d}\boldsymbol{x} = 0 \quad \forall g \in P^0(K),\ \forall v \in P^1(K), \tag{12}$$

$$\int_{l_{i,K}} g(v - \Pi_h^* v)\mathrm{d}s = 0 \quad \forall g \in P^0(l_{i,K}),\ \forall v \in P^1(l_{i,K}),\ \forall l_{i,K} \subset \partial K. \tag{13}$$

Here, $P^0(K)$ and $P^1(K)$ are the constant space and linear space on K.

Proof. Denote the vertices of K by $P_i,\ i = 1, \ldots, d+1$. For any $v \in P^1(K)$, we have

$$\int_K v\mathrm{d}\boldsymbol{x} = \sum_{i=1}^{d+1} \left(\frac{|K|}{d+1}\, v(P_i) \right) = \sum_{i=1}^{d+1} v(P_i) \int_{K \cap K^*_{P_i}} 1\mathrm{d}\boldsymbol{x} = \int_K \Pi_h^* v \ \mathrm{d}\boldsymbol{x},$$

which implies

$$\int_K (v - \Pi_h^* v)\, \mathrm{d}\boldsymbol{x} = 0.$$

Since $g \in P^0(K)$ is constant on K, the above equality gives (12).

Similarly, we can obtain (13). □

We consider the piecewise linear and piecewise constant approximations of $\mathbb{D}$ on $\mathcal{T}_h$ as

$$\mathbb{D}_1(\boldsymbol{x}) = \sum_{j=1}^{N_v} \mathbb{D}(P_j)\phi_j(\boldsymbol{x}),$$

$$\mathbb{D}_0(\boldsymbol{x}) = \frac{1}{|K|} \int_K \mathbb{D}(\tilde{\boldsymbol{x}}) d\tilde{\boldsymbol{x}} \quad \forall \boldsymbol{x} \in K,\ K \in \mathcal{T}_h.$$

Then from (12) and (13) we have

$$\sum_{K \in \mathcal{T}_h} \int_K \nabla \cdot (\mathbb{D}_1 \nabla v^h)(v^h - \Pi_h^* v^h)\mathrm{d}\boldsymbol{x} = 0 \quad \forall v^h \in U^h, \tag{14}$$

$$\sum_{K \in \mathcal{T}_h} \int_{\partial K} (\mathbb{D}_0 \nabla v^h) \cdot \mathbf{n}\, (v^h - \Pi_h^* v^h)\mathrm{d}s = 0 \quad \forall v^h \in U^h, \tag{15}$$

where $\mathbb{D}_0|_{\partial K}$ is understood as $\mathbb{D}_0|_{\partial K} = \mathbb{D}_0|_K$.

3. Conditioning of the stiffness matrix

In this section we study the conditioning of the stiffness matrix A_{FV} of linear FVEM. As shown in the previous section, A_{FV} is generally nonsymmetric for a non-piecewise-constant diffusion matrix. It is well known that a condition number in the standard definition does not provide much information for the convergence of iterative methods for nonsymmetric systems. On the other hand, when its symmetric part, $(A_{FV} + A_{FV}^T)/2$, is positive definite, which is to be shown later in this section, the convergence of the generalized minimal residual method (GMRES) is given by Eisenstat et al. [16] as

$$\|r_n\|_2 \le \left(1 - \frac{\lambda_{\min}^2((A_{FV} + A_{FV}^T)/2)}{\sigma_{\max}^2(A_{FV})} \right)^{n/2} \|r_0\|_2, \tag{16}$$

where $\sigma_{\max}(A_{FV})$ is the largest singular value of A_{FV}, $\lambda_{\min}((A_{FV}+A_{FV}^T)/2)$ is the minimal eigenvalue of the symmetric part, r_n is the residual of the corresponding linear system at the n-th iterate, and $\|\cdot\|_2$ stands for the matrix or vector 2-norm. From this, we can consider the "condition number"

$$\kappa(A_{FV}) = \frac{\sigma_{\max}(A_{FV})}{\lambda_{\min}((A_{FV}+A_{FV}^T)/2)}. \tag{17}$$

This definition reduces to the standard definition of the condition number (in 2-norm) for symmetric matrices. For notational simplicity and without causing confusion, we use the standard notation for this definition here and will hereafter simply refer to this as the condition number of A_{FV}.

In the following we shall show that the symmetric part of A_{FV} is positive definite when the mesh is sufficiently fine. We shall also establish an upper bound for $\sigma_{\max}(A_{FV})$ and a lower bound for $\lambda_{\min}((A_{FV}+A_{FV}^T)/2)$. Similar bounds will be obtained for the situation with the Jacobi (diagonal) preconditioning. As an additional benefit, the bounds will be used to reveal the effects of the interplay between the mesh geometry and the diffusion matrix on the conditioning of A_{FV}.

In our analysis, we use results for the conditioning of the stiffness matrix (A_{FE}) of a linear finite element approximation of (1). This topic has been studied by a number of researchers; e.g., see [1, 4, 15, 19, 25, 29, 39, 44]. Recall that the entries of A_{FE} are given by

$$a_{ij}^{FE} = \int_\Omega (\mathbb{D}\nabla\phi_j)\cdot\nabla\phi_i \mathrm{d}\boldsymbol{x} = \sum_{K\in\omega_{ij}} \int_K (\mathbb{D}\nabla\phi_j)\cdot\nabla\phi_i \mathrm{d}\boldsymbol{x}, \quad i,j=1,\dots,N_{vi}, \tag{18}$$

and A_{FE} is symmetric and positive definite for any diffusion matrix.

Denote the set of the indices of the neighboring vertices of P_j (excluding P_j) by $\mathcal{N}_j^0$ and define $\mathcal{N}_j = \{j\}\cup\mathcal{N}_j^0$. Let $p_{\mathcal{N}_j}$ be the number of the elements (indices of points) in $\mathcal{N}_j$ and let $p_{\max} = \max_{1\le j\le N_{vi}} p_{\mathcal{N}_j}$. Let

$$\begin{cases} C_0 = \frac{\sqrt{p_{\max}}}{\underline{d}}, \qquad C_{\tilde{\nabla}} = \frac{d}{d+1}\left(\frac{\sqrt{d+1}}{d!}\right)^{\frac{2}{d}}, \\ C_{\mathbb{D},K} = d^2\,(h_K\,|\mathbb{D}|_{2,\infty,K} + |\mathbb{D}|_{1,\infty,K}), \\ H_h = \max\limits_{K\in\mathcal{T}_h} C_{\mathbb{D},K}\,h_K. \end{cases} \tag{19}$$

Notice that $H_h\to 0$ as $h\equiv\max_K h_K\to 0$.

Lemma 3.1. *There holds*

$$|a_{ij}^{FV} - a_{ij}^{FE}| \le \sum_{K\in\omega_{ij}} C_{\mathbb{D},K}\,|K|^{1/2}\,|\phi_j|_{1,K} \quad \forall i,j=1,\dots,N_{vi}. \tag{20}$$

Proof. Using the definitions of a_{ij}^{FV} and a_{ij}^{FE} and the divergence theorem, we have

$$\begin{aligned}
a_{ij}^{FV} &= -\int_{\partial K_{P_i}^*} (\mathbb{D}\nabla\phi_j)\cdot\mathbf{n}\mathrm{d}s = -\sum_{K^*\in\mathcal{T}_h^*}\int_{\partial K^*}(\mathbb{D}\nabla\phi_j)\cdot\mathbf{n}\,\Pi_h^*\phi_i\mathrm{d}s \\
&= -\sum_{K\in\omega_{ij}}\sum_{K^*\in\mathcal{T}_h^*}\int_{\partial K^*\cap K}(\mathbb{D}\nabla\phi_j)\cdot\mathbf{n}\,\Pi_h^*\phi_i\mathrm{d}s \\
&= \sum_{K\in\omega_{ij}}\int_{\partial K}(\mathbb{D}\nabla\phi_j)\cdot\mathbf{n}\,\Pi_h^*\phi_i\mathrm{d}s - \sum_{K\in\omega_{ij}}\int_K \nabla\cdot(\mathbb{D}\nabla\phi_j)\,\Pi_h^*\phi_i\mathrm{d}\boldsymbol{x}, \\
a_{ij}^{FE} &= \sum_{K\in\omega_{ij}}\int_K(\mathbb{D}\nabla\phi_j)\cdot\nabla\phi_i\mathrm{d}\boldsymbol{x} \\
&= \sum_{K\in\omega_{ij}}\int_{\partial K}(\mathbb{D}\nabla\phi_j)\cdot\mathbf{n}\,\phi_i\mathrm{d}s - \sum_{K\in\omega_{ij}}\int_K \nabla\cdot(\mathbb{D}\nabla\phi_j)\,\phi_i\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Noticing that $|\phi_i - \Pi_h^*\phi_i| \le 1$, we get

$$|\phi_i - \Pi_h^*\phi_i|_{0,K} = \left(\int_K(\phi_i-\Pi_h^*\phi_i)^2\,\mathrm{d}\boldsymbol{x}\right)^{1/2} \le \left(\int_K 1\,\mathrm{d}\boldsymbol{x}\right)^{1/2} \le |K|^{1/2}.$$

Using the orthogonality properties (14) and (15), the trace theorem, and the fact that $\nabla\phi_j$ is constant in each element, we have

$$\begin{aligned}
|a_{ij}^{FV} - a_{ij}^{FE}| &= |\sum_{K\in\omega_{ij}}\int_{\partial K}(\mathbb{D}\nabla\phi_j)\cdot\mathbf{n}\,(\phi_i-\Pi_h^*\phi_i)\mathrm{d}s \\
&\quad - \sum_{K\in\omega_{ij}}\int_K\nabla\cdot(\mathbb{D}\nabla\phi_j)\,(\phi_i-\Pi_h^*\phi_i)\mathrm{d}\boldsymbol{x}| \\
&\le |\sum_{K\in\omega_{ij}}\int_K\nabla\cdot((\mathbb{D}-\mathbb{D}_1)\nabla\phi_j)(\phi_i-\Pi_h^*\phi_i)\mathrm{d}\boldsymbol{x}| \\
&\quad + |\sum_{K\in\omega_{ij}}\int_{\partial K}((\mathbb{D}-\mathbb{D}_0)\nabla\phi_j)\cdot\mathbf{n}\,(\phi_i-\Pi_h^*\phi_i)\mathrm{d}s| \\
&\le \sum_{K\in\omega_{ij}}\left(d^2\,h_K\,|\mathbb{D}|_{2,\infty,K}\,|\phi_j|_{1,K}|\phi_i-\Pi_h^*\phi_i|_{0,K}\right) \\
&\quad + \sum_{K\in\omega_{ij}}\left(d^2\,|\mathbb{D}|_{1,\infty,K}\,|\phi_j|_{1,K}|\phi_i-\Pi_h^*\phi_i|_{0,K}\right) \\
&\le \sum_{K\in\omega_{ij}} C_{\mathbb{D},K}\,|K|^{1/2}\,|\phi_j|_{1,K},
\end{aligned}$$

which gives (20). □

Lemma 3.2. *Let $a(\cdot,\cdot)$ be the bilinear form of FEM associated with the BVP* (1) *and let $a_h(\cdot,\cdot)$ be the bilinear form of FVEM defined in* (4)*. Then,*

$$|a_h(u^h,\Pi_h^*u^h) - a(u^h,u^h)| \le H_h\,|u^h|_{1,\Omega}^2 \quad \forall u^h\in U^h. \tag{21}$$

Proof. The proof is similar to that of Lemma 3.1. Indeed, for any $u^h \in U^h$, from the trace theorem and the orthogonality properties (14) and (15) we have

$$\begin{aligned}
|a_h(u^h, \Pi_h^* u^h) - a(u^h, u^h)| &= |\sum_{K\in\mathcal{T}_h} \int_K \nabla\cdot(\mathbb{D}\nabla u^h)(u^h - \Pi_h^* u^h)\mathrm{d}\boldsymbol{x} \\
&\quad - \sum_{K\in\mathcal{T}_h} \int_{\partial K} (\mathbb{D}\nabla u^h)\cdot\mathbf{n}\,(u^h - \Pi_h^* u^h)\mathrm{d}s| \\
&\le |\sum_{K\in\mathcal{T}_h} \int_K \nabla\cdot((\mathbb{D}-\mathbb{D}_1)\nabla u^h)(u^h - \Pi_h^* u^h)\mathrm{d}\boldsymbol{x}| \\
&\quad + |\sum_{K\in\mathcal{T}_h} \int_{\partial K} ((\mathbb{D}-\mathbb{D}_0)\nabla u^h)\cdot\mathbf{n}\,(u^h - \Pi_h^* u^h)\mathrm{d}s| \\
&\le \sum_{K\in\mathcal{T}_h} (d^2 h_K^2 |\mathbb{D}|_{2,\infty,K}\,|u^h|_{1,K}^2 + d^2 h_K |\mathbb{D}|_{1,\infty,K}\,|u^h|_{1,K}^2) \\
&\le \sum_{K\in\mathcal{T}_h} \left(C_{\mathbb{D},K} h_K |u^h|_{1,K}^2\right) \\
&\le H_h\,|u^h|_{1,\Omega}^2.
\end{aligned}$$

□

These two lemmas indicate that A_{FV} and A_{FE} are "close" when the mesh is sufficiently fine. Thus, we can establish properties of A_{FV} via estimating the difference between A_{FV} and A_{FE}. Moreover, it is interesting to point out that a result similar to Lemma 3.2 has been obtained in [40] for isotropic meshes.

3.1. **Largest singular value of the stiffness matrix.** We assume that the reference element $\hat{K}$ has been chosen to be equilateral and have unit diameter. Denote the affine mapping between $\hat{K}$ and element K by F_K and its Jacobian matrix by F_K'.

Theorem 3.1. *Assume that the mesh is sufficiently fine so that $H_h < \underline{d}$, where $\underline{d}$ is the minimum eigenvalue of $\mathbb{D}$ (cf. (2)). Then, the largest singular value of the stiffness matrix $A_{FV} = (a_{ij}^{FV})$ for the linear FVEM approximation of BVP (1) is bounded above by*

$$\sigma_{\max}(A_{FV}) \;\le\; C_{\tilde{\nabla}}(d+1)(1+C_0 H_h)\max_j \sum_{K\in\omega_j} |K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2, \tag{22}$$

$$\sigma_{\max}(S^{-1}A_{FV}S^{-1}) \;\le\; \frac{(1+C_0 H_h)}{(1-\underline{d}^{-1}H_h)}(d+1), \tag{23}$$

where S is the Jacobi preconditioner for A_{FV}, i.e., $S = (A_{FV}^D)^{1/2}$, with A_{FV}^D being the diagonal part of A_{FV}.

Proof. Let $A_\delta = A_{FV} - A_{FE}$. Then, from Hölder's inequality and Lemma 3.1 we have

$$\begin{aligned}\|A_\delta \boldsymbol{v}\|^2 &= \sum_{i=1}^{N_{vi}} \left(\sum_{j\in\mathcal{N}_i} v_j (a_{ij}^{FV} - a_{ij}^{FE}) \right)^2 \\ &\le \sum_{i=1}^{N_{vi}} p_{\mathcal{N}_i} \sum_{j\in\mathcal{N}_i} v_j^2 (a_{ij}^{FV} - a_{ij}^{FE})^2 \\ &\le p_{\max} \sum_{j=1}^{N_{vi}} v_j^2 \sum_{i\in\mathcal{N}_j} (a_{ij}^{FV} - a_{ij}^{FE})^2 \\ &\le p_{\max} \sum_{j=1}^{N_{vi}} v_j^2 \sum_{i\in\mathcal{N}_j} \left(\sum_{K\in\omega_{ij}} C_{\mathbb{D},K} |K|^{1/2} |\phi_j|_{1,K} \right)^2, \end{aligned} \tag{24}$$

where $p_{\mathcal{N}_i}$ is the number of the elements (indices of points) in $\mathcal{N}_i$ and $p_{\max}$ is the maximum value of all $p_{\mathcal{N}_i}$ (defined upon (19)).

Next we establish a lower bound for a_{jj}^{FE}. We have

$$\begin{aligned} a_{jj}^{FE} &= \sum_{K\in\omega_j} \int_K (\mathbb{D}\nabla\phi_j)\cdot\nabla\phi_j \mathrm{d}\boldsymbol{x} \\ &\ge \underline{d} \sum_{K\in\omega_j} \int_K (\nabla\phi_j)\cdot\nabla\phi_j \mathrm{d}\boldsymbol{x} = \underline{d} \sum_{K\in\omega_j} |\phi_j|_{1,K}^2. \end{aligned} \tag{25}$$

We observe that when going through all elements in $\mathcal{N}_j$, each mesh element in ω_j will be encountered $(d+1)$ times (due to the fact that each element has $(d+1)$ vertices). Then, from Jensen's inequality we have

$$\begin{aligned}(a_{jj}^{FE})^2 &\ge \underline{d}^2 \left(\sum_{K\in\omega_j} |\phi_j|_{1,K}^2 \right)^2 = \underline{d}^2 \left(\frac{1}{d+1} \sum_{i\in\mathcal{N}_j} \sum_{K\in\omega_{ij}} |\phi_j|_{1,K}^2 \right)^2 \\ &\ge \frac{\underline{d}^2}{(d+1)^2} \sum_{i\in\mathcal{N}_j} \left(\sum_{K\in\omega_{ij}} |\phi_j|_{1,K}^2 \right)^2. \end{aligned} \tag{26}$$

Moreover,

$$|\phi_j|_{1,K} = \left(\int_K |\nabla\phi_j|^2 \mathrm{d}\boldsymbol{x} \right)^{1/2} \ge \left(\int_K (\frac{1}{h_K})^2 \mathrm{d}\boldsymbol{x} \right)^{1/2} = \frac{1}{h_K} |K|^{1/2}. \tag{27}$$

Denoting the diagonal part of A_{FE} by A_{FE}^D, and combining (24), we have

$$\sup_{\boldsymbol{v}\neq 0} \frac{\|A_\delta \boldsymbol{v}\|^2}{\|A_{FE}^D \boldsymbol{v}\|^2} \le \sup_{\boldsymbol{v}\neq 0} \frac{p_{\max} \sum_{j=1}^{N_{vi}} v_j^2 \sum_{i\in\mathcal{N}_j} \beta_{ij}^2}{\sum_{j=1}^{N_{vi}} v_j^2 (a_{jj}^{FE})^2} \le p_{\max} \max_{1\le j\le N_{vi}} \frac{\sum_{i\in\mathcal{N}_j} \beta_{ij}^2}{(a_{jj}^{FE})^2},$$

where $\alpha_K = C_{\mathbb{D},K}\,|K|^{1/2}$ and $\beta_{ij} = \sum_{K\in\omega_{ij}} \alpha_K |\phi_j|_{1,K}$. And with (26), we have

$$\begin{aligned}
\frac{\sum_{i\in\mathcal{N}_j}\beta_{ij}^2}{(a_{jj}^{FE})^2} &\le \frac{(d+1)^2}{\underline{d}^2}\frac{\sum_{i\in\mathcal{N}_j}\left(\sum_{K\in\omega_{ij}}\alpha_K\,|\phi_j|_{1,K}\right)^2}{\sum_{i\in\mathcal{N}_j}\left(\sum_{K\in\omega_{ij}}|\phi_j|_{1,K}^2\right)^2}\\
&\le \frac{(d+1)^2}{\underline{d}^2}\max_{i\in\mathcal{N}_j}\left(\frac{\sum_{K\in\omega_{ij}}\alpha_K\,|\phi_j|_{1,K}}{\sum_{K\in\omega_{ij}}|\phi_j|_{1,K}^2}\right)^2\\
&\le \frac{(d+1)^2}{\underline{d}^2}\max_{i\in\mathcal{N}_j}\left(\max_{K\in\omega_{ij}}\frac{\alpha_K\,|\phi_j|_{1,K}}{|\phi_j|_{1,K}^2}\right)^2.
\end{aligned}$$

So, with (27), we have

$$\begin{aligned}
\sup_{\boldsymbol{v}\neq 0}\frac{\|A_\delta\boldsymbol{v}\|^2}{\|A_{FE}^D\boldsymbol{v}\|^2} &\le p_{\max}\frac{(d+1)^2}{\underline{d}^2}\max_{1\le j\le N_{vi}}\max_{i\in\mathcal{N}_j}\left(\max_{K\in\omega_{ij}}\frac{\alpha_K}{|K|^{1/2}h_K^{-1}}\right)^2\\
&= p_{\max}\frac{(d+1)^2}{\underline{d}^2}\left(\max_{K\in\mathcal{T}_h}\frac{C_{\mathbb{D},K}\,|K|^{1/2}}{\frac{1}{h_K}\,|K|^{1/2}}\right)^2\\
&\le (d+1)^2C_0^2H_h^2, \qquad (28)
\end{aligned}$$

where C_0 and H_h are defined in (19). From this we have

$$\sup_{\boldsymbol{v}\neq 0}\frac{\|A_\delta\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2} \le \sup_{\boldsymbol{v}\neq 0}\frac{\|A_\delta\boldsymbol{v}\|^2}{\|A_{FE}^D\boldsymbol{v}\|^2}\cdot\sup_{\boldsymbol{v}\neq 0}\frac{\|A_{FE}^D\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2} \le (d+1)^2C_0^2H_h^2(\max_j a_{jj}^{FE})^2.$$

Then,

$$\begin{aligned}
&\sup_{\boldsymbol{v}\neq 0}\frac{\|A_{FV}\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2}\\
=\;&\sup_{\boldsymbol{v}\neq 0}\frac{\|A_{FE}\boldsymbol{v}+A_\delta\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2}\\
\le\;&\sup_{\boldsymbol{v}\neq 0}\frac{\|A_{FE}\boldsymbol{v}\|^2+2\|A_{FE}\boldsymbol{v}\|\,\|A_\delta\boldsymbol{v}\|+\|A_\delta\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2}\\
\le\;&\sup_{\boldsymbol{v}\neq 0}\frac{\|A_{FE}\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2}+2\sup_{\boldsymbol{v}\neq 0}\frac{\|A_{FE}\boldsymbol{v}\|}{\|\boldsymbol{v}\|}\cdot\sup_{\boldsymbol{v}\neq 0}\frac{\|A_\delta\boldsymbol{v}\|}{\|\boldsymbol{v}\|}+\sup_{\boldsymbol{v}\neq 0}\frac{\|A_\delta\boldsymbol{v}\|^2}{\|\boldsymbol{v}\|^2}\\
\le\;&(d+1)^2(\max_j a_{jj}^{FE})^2+2\,(d+1)\max_j a_{jj}^{FE}\cdot(d+1)C_0H_h\max_j a_{jj}^{FE}\\
&+((d+1)C_0H_h)^2(\max_j a_{jj}^{FE})^2\\
\le\;&(1+C_0H_h)^2(d+1)^2(\max_j a_{jj}^{FE})^2,
\end{aligned}$$

which gives

$$\sigma_{\max}(A_{FV}) = \sup_{\boldsymbol{v}\neq 0}\frac{\|A_{FV}\boldsymbol{v}\|}{\|\boldsymbol{v}\|} \le (1+C_0H_h)(d+1)\max_j a_{jj}^{FE}. \qquad (29)$$

From [25, Lemma 2.5] we have

$$\max_j a_{jj}^{FE} \le C_{\tilde{\nabla}}\max_j\sum_{K\in\omega_j}|K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2.$$

Combining this with (29) we obtain (22).

For (23), from Lemma 3.1, (25), and (27) we have

$$\frac{|(a_{jj}^{FV} - a_{jj}^{FE})|}{|a_{jj}^{FE}|} \leq \underline{d}^{-1} H_h,$$

which implies

$$|a_{jj}^{FE}| \leq \frac{|a_{jj}^{FV}|}{1 - \underline{d}^{-1} H_h}. \tag{30}$$

Combining this with (29), we get

$$\sigma_{\max}(A_{FV}) \leq \frac{(1 + C_0 H_h)}{1 - \underline{d}^{-1} H_h} (d+1) \max_j a_{jj}^{FV}.$$

Applying the same procedure for any diagonal scaling $S = (s_j)$ we can obtain

$$\max_j (s_j^{-2} a_{jj}^{FV}) \leq \sigma_{\max}(S^{-1} A_{FV} S^{-1}) \leq \frac{(1 + C_0 H_h)}{(1 - \underline{d}^{-1} H_h)} (d+1) \max_j (s_j^{-2} a_{jj}^{FV}).$$

For the Jacobi preconditioning we have $s_j^2 = a_{jj}^{FV}$, which gives estimate (23). □

3.2. **Smallest eigenvalue of** $(A_{FV} + A_{FV}^T)/2$**.**

Lemma 3.3. *A_{FV} and $(A_{FV} + A_{FV}^T)/2$ are positive definite when the mesh is sufficiently fine so that $H_h < \underline{d}$.*

Proof. From (2) we have

$$a(u^h, u^h) = \sum_{K \in \mathcal{T}_h} \int_K \nabla u^h \cdot (\mathbb{D} \nabla u^h) \mathrm{d}\boldsymbol{x} \geq \underline{d} \, |u^h|_{1,\Omega}^2. \tag{31}$$

Then, from Lemma 3.2 we have

$$\begin{aligned} a_h(u^h, \Pi_h^* u^h) &\geq a(u^h, u^h) - |a_h(u^h, \Pi_h^* u^h) - a(u^h, u^h)| \\ &\geq (\underline{d} - H_h) |u^h|_{1,\Omega}^2. \end{aligned} \tag{32}$$

From Poincare's inequality, there exists a constant $\gamma > 0$ such that

$$\boldsymbol{u}^T \frac{A_{FV} + A_{FV}^T}{2} \boldsymbol{u} = \boldsymbol{u}^T A_{FV} \boldsymbol{u} = a_h(u^h, \Pi_h^* u^h) \geq \gamma (\underline{d} - H_h) ||u^h||_{L^2(\Omega)}^2.$$

Thus, A_{FV} and $(A_{FV} + A_{FV}^T)/2$ are positive definite when $\underline{d} > H_h$. □

Theorem 3.2. *Assume that the mesh is sufficiently fine so that $H_h < \underline{d}$. The smallest eigenvalue of $(A_{FV} + A_{FV}^T)/2$ for the linear FVEM approximation of BVP (1) is bounded from below by*

(33)

$$\lambda_{\min}\left(\frac{A_{FV} + A_{FV}^T}{2}\right) \geq \frac{C \underline{d}}{N} \div \begin{cases} 1 & \text{for} \quad d = 1, \\ (1 - \underline{d}^{-1} H_h)(1 + \ln(\frac{|\overline{K}|}{|K_{\min}|})) & \text{for} \quad d = 2, \\ (1 - \underline{d}^{-1} H_h) \Big(\frac{1}{N} \sum_{K \in \mathcal{T}_h} \Big(\frac{|\overline{K}|}{|K|} \Big)^{\frac{d-2}{2}} \Big)^{\frac{2}{d}} & \text{for} \quad d \geq 3, \end{cases}$$

where $|\overline{K}| = \frac{1}{N}|\Omega|$ is the average element size and C is a constant independent of the mesh and the diffusion matrix. Moreover, the smallest singular value of the

diagonally (Jacobi) preconditioned stiffness matrix is bounded from below by

$$\lambda_{\min}(S^{-1}\frac{A_{FV}+A_{FV}^T}{2}S^{-1}) \tag{34}$$
$$\geq \frac{C}{N^{\frac{2}{d}}} \div \begin{cases} \left(\frac{1}{N\underline{d}}\sum_{K\in\mathcal{T}_h}\mathbb{D}(\boldsymbol{x}_K)\frac{|\overline{K}|}{|K|}\right) & \text{for} \quad d=1 \\ \left(\frac{1}{N(\underline{d}-H_h)}\sum_{K\in\mathcal{T}_h}|K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2\right) & \\ \quad\cdot\left(1+\left|\ln\frac{\max_{K\in\mathcal{T}_h}\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2}{\sum_{K\in\mathcal{T}_h}|K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2}\right|\right) & \text{for} \quad d=2 \\ \left(\frac{1}{N(\underline{d}-H_h)^{\frac{d}{2}}}\sum_{K\in\mathcal{T}_h}|K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2^{\frac{d}{2}}\right)^{\frac{2}{d}} & \text{for} \quad d\geq 3. \end{cases}$$

Proof. The proof of this theorem is similar to that of Lemma 5.1 of [29] for linear finite element approximation. For completeness, we give the detail of the proof here.

As in [29], we need to treat the cases with $d=1$, $d=2$, and $d\geq 3$ separately since the proof is based on Sobolev's inequality [20, Theorem 7.10] which has different forms in these cases. In the following, the function $u^h\in U^h$ and its vector form $\boldsymbol{u}=(u_1,\ldots,u_{N_{vi}})^T$ are used synonymously.

Case $d=1$. In one dimension, it is known (e.g., see [30]) that

$$\boldsymbol{u}^T A_{FV}\boldsymbol{u} = a_h(u^h,\Pi_h^* u^h) \geq \underline{d}\,|u_h|_{1,\Omega}^2.$$

From Sobolev's inequality and the equivalence of vector norms, we have

$$\begin{aligned}\boldsymbol{u}^T\frac{A_{FV}+A_{FV}^T}{2}\boldsymbol{u} &= \boldsymbol{u}^T A_{FV}\boldsymbol{u} \geq \underline{d}\,|u^h|_{1,\Omega}^2 \geq \underline{d}\,C_S\,|\Omega|^{-1}\sup_{\Omega}|u^h|^2 \\ &= \underline{d}\,C_S\,|\Omega|^{-1}\max_j u_j^2 \geq \underline{d}\,C_S\,|\Omega|^{-1}N^{-1}\boldsymbol{u}^T\boldsymbol{u},\end{aligned}$$

where C_S is the constant associated with Sobolev's inequality. Thus, we have

$$\lambda_{\min}((A_{FV}+A_{FV}^T)/2) \geq C\underline{d}N^{-1},$$

which gives (33) (with $d=1$).

With diagonal scaling, we have

$$\boldsymbol{u}^T S^{-1}\frac{A_{FV}+A_{FV}^T}{2}S^{-1}\boldsymbol{u} \geq C\underline{d}\max_j s_j^{-2}u_j^2 \geq C\,\underline{d}\,\frac{\sum_j u_j^2}{\sum_j s_j^2} \geq C\,\underline{d}\,\frac{\boldsymbol{u}^T\boldsymbol{u}}{\sum_j s_j^2}. \tag{35}$$

In one dimension, $|\nabla\phi_j| = |K|^{-1}$ when restricted in K. From this and noticing that ω_j contains at most two elements, we have

$$\begin{aligned}\sum_j s_j^2 = \sum_j A_{jj}^{FV} &= \sum_j\sum_{K\in\omega_j} -\mathbf{n}_{\boldsymbol{x}_K}\cdot(\mathbb{D}(\boldsymbol{x}_K)\nabla\phi_j) \\ &= \sum_j\sum_{K\in\omega_j}\frac{\mathbb{D}(\boldsymbol{x}_K)}{|K|} \leq 2\sum_{K\in\mathcal{T}_h}\frac{\mathbb{D}(\boldsymbol{x}_K)}{|K|},\end{aligned}$$

where $\boldsymbol{x}_K$ denotes the centroid of K and $\mathbf{n}_{\boldsymbol{x}_K}$ is the outward normal vector from P_j to $\boldsymbol{x}_K$. Substituting this into (35) we get (34) (for $d=1$).

Case $d = 2$. From the proof of Lemma 5.1 of [29], we have

$$|u^h|_{1,\Omega}^2 \geq Cq^{-1}\left(\sum_{K\in\mathcal{T}_h}\alpha_K^{\frac{q}{q-2}}\right)^{-\frac{q-2}{q}}\left(\sum_j u_j^2\sum_{K\in\omega_j}\alpha_K|K|^{\frac{2}{q}}\right), \tag{36}$$

where $\{\alpha_K,\ K\in\mathcal{T}_h\}$ is an arbitrary set of not-all-zero nonnegative numbers and $q>2$ is an arbitrary constant. Taking $\alpha_K = |K|^{-2/q}$ gives

$$|u^h|_{1,\Omega}^2 \geq Cq^{-1}\left(\sum_{K\in\mathcal{T}_h}|K|^{-\frac{2}{q-2}}\right)^{-\frac{q-2}{q}}\sum_j u_j^2. \tag{37}$$

Then from (32) and the above inequality we have

$$\begin{aligned}
\boldsymbol{u}^T\frac{A_{FV}+A_{FV}^T}{2}\boldsymbol{u} &= \boldsymbol{u}^T A_{FV}\boldsymbol{u} = a_h(u^h,\Pi_h^* u^h)\\
&\geq (\underline{d}-H_h)|u^h|_{1,\Omega}^2\\
&\geq C(\underline{d}-H_h)q^{-1}\left(\sum_{K\in\mathcal{T}_h}|K|^{-\frac{2}{q-2}}\right)^{-\frac{q-2}{q}}\sum_j u_j^2\\
&\geq C(\underline{d}-H_h)q^{-1}\left(N|K_{\min}|^{-\frac{2}{q-2}}\right)^{-\frac{q-2}{q}}\sum_j u_j^2\\
&= C(\underline{d}-H_h)N^{-1}\left[q^{-1}(N\,|K_{\min}|)^{\frac{2}{q}}\right]\sum_j u_j^2, \qquad (38)
\end{aligned}$$

where $K_{\min}$ denotes the element with the minimal area. The above bound can be maximized for $q = \max\{2, |\ln(N\,|K_{\min}|)|\}$ (with $q=2$ being viewed as the limiting case $q\to 2^+$) with

$$q^{-1}(N\,|K_{\min}|)^{\frac{2}{q}} \geq \frac{C}{1+|\ln(N\,|K_{\min}|)|}.$$

Substituting this into (38) and using the definition of the average element size, we obtain (33) (with $d=2$).

With diagonal scaling, we have

$$\begin{aligned}
&\boldsymbol{u}^T S^{-1}\frac{A_{FV}+A_{FV}^T}{2}S^{-1}\boldsymbol{u}\\
&\quad\geq C(\underline{d}-H_h)\frac{1}{q}\left(\sum_{K\in\mathcal{T}_h}\alpha_K^{\frac{q}{q-2}}\right)^{-\frac{q-2}{q}}\left(\sum_j u_j^2 s_j^{-2}\sum_{K\in\omega_j}\alpha_K|K|^{\frac{2}{q}}\right).
\end{aligned}$$

For the Jacobi preconditioning $s_j^2 = a_{jj}^{FV}$. Letting $u^h=\phi_j$ in (21), we have

$$\begin{aligned}
s_j^2 &= a_{jj}^{FV} = a_{jj}^{FE} + (a_{jj}^{FV}-a_{jj}^{FE})\\
&\geq a(\phi_j,\phi_j) - |a_h(\phi_j,\Pi_h^*\phi_j) - a(\phi_j,\phi_j)|\\
&\geq \sum_{K\in\omega_j}|K|\,|\nabla\phi_j|\,|\nabla\phi_j|\left(\frac{\nabla\phi_j}{|\nabla\phi_j|}\cdot(\mathbb{D}_K\frac{\nabla\phi_j}{|\nabla\phi_j|}) - H_h\right).
\end{aligned}$$

Take

$$\alpha_K = |K|^{\frac{q-2}{q}} \sum_{i_K=1}^{d+1} \nabla\phi_{i_K} \cdot (\mathbb{D}_K \nabla\phi_{i_K})$$
$$= |K|^{\frac{q-2}{q}} \sum_{i_K=1}^{d+1} \hat{\nabla}\hat{\phi}_{i_K} \cdot ((F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\hat{\nabla}\hat{\phi}_{i_K}),$$

where $\hat{\phi}_i$ is a linear basis function and $\hat{\nabla}$ is the gradient operator on the reference element $\hat{K}$. It is not difficult to show that

$$s_j^{-2} \sum_{K\in\omega_j} \alpha_K |K|^{\frac{2}{q}} \geq 1, \qquad \alpha_K \leq (d+1) C_{\hat{\phi}} |K|^{\frac{q-2}{q}} \|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2,$$

where $C_{\hat{\phi}} = \max_{i_K=1,\ldots,d+1} \|\hat{\nabla}\hat{\phi}_{i_K}\|^2$. With these and choosing the value for the index q in a similar manner as for the case without scaling we obtain (34) (for $d = 2$).

Case $d \geq 3$. Following a similar procedure as for the $d = 2$ case, we have

$$\boldsymbol{u}^T \frac{A_{FV} + A_{FV}^T}{2} \boldsymbol{u} \geq C(\underline{d} - H_h) \left(\sum_{k\in\mathcal{T}_h} \alpha_K^{\frac{d}{2}} \right)^{-\frac{2}{d}} \sum_j u_j^2 \sum_{K\in\omega_j} \alpha_K |K|^{\frac{d-2}{d}}.$$

Choosing $\alpha_K = |K|^{-\frac{d-2}{d}}$ gives

$$\boldsymbol{u}^T \frac{A_{FV} + A_{FV}^T}{2} \boldsymbol{u} = C(\underline{d} - H_h) \left(\sum_{k\in\mathcal{T}_h} |K|^{\frac{2-d}{2}} \right)^{-\frac{2}{d}} \sum_j u_j^2.$$

The estimate (34) for $d \geq 3$ follows from this and the definition of the average element size.

The bound for the diagonally scaled stiffness matrix is obtained by choosing

$$\alpha_K = |K|^{\frac{2}{d}} \sum_{i_K=1}^{d+1} \hat{\nabla}\hat{\phi}_{i_K} \cdot ((F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\hat{\nabla}\hat{\phi}_{i_K}).$$

□

3.3. **Condition number of the stiffness matrix.**

Theorem 3.3. *The condition number of the stiffness matrix for the linear finite volume element approximation of homogeneous BVP* (1) *is bounded by*

$$\kappa(A_{FV}) \leq C\,(1 + C_0 H_h)\, N^{\frac{2}{d}} \cdot \left(\frac{1}{\underline{d}\, N^{\frac{2-d}{d}}} \max_j \sum_{K\in\omega_j} |K|\, \|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2 \right)$$
$$\times \begin{cases} 1 & \text{for} \quad d = 1, \\ (1 - \underline{d}^{-1} H_h)(1 + \ln(\frac{|\overline{K}|}{|K_{\min}|})) & \text{for} \quad d = 2, \\ (1 - \underline{d}^{-1} H_h)\Big(\frac{1}{N} \sum_{K\in\mathcal{T}_h} \Big(\frac{|\overline{K}|}{|K|}\Big)^{\frac{d-2}{2}}\Big)^{\frac{2}{d}} & \text{for} \quad d \geq 3, \end{cases} \tag{39}$$

where $H_h = \max_{K\in\mathcal{T}_h} d^2(h_K^2|\mathbb{D}|_{2,\infty,K} + h_K|\mathbb{D}|_{1,\infty,K})$ *and* $|\overline{K}| = \frac{1}{N}|\Omega|$ *is the average element size. With the diagonally (Jacobi) preconditioning, the "condition number" of the stiffness matrix is bounded by*

$$\kappa(S^{-1}A_{FV}S^{-1}) \le \frac{C\,(1+C_0H_h)\,N^{\frac{2}{d}}}{(1-\underline{d}^{-1}H_h)} \times \begin{cases} \left(\frac{1}{N\underline{d}}\sum_{K\in\mathcal{T}_h}\mathbb{D}(\boldsymbol{x}_K)\frac{|\overline{K}|}{|K|}\right) & \text{for} \quad d=1, \\ \left(\frac{1}{N(\underline{d}-H_h)}\sum_{K\in\mathcal{T}_h}|K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2\right) & \\ \quad\cdot\left(1+\left|\ln\frac{\max_{K\in\mathcal{T}_h}\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2}{\sum_{K\in\mathcal{T}_h}|K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2}\right|\right) & \text{for} \quad d=2, \\ \left(\frac{1}{N(\underline{d}-H_h)^{\frac{d}{2}}}\sum_{K\in\mathcal{T}_h}|K|\,\|(F_K')^{-1}\mathbb{D}_K(F_K')^{-T}\|_2^{\frac{d}{2}}\right)^{\frac{2}{d}} & \text{for} \quad d\ge 3. \end{cases} \tag{40}$$

Proof. The conclusions follow from Theorems 3.1 and 3.2. □

The upper bounds in the above theorem also show the effects of the interplay between the mesh geometry and the diffusion matrix. To see this, we consider $\mathbb{D}^{-1}$-uniform meshes (a special case of $\mathbb{M}$-uniform meshes) that are defined essentially as uniform meshes in the metric specified by $\mathbb{D}^{-1}$. It is known (e.g., see [26]) that a $\mathbb{D}^{-1}$-uniform mesh $\mathcal{T}_h$ satisfies

$$(F_K')^{-1}\mathbb{D}_K(F_K')^{-T} = h_{\mathbb{D}^{-1}}^{-2} I \quad \forall K\in\mathcal{T}_h, \tag{41}$$

where $h_{\mathbb{D}^{-1}}$ is the average element size in metric $\mathbb{D}^{-1}$, i.e.,

$$h_{\mathbb{D}^{-1}} = \left(\frac{1}{N}\sum_{K\in\mathcal{T}_h}|K|\,\det(\mathbb{D}_K)^{-\frac{1}{2}}\right)^{\frac{1}{d}}.$$

From (2) it is not difficult to see

$$\frac{|\Omega|^{\frac{1}{d}}}{\sqrt{\overline{d}}} \le N^{\frac{1}{d}}h_{\mathbb{D}^{-1}} \le \frac{|\Omega|^{\frac{1}{d}}}{\sqrt{\underline{d}}}.$$

Then, for a $\mathbb{D}^{-1}$-uniform mesh, combining (41) with the above theorem we have

$$\kappa(A_{FV}) \le C\,(1+C_0H_h)\,N^{\frac{2}{d}}\cdot\left(N\max_j|\omega_j|\right) \times \begin{cases} 1 & \text{for} \quad d=1, \\ (1-\underline{d}^{-1}H_h)(1+\ln(\frac{|\overline{K}|}{|K_{\min}|})) & \text{for} \quad d=2, \\ (1-\underline{d}^{-1}H_h)\left(\frac{1}{N}\sum_{K\in\mathcal{T}_h}\left(\frac{|\overline{K}|}{|K|}\right)^{\frac{d-2}{2}}\right)^{\frac{2}{d}} & \text{for} \quad d\ge 3, \end{cases} \tag{42}$$

$$\kappa(S^{-1}A_{FV}S^{-1}) \le \frac{C\,(1+C_0H_h)^2\,N^{\frac{2}{d}}}{(1-\underline{d}^{-1}H_h)^2}, \tag{43}$$

where C is a constant which depends on $\mathbb{D}$ but not on the mesh. From (42) we can see that the mesh nonuniformity (in the Euclidean metric) can still have significant effects on the conditioning of the stiffness matrix even for $\mathbb{D}^{-1}$-uniform meshes. Since a mesh cannot in general be uniform in the Euclidean metric and the metric

$\mathbb{D}^{-1}$ simultaneously, mesh nonuniformity will have effects on the conditioning of the stiffness matrix. On the other hand, the situation is different for Jacobi preconditioning. The estimate (43) shows that the effect of mesh nonuniformity in the Euclidean metric is totally eliminated by the preconditioning. In fact, the bound is almost the same as that for the Laplace operator on a uniform mesh.

The above analysis shows the importance of using Jacobi preconditioning and having a mesh that is uniform in the metric specified by the inverse of the diffusion matrix. Moreover, it is consistent with those of [23, 25, 27, 29] for linear finite element discretization. For example, the main difference between the bounds in Theorem 3.3 here and those in Theorem 5.2 of [29] lies in the factors containing H_h that tend to 1 (or $\underline{d}$) as $H_h \to 0$ or the mesh is refined. The observation is also consistent with Lemma 3.2 which states that the difference between the linear forms of FVEM and FEM is $\mathcal{O}(H_h)$.

4. Conditioning of the mass matrix

In this section we discuss the mathematical properties for the mass matrix. Although this is a topic not directly related to the FVEM solution of boundary value problems, it is useful for the FVEM solution of time dependent and eigenvalue problems; e.g., see [23, 25] for finite element discretization. Moreover, it is theoretically interesting to know how the interplay between the mesh geometry and the diffusion matrix affects the conditioning of the mass matrix.

The entries of the mass matrix $M = (M_{ij})$ are given by

$$\begin{aligned} M_{ij} &= \int_{K^*_{P_i}} \phi_j \mathrm{d}\boldsymbol{x} = \sum_{K\in\omega_{ij}} |K| \int_{K\cap K^*_{P_i}} \phi_j \mathrm{d}\boldsymbol{x} \\ &= \begin{cases} m_1\,|\omega_i|, & i=j, \\ m_2\,|\omega_{ij}|, & i\neq j, \end{cases} \qquad i,j=1,\ldots,N_{vi}, \end{aligned} \tag{44}$$

where m_1 and m_2 are given by (see Appendix A for the derivation)

$$\begin{aligned} m_1 &= \frac{1}{(d+1)^3}\left(1+(d+1)\sum_{i=1}^{d}\frac{1}{i}\right), \\ m_2 &= \frac{1}{d(d+1)^3}\left(d^2+2d-(d+1)\sum_{i=1}^{d}\frac{1}{i}\right). \end{aligned} \tag{45}$$

For $d = 1$, 2, and 3, we have

$$m_1 = \begin{cases} 3/8, & d=1, \\ 11/54, & d=2, \\ 25/192, & d=3, \end{cases} \qquad m_2 = \begin{cases} 1/8, & d=1, \\ 7/108, & d=2, \\ 23/576, & d=3. \end{cases}$$

Obviously, M is symmetric. Moreover, denote the local mass matrix on K by M_K and that on $\hat{K}$ by $M_{\hat{K}}$. Then,

$$(M_{\hat{K}})_{ij} = \int_{\hat{K}^*_{\hat{P}_i}\cap\hat{K}} \hat{\phi}_j \mathrm{d}\boldsymbol{\xi} = \begin{cases} m_1, & i=j, \\ m_2, & i\neq j, \end{cases} \qquad i,j=1,\ldots,d+1.$$

It can also be shown that

$$(m_1 - m_2)I = \frac{1}{(d+1)^2}\left((d+1)\sum_{i=1}^{d}\frac{1}{i} - d\right) I \tag{46}$$
$$\leq M_{\hat{K}} \leq (m_1 + dm_2)I = \frac{1}{d+1}I.$$

Then, for any vector $\boldsymbol{u}$, letting $\boldsymbol{u}_K$ be the restriction of the vector $\boldsymbol{u}$ on K we have

$$\begin{aligned}\boldsymbol{u}^T M\boldsymbol{u} &= \sum_{K\in\mathcal{T}_h} \boldsymbol{u}_K^T M_K \boldsymbol{u}_K = \sum_{K\in\mathcal{T}_h} |K|\, \boldsymbol{u}_K^T M_{\hat{K}} \boldsymbol{u}_K \\ &\geq \sum_{K\in\mathcal{T}_h} (m_1 - m_2)\,|K|\,\|\boldsymbol{u}_K\|^2 = \sum_i \left((m_1 - m_2)\,u_i^2 \sum_{K\in\omega_i} |K|\right) \\ &\geq (m_1 - m_2)\,\|\boldsymbol{u}\|^2 \min_i |\omega_i|. \qquad (47)\end{aligned}$$

Thus, M is also positive definite.

4.1. **Condition number of the mass matrix.**

Theorem 4.1. *The condition number of the mass matrix for the linear FVEM on a simplicial mesh is bounded by*

$$\frac{|\omega_{\max}|}{|\omega_{\min}|} \leq \kappa(M) \leq \frac{1}{(d+1)(m_1 - m_2)}\frac{|\omega_{\max}|}{|\omega_{\min}|}, \tag{48}$$

where $|\omega_{\max}| = \max_j |\omega_j|$ *and* $|\omega_{\min}| = \min_j |\omega_j|$.

Proof. From (46), we have

$$\begin{aligned}\boldsymbol{u}^T M\boldsymbol{u} &= \sum_{K\in\mathcal{T}_h} \boldsymbol{u}_K^T M_K \boldsymbol{u}_K = \sum_{K\in\mathcal{T}_h} |K| \boldsymbol{u}_K^T M_{\hat{K}} \boldsymbol{u}_K \\ &\leq (m_1 + d\,m_2) \sum_{K\in\mathcal{T}_h} |K| \|\boldsymbol{u}_K\|_2^2 \\ &= \frac{1}{d+1} \sum_{K\in\mathcal{T}_h} |K| \|\boldsymbol{u}_K\|_2^2.\end{aligned}$$

Rearranging the sum on the right-hand side according to the vertices and using (44), we get

$$\boldsymbol{u}^T M\boldsymbol{u} \leq \frac{1}{d+1} \sum_{K\in\mathcal{T}_h} |K| \|\boldsymbol{u}_K\|_2^2 = \frac{1}{d+1} \sum_i u_i^2 |\omega_i| \tag{49}$$
$$= \frac{1}{m_1(d+1)} \sum_i u_i^2 M_{ii},$$

which implies

$$\lambda_{\max}(M) \leq \frac{1}{m_1(d+1)} \max_i M_{ii}.$$

Similarly, we have

$$\lambda_{\min}(M) \geq \frac{m_1 - m_2}{m_1} \min_i M_{ii}. \tag{50}$$

Moreover, it is not difficult to see that

$$\lambda_{\max}(M) \geq \max_i M_{ii}, \qquad \lambda_{\min}(M) \leq \min_i M_{ii}.$$

Combining the above estimates gives rise to

$$\max_i M_{ii} \le \lambda_{\max}(M) \le \frac{1}{m_1(d+1)} \max_i M_{ii},$$
$$\frac{m_1 - m_2}{m_1} \min_i M_{ii} \le \lambda_{\min}(M) \le \min_i M_{ii},$$

which lead to

$$\frac{\max_i M_{ii}}{\min_i M_{ii}} \le \kappa(M) \le \frac{1}{(m_1 - m_2)(d+1)} \frac{\max_i M_{ii}}{\min_i M_{ii}}.$$

From (44) we obtain (48). □

The theorem shows that $\kappa(M) = \mathcal{O}(1)$ when the mesh is uniform or close to being uniform. However, when the mesh is nonuniform, the condition number of M can be very large.

4.2. **Diagonal scaling for the mass matrix.** For any diagonal scaling $S = (s_i)$, like Theorem 4.1 we can obtain

$$\frac{\max_i s_i^{-2} M_{ii}}{\min_i s_i^{-2} M_{ii}} \le \kappa(S^{-1} M S^{-1}) \le \frac{1}{(m_1 - m_2)(d+1)} \frac{\max_i s_i^{-2} M_{ii}}{\min_i s_i^{-2} M_{ii}}. \tag{51}$$

For the Jacobi preconditioning $s_i^2 = M_{ii}$, we have the following theorem.

Theorem 4.2. *The condition number of the Jacobi preconditioned FVEM mass matrix with a simplicial mesh has a mesh-independent bound,*

$$\kappa(S^{-1} M S^{-1}) \le \frac{1}{(m_1 - m_2)(d+1)}.$$

The results in Theorems 4.1 and 4.2 are similar to those results for linear FEM; e.g., see [29, 39].

4.3. **Diagonal and lump of the mass matrix.**

Lemma 4.1. *The linear FVEM mass matrix M and its diagonal part M_D satisfy*

$$\frac{m_1 - m_2}{m_1} M_D \le M \le \frac{1}{m_1(d+1)} M_D, \tag{52}$$

where the less-than-or-equal sign is in the sense of seminegative definiteness.

Proof. This follows from (49) and (50) directly. □

Lemma 4.2. *Let M_{lump} be the lumped linear FVEM mass matrix defined through*

$$M_{ii,lump} = \int_{K^*_{P_i}} \sum_{j=1}^{N_{vi}} \phi_j(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \quad i = 1, \ldots, N_{vi}.$$

Then

$$m_1 |\omega_i| \le M_{ii,lump} \le \frac{|\omega_i|}{d+1}. \tag{53}$$

Proof. Since

$$\phi_i(\boldsymbol{x}) \le \sum_{j=1}^{N_{vi}} \phi_j(\boldsymbol{x}) \le 1,$$

with (44) we have

$$M_{ii,lump} \geq \int_{K^*_{P_i}} \phi_i(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = M_{ii} = m_1 |\omega_i|$$

and

$$M_{ii,lump} \leq \int_{K^*_{P_i}} 1 \mathrm{d}\boldsymbol{x} = \frac{|\omega_i|}{d+1}.$$

□

Lemma 4.3. *The linear FVEM mass matrix M and the lumped mass matrix M_{lump} satisfy*

$$(d+1)(m_1 - m_2) M_{lump} \leq M \leq \frac{1}{m_1(d+1)} M_{lump}.$$

Proof. Since $M_D \leq M_{lump}$, we get the upper bound directly from (52). The lower bound in (52) together with the upper bound in (53) give the lower bound

$$\begin{aligned} M &\geq \frac{m_1 - m_2}{m_1} M_D = \frac{m_1 - m_2}{m_1} m_1 \mathrm{diag}(|\omega_1|, \ldots, |\omega_{N_{vi}}|) \\ &\geq (d+1)(m_1 - m_2) M_{lump}. \end{aligned}$$

□

5. Numerical examples

In this section we present numerical results for a selection of d-dimensional ($d = 1, 2, 3$) examples to illustrate the theoretical results obtained in the previous sections. Note that all bounds on the smallest eigenvalue $\lambda_{\min}((A_{FV} + A_{FV}^T)/2)$ (cf. Theorem 3.2) contain a constant C. We obtain its value by calibrating the bound with uniform meshes through comparing the exact and estimated values. For the largest singular value $\sigma_{\max}(A_{FV})$ we use explicit bounds (22) and (23) where analytical expressions are available for the constants. Predefined meshes are used to demonstrate the influence of the number and shape of mesh elements on the condition number of the stiffness matrix and to verify the improvement achieved with the diagonal scaling. The first three examples are adopted from [29]. The results presented here for these examples are comparable with those obtained in [29] with a linear finite element discretization.

Example 5.1. This is a one-dimensional example with $\mathbb{D} = 1 + \exp(x^5)$ and a mesh given by Chebyshev nodes in the interval [0, 1],

$$x_i = \frac{1}{2}\left(1 - \cos\frac{\pi(2i-1)}{2(N-1)}\right), \quad i = 1, \ldots, N-1.$$

The exact condition number of the stiffness matrix and its estimates (39) and (40) are shown in Figure 2(a). The exact $\sigma_{\max}(A_{FV})$ and $\lambda_{\min}((A_{FV} + A_{FV}^T)/2)$ and their estimates (22) and (33) are shown in Figure 2(b). The results show that the estimates have the same asymptotic order as the corresponding exact values as N increases. Moreover, they show that the Jacobi preconditioning has significant impacts on the condition number. Not only is $\kappa(S^{-1}A_{FV}S^{-1})$ significantly lower than $\kappa(A_{FV})$ but also it has a lower order than the latter does as N increases. □

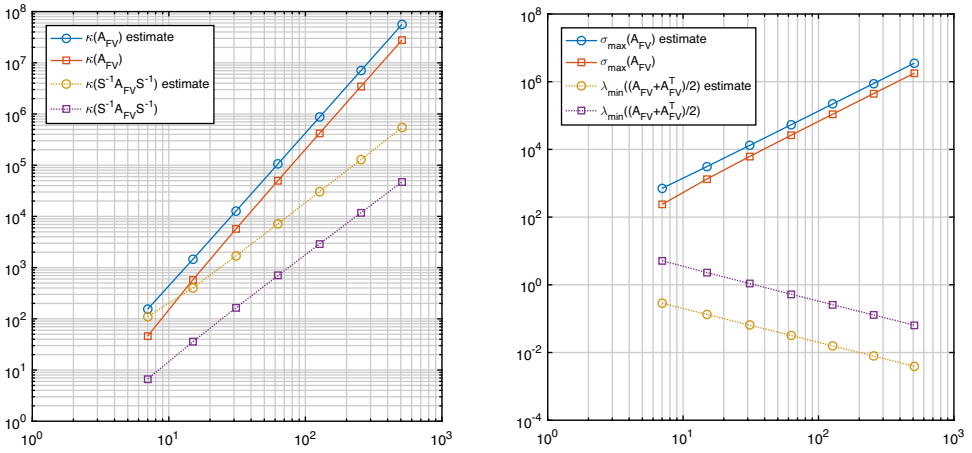


FIGURE 2. Example 5.1. Exact and estimated condition number (left) and exact and estimated greatest singular value (eigenvalue) (right) of the stiffness matrix as a function of N ($d = 1$).

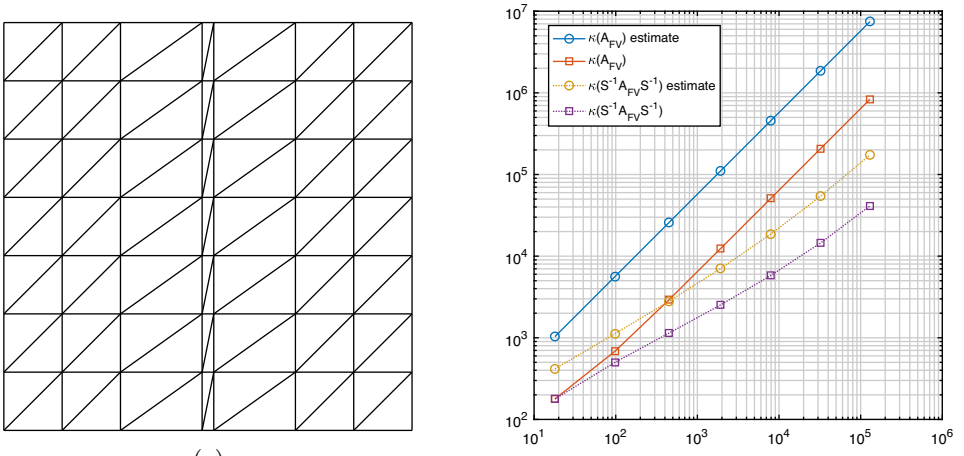
Example 5.2. In this two-dimensional example, $\mathbb{D} = I$, $\Omega = (0,1) \times (0,1)$, and a mesh (cf. Figure 3(a)) with $O(N^{1/2})$ skew elements and a maximum element aspect ratio of 125 : 1 are used. The condition number and its estimate are shown in Figure 3(b) as functions of N. One can see that both the exact values and the estimates have the same asymptotic order as N increases. One can also see that the condition number with scaling is significantly smaller than that without scaling and the asymptotic order of the former is also smaller than that of the latter. □

Example 5.3. In this three-dimensional example, $\mathbb{D} = I$, Ω is the unit cube, and a mesh shown in Figure 4(a) and having $O(N^{2/3})$ skew elements with a maximum aspect ratio of 125 : 1 is used. The results are shown in Figure 4(b) as N increases. We can see that scaling not only reduces the condition number significantly but also lowers the asymptotic order in N. Moreover, the bound (39) and $\kappa(A_{FV})$ have the same asymptotic order. However, the order of the bound (40) in N is slightly higher than that of $\kappa(S^{-1}A_{FV}S^{-1})$. Similar trends have been observed for a linear finite element discretization in [29]. □

Example 5.4. The setting of this example is essentially the same as that in Example 5.2 except that the size of the mesh is fixed at $N = 32258$ but its maximum aspect ratio of elements increases and that the diffusion matrix is chosen as

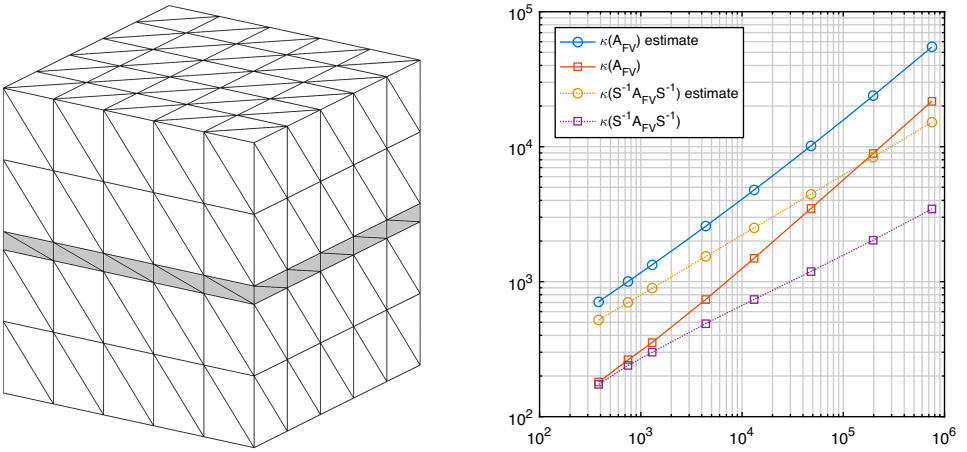
$$\mathbb{D} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix},$$

where $\theta = \pi \sin(x) \cos(y)$. An example of mesh and the condition number of the stiffness matrix and its estimate are shown in Figure 5. The results show that the condition number and its estimate are essentially linear functions of the maximum element aspect ratio. Moreover, the condition number is much smaller with scaling than without scaling.



(a) (b)

FIGURE 3. Example 5.2. (a) A mesh example with a maximum element aspect ratio of 125:1. (b) Exact and estimate condition numbers as functions of N.

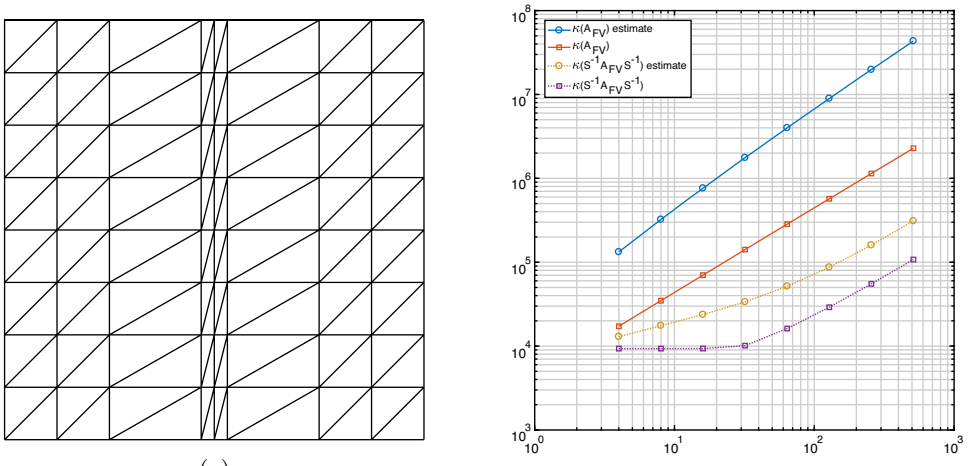


(a) (b)

FIGURE 4. Example 5.3. (a) A mesh example with a maximum element aspect ratio of 125:1. (b) Exact and estimate condition numbers as functions of N.

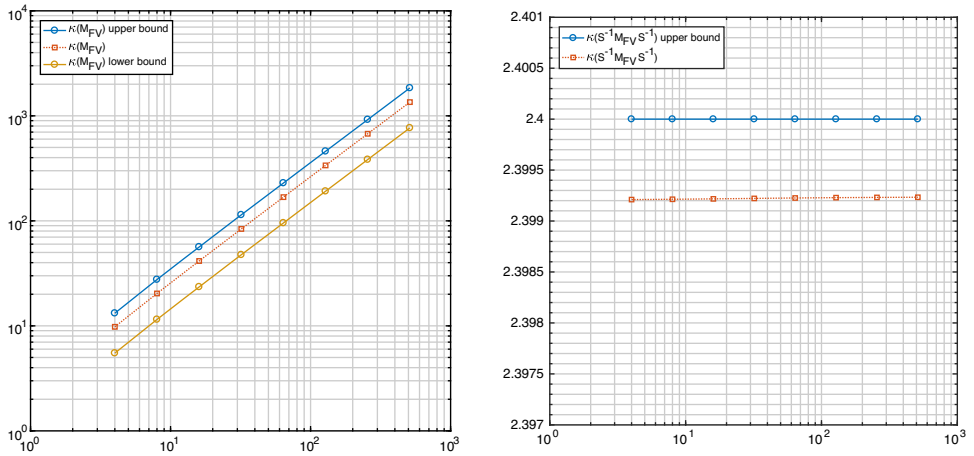
The condition number and its bounds for the mass matrix are shown in Figure 6. Without scaling, they increase linearly with the maximum element aspect ratio. On the contrary, they stay constant when the Jacobi scaling is used. This is consistent with Theorem 4.2.

A comparison of the condition numbers of A_{FV} and A_{FE} (with and without scaling) as N increases is shown in Figure 7. It can be seen that $\kappa(A_{FV})$ and $\kappa(A_{FE})$ are almost indistinguishable. The same goes for $\kappa(S^{-1}A_{FV}S^{-1})$ and $\kappa(S_{FE}^{-1}A_{FE}S_{FE}^{-1})$.



(a) (b)

FIGURE 5. Example 5.4. (a) The predefined meshes. (b) Exact and estimate condition numbers as functions of the maximum element aspect ratio when the size of the mesh is fixed at $N = 32258$.



(a) (b)

FIGURE 6. Example 5.4. The condition number of the mass matrix and its bounds are shown as functions of the maximum element aspect ratio when the size of the mesh is fixed at $N = 32258$. (a) Without Jacobi preconditioning. (b) With Jacobi preconditioning.

We apply the generalized minimum residual (GMRES) method to the solution of the algebraic system resulting from the FV discretization of (1), where f is chosen such that the exact solution of the BVP is given by $u = \sin(\pi x)\sin(\pi y)$. GMRES is employed without restart and with a zero vector as the initial guess. A mesh as in Figure 5(a) is used with the maximum element aspect ratio being set to be $1/125$. The number of GMRES iterations required to reduce the 2-norm of the

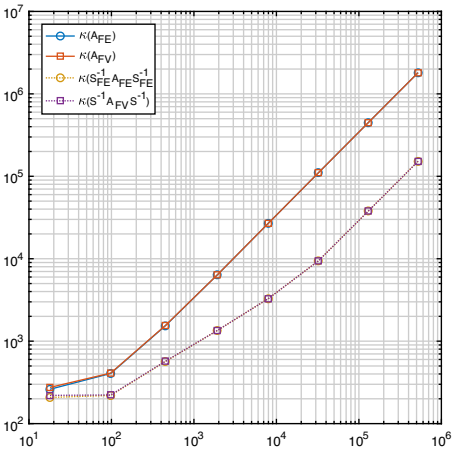


FIGURE 7. Example 5.4. The condition numbers of A_{FV} and A_{FE} (with and without scaling) are shown as functions of the mesh size (N). The maximum element aspect ratio for the meshes is kept to be 1/125. $\kappa(A_{FV})$ and $\kappa(A_{FE})$ are almost indistinguishable. The same goes for $\kappa(S^{-1}A_{FV}S^{-1})$ and $\kappa(S_{FE}^{-1}A_{FE}S_{FE}^{-1})$.

relative residual to or less than 10^{-6} is reported in Table 1. It can be seen that the Jacobi preconditioning helps reduce the number of iterations for this example.

TABLE 1. Example 5.4. The number of GMRES iterations required to reduce the 2-norm of the relative residual to or less than 10^{-6} for solving FVEM algebraic systems with and without scaling.

N	$\kappa(A_{FV})$	Iteration	$\kappa(S^{-1}A_{FV}S^{-1})$	Iteration
18	2.75e+2	7	2.20e+2	7
98	4.10e+2	33	2.23e+2	33
450	1.54e+3	85	5.70e+2	98
1922	6.38e+3	186	1.35e+3	183
7938	2.68e+4	395	3.27e+3	355
32258	1.10e+5	816	9.41e+3	623
130050	4.48e+5	1658	3.79e+4	1239

6. CONCLUSIONS

In the previous sections we have studied the conditioning of the stiffness matrix A_{FV} of the linear finite volume element discretization of the boundary value problem (1) with general simplicial meshes. Since A_{FV} is nonsymmetric in general, we define its condition number (17) as the ratio of $\sigma_{\max}(A_{FV})$, the maximum singular value, to $\lambda_{\min}((A_{FV} + A_{FV}^T)/2)$, the minimum eigenvalue of its symmetric part, in view of the convergence of GMRES (cf. (16)). The situations with and without Jacobi preconditioning have been considered. An upper bound on the maximum

singular value and a lower bound on the minimum eigenvalue of the symmetric part have been obtained in Theorems 3.1 and 3.2, respectively, and an upper bound on the condition number has been obtained in Theorem 3.3.

It is noted that those theoretical results have been obtained for a general diffusion matrix $\mathbb{D}$ and a sufficiently fine, arbitrary simplicial mesh in any dimension. They not only provide a bound on the condition number of the stiffness matrix but also shed light on the effects of the interplay between the diffusion matrix and the mesh geometry. Particularly, the bounds reveal that without scaling, the condition number is affected by the number of the elements N, the mesh nonuniformity in the Euclidean metric, and the mesh nonuniformity in the metric specified by $\mathbb{D}^{-1}$. For meshes that are uniform in $\mathbb{D}^{-1}$, the last factor will be eliminated but the mesh nonuniformity in the Euclidean metric still plays a role; see (42). On the other hand, the analysis shows that the effects by the mesh nonuniformity in the Euclidean metric can be eliminated by scaling. For the situation with scaling and a $\mathbb{D}^{-1}$-uniform mesh, the condition number depends only on the number of the elements (cf. (43)). Numerical examples confirm the above analysis.

A similar analysis has been carried out for the mass matrix in §4. The main results are stated in Theorems 4.1 and 4.2. They show that the condition number of the mass matrix for the linear FVEM discretization depends only on the mesh nonuniformity in the Euclidean metric and scaling can effectively eliminate its effects.

It is remarked that the results and observations made in this work are comparable and consistent with those in [29] for a linear finite element discretization of (1). The only noticeable difference is that the assumption of the mesh being sufficiently fine is needed in the current analysis. This is not surprising since FVEM generally does not preserve the symmetry of the underlying differential operator. Moreover, when the mesh is sufficiently fine, roughly speaking, both the FVEM and FEM discretizations are close to the differential operator and thus should exhibit similar behaviors. In this spirit, it is expected that the analysis in this work can be extended to higher-order FVEMs without major modifications; see [24] for studies for higher-order FEMs.

Appendix A: The expressions for m_1 and m_2

To obtain the values of m_1 and m_2 for general d dimensions, we consider K to be a right simplex as shown in Figure 8 for two and three dimensions. The dual element $K^*_{P_i}$ restricted to the primary element K is a polyhedron with $2d$ faces (see the polyhedron $P_1M_1M_0M_2$ in Figure 8(a) and the polyhedron $P_1M_1M_2M_3M_4M_5M_6M_0$ in Figure 8(b)). We now consider $d = 2$, $d = 3$, and a general d case separately.

The $d = 2$ case. Consider the triangle $K = \triangle P_1P_2P_3$ in Figure 8(a), where $P_1 = (0, 0)$, $P_2 = (1, 0)$, and $P_3 = (1, 1)$. Denote the midpoints of P_1P_2 and P_1P_3 by M_1 and M_2, respectively, and the barycenter of K by M_0. Then $\phi_{P_1}|_K = 1 - x$. It is not difficult to see that

$$\int_{K^*_{P_i}\cap K} \phi_{P_1} \mathrm{d}\boldsymbol{x} = 2\int_{K_0} \phi_{P_1} \mathrm{d}\boldsymbol{x} = m_1|K| = \frac{1}{2}m_1. \tag{54}$$

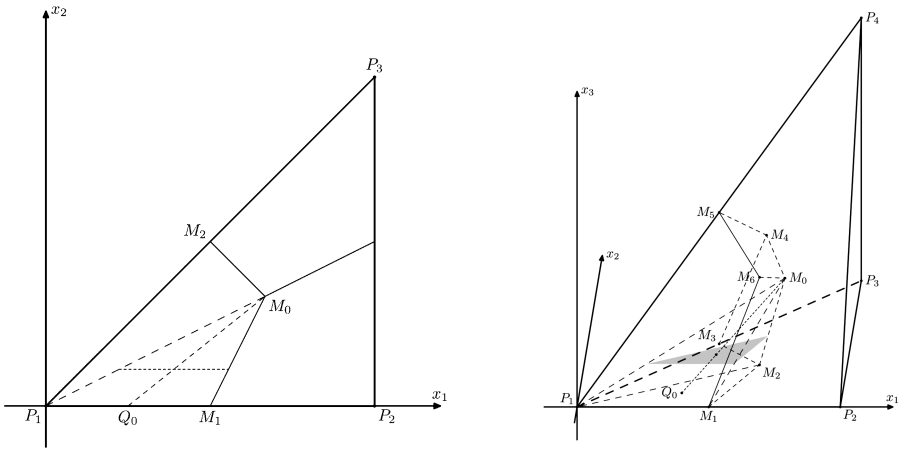


FIGURE 8. (a) A right triangle in two dimensions. (b) A right tetrahedron in three dimensions.

Here, $K_0 = \triangle P_1 M_1 M_0$. Since $\phi_{P_1}|_K$ is a linear function, the integral of ϕ_{P_1} on P_1M_1 equals $\phi_{P_1}(Q_0)$ multiplied by the length of P_1M_1. Thus,

$$\begin{aligned}\int_{K_0} \phi_{P_1} \mathrm{d}\boldsymbol{x} &= \int_{y(Q_0)}^{y(M_0)} \left(\frac{y - y(Q_0)}{y(M_0) - y(Q_0)} (\phi_{P_1}(M_0) - \phi_{P_1}(Q_0)) + \phi_{P_1}(Q_0) \right) \\ &\qquad \cdot \frac{y(M_0) - y}{y(M_0) - y(Q_0)} |P_1 M_1| \, \mathrm{d}y \\ &= \int_0^{\frac{1}{3}} \left(\frac{y}{(\frac{1}{3} - 0)} ((1 - \frac{2}{3}) - (1 - \frac{1}{4})) + (1 - \frac{1}{4}) \right) \frac{(\frac{1}{3} - y)}{(\frac{1}{3} - 0)} \frac{1}{2} \, \mathrm{d}y \\ &= \frac{11}{216}.\end{aligned}$$

Combining this with (54), we have

$$m_1 = 4 \int_{K_0} \phi_{P_1} \mathrm{d}\boldsymbol{x} = \frac{11}{54}.$$

On the other hand,

$$|K^*_{P_i} \cap K| = \frac{1}{3}|K| = \int_{K^*_{P_i} \cap K} \sum_{j=1,2,3} \phi_{P_j} \mathrm{d}\boldsymbol{x} = (m_1 + 2m_2)|K|.$$

Then,

$$m_2 = \frac{7}{108}.$$

The $d = 3$ case. Consider the tetrahedron $K = P_1P_2P_3P_4$ in Figure 8(b), where $P_1 = (0, 0, 0)$, $P_2 = (1, 0, 0)$, $P_3 = (1, 1, 0)$, and $P_4 = (1, 1, 1)$. Denote the midpoints of P_1P_2, P_1P_3, and P_1P_4 by M_1, M_3, and M_5, respectively, and the barycenters of the corresponding faces of K by M_2, M_4, and M_6. Let M_0 be the centroid of K and let Q_0 be the barycenter of $\triangle P_1M_1M_2$. Then $\phi_{P_1}|_K = 1 - x$. It is not difficult to see that

$$\frac{1}{6}m_1 = m_1|K| = \int_{K^*_{P_i}\cap K} \phi_{P_1}\,\mathrm{d}\boldsymbol{x} = 3 \times (2\int_{K_0} \phi_{P_1}\,\mathrm{d}\boldsymbol{x}). \tag{55}$$

Here, K_0 is the tetrahedron formed by the vertices P_1, M_1, M_2, and M_0. Since $\phi_{P_1}|_K$ is a linear function, the integral of ϕ_{P_1} on $\triangle P_1M_1M_2$ equals $\phi_{P_1}(Q_0)$ multiplied by the area of $\triangle P_1M_1M_2$. Thus,

$$\begin{aligned}
&\int_{K_0} \phi_{P_1}\,\mathrm{d}\boldsymbol{x}\\
=\ &\int_{z(Q_0)}^{z(M_0)} \Big(\frac{z - z(Q_0)}{z(M_0) - z(Q_0)}(\phi_{P_1}(M_0) - \phi_{P_1}(Q_0)) + \phi_{P_1}(Q_0)\Big)\\
&\qquad \Big(\frac{z(M_0) - z}{z(M_0) - z(Q_0)}\Big)^2 |\triangle P_1M_1M_2|\,\mathrm{d}z\\
=\ &\int_0^{\frac{1}{4}} \Big(\frac{z}{\frac{1}{4} - 0}\big((1 - \frac{3}{4}) - (1 - \frac{\frac{1}{2} + \frac{2}{3}}{3})\big) + (1 - \frac{\frac{1}{2} + \frac{2}{3}}{3})\Big) \Big(\frac{\frac{1}{4} - z}{\frac{1}{4} - 0}\Big)^2 \frac{1}{6}\,\mathrm{d}z\\
=\ &\frac{25}{6912}.
\end{aligned}$$

From (55), we obtain

$$m_1 = 36\int_{K_0} \phi_{P_1}\,\mathrm{d}\boldsymbol{x} = \frac{25}{192}.$$

On the other hand,

$$|K^*_{P_i} \cap K| = \frac{1}{4}|K| = \int_{K^*_{P_i}\cap K} \sum_{j=1}^{4} \phi_{P_j}\,\mathrm{d}\boldsymbol{x} = (m_1 + 3m_2)|K|.$$

Then,

$$m_2 = \frac{23}{576}.$$

The general d case. A similar procedure can be used in general d dimensions. We have

$$\begin{aligned}
&\int_{K^*_{P_i}\cap K}\phi_{P_1}\mathrm{d}\boldsymbol{x} = d\int_{K_0}\phi_{P_1}\mathrm{d}\boldsymbol{x}\\
&= d\int_{x_d(Q_0)}^{x_d(M_0)}\left(\frac{x_d-x_d(Q_0)}{x_d(M_0)-x_d(Q_0)}(\phi_{P_1}(M_0)-\phi_{P_1}(Q_0))+\phi_{P_1}(Q_0)\right)\\
&\qquad\cdot\Big(\frac{x_d(M_0)-x_d}{x_d(M_0)-x_d(Q_0)}\Big)^{d-1}S_{K_{d-1}}\,\mathrm{d}x_d\\
&= d\int_0^{\frac{1}{d+1}}\left(\frac{x_d}{\frac{1}{d+1}-0}\big(\,(1-\frac{d}{d+1})-(1-\frac{\sum_{j=1}^d(1-\frac{1}{j})}{d}\big)\big)+(1-\frac{\sum_{j=1}^d(1-\frac{1}{j})}{d}\big)\right)\\
&\qquad\cdot\Big(\frac{\frac{1}{d+1}-x_d}{\frac{1}{d+1}-0}\Big)^{d-1}|K|\,\mathrm{d}x_d\\
&= d\int_0^{\frac{1}{d+1}}\left(x_d(d+1)\Big(\frac{1}{d+1}+\frac{\sum_{j=1}^d\frac{1}{j}}{d}\Big)+\frac{\sum_{j=1}^d\frac{1}{j}}{d}\right)\Big(\frac{\frac{1}{d+1}-x_d}{\frac{1}{d+1}-0}\Big)^{d-1}|K|\,\mathrm{d}x_d\\
&= \frac{1}{(d+1)^3}\left(1+(d+1)\sum_{i=1}^d\frac{1}{i}\right)|K|.
\end{aligned}$$

Here, $S_{K_{d-1}}$ is the $(d-1)$-dimensional measure of the face of $K^*_{P_i}\cap K$ restricted on $x_d=0$, which is equal to $|K|$ in the current situation, and K_0 denotes the polyhedron bounded by the face of $K^*_{P_i}\cap K$ restricted on $x_d=0$ and M_0, whose $(d-1)$-dimensional measure is $|K|/d$.[1] Thus,

$$m_1 = \frac{1}{(d+1)^3}\left(1+(d+1)\sum_{i=1}^d\frac{1}{i}\right). \tag{56}$$

Moreover, we have

$$|K^*_{P_i}\cap K| = \frac{1}{d+1}|K| = \int_{K^*_{P_i}\cap K}\sum_{j=1}^{d+1}\phi_{P_j}\mathrm{d}\boldsymbol{x} = (m_1+dm_2)|K|.$$

Combining this with (56), we obtain (45).

Acknowledgment

The first-named author is thankful to the Department of Mathematics of the University of Kansas for the hospitality during his visit.

References

[1] M. Ainsworth, W. McLean, and T. Tran, *The conditioning of boundary element equations on locally refined meshes and preconditioning by diagonal scaling*, SIAM J. Numer. Anal. **36** (1999), no. 6, 1901–1932, DOI 10.1137/S0036142997330809. MR1712149

[2] M. Ainsworth, B. McLean, and T. Tran, *Diagonal scaling of stiffness matrices in the Galerkin boundary element method*, ANZIAM J. **42** (2000), no. 1, 141–150, DOI 10.1017/S1446181100011676. Papers in honour of David Elliott on the occasion of his sixty-fifth birthday. MR1783377

[1] For simplicity, in the 3-dimensional case, K_0 denotes the tetrahedron $P_1P_2P_3P_4$, i.e., half of the polyhedron, which is bounded by the face of $K^*_{P_i}\cap K$ restricted on $x_d=0$ and M_0.

[3] R. E. Bank and D. J. Rose, *Some error estimates for the box method*, SIAM J. Numer. Anal. **24** (1987), no. 4, 777–787, DOI 10.1137/0724050. MR899703

[4] R. E. Bank and L. R. Scott, *On the conditioning of finite element equations with highly refined meshes*, SIAM J. Numer. Anal. **26** (1989), no. 6, 1383–1394, DOI 10.1137/0726080. MR1025094

[5] T. Barth and M. Ohlberger, *Finite Volume Methods: Foundation and Analysis*, Vol. 1, Chap. 15, John Wiley & Sons, 2004, 1–57.

[6] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Texts in Applied Mathematics, vol. 15, Springer-Verlag, New York, 1994. MR1278258

[7] C. Bi and V. Ginting, *Two-grid finite volume element method for linear and nonlinear elliptic problems*, Numer. Math. **108** (2007), no. 2, 177–198, DOI 10.1007/s00211-007-0115-9. MR2358002

[8] Z. Q. Cai, J. Mandel, and S. 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

[9] Z. Q. Cai, *On the finite volume element method*, Numer. Math. **58** (1991), no. 7, 713–735, DOI 10.1007/BF01385651. MR1090257

[10] W. Cao, Z. Zhang, and Q. Zou, *Is $2k$-conjecture valid for finite volume methods?*, SIAM J. Numer. Anal. **53** (2015), no. 2, 942–962, DOI 10.1137/130936178. MR3328149

[11] L. 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

[12] Z. Chen, Y. Xu, and Y. Zhang, *A construction of higher-order finite volume methods*, Math. Comp. **84** (2015), no. 292, 599–628, DOI 10.1090/S0025-5718-2014-02881-0. MR3290957

[13] Z. Chen, J. Wu, and Y. 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

[14] S.-H. Chou and X. Ye, *Unified analysis of finite volume methods for second order elliptic problems*, SIAM J. Numer. Anal. **45** (2007), no. 4, 1639–1653, DOI 10.1137/050643994. MR2338403

[15] Q. Du, D. Wang, and L. Zhu, *On mesh geometry and stiffness matrix conditioning for general finite element spaces*, SIAM J. Numer. Anal. **47** (2009), no. 2, 1421–1444, DOI 10.1137/080718486. MR2497335

[16] S. C. Eisenstat, H. C. Elman, and M. H. Schultz, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal. **20** (1983), no. 2, 345–357, DOI 10.1137/0720023. MR694523

[17] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Applied Mathematical Sciences, vol. 159, Springer-Verlag, New York, 2004. MR2050138

[18] R. E. Ewing, T. Lin, and Y. 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

[19] I. Fried, *Bounds on the spectral and maximum norms of the finite element stiffness, flexibility and mass matrices* (English, with Russian summary), Internat. J. Solids and Structures **9** (1973), 1013–1034. MR0345400

[20] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition. MR1814364

[21] I. G. Graham and W. McLean, *Anisotropic mesh refinement: the conditioning of Galerkin boundary element matrices and simple preconditioners*, SIAM J. Numer. Anal. **44** (2006), no. 4, 1487–1513, DOI 10.1137/040621247. MR2257114

[22] W. Hackbusch, *On first and second order box schemes* (English, with German summary), Computing **41** (1989), no. 4, 277–296, DOI 10.1007/BF02241218. MR993825

[23] W. Huang, *Sign-preserving of principal eigenfunctions in P1 finite element approximation of eigenvalue problems of second-order elliptic operators*, J. Comput. Phys. **274** (2014), 230–244, DOI 10.1016/j.jcp.2014.06.012. MR3231765

[24] W. Huang, L. Kamenski, and J. Lang, *Stability of explicit Runge-Kutta methods for high order finite element approximation of linear parabolic equations*, in Numerical Mathematics and Advanced Applications (Proceedings of the 2013 European Numerical Mathematics and Advanced Applications Conference ENUMATH-2013, Lausanne, Switzerland, August 26–30,

2013), Lecture Notes in Computational Science and Engineering, vol. 103, Springer, Cham, 2015, pp. 165–173.

[25] W. Huang, L. Kamenski, and J. Lang, *Stability of explicit one-step methods for P1-finite element approximation of linear diffusion equations on anisotropic meshes*, SIAM J. Numer. Anal. **54** (2016), no. 3, 1612–1634, DOI 10.1137/130949531. MR3505307

[26] W. Huang and R. D. Russell, *Adaptive Moving Mesh Methods*, Applied Mathematical Sciences, vol. 174, Springer, New York, 2011. MR2722625

[27] L. Kamenski and W. Huang, *A study on the conditioning of finite element equations with arbitrary anisotropic meshes via a density function approach*, J. Math. Study **47** (2014), no. 2, 151–172. MR3260337

[28] L. Kamenski, W. Huang, and J. Lang, *Conditioning of implicit Runge-Kutta integration of finite element approximation of linear diffusion equations on anisotropic meshes*, (submitted), 2017.

[29] L. Kamenski, W. Huang, and H. Xu, *Conditioning of finite element equations with arbitrary anisotropic meshes*, Math. Comp. **83** (2014), no. 289, 2187–2211, DOI 10.1090/S0025-5718-2014-02822-6. MR3223329

[30] R. Li, Z. Chen, and W. Wu, *Generalized Difference Methods for Differential Equations: Numerical Analysis of Finite Volume Methods*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 226, Marcel Dekker, Inc., New York, 2000. MR1731376

[31] Y. Li, S. Shu, Y. Xu, and Q. 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

[32] F. 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

[33] Y. Lin, M. Yang, and Q. Zou, *L^2 error estimates for a class of any order finite volume schemes over quadrilateral meshes*, SIAM J. Numer. Anal. **53** (2015), no. 4, 2009–2029, DOI 10.1137/140963121. MR3384836

[34] J. Lv and Y. Li, *Optimal biquadratic finite volume element methods on quadrilateral meshes*, SIAM J. Numer. Anal. **50** (2012), no. 5, 2379–2399, DOI 10.1137/100805881. MR3022223

[35] T. 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

[36] J. R. Shewchuk, *What is a good linear element?—Interpolation, conditioning, and quality measures*, in Proceedings of the 11th International Meshing Roundtable, Sandia National Laboratories, Albuquerque, NM, 2002, 115–126.

[37] X. Wang and Y. Li, *L^2 error estimates for high order finite volume methods on triangular meshes*, SIAM J. Numer. Anal. **54** (2016), no. 5, 2729–2749, DOI 10.1137/140988486. MR3544655

[38] X. Wang and Y. Li, *Superconvergence of quadratic finite volume method on triangular meshes*, J. Comput. Appl. Math. **348** (2019), 181–199, DOI 10.1016/j.cam.2018.08.025. MR3886669

[39] A. J. Wathen, *Realistic eigenvalue bounds for the Galerkin mass matrix*, IMA J. Numer. Anal. **7** (1987), no. 4, 449–457, DOI 10.1093/imanum/7.4.449. MR968517

[40] J. Xu and Q. 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

[41] M. Yang, C. Bi, and J. Liu, *Postprocessing of a finite volume element method for semilinear parabolic problems*, M2AN Math. Model. Numer. Anal. **43** (2009), no. 5, 957–971, DOI 10.1051/m2an/2009017. MR2559740

[42] Z. Zhang and Q. Zou, *Vertex-centered finite volume schemes of any order over quadrilateral meshes for elliptic boundary value problems*, Numer. Math. **130** (2015), no. 2, 363–393, DOI 10.1007/s00211-014-0664-7. MR3343929

[43] L. Zhu and Q. Du, *Mesh-dependent stability for finite element approximations of parabolic equations with mass lumping*, J. Comput. Appl. Math. **236** (2011), no. 5, 801–811, DOI 10.1016/j.cam.2011.05.030. MR2853505

[44] L. Zhu and Q. Du, *Mesh dependent stability and condition number estimates for finite element approximations of parabolic problems*, Math. Comp. **83** (2014), no. 285, 37–64, DOI 10.1090/S0025-5718-2013-02703-2. MR3120581

School of Mathematics, Jilin University, Changchun 130012, People's Republic of China
Email address: wxjldx@jlu.edu.cn

Department of Mathematics, The University of Kansas, Lawrence, Kansas 66045
Email address: whuang@ku.edu

School of Mathematics, Jilin University, Changchun 130012, People's Republic of China
Email address: yonghai@jlu.edu.cn