MATHEMATICS OF COMPUTATION
Volume 83, Number 285, January 2014, Pages 37–64
S 0025-5718(2013)02703-2
Article electronically published on May 2, 2013

MESH DEPENDENT STABILITY AND CONDITION NUMBER ESTIMATES FOR FINITE ELEMENT APPROXIMATIONS OF PARABOLIC PROBLEMS

LIYONG ZHU AND QIANG DU

Abstract. In this paper, we discuss the effects of spatial simplicial meshes on the stability and the conditioning of fully discrete approximations of a parabolic equation using a general finite element discretization in space with explicit or implicit marching in time. Based on the new mesh dependent bounds on extreme eigenvalues of general finite element systems defined for simplicial meshes, we derive a new time step size condition for the explicit time integration schemes presented, which provides more precise dependence not only on mesh size but also on mesh shape. For the implicit time integration schemes, some explicit mesh-dependent estimates of the spectral condition number of the resulting linear systems are also established. Our results provide guidance to the studies of numerical stability for parabolic problems when using spatially unstructured adaptive and/or possibly anisotropic meshes.

1. Introduction

In this paper, we are concerned with fully discrete approximations of parabolic equations using a general finite element discretization in space and an explicit or implicit Euler marching scheme in time. Such approximations have been extensively studied over the years and are widely used in practice [24, 27]. It is well known that for an explicit scheme, some time step size constraints closely related to spatial meshes are required for the stable integration in time, while for an implicit scheme, condition numbers of resulting linear algebraic systems are also dependent on mesh geometry.

In more recent years, the use of adaptive time steps coupled with spatially adaptive unstructured meshes has become increasingly popular. Yet, in such a context, there has not been rigorous analysis on effects of unstructured spatial simplicial meshes on the stability and the conditioning of fully discrete approximations of parabolic equations with a general spatial finite element discretization. Our current understanding has largely been based on the limited analysis for the linear element on quasi-uniform and shape-regular meshes which has been known for

Received by the editor January 5, 2011 and, in revised form, November 16, 2011 and April 15, 2012.

2010 *Mathematics Subject Classification.* Primary 65N30, 65F10.

Key words and phrases. Stable time step size, condition number, mesh quality, finite element method, unstructured mesh, parabolic problem.

The first author is supported in part by the National Natural Science Foundation of China (Nos.11001007, 91130019) and Research Fund for the Doctoral Program of Higher Education of China (No. 20101102120031) and ISTCP of China (No. 2010DFR00700).

The second author is supported in part by NSF DMS-1016073. Part of this work was completed during this author's visit to the Beijing Computational Science Research Center, China.

decades [2, 6, 23, 24]. Most of the known results are imprecise concerning mesh dependence and they are not directly applicable to highly unstructured adaptive and perhaps anisotropic simplicial meshes. As a continuation of the series of works presented in [9, 10, 26], the goal of the study undertaken here is to explore, for general linear second order in space parabolic problems and general spatial finite element spaces, more precise relations among the mesh geometry, the conditioning of resulting finite element linear systems of equations for implicit schemes and the stable time step size for explicit time integration schemes. In light of the growing use of adaptive spatial meshes and time steps in finite element methods, it is not only theoretically interesting but practically useful to have a better understanding of such relations.

As discussed in [9, 10, 26], an integrated finite element methodology often involves the generation and optimization of a geometric mesh, the assembly of a discrete algebraic system using a finite element basis, the solution of such a system by some algebraic solvers and the subsequent analysis of the numerical results. There is thus considerable interest in understanding the interplay between the various components in order to improve the overall performance of finite element simulations. Historically, connections have been made between the performance of algebraic solvers (or condition numbers of global stiffness matrices) and general unstructured meshes [3, 4, 5, 7, 8, 9, 11, 12, 16, 17, 18, 21, 22]. Recently in [10], a more precise relation is established between mesh geometry and spectral condition numbers of stiffness matrices for some typical second order elliptic equations discretized by general finite element methods based on unstructured simplicial meshes in any space dimension, which shed new light on the development of mesh generation strategies and algebraic solvers for finite element methods. In terms of parabolic problems, although there exist some explorations of the relationship between mesh geometry and stable time step sizes for explicit time integration schemes in the finite element literature [19, 20, 22], precise and systematic descriptions of such relations for general finite element spaces remain to be developed. Similarly, for implicit time integration schemes, very few precise relations were known between mesh geometry and spectral condition numbers of stiffness matrices. Thus, it is natural to see if the framework developed in [10] for elliptic problems can be extended to the case of parabolic equations.

It turns out that a key and new ingredient to be studied in the context of parabolic equations is the so-called preconditioned stiffness matrix with the mass matrix being the preconditioner. This presents challenges that are significantly different from those studies given for elliptic problems where only stiffness matrices need to be analyzed. In this work, we not only adopt the techniques given in [13, 14] and further developed in [10] to estimate the eigenvalues of the mass and stiffness matrices, but also provide some new estimates on preconditioned stiffness matrices. Such estimates, coupled with the elegant trace formula and the framework on the mesh dependence given in [10], allow us to derive more precise bounds, with respect to the mesh geometry, on stable time step sizes for an explicit time integration scheme with a general finite element spatial discretization of parabolic problems. These bounds are explicitly expressed by some universal mesh geometric quantities which are dependent not only on the mesh size but also on the mesh shape. Such estimates, to our knowledge, have not been presented before in the literature. Another contribution of this paper is to establish refined relationships between the spectral

condition number of the resulting matrix and the mesh geometry for an implicit time integration scheme with a general finite element discretization of parabolic problems. These analytical results offer theoretical guidance to the further studies of both linear algebraic solvers and unstructured geometric meshing. The theoretical analysis is also complemented by numerical experiments as further validation.

This paper is organized as follows. In section 2, we recall basic finite element spatial discretizations of parabolic equations and describe the main issues we are concerned with. In section 3, some existing results on extreme eigenvalue estimates for general finite element spaces are briefly stated and some new estimates on eigenvalues of the preconditioned stiffness matrix are given. In section 4, we derive mesh dependent time step size conditions for the stability of an explicit time integration scheme, while in section 5, the mesh-dependent spectral condition number estimates are established for an implicit time integration scheme. In section 6, numerical examples are provided to substantiate the theoretical analysis. A final conclusion is given in section 7.

2. Finite element spatial discretizations of parabolic problems

We first introduce the model equation and its finite element spatial discretization, along with explicit and implicit time integration schemes.

Given an open bounded convex domain $\Omega \in \mathbb{R}^d$ with a Lipschitz-continuous boundary and a finite time interval $(0, T]$, consider the following time-dependent parabolic equation with the solution $u = u(\mathbf{x}, t)$:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \displaystyle\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}(a_{ij}\frac{\partial u}{\partial x_j}) + f, & \text{in} \quad \Omega \times (0, T], \\ u(\mathbf{x}, t) = 0, & \text{on} \quad \partial\Omega \times (0, T], \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \text{in} \quad \Omega, \end{cases} \tag{2.1}$$

where entries of the coefficient matrix $\tilde{A} = (a_{ij})_{i,j=1}^{d}$ and the right-hand side function f are assumed to be constants in time, and furthermore, $\tilde{A}$ is assumed to be symmetric positive definite, and uniformly bounded in $\overline{\Omega}$. The independence of $\tilde{A}$ and f on the time variable simplifies the notation, though much of our discussion remains valid for time-dependent coefficients and time-dependent right-hand side functions as well. We also choose to work with the homogeneous Dirichlet boundary condition for simplicity. We point out that here the given diffusion matrix $\tilde{A}$ and the right-hand side function f may admit discontinuities in both space and time. Thus, our discussion is applicable to many interesting applications such as composite materials.

The finite element method is employed in the spatial discretization. Let $\mathcal{T}$ denote a finite element mesh (a simplicial mesh for much of our discussion). An appropriate finite element space, $V_h \subset H_0^1(\Omega)$, with suitably chosen nodal basis functions $\{\phi_j\}_{j=1}^{N}$ may then be employed to discretize the above continuous problem, resulting in a finite element approximation in the form of a system of time-dependent ordinary differential equations such as

$$\int_\Omega \frac{\partial u_h(\mathbf{x}, t)}{\partial t} v_h \, \mathrm{d}\mathbf{x} + a_\Omega(u_h, v_h) = \int_\Omega f v_h \, \mathrm{d}\mathbf{x}, \quad \forall\, v_h \in V_h \,, \tag{2.2}$$

where the bilinear form $a_\Omega = a_\Omega(u, v)$ is given by

$$a_\Omega(u,v) := \int_\Omega \sum_{i,j=1}^{d} (a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j}) \, \mathrm{d}\mathbf{x}, \tag{2.3}$$

for any $u, v \in H_0^1(\Omega)$. To solve the above ODE system, we further consider the two most popular first order time integration schemes: the forward and the backward Euler time discretizations. First, the explicit forward Euler time integration scheme can be written as

$$\int_\Omega \frac{u_h^{i+1} - u_h^i}{\Delta t_i} v_h \, \mathrm{d}\mathbf{x} + a_\Omega(u_h^i, v_h) = \int_\Omega f v_h \, \mathrm{d}\mathbf{x}, \ \forall \ v_h \in V_h, \tag{2.4}$$

where Δt_i is a time step size, and u_h^i is the approximation solution of $u(\mathbf{x}, t)$ at time t_i. For any $u_h \in V_h$, we denote by $\mathbf{u}_i$ the vector containing the coordinates of u_h^i with respect to the basis $\{\phi_j\}$, so that $u_h^i = \sum(\mathbf{u}_i)_j \phi_j$. We also use a similar notation for the right-hand side term f.

In matrix notation, the discrete solution $\mathbf{u}_{i+1}$ satisfies

$$M\mathbf{u}_{i+1} = (M - \Delta t_i K)\mathbf{u}_i + \Delta t_i \mathbf{f}, \quad i = 0, 1, 2, 3, \ldots, \tag{2.5}$$

where K and M are respectively $N \times N$ stiffness and mass matrices, generated by the finite element basis $\{\phi_i\}$, that is,

$$K = (k_{ij}), \ k_{ij} = a_\Omega(\phi_i, \phi_j) \quad \text{and} \quad M = (m_{ij}), \ m_{ij} = \int_\Omega \phi_i \phi_j \, \mathrm{d}\mathbf{x}. \tag{2.6}$$

Obviously, K and M are both symmetric and positive definite.

Similarly, the implicit backward Euler scheme of (2.2) is given by

$$\int_\Omega \frac{u_h^{i+1} - u_h^i}{\Delta t_i} v_h \, \mathrm{d}\mathbf{x} \quad + \quad a_\Omega(u_h^{i+1}, v_h) = \int_\Omega f v_h \, \mathrm{d}\mathbf{x}, \ \forall \ v_h \in V_h, \tag{2.7}$$

or, in the matrix form,

$$(M + \Delta t_i K)\mathbf{u}_{i+1} = M\mathbf{u}_i + \Delta t_i \mathbf{f}, \quad i \geq 0 \, . \tag{2.8}$$

We note that for problems with time-dependent coefficients, the only complication is that K depends on the time variable. Since our discussions focus mostly on properties of (2.4) and (2.7) at a single step, they can be easily adapted to treat time-dependent coefficients as well. Some conditions on the diffusion coefficients are needed. We assume that the diffusion coefficient matrix satisfies

$$0 < \beta_1^\tau I \leq \tilde{A}^\tau(\mathbf{x}) \leq \beta_2^\tau I \tag{2.9}$$

uniformly for $\mathbf{x} \in \tau$ for some positive constants β_1^τ and β_2^τ, where τ is a generic simplicial element in $\mathcal{T}$ and $\tilde{A}^\tau(\mathbf{x})$ is the restriction of the coefficient matrix $\tilde{A}$ on τ, I is the standard identity matrix

The explicit integration scheme (2.4) is a widely used scheme because the resulting linear system can often be readily solved, especially when M is approximated by a diagonal matrix, a technique commonly called *mass lumping* [23]. However, the time step size Δt_i must be restricted in order to satisfy the stability condition associated with the explicit scheme. Obviously, to understand what the important factors are that affect the choice of permissible time step size can be important for the efficient implementation of the explicit scheme.

Let $\mathbf{e}_i = \mathbf{u}_i^* - \mathbf{u}_i$ with both $\mathbf{u}_i^*$ and $\mathbf{u}_i$ computed by (2.5) with different initial conditions. It is easy to see that

$$\mathbf{e}_{i+1} = (I - \Delta t_i M^{-1} K)\mathbf{e}_i \ . \tag{2.10}$$

For a vector $X \in \mathbb{R}^N$, let $'$ denote the standard transpose operation of vectors. We use $\lambda_{\max}(A)$ ($\lambda_{\min}(A)$) to denote the maximum (minimum) eigenvalue of a matrix A. Thus, to ensure the stability condition that the norm of $\mathbf{e}_{i+1}$ is bounded by that of $\mathbf{e}_i$, being in either one of the following norms

$$\|\mathbf{e}_i\|_{L^2(\Omega)} = (\mathbf{e}_i' M \mathbf{e}_i)^{1/2}; \quad \|\mathbf{e}_i\|_{H^1(\Omega)} = (\mathbf{e}_i' K \mathbf{e}_i)^{1/2}, \tag{2.11}$$

we have the well-known stability condition that $|\lambda_{\max}(I - \Delta t_i M^{-1} K)| < 1$ or equivalently,

$$\Delta t_i < \frac{2}{\lambda_{\max}(M^{-1}K)}. \tag{2.12}$$

Here, we refer $M^{-1}K$ as the preconditioned stiffness matrix (by the mass matrix). Thus, it remains to give some sharp and exact estimates of $\lambda_{\max}(M^{-1}K)$ for general finite element spaces to get precise time step size constraints.[1]

Similarly, for the implicit scheme (2.7), it is unconditionally stable so that no constraint on Δt_i is required. But, at each time step, a linear algebraic system has to be solved to find the numerical solution at the new time step. The corresponding coefficient matrices are either $M + \Delta t_i K$ or $I + \Delta t_i M^{-1} K$, so that their condition number estimates also depend crucially on the extreme eigenvalues of M, K and the preconditioned stiffness matrix $M^{-1}K$. It is worthwhile to point out that while we only focus on the scheme (2.7), similar discussions on condition number estimates for linear systems associated with other implicit schemes, such as the Crank-Nicolson scheme, would also depend on the estimates for $M^{-1}K$.

Before ending the discussion on the background, let us introduce some notations and assumptions on the finite element spaces under consideration. While the same assumptions have been used in [10], we phrase them in more precise terms below.

For any simplicial element $\tau \in \mathcal{T}$, we let $|\tau|$ be its volume, $\{|A_i|\}_{i=1}^{d+1}$ the areas (volumes) of its $d-1$ dimensional faces, and $(b_1, b_2, ..., b_{d+1})$ the barycentric coordinates on τ so that τ is affinely mapped into a reference simplex described by

$$\tau_0 = \{(b_1, b_2, ..., b_{d+1}) \mid b_i \geq 0, \sum b_j = 1\} \ .$$

We also use $\{L_i(\{b_j\})\}_{i=1}^n$ to denote a general form of the nodal basis of the finite element space whose restriction on any $\tau \in \mathcal{T}$ is given, via the mapping to the reference element τ_0, by the same nodal basis on τ_0. The finite element space is given by functions whose restrictions on $\tau \in \mathcal{T}$ are linear combinations of $\{L_i\}$, subject to appropriate boundary conditions on the boundary nodes. In addition, the following assumptions about the nodal basis on any $\tau \in \mathcal{T}$ are made:

- A1. A function in the nodal basis on τ is a multi-variable polynomial of $\{b_j\}$ with the coefficients being independent of the element geometry.
- A2. The nodal basis on τ_0 (thus on any τ), being a set of functions, is invariant with respect to the permutation of the vertices of τ_0 (or τ, correspondingly).

[1] We do not consider the maximum norm stability nor the discrete maximum principle here for which a large number of works exist in the literature; see for instance [24].

In the above assumptions, (A1) implies that the only geometric dependence of a particular basis function is exclusively through its dependence on $\{b_j\}_{j=1}^{d+1}$. It also limits the finite element spaces to be that involving only piecewise polynomials but a similar theory can also be developed for more general spaces that involve non-polynomial basis functions. The assumption (A2) is referred to as a permutation invariance property [10]. These assumptions are satisfied by many typical finite element spaces including the classical standard Lagrange finite element spaces of any order, and other exotic spaces like the enrichment of the conforming linear element with bubble functions or stabilized finite element spaces [1]. Some simple examples that satisfy (A1)-(A2) include the conventional linear element with basis $\{b_j\}_{j=1}^{d+1}$ and the quadratic element with basis $\{b_j(2b_j-1)\}_{j=1}^{d+1} \cup \{4b_jb_k\}_{j\neq k}$, as well as the linear element with a bubble $\{b_j\}_{j=1}^{d+1} \cup \{d^d b_1 b_2 \cdots b_{d+1}\}$. In general, given an arbitrary finite element space, so long as the nodal basis function construction on a simplicial element is based on mapping from that defined on the reference element via the barycentric coordinates and their partial products only, it is then easy to check (A1)-(A2) through a direct verification about the permutation invariance of the set of basis functions on τ_0 with respect to the vertices.

Under the above assumptions (A1)-(A2) on the nodal basis, we now present some important constants. We first introduce an algebraic constant γ_n^d as in [10], which depends only on the nodal basis in the reference element τ_0:

$$\gamma_n^d = \frac{(d-1)!}{d^2(d+1)} \sum_{m=1}^{n} \sum_{j=1}^{d+1} \sum_{k=1}^{d+1} \int_{\tau_0} \left(\frac{\partial L_m}{\partial b_j} - \frac{\partial L_m}{\partial b_k} \right)^2 \mathrm{d}\mathbf{x} \,. \tag{2.13}$$

Note that the volume of τ_0 is given by $|\tau_0| = 1/d!$. Let us also introduce a geometric constant $\mathcal{Q}_d(\tau)$:

$$\mathcal{Q}_d(\tau) = \frac{1}{|\tau|} \sum_{i=1}^{d+1} |A_i|^2 \,, \tag{2.14}$$

which is purely a mesh-dependent constant that is independent of the finite element basis selections [10, 22]. For a given element τ, $\mathcal{Q}_d(\tau)/|\tau|$ represents the sum of the squares of the reciprocals of the altitudes of τ in all directions, which is an important geometric quantity to be used later in the stability estimates. To provide some motivation on the above definitions, we recall that a new trace formula has been established in [10], namely, the trace of the element stiffness matrix K_τ associated with the discretization of the Laplacian operator is in fact the product of an algebraic constant γ_n^d and a geometric factor $\mathcal{Q}_d(\tau)$, under assumptions (A1)-(A2). Thus, γ_n^d can be defined equivalently as an algebraic constant representing the trace of K_{τ_0}, divided by the factor $(d+1)!$. This new formula has been the key to a better understanding of the relation between spectral properties of the stiffness matrix and the mesh geometry [10]. Such a fact is again useful for estimating the preconditioned stiffness matrix here.

3. Estimates for the extreme eigenvalues

Estimates on eigenvalues of stiffness and mass matrices have been provided in standard finite element texts [2], although their precise dependence on the finite element meshes has only been studied more recently [10, 22]. In this section, we recall briefly the extreme eigenvalue estimates on the element stiffness (mass) matrix

of a general finite element discretization of a related elliptic equation given in [10] which extended some earlier works of Fried [13, 14]. Then, we derive some new estimates on the extreme eigenvalues of the preconditioned stiffness matrix $M^{-1}K$. These results form the basis of discussions on parabolic systems.

3.1. **Known eigenvalues estimates for element stiffness and mass matrices.** On a (simplicial) element $\tau \in \mathcal{T}$, we denote the element matrices corresponding to K and M by K_τ and M_τ, respectively, with n being their dimensions, which corresponds to the degree of freedom or the number of nodal basis functions associated with the element τ. A geometrically precise estimate for the maximum eigenvalue of the element stiffness matrix is given in [10]:

Lemma 3.1. *For any general finite element spaces defined on a simplicial mesh $\mathcal{T}$ with the nodal basis on any d-dimensional simplex $\tau \in \mathcal{T}$ satisfying the assumptions (A1)-(A2) specified above, we have the following estimate for the maximum eigenvalue of the element stiffness matrix K_τ,*

$$\frac{\beta_1^\tau \gamma_n^d}{n-1} \mathcal{Q}_d(\tau) \le \lambda_{\max}(K_\tau) \le \beta_2^\tau \gamma_n^d \mathcal{Q}_d(\tau) \,, \tag{3.1}$$

where n is the dimension of the local finite element space, γ_n^d and $\mathcal{Q}_d(\tau)$ are given by (2.13) *and* (2.14)*, and $\beta_1^\tau, \beta_2^\tau$ are defined by* (2.9)*.*

As in [10], it is important to note that γ_n^d is a positive constant that depends only on the corresponding basis functions on the reference simplex τ_0 and is independent of the geometry of the particular element τ while the other factor $\mathcal{Q}_d(\tau)$ is completely independent of the choice of the finite element spaces as long as they satisfy (A1)-(A2). The above estimate does provide a precise control on the contribution due to the mesh geometry on the maximum eigenvalue of the element stiffness matrix which is crucial for estimating stable time steps and condition numbers of globally assembled matrices for parabolic problems.

For the mass matrix, its extreme eigenvalues have also been carefully studied before, for example, in [25] and [10]. Here, we state a result based on the use of local nodal basis. One may find detailed computation in [10].

Lemma 3.2. *For the mass matrix defined by* (2.6)*, the corresponding element mass matrix M_τ on an element τ satisfies*

$$M_\tau = \frac{|\tau|}{|\tau_0|} M_{\tau_0} \,, \tag{3.2}$$

where $|\tau_0|$ is the volume of the reference simplex τ_0 and M_{τ_0} is the element mass matrix on τ_0. Consequently,

$$\min_{\tau\in\mathcal{T}} \lambda_{\min}(M_\tau) = \delta_n \min_{\tau\in\mathcal{T}} |\tau| \,, \quad \max_{\tau\in\mathcal{T}} \lambda_{\max}(M_\tau) = \sigma_n \max_{\tau\in\mathcal{T}} |\tau| \,, \tag{3.3}$$

where δ_n and σ_n are two constants given by

$$\delta_n = \frac{1}{|\tau_0|} \lambda_{\min}(M_{\tau_0}) \,, \quad \sigma_n = \frac{1}{|\tau_0|} \lambda_{\max}(M_{\tau_0}) \,. \tag{3.4}$$

Note that the constants δ_n and σ_n are independent of the particular element τ but only dependent on τ_0 and the corresponding local finite element basis.

3.2. **New estimates for the maximum eigenvalue of $M^{-1}K$.** In this subsection, we estimate the maximum eigenvalue of the preconditioned stiffness matrix $M^{-1}K$, which is a key result for understanding the desired properties of the time dependent problems.

For a vector $X \in \mathbb{R}^N$ and any element τ, we use X_τ to denote the portion of X associated with τ. Then using the standard Rayleigh quotient argument, we have

$$\begin{aligned}
\lambda_{\max}(M^{-1}K) &= \max_{X\in\mathbb{R}^N, X\neq \mathbf{0}} \Big\{ \frac{X'KX}{X'MX} \Big\} = \max_{X\in\mathbb{R}^N, X\neq \mathbf{0}} \Big\{ \frac{\sum\limits_\tau X'_\tau K_\tau X_\tau}{\sum\limits_\tau X'_\tau M_\tau X_\tau} \Big\} \\
&= \max_{X\in\mathbb{R}^N, X\neq \mathbf{0}} \Big\{ \frac{\sum\limits_\tau \frac{X'_\tau K_\tau X_\tau}{X'_\tau M_\tau X_\tau} X'_\tau M_\tau X_\tau}{\sum\limits_\tau X'_\tau M_\tau X_\tau} \Big\} \\
&\le \max_{X\in\mathbb{R}^N, X\neq \mathbf{0}} \Big\{ \frac{\sum\limits_\tau \lambda_{\max}(M_\tau^{-1}K_\tau) X'_\tau M_\tau X_\tau}{\sum\limits_\tau X'_\tau M_\tau X_\tau} \Big\} \qquad (3.5) \\
&\le \max_\tau \{\lambda_{\max}(M_\tau^{-1}K_\tau)\}.
\end{aligned}$$

We state the above result as the following lemma.

Lemma 3.3. *The maximum eigenvalue of $M^{-1}K$ has the estimate:*

$$\lambda_{\max}(M^{-1}K) \le \max_\tau \{\lambda_{\max}(M_\tau^{-1}K_\tau)\}. \qquad (3.6)$$

In order to derive a lower bound of $\lambda_{\max}(M^{-1}K)$, we introduce an assumption referred to here as the *gradual volume change condition*: there exists a positive constant c_1 such that, for any element $\tau \in \mathcal{T}$ with the adjacent elements $\tilde{\tau}_i$ ($i = 1, 2, 3, ..., p_\tau$), we have

$$|\tilde{\tau}_i|/|\tau| \le c_1, \quad \forall\, 1 \le i \le p_\tau, \qquad (3.7)$$

where p_τ is the number of elements adjacent to τ. Let us point out that this *gradual volume change condition* is very weak, and is easily satisfied by most commonly used meshes in finite element methods. For convenience, we denote the lumped region of τ with its adjacent elements by $\mathrm{T}_\tau = \bigcup_{i=1}^{p_\tau} \tilde{\tau}_i$.

Lemma 3.4. *If a mesh $\mathcal{T}$ satisfies the* gradual volume change condition *as defined by* (3.7)*, the maximum eigenvalue of $M^{-1}K$ has the estimate*

$$\frac{\max\limits_\tau \{\lambda_{\max}(M_\tau^{-1}K_\tau)\}}{(1+\kappa(M_0)c_1P_*)} \le \lambda_{\max}(M^{-1}K), \qquad (3.8)$$

where P_* is the maximum number of elements meeting at a nodal point, and the spectral condition number is $\kappa(M_0) = \lambda_{\max}^{M_0}/\lambda_{\min}^{M_0}$.

Proof. We take a special vector $\hat{X} \in \mathrm{R}^N$ such that a portion of it is the eigenvector corresponding to the eigenvalue $\max\limits_\tau\{\lambda_{\max}(M_\tau^{-1}K_\tau)\}$ and the rest zero. Here we denote the corresponding element where the above maximum is reached by τ^*. By

Lemma 3.2 and the standard Rayleigh quotient argument, we get that

$$\begin{aligned}\sum_{\tilde{\tau}\in T_{\tau^*}} \hat{X}'_{\tilde{\tau}} M_{\tilde{\tau}} \hat{X}_{\tilde{\tau}} &\le \sum_{\tilde{\tau}\in T_{\tau^*}} \lambda_{\max}^{M_{\tilde{\tau}}} \hat{X}'_{\tilde{\tau}} \hat{X}_{\tilde{\tau}} = \sum_{\tilde{\tau}\in T_{\tau^*}} \lambda_{\max}^{M_0}(|\tilde{\tau}|/|\tau_0|)\hat{X}'_{\tilde{\tau}} \hat{X}_{\tilde{\tau}} \\ &\le \lambda_{\max}^{M_0} c_1(|\tau^*|/|\tau_0|) \sum_{\tilde{\tau}} \hat{X}'_{\tilde{\tau}} \hat{X}_{\tilde{\tau}} \le \lambda_{\max}^{M_0} c_1(|\tau^*|/|\tau_0|) P_* \hat{X}'_{\tau^*} \hat{X}_{\tau^*} \\ &\le \kappa(M_0) c_1 P_*(|\tau^*|/|\tau_0|)\hat{X}'_{\tau^*} M_0 \hat{X}_{\tau^*} = (\kappa(M_0)c_1 P_*)\hat{X}'_{\tau^*} M_{\tau^*} \hat{X}_{\tau^*}.\end{aligned} \tag{3.9}$$

Thus, we have from this that

$$\begin{aligned}\lambda_{\max}(M^{-1}K) &= \max_{X\in\mathbb{R}^N, X\neq \mathbf{0}} \Big\{ \frac{\sum_{\tau} \frac{X'_\tau K_\tau X_\tau}{X'_\tau M_\tau X_\tau} X'_\tau M_\tau X_\tau}{\sum_{\tau} X'_\tau M_\tau X_\tau} \Big\} \ge \Big\{ \frac{\sum_{\tau} \frac{\hat{X}'_\tau K_\tau \hat{X}_\tau}{\hat{X}'_\tau M_\tau \hat{X}_\tau} \hat{X}'_\tau M_\tau \hat{X}_\tau}{\sum_{\tau} \hat{X}'_\tau M_\tau \hat{X}_\tau} \Big\} \\ &= \Big\{ \frac{\sum_{\tilde{\tau}\in T_{\tau^*}\cup\tau^*} \frac{\hat{X}'_{\tilde{\tau}} K_{\tilde{\tau}} \hat{X}_{\tilde{\tau}}}{\hat{X}'_{\tilde{\tau}} M_{\tilde{\tau}} \hat{X}_{\tilde{\tau}}} \hat{X}'_{\tilde{\tau}} M_{\tilde{\tau}} \hat{X}_{\tilde{\tau}}}{\sum_{\tilde{\tau}\in T_{\tau^*}\cup\tau^*} \hat{X}'_{\tilde{\tau}} M_{\tilde{\tau}} \hat{X}_{\tilde{\tau}}} \Big\} \ge \Big\{ \frac{\frac{\hat{X}'_{\tau^*} K_{\tau^*} \hat{X}_{\tau^*}}{\hat{X}'_{\tau^*} M_{\tau^*} \hat{X}_{\tau^*}} \hat{X}'_{\tau^*} M_{\tau^*} \hat{X}_{\tau^*}}{\sum_{\tilde{\tau}\in T_{\tau^*}\cup\tau^*} \hat{X}'_{\tilde{\tau}} M_{\tilde{\tau}} \hat{X}_{\tilde{\tau}}} \Big\} \\ &\ge \frac{\max_{\tau}\{\lambda_{\max}(M_\tau^{-1}K_\tau)\}}{(1+\kappa(M_0)c_1 P_*)}.\end{aligned}$$

This completes the proof of Lemma (3.4). □

Since M_τ is symmetric positive definite, we have

$$\Lambda(M_\tau^{-1}K_\tau) = \Lambda(M_\tau^{1/2} M_\tau^{-1} K_\tau M_\tau^{-1/2}) = \Lambda(M_\tau^{-1/2} K_\tau M_\tau^{-1/2}),$$

where the notation $\Lambda(A)$ represents the set of all eigenvalues of a matrix A. Hence, the key is to estimate eigenvalues of $M_\tau^{-1/2} K_\tau M_\tau^{-1/2}$. For simplicity, let $P = M_\tau^{-1/2}$. Obviously, P is also a symmetric and positive definite matrix. Thus,

$$\begin{aligned}\lambda_{\max}(PK_\tau P) &= \max_{Y\in\mathbb{R}^n, Y\neq\mathbf{0}} \Big\{ \frac{Y'(PK_\tau P)Y}{Y'Y} \Big\} \\ &= \max_{Y\in\mathbb{R}^n, Y\neq\mathbf{0}} \Big\{ \frac{(PY)'K_\tau(PY)}{(PY)'(PY)} \cdot \frac{Y'M_\tau^{-1}Y}{Y'Y} \Big\} \\ &\le \lambda_{\max}(K_\tau)\lambda_{\max}(M_\tau^{-1}) = \frac{\lambda_{\max}(K_\tau)}{\lambda_{\min}(M_\tau)}.\end{aligned} \tag{3.10}$$

On the other hand, with the Rayleigh quotient, taking the special vector $Y = \tilde{Y}$ such that $P\tilde{Y}$ is an eigenvector corresponding to the maximum eigenvalue of K_τ, then we can get,

$$\begin{aligned}\lambda_{\max}(PK_\tau P) = \lambda_{\max}(P'K_\tau P) &= \max_{Y\in\mathbb{R}^n, Y\neq\mathbf{0}} \Big\{ \frac{Y'(P'K_\tau P)Y}{Y'Y} \Big\} \\ &= \max_{Y\in\mathbb{R}^n, Y\neq\mathbf{0}} \Big\{ \frac{(PY)'K_\tau(PY)}{(PY)'(PY)} \frac{Y'P'PY}{Y'Y} \Big\} \\ &\ge \frac{(P\tilde{Y})'K_\tau(P\tilde{Y})}{(P\tilde{Y})'(P\tilde{Y})} \frac{\tilde{Y}'P'P\tilde{Y}}{\tilde{Y}'\tilde{Y}} = \lambda_{\max}(K_\tau)\frac{\tilde{Y}'P'P\tilde{Y}}{\tilde{Y}'\tilde{Y}} \\ &\ge \lambda_{\max}(K_\tau)\lambda_{\min}(M_\tau^{-1}) = \lambda_{\max}(K_\tau)/\lambda_{\max}(M_\tau).\end{aligned} \tag{3.11}$$

With (3.6), (3.8), (3.10), (3.11), (3.1) and (3.3), we can immediately get the following estimates for the maximum eigenvalue of $M^{-1}K$.

Theorem 3.1. *For any general finite element space defined on a simplicial mesh $\mathcal{T}$ with the nodal basis on any d-dimensional simplex $\tau \in \mathcal{T}$ satisfying assumptions (A1)-(A2) specified before, the maximum eigenvalue of the matrix $M^{-1}K$ has the estimate*

$$\lambda_{\max}(M^{-1}K) \le \tilde{\gamma}_n^d \max_{\tau\in\mathcal{T}}\{\frac{\beta_2^\tau \mathcal{Q}_d(\tau)}{|\tau|}\}, \tag{3.12}$$

furthermore, when the mesh $\mathcal{T}$ satisfies the gradual volume change condition (3.7)*,*

$$\hat{\gamma}_n^d \max_{\tau\in\mathcal{T}}\{\frac{\beta_1^\tau \mathcal{Q}_d(\tau)}{|\tau|}\} \le \lambda_{\max}(M^{-1}K), \tag{3.13}$$

where n is the cardinality of the set of local nodal basis, and $\hat{\gamma}_n^d, \tilde{\gamma}_n^d$ are given by

$$\hat{\gamma}_n^d = \frac{\lambda_{\max}^{-1}(M_{\tau_0})}{(n-1)(1+c_1P_*\kappa(M_0))}|\tau_0|\gamma_n^d, \quad \tilde{\gamma}_n^d = \lambda_{\min}^{-1}(M_{\tau_0})|\tau_0|\gamma_n^d. \tag{3.14}$$

Here, let us point out that the constants $\tilde{\gamma}_n^d$ and $\hat{\gamma}_n^d$ are independent of the particular element τ, but are only dependent on τ_0, the corresponding local finite element basis and the global mesh parameters (c_1 and P_*). The other factors in the bounds, which are dependent on mesh geometry and equations, are completely independent of the choice of the finite element spaces. As pointed out previously, for a given element τ, $\mathcal{Q}_d(\tau)/|\tau|$ represents the sum of the squares of the reciprocals of the altitudes of τ in all directions geometrically. When the mesh $\mathcal{T}$ satisfies the *gradual volume change condition* as defined by (3.7), both the upper and lower bounds of $\lambda_{\max}(M^{-1}K)$ have the same mesh-dependent factor, $\max_{\tau\in\mathcal{T}}\{\mathcal{Q}_d(\tau)/|\tau|\}$, which implies that the estimates, in the isotropic case, are very precise and sharp with respect to the mesh geometry.

4. Mesh dependent stability condition for the explicit scheme

In this section, we apply the general estimates in the previous sections to derive the mesh dependent permissible time step sizes for the explicit time integration scheme (2.4) for parabolic problems. Some of these estimates are widely known and are consistent with the popular understanding in the finite element and meshing community, while others are interesting on their own.

4.1. **Mesh-dependent stability condition.**

With the estimates in the previous sections, we have the following theorem on the mesh-dependent stability condition.

Theorem 4.1. *For the explicit time integration scheme* (2.4) *discretized by finite element space defined on a simplicial mesh $\mathcal{T}$ with the nodal basis on any d-dimensional simplex $\tau \in \mathcal{T}$ satisfying the assumptions specified above, the stability condition for the time step size Δt_i is given by*

$$\Delta t_i < \frac{2}{\tilde{\gamma}_n^d} \min_{\tau\in\mathcal{T}} \frac{|\tau|}{\beta_2^\tau \mathcal{Q}_d(\tau)}\ , \tag{4.1}$$

where n is the cardinality of the set of local nodal basis functions, $\mathcal{Q}_d(\tau)$ and $\tilde{\gamma}_n^d$ are respectively defined by (2.14) *and* (3.14)*, and β_2^τ is defined in* (2.9)*.*

Theorem 4.1 directly follows from the estimates (2.12) and (3.12). We give some remarks on the above stability condition.

Remark 4.1. In [22], some preliminary discussion on mesh-dependent stable time step size was provided for the simple linear finite element. The result here is new in two aspects: first, our result is for general finite element spaces, and second, we do not need to use the *mass lumping*. The latter case is of special interest in practice which is examined more closely in a separate work [26] where spectral estimates involving lumped diagonal mass matrices can be more directly derived than those involving general non-diagonal mass matrices as being considered here.

Remark 4.2. The stable time step size condition given by (4.1) is very elegant: it implies that the stable time step size of the explicit differencing is a product of two factors, with one being $2/\tilde{\gamma}_n^d$ which is completely independent of the geometry of the particular element τ, and the other being $\min_{\tau\in\mathcal{T}}\{|\tau|/(\beta_2^{\mathcal{T}}\mathcal{Q}_d(\tau))\}$, which is completely independent of the choice of the finite element spaces. This allows us to explore the precise relationship between the stable time step size condition and the mesh geometry in various special cases, as discussed later.

4.2. **Relating time step size constraint to mesh geometry.** We now explore the stability condition (4.1) for a variety of commonly encountered meshes.

Let us first discuss the effect of the element size on the stable time step size when the shape of element remains regular. That is, we consider a simplicial mesh $\mathcal{T}$ with all the simplices $\tau \in \mathcal{T}$ satisfying the shape regular assumption:

$$\rho_1|\tau|^{2-2/d} \le \sum_i |A_i|^2 \le \rho_2|\tau|^{2-2/d}\,, \quad \forall\, \tau \in \mathcal{T} \tag{4.2}$$

for some positive constants ρ_1 and ρ_2, independent of τ. Here, for any d-dimensional simplex τ, $|\tau|$ is its volume and $\{|A_i|\}_{i=1}^{d+1}$ are the areas (volumes) of its $d-1$-dimensional faces. We refer to such meshes as *shape regular* (as commonly called in the finite element literature [6]). Meanwhile, we refer to a simplicial mesh $\mathcal{T}$ as being *quasi-volume-uniform* if

$$\min_{\tau\in\mathcal{T}}|\tau| \ge \rho_3 \max_{\tau\in\mathcal{T}}|\tau| \tag{4.3}$$

holds for some positive constant ρ_3, independent of τ [10]. Note that for the shape regular meshes, the notion of quasi-volume-uniform is equivalent to the conventional quasi-uniform triangulation assumption characterized by a bounded ratio of the largest and smallest element sizes [6, 23].

For the shape-regular and quasi-volume-uniform meshes, (4.1) recovers a most widely known estimate [22].

Corollary 4.1. *For the explicit time integration scheme* (2.4) *discretized by a finite element space which satisfies (A1)-(A2) and is defined on a quasi-volume-uniform mesh, in the sense of* (4.3)*, d-dimensional simplicial mesh with h being the mesh parameter (diameter of the largest simplex), if we further assume that all simplices are shape regular in the sense of* (4.2)*, then it is stable if the time step size satisfies*

$$\Delta t_i < c_{(n,d)}^{(1)} h^2 \min_{\tau}\{1/\beta_2^{\mathcal{T}}\}, \tag{4.4}$$

for some constant $c_{(n,d)}^{(1)}$ which is dependent upon the finite element basis defined on the reference element τ_0 and the space dimension d, but is independent of mesh geometry.

Corollary 4.1 follows directly from (2.14), (4.1), (4.2) and (4.3) and it also indicates that, for a uniform mesh of well-shaped elements, the maximum permissible time step size Δt_i for problems with isotropic diffusion coefficients is roughly proportional to h^2. Hence, smaller elements require a smaller time step size and more computation. The result in Corollary 4.1 for the linear finite element case with *mass lumping* is widely known in the finite element and meshing community [20, 22], while the same result on general finite element spaces is rarely provided.

Now, we consider the mesh with well-shaped elements which satisfies the shape regular condition (4.2) but does not necessarily satisfy the quasi-volume-uniform condition (4.3). Such cases often appear in the isotropic adaptive finite element computation. In this case, we have the following corollary.

Corollary 4.2. *For the explicit time integration scheme* (2.4) *discretized by a finite element space which satisfies (A1)-(A2) and is defined on a shape regular d-dimensional simplicial mesh, then it is stable if the time step size satisfies*

$$\Delta t_i < c^{(2)}_{(n,d)} \min_{\tau\in\mathcal{T}}\{|\tau|^{2/d}/\beta_2^\tau\} \tag{4.5}$$

for some constant $c^{(2)}_{(n,d)}$ which is dependent on the finite element basis on the reference element τ_0 and dimension d, but is independent of mesh geometry.

From the above corollary, for the diffusion problems with isotropic coefficients, we can see that the time step size is usually dictated by the size of the *scaled* element volume in a non-uniform mesh with well-shaped elements. The scaling effect is introduced by the appearance of the factor β_2^τ. Thus, it helps to maintain a relatively larger stable time step size if the smaller mesh elements happen to correspond to smaller values of the diffusion coefficients on these elements.

Next, we consider an anisotropic but quasi-volume-uniform mesh which satisfies the condition (4.3) but not (4.2). For $\mathcal{Q}_d(\tau)$, the following explicit expressions are given in [10],

$$\mathcal{Q}_2(\tau) = 4\sum_{i=1}^{3}\cot\theta_i^\tau\ , \qquad \mathcal{Q}_3(\tau) = 3\sum_{i=1}^{4}\sum_{j=1,j\neq i}^{4} l_{ij}\cot\theta_{ij}^\tau\ , \tag{4.6}$$

where θ_i^τ is the triangle interior angle and θ_{ij}^τ is the dihedral angle on the edge E_{ij} with the length l_{ij}. Then, by (4.1), (4.3), and (4.6), we have the following result.

Corollary 4.3. *For the explicit time integration scheme* (2.4) *discretized by a finite element space which satisfies (A1)-(A2) and is defined on a quasi-volume-uniform mesh in the sense of* (4.3) *d-dimensional simplicial mesh, then it is stable if the time step size satisfies*

$$\Delta t_i < c^{(3)}_{(n,2)} \min_{\tau\in\mathcal{T}}\{|\tau|(\beta_2^\tau\sum_{i=1}^{3}\cot\theta_i^\tau)^{-1}\},\ \textit{when } d=2, \tag{4.7}$$

$$\Delta t_i < c^{(4)}_{(n,3)} \min_{\tau\in\mathcal{T}}\{|\tau|(\beta_2^\tau\sum_{i=1}^{4}\sum_{j=1,j\neq i}^{4} l_{ij}\cot\theta_{ij}^\tau)^{-1}\},\ \textit{when } d=3, \tag{4.8}$$

for some constants $c^{(3)}_{(n,2)}$ and $c^{(4)}_{(n,3)}$ which are dependent on the finite element basis on the reference element τ_0 and dimension d, but are independent of element geometry.

The above corollary indicates that, for the problems with isotropic diffusion coefficients and quasi-volume-uniform meshes, if anisotropic elements are used or the shape-regularity cannot be met so that there are elements with small angles in two dimensions and elements with small dihedral angles in three dimensions, then conditions (4.7) and (4.8) imply that the explicit time marching is required to take even smaller time step sizes to ensure stability. While for the problems with anisotropic diffusion coefficients, a relatively larger stable time step size may be taken by choosing a suitably aligned mesh.

In summary, we see that in practice, both the spatial mesh size and mesh shape, properly scaled by the diffusion coefficients, may affect the stability conditions on the time step size as predicted by the more general estimate (4.1).

5. Mesh dependent condition number estimates for the implicit scheme

In the section, we derive some precise and mesh-dependent estimates of the condition number of the resulting matrix for the implicit scheme, which is helpful for us to improve the efficiency and accuracy of the solution to the linear system by mesh optimization and mesh-based pre-conditioning. We first present some general estimates on the global mass/stiffness matrices given in [13, 14] which are used in the later condition number estimates.

5.1. **Estimates for the global mass/stiffness matrices.** For the stiffness and mass matrices K and M defined by (2.6), their eigenvalues are denoted by $\{\lambda_i^K\}_{i=1}^N$ and $\{\lambda_i^M\}_{i=1}^N$, respectively, which are ordered by

$$\lambda_1^K \le \lambda_2^K \le \cdots \le \lambda_N^K, \qquad \lambda_1^M \le \lambda_2^M \le \cdots \le \lambda_N^M .$$

In this notation, λ_1^K and λ_1^M are the minimum eigenvalues of K and M, and λ_N^K and λ_N^M are the maximum eigenvalues, respectively. Similarly, we use $\{\lambda_i^{K_\tau}\}_{i=1}^n$ and $\{\lambda_i^{M_\tau}\}_{i=1}^n$ to denote the eigenvalues of K_τ and M_τ, respectively, which are also ordered by

$$\lambda_1^{K_\tau} \le \lambda_2^{K_\tau} \le \cdots \le \lambda_n^{K_\tau}, \qquad \lambda_1^{M_\tau} \le \lambda_2^{M_\tau} \le \cdots \le \lambda_n^{M_\tau} .$$

In [13, 14], some estimates for the extreme eigenvalues of the global stiffness (mass) matrix in relation to that of the element stiffness (mass) matrix were given:

$$\max_{\tau\in\mathcal{T}}(\lambda_n^{K_\tau}) \le \lambda_N^K \le P_* \max_{\tau\in\mathcal{T}}(\lambda_n^{K_\tau}), \quad \lambda_1^* \min_{\tau\in\mathcal{T}}(\lambda_1^{\tilde{M}_\tau}) \le \lambda_1^K \le \hat{\lambda}_1^* P_* \max_{\tau\in\mathcal{T}}(\lambda_n^{\tilde{M}_\tau}), \tag{5.1}$$

$$\min_{\tau\in\mathcal{T}}(\lambda_n^{M_\tau}) \le \lambda_1^M \le P_* \max_{\tau\in\mathcal{T}}(\lambda_n^{M_\tau}), \quad \max_{\tau\in\mathcal{T}}(\lambda_n^{M_\tau}) \le \lambda_N^M \le P_* \max_{\tau\in\mathcal{T}}(\lambda_n^{M_\tau}), \tag{5.2}$$

where P_* is the maximum number of elements meeting at a nodal point[2], λ_1^* is the smallest eigenvalue corresponding to the generalized eigenvalue problem

$$\begin{cases} -\sum_{i,j=1}^{d} \dfrac{\partial}{\partial x_i}(a_{ij}\dfrac{\partial u}{\partial x_j}) = \lambda\rho u, & \text{in} \quad \Omega, \\ u = 0, & \text{on} \quad \partial\Omega, \end{cases} \tag{5.3}$$

with the eigenfunction u being non-zero, and a density function $\rho = \rho(\mathbf{x})$ which can be taken to be the unit constant in most cases, but with a non-uniform density, sharper estimates can be established for highly non-uniform meshes. Here, $\hat{\lambda}_1^*$ is

[2]In this work, we always consider meshes with uniformly bounded P_*.

the corresponding finite element smallest eigenvalue computed from $KV = \lambda\tilde{m}V$ and the generalized mass matrix $\tilde{M} = (\tilde{m}_{ij})$ is as defined by $\tilde{m}_{ij} = \zeta_n\rho\phi_i\phi_j\,dx$, and $\tilde{M}_\tau$ is the mass matrix on element τ.

For a non-uniform density function ρ, by using (3.3), simple calculations give

$$\delta_n \min_{\tau\in\mathcal{T}}\{\rho^\tau_{\min}|\tau|\} \le \min_{\tau\in\mathcal{T}} \lambda_1^{\tilde{M}_\tau} \le \delta_n \min_{\tau\in\mathcal{T}}\{\rho^\tau_{\max}|\tau|\}\ , \tag{5.4}$$

$$\sigma_n \max_{\tau\in\mathcal{T}}\{\rho^\tau_{\min}|\tau|\} \le \max_{\tau\in\mathcal{T}} \lambda_n^{\tilde{M}_\tau} \le \sigma_n \max_{\tau\in\mathcal{T}}\{\rho^\tau_{\max}|\tau|\}\ , \tag{5.5}$$

where $\rho^\tau_{\min}$ and $\rho^\tau_{\max}$ are the minimum and maximum values of ρ on element τ, and δ_n, σ_n are given by (3.4). The smallest exact eigenvalue λ_1^* can be regarded as a constant that depends only on intrinsic properties of the continuous problem, but does not depend on discretization parameters.

5.2. **The mesh dependent condition number estimates.** As discussed earlier, the implicit scheme requires the numerical solution of a linear system with the coefficient matrix $(M+\Delta t_i K)$. In order to derive mesh dependent condition number estimates, we first present a preliminary lemma

Lemma 5.1. *Let $\lambda_1(M + \Delta t_i K) \le \lambda_2(M + \Delta t_i K) \le \cdots \le \lambda_N(M + \Delta t_i K)$ be all the eigenvaules of $(M + \Delta t_i K)$. Then, we have*

$$\begin{aligned}(\min_{\tau\in\mathcal{T}}(\lambda_1^{M_\tau}) + \Delta t_i\lambda_1^* \min_{\tau\in\mathcal{T}}(\lambda_1^{\tilde{M}_\tau})) \le\ & \lambda_1(M + \Delta t_i K) \le P_*(\max_{\tau\in\mathcal{T}}(\lambda_n^{M_\tau}) \\ & + \Delta t_i\hat{\lambda}_1^* \max_{\tau\in\mathcal{T}}(\lambda_n^{\tilde{M}_\tau}))\ ,\end{aligned} \tag{5.6}$$

$$\max_{\tau\in T}(\lambda_1^{M_\tau} + \Delta t_i\lambda_n^{K_\tau}) \le \lambda_N(M + \Delta t_i K) \le P_* \max_{\tau\in\mathcal{T}}(\lambda_n^{M_\tau} + \Delta t_i\lambda_n^{K_\tau}), \tag{5.7}$$

where K_τ/M_τ is the element stiffness/mass matrix on the element τ, $\tilde{M}_\tau$ is the element mass matrix associated with the generalized mass matrix $\tilde{M}$, and P_* is the maximum number of elements meeting at a single vertex, $\lambda_1^*(\hat{\lambda}_1^*)$ is the smallest exact (numerical) eigenvalue of the eigenvalue problem (5.3).

Proof. Given any nonzero vector $X \in \mathbb{R}^N$, we have from the ellipticity that

$$X'KX \ge \lambda_1^* X'\tilde{M}X,$$

which yields that

$$\frac{X'(M + \Delta t_i K)X}{X'X} \ge \frac{X'MX + \Delta t_i\lambda_1^* X'\tilde{M}X}{X'X} \ge (\lambda_1^M + \Delta t_i\lambda_1^*\lambda_1^{\tilde{M}}).$$

Taking a special vector $X = \hat{X}$ such that it is the eigenvector corresponding to $\lambda_1(K)$ and using the Rayleigh quotient argument, we can arrive at

$$\lambda_1(M + \Delta t_i K) \quad \le \quad \lambda_N^M + \Delta t_i\lambda_1^K. \tag{5.8}$$

Then we may apply the results of (5.1) and (5.2) to get (5.6).

For the proof of (5.7), we follow the technique used for (5.1)-(5.2) in [13, 14]. Namely, for a unit vector $X \in \mathbb{R}^N$ and any element τ, we let x_τ denote the portion

of X associated with τ. We then decompose the following quadratic form into the sum of contributions on each element:

$$X'(M+\Delta t_i K)X = \sum_\tau x'_\tau (M_\tau + \Delta t_i K_\tau) x_\tau$$
$$\leq \sum_\tau \lambda_n^{(M_\tau+\Delta t_i K_\tau)} x'_\tau x_\tau \leq \max_\tau \lambda_n^{(M_\tau+\Delta t_i K_\tau)} \sum_\tau x'_\tau x_\tau .$$

Obviously, we have $\lambda_n^{(M_\tau+\Delta t_i K_\tau)} \leq \lambda_n^{M_\tau} + \Delta t_i \lambda_n^{K_\tau}$, so by the inequality $\sum_\tau x'_\tau x_\tau \leq P_*$, we get the right-hand side of inequality (5.7) from the standard Rayleigh quotient argument. On the other hand, to derive the lower bound on the maximum eigenvalue, we select a special vector $X = \hat{X}$ such that a portion of it is the eigenvector corresponding to $\max_\tau\{\lambda_{\max}(M_\tau + \Delta t_i K_\tau)\}$ and the rest zero. Then, we have

$$\lambda_N(M+\Delta t_i K) \geq \frac{X'(M+\Delta t_i K)X}{X'X} = \frac{\sum_\tau x'_\tau (M_\tau + \Delta t_i K_\tau) x_\tau}{X'X}$$
$$\geq \frac{\sum_\tau \hat{x}'_\tau (M_\tau + \Delta t_i K_\tau) \hat{x}_\tau}{\hat{X}'\hat{X}} \geq \max_\tau\{\lambda_{\max}(M_\tau + \Delta t_i K_\tau)\}. \tag{5.9}$$

Furthermore, taking a special vector $x_\tau = \tilde{x}_\tau$ such that it is the eigenvector corresponding to $\lambda_{\max}^{K_\tau}$, the standard Rayleigh quotient argument gives

$$\lambda_{\max}(M_\tau + \Delta t_i K_\tau) \geq \lambda_{\min}(M_\tau) + \Delta t_i \lambda_{\max}(K_\tau). \tag{5.10}$$

The left-hand side of inequality (5.7) follows directly from (5.9) and (5.10). This completes the proof of lemma 5.1. □

Then, we have the following mesh-dependent estimates of the spectral condition number $\kappa(M+\Delta t_i K)$.

Theorem 5.1. *For the implicit time integration scheme* (2.7) *discretized by a general finite element space defined on a d-dimensional simplicial mesh, let (A1)-(A2) be satisfied and let $M+\Delta t_i K$ be the coefficient matrix of the resulting linear system, then its condition number satisfies*

$$\frac{\max_{\tau\in\mathcal{T}} \left\{\delta_n|\tau| + \Delta t_i \gamma_n^d \beta_1^\tau \mathcal{Q}_d(\tau)/(n-1)\right\}}{P_*\sigma_n(\max_{\tau\in\mathcal{T}}|\tau| + \Delta t_i \hat{\lambda}_1^* \max_{\tau\in\mathcal{T}}\{\rho_{\max}^\tau|\tau|\})} \leq \kappa(M+\Delta t_i K)$$
$$\leq \frac{P_* \max_{\tau\in\mathcal{T}} \left\{\sigma_n|\tau| + \Delta t_i \gamma_n^d \beta_2^\tau \mathcal{Q}_d(\tau)\right\}}{\delta_n(\min_{\tau\in\mathcal{T}}|\tau| + \Delta t_i \lambda_1^* \min_{\tau\in\mathcal{T}}\{\rho_{\min}^\tau|\tau|\})}, \tag{5.11}$$

where $\mathcal{Q}_d(\tau)$ and the constants δ_n, σ_n are respectively defined by (2.14) *and* (3.4), *β_2^τ and γ_n^d are, respectively, defined by* (2.9) *and* (2.13), *and P_* is the maximum number of elements meeting at a single vertex, $\lambda_1^*(\hat{\lambda}_1^*)$ is the smallest exact (numerical) eigenvalue of the eigenvalue problem* (5.3).

Proof. The result follows immediately from estimates (3.1), (3.3), (5.1), (5.2), (5.4), (5.5), and Lemma 5.1. □

The results in Theorem 5.1 show that the condition number is not only dependent on the mesh geometry but also on the time step size Δt_i. When the time step size Δt_i is relatively small, the mesh shape has little effect on the condition number of the coefficient matrix since the condition number in this case is determined by the

term $(\max_{\tau\in\mathcal{T}}|\tau|)/(\min_{\tau\in\mathcal{T}}|\tau|)$. In particular, for a quasi-volume-uniform mesh in the sense of (4.3), the condition number would be independent of the mesh geometry.

When the time step size Δt_i is large, we see that $\kappa(M+\Delta t_i K)$ is essentially dominated by $\kappa(K)$, and one would encounter the same type of ill-conditioning for the system resulted from the implicit time integration of a parabolic equation as that corresponding to the associated elliptic problem. That is, on a typical quasi-uniform and shape regular d-dimensional simplicial mesh with h being the mesh parameter (diameter of the largest simplex), $\kappa(M+\Delta t_i K)$ is dominated by $C_n^d h^{-2}$ as $h\to 0$ [10], where the constant C_n^d is dependent on the finite element basis on the reference element τ_0, the space dimension d and the diffusion coefficients of the continuous problem, but it is independent of h. Meanwhile, on the anisotropic mesh that satisfies the quasi-volume-uniform condition (4.3) but not the shape regular condition (4.2), with (4.6) and (5.11), simple calculation gives that $\kappa(M+\Delta t_i K)$ is dominated by

$$C_{(n,2)}\max_{\tau\in\mathcal{T}}(\beta_2^\tau\sum_{i=1}^{3}\cot\theta_i^\tau)\max_{\tau\in\mathcal{T}}(|\tau|^{-1}),\quad \text{for } d=2 \tag{5.12}$$

or

$$C_{(n,3)}\max_{\tau\in\mathcal{T}}(\beta_2^\tau\sum_{i=1}^{4}\sum_{j=1,j\neq i}^{4}l_{ij}\cot\theta_{ij}^\tau)\max_{\tau\in\mathcal{T}}(|\tau|^{-1})\ ,\ \text{for } d=3, \tag{5.13}$$

where constants $C_{(n,2)}$ and $C_{(n,3)}$ are dependent on the finite element basis on the reference element τ_0, dimension d, and the parabolic problem, but are independent of the mesh geometry. In this case, a small angle in the two-dimensional case and a small dihedral angle in the three-dimensional case would make the term $\max_{t\in\tau}\mathcal{Q}_\tau$ extremely large, which would incur bad conditioning.

Other cases of a non-uniform mesh with irregular elements could also be discussed, following the results in Theorem 5.1. The estimates in Theorem 5.1 give an explicit relationship between the spectral condition number and the mesh geometry, which is very precise with respect to the latter.

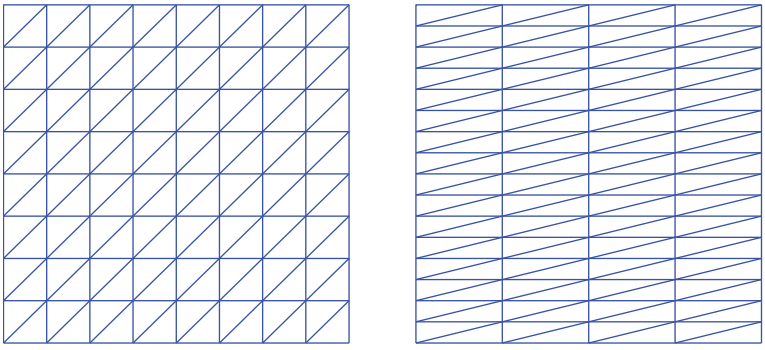
6. Numerical experiments

In this section, we present some numerical experiments by implementing both linear and quadratic Lagrange finite elements on a spatial domain $\Omega=[0,1]^2$. Test runs are performed on several sets of meshes, as shown in Figures 6.1 and 6.2, to substantiate the theoretical results obtained in previous sections. These meshes are artificially generated. While they are relatively simple, they serve to illustrate a number of key geometric features related to our analysis: effects due to uniform and non-uniform element areas, regular and irregular shapes, and their combinations. The time discretization is implemented by either the explicit scheme or the implicit scheme with a step size that is uniform in time. Different values of the time step size are taken in different test runs for comparison purposes.

For simplicity, we focus on the isotropic heat conduction equation

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+f,$$

with a homogeneous boundary condition $u(x,y,t)=0$ on $\partial\Omega$.



(a) (b)

FIGURE 6.1. Uniform triangular meshes with isosceles right triangles (a) and right triangles with small angles (b).

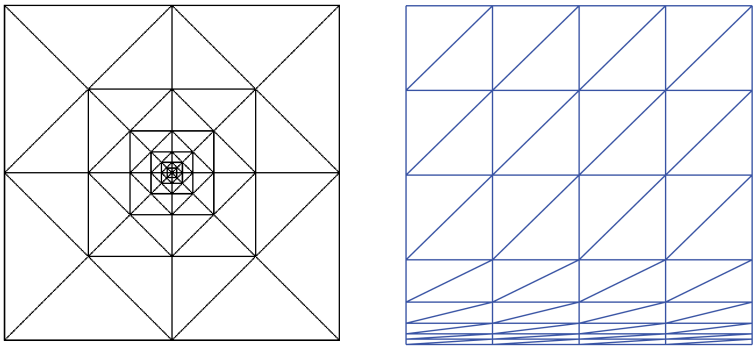


FIGURE 6.2. Adaptive triangular meshes with isosceles right triangles (a) and right triangles with small angles (b).

6.1. **Verifying the dependence of stability limit on mesh geometry.** In this subsection, we take the meshes shown in Figures 6.1 and 6.2 to examine the effects of the mesh geometry on the stability limit in the explicit scheme. Here we use Δt and Δt_l to denote the stability limits estimated by formulae (2.12) and (4.1), respectively. The values of Δt are computed by (2.12) with the help of eigenvalue routines in Matlab 7.0.

In order to get the analytic stability limit Δt_l, we need to calculate the term on the right-hand side of (4.1). With the definition of $\tilde{\gamma}_n^d$, a simple calculation gives that, for the linear element, $\tilde{\gamma}_3^2 = 6$, while for the quadratic element, $\tilde{\gamma}_6^2 \approx 209.3157$.

For the isosceles right triangle elements as shown in Figure 6.1(a), we denote the isosceles length by a so that

$$\frac{|\tau|}{\mathcal{Q}_d(\tau)} = \frac{|\tau|^2}{\sum_{i=1}^3 l_i^2} = \frac{(1/2a^2)^2}{4a^2} = \frac{a^2}{16}.$$

Then, by (4.1), we have

$$\Delta t_l < \frac{2}{6} \times \frac{a^2}{16} = \frac{a^2}{48}. \tag{6.1}$$

Similarly, for the quadratic element,

$$\Delta t_l < \frac{2}{\tilde{\gamma}_6^2} \times \frac{a^2}{16} \approx (5.971841\text{E-}4)a^2. \tag{6.2}$$

For the right triangle element with small angle as shown in Figure 6.1(b), with the linear element, simple calculations give

$$\frac{|\tau|}{\mathcal{Q}_d(\tau)} = \frac{|\tau|^2}{\sum_{i=1}^3 l_i^2} = \frac{s^2}{(8s/\sin(2\theta))} = \frac{s\sin(2\theta)}{8},$$

where s is the element area and θ is the smallest interior angle in the right triangle. Thus, by (4.1), we have

$$\Delta t_l < \frac{2}{6} \times \frac{s\sin(2\theta)}{8} = \frac{\sin(2\theta)s}{24}. \tag{6.3}$$

Similarly, for the quadratic element, we can get

$$\Delta t_l < \frac{2}{\tilde{\gamma}_6^2} \times \frac{\sin(2\theta)s}{8} \approx (1.194366\text{E-}3)\sin(2\theta)s. \tag{6.4}$$

Some uniform and shape regular meshes (of different sizes), as shown in Figure 6.1(a), are used first to examine the effect of element size on the stability limit. The stability limit Δt, Δt_l, the ratio of change $R(\Delta t) = \Delta t^h/\Delta t^{h/2}$ with respect to Δt, and the ratio of $\Delta t/\Delta t_l$ in different cases are summarized in Tables 6.1 and 6.2 for the linear and quadratic elements, respectively. Here, the $N \times N$ mesh means that the length of domain in every axis direction is divided uniformly into N parts (leading to $2N^2$ total triangle elements). From the results in Tables 6.1 and 6.2, we can see that the ratios of change $R(\Delta t)$ nearly remain a constant 4, which says that the stability limit becomes one fourth of the original one when the mesh size gets halved. This tells us that, for the explicit calculation on the uniform regular mesh, the stability limit is proportional to the square of the typical mesh size (or the area of the typical element), which confirms Corollary 4.1. Obviously, a smaller mesh size will result in a smaller stability limit in this case, which thus leads to more computations.

TABLE 6.1. The stability limit for the linear element on the uniform shape regular mesh

mesh	8×8	16×16	32×32	64×64
Δt	4.0608E-3	9.8604E-4	2.4473E-4	6.1072E-5
Δt_l	3.2552E-4	8.1378E-5	2.0345E-5	5.0863E-6
$R(\Delta t)$	-	4.12	4.03	4.01
$\Delta t/\Delta t_l$	12.47	12.12	12.02	12.01

In the second example, uniform shape irregular meshes, as shown in Figure 6.1(b), are taken to explore the effect of the element shape on the stability limit. In this case, the total number of elements is fixed at 2048, thus the area of each element remains to be a constant of 1/2048=4.8828125E-4. The element shape is changing along with the smallest interior angle θ. The stability limit Δt, Δt_l, the ratio $\Delta t/\Delta t_l$ on the different meshes with different θ are summarized in Tables 6.3 and 6.4 for the linear and quadratic elements, respectively. These results show that the

TABLE 6.2. The stability limit for the quadratic element on the uniform shape regular mesh

mesh	8×8	16×16	32×32	64×64
Δt	2.5058E-4	6.1138E-5	1.5188E-5	3.7908E-6
Δt_l	9.3312E-6	2.3327E-6	5.8318E-7	1.4579E-7
$R(\Delta t)$	-	4.10	4.03	4.01
$\Delta t/\Delta t_l$	26.85	26.21	26.04	26.00

TABLE 6.3. The stability limit for the linear element on the uniform shape irregular mesh

mesh	32×32	16×64	8×128	4×256
θ	0.250000π	0.077979π	0.019869π	0.004973π
Δt	2.4473E-4	1.1502E-4	3.0408E-5	7.6281E-6
Δt_l	2.0345E-5	9.5740E-6	2.5332E-6	6.3560E-7
$\Delta t/\Delta t_l$	12.03	12.01	12.00	12.00

TABLE 6.4. The stability limit for the quadratic element on the uniform shape irregular mesh

mesh	32×32	16×64	8×128	4×256
θ	0.250000π	0.077979π	0.019869π	0.004973π
Δt	1.5188E-5	6.7329E-6	1.7626E-6	4.5234E-7
Δt_l	5.8316E-7	2.7444E-7	7.2616E-8	1.8219E-8
$\Delta t/\Delta t_l$	26.04	24.53	24.27	24.83

stability limit decreases along with the smallest interior angle of element becomes small. In order to demonstrate the relationship more closely, in Figure 6.3 (left), we plot the curves of the stability limit with respect to $\sin(2\theta)$, respectively, for both linear and quadratic elements. The perfect linear behavior implies that the stability limit is proportional to $\sin(2\theta)$, which is consistent with the prediction of estimate (4.7) given in the previous section.

In the third example, we test a slightly more complicated adaptive (non-uniform) mesh with shape regular elements as shown in Figure 6.2(a). This may represent a typical situation in adaptive finite element computation. In such a mesh, all elements are similar isosceles right triangles. We denote the level of refinement by $2k$, and then the right edge length of the smallest element is $1/2^k$ and the smallest area is $2^{-(2k+1)}$; in this case, $\mathcal{Q}_d(\tau) = 8$. Thus, by (4.1), for the linear element, we have

$$\Delta t_l < \frac{2}{6} \times \frac{\min_{\tau \in \mathcal{T}} |\tau|}{8} = \frac{1}{24} \times (\frac{1}{2} \times \frac{1}{2^k} \times \frac{1}{2^k}) = \frac{2^{-(2k+1)}}{24},$$

and for the quadratic element,

$$\Delta t_l < \frac{2}{\tilde{\gamma}_6^2} \times \frac{\min_{\tau \in \mathcal{T}} |\tau|}{8} \approx (1.1944\text{E}-3) \times 2^{-(2k+1)}.$$

The stability limit Δt, Δt_l and the ratio $\Delta t/\Delta t_l$, are summarized in Tables 6.5 and 6.6 for both the linear and quadratic elements, respectively. The results in Tables 6.5 and 6.6 show the stability limit decreases along with the refinement

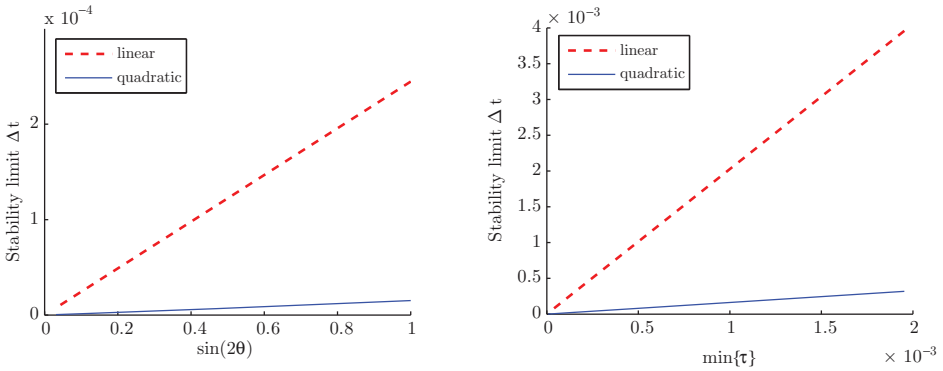


FIGURE 6.3. Plots against $\sin(2\theta)$ of the stability limit Δt for both the linear and quadratic elements on uniform shape irregular mesh (left); Plots against the smallest element area of the stability limit Δt for both the linear and quadratic elements on adaptive shape regular mesh (right).

level becoming large. In Figure 6.3 (right), we plot the curves of the stability limit Δt with respect to the smallest element area for both the linear and quadratic elements, respectively. The perfect linear behavior tells us the stability limit on such an adaptive mesh is decided by the area of the smallest element. These numerical results verify the theoretical results in Theorem 4.1.

TABLE 6.5. The stability limit for the linear element on the adaptive shape regular mesh

mesh (2k)	6	8	10	12	14
Δt	3.9583E-3	9.8957E-4	2.4739E-4	6.1848E-5	1.5462E-5
Δt_l	3.2552E-4	8.1384E-5	2.0345E-5	5.0862E-6	1.2716E-6
$\Delta t/\Delta t_l$	12.16	12.16	12.16	12.16	12.16

TABLE 6.6. The stability limit for the quadratic element on the adaptive shape regular mesh

mesh(2k)	6	8	10	12	14
Δt	3.1844E-4	7.9610E-5	1.9902E-5	4.9756E-6	1.2439E-6
Δt_l	9.3310E-6	2.3327E-6	5.8318E-7	1.4580E-7	3.6450E-8
$\Delta t/\Delta t_l$	34.13	34.13	34.13	34.13	34.13

In the fourth example, another non-uniform adaptive mesh with irregular element shapes, as shown in Figure 6.2(b), is used for the purpose of examining the dependence of the stability limit on the mesh geometry. In this triangulation, the elements in the upper-half domain are isosceles right triangles, while the elements in the lower-half domain are right triangles with a short right edge b of varying sizes and a fixed long right edge $a = 1/4$. The $4 \times N$ mesh means that the length of the domain in the y-axis direction is divided adaptively into N parts. Here the upper-half portion of domain in the y-axis direction are divided into 4 parts uniformly,

while its lower-half portion is divided into $N-4$ parts by the bisection method. And the shortest edge length of elements on the $4 \times N$ mesh is $1/2^N$. In this case, $|\tau|/\mathcal{Q}_d(\tau) = b^2/(8+8(b/a)^2)$, which is a monotonically increasing function of b. Thus, by (4.1),

$$\Delta t < \frac{2}{6} \times \frac{1}{8} \times \left(\frac{b_{\min}^2}{1+(b_{\min}/a)^2}\right) = \frac{b_{\min}^2}{24(1+(b_{\min}/a)^2)}.$$

Similarly, for the quadratic element case, we have

$$\Delta t < \frac{2}{\tilde{\gamma}_6^2} \times \frac{1}{8} \times \left(\frac{b_{\min}^2}{1+(b_{\min}/a)^2}\right) \approx \frac{b_{\min}^2 1.1944\text{E}-3}{1+(b_{\min}/a)^2}.$$

In Tables 6.7 and 6.8, the numerical stability limits Δt, Δt_l, the ratio of change $R(\Delta t) = \Delta t^N/\Delta t^{(N+2)}$ and the ratio $\Delta t/\Delta t_l$ for both the linear and quadratic elements are presented. We see that both the ratios of change in Tables 6.7 and 6.8 are nearly the same constant of 16. Since $b_{\min} = 1/2^N$, the above results indicate that the stability limit become one sixteenth of the original one when the shortest edge length $b_{\min}$ becomes one fourth of the original one. This implies that the stability limit is proportional to the square of the shortest edge length of element in this case. Thus, a smaller $b_{\min}$ will incur a smaller stability limit.

TABLE 6.7. The stability limit for the linear element on the adaptive shape irregular mesh

mesh	4× 8	4×10	4×12	4×14	4×16
Δt	1.9714E-4	1.2384E-5	7.7423E-7	4.8390E-8	3.0244E-9
Δt_l	1.0133E-5	6.3563E-7	3.9736E-8	2.4835E-9	1.5522E-10
$R(\Delta t)$	-	15.92	16.00	16.00	16.00
$\Delta t/\Delta t_l$	19.46	19.48	19.48	19.48	19.48

TABLE 6.8. The stability limit for the quadratic element on the adaptive shape irregular mesh

mesh	4× 8	4×10	4×12	4×14	4×16
Δt	9.5291E-6	5.9732E-7	3.7339E-8	2.3337E-9	1.4586E-10
Δt_l	2.9047E-7	1.8221E-8	1.1391E-9	7.1192E-11	4.4494E-12
$R(\Delta t)$	-	15.95	16.00	16.00	16.00
$\Delta t/\Delta t_l$	32.81	32.78	32.78	32.78	32.78

In the above four examples, the ratio of $\Delta t/\Delta t_l$ is also presented in Tables 6.1–6.8. We can see that the ratio of $\Delta t/\Delta t_l$ almost remains a constant for each case, which implies that the obtained bounds are sharp, with respect to the mesh geometry, if we ignore the constant factors that are independent of mesh geometry. Naturally, we see that, given the ratios being much larger than 1 in some cases, there is room for further improvement on estimating such mesh independent factors.

6.2. **Verifying the relation between the condition number and the mesh geometry.** In this subsection, the same four sets of meshes, as shown in Figures 6.1 and 6.2, are taken to evaluate the theoretically established relationship between the condition number of the resulting matrix and the mesh geometry in the implicit calculation (2.7). The results are summarized in the Tables 6.9–6.12.

TABLE 6.9. The condition number κ on the uniform shape regular mesh

(a) The linear element case with Δt_1=5.0E-6 and Δt_2=1.0E-1

mesh	8×8	16×16	32×32	64×64
$\kappa(M)$	1.0000	1.0000	1.0000	1.0000
$\kappa(K)$	25.2741	1.0309E+2	4.1435E+2	1.6594E+3
$\kappa(M+\Delta t_1 K)$	1.0024	1.0100	1.0408	1.1636
$\kappa(M+\Delta t_2 K)$	17.0419	68.6862	2.7528E+2	1.1017E+3
$\log(R(\Delta t_2))$	-	2.01	2.00	2.00

(b) The quadratic element case with Δt_1=5.0E-7 and Δt_2=1.0E-1

mesh	8×8	16×16	32×32	64×64
$\kappa(M)$	5.5446	5.7552	5.8237	5.8443
$\kappa(K)$	1.3724E+2	5.5224E+2	2.2123E+3	8.8525E+3
$\kappa(M+\Delta t_1 K)$	5.5258	5.6966	5.6406	5.2110
$\kappa(M+\Delta t_2 K)$	91.4343	3.6689E+2	1.4687E+3	5.8761E+3
$\log(R(\Delta t_2))$	-	2.00	2.00	1.99

TABLE 6.10. The condition number κ on the uniform shape irregular mesh

(a) The linear element case with Δt_1=1.0E-8 and Δt_2=1.0E-1

mesh	32×32	16×64	8×128	4×256
θ	0.25π	0.077979π	0.019869π	0.004973π
$\kappa(M)$	1.0000	1.0000	1.0000	1.0000
$\kappa(K)$	414.35	8.8241E+2	3.3536E+3	1.3626E+4
$\kappa(M+\Delta t_1 K)$	1.0001	1.0002	1.0006	1.0026
$\kappa(M+\Delta t_2 K)$	275.28	5.8570E+2	2.2214E+3	8.9665E+3

(b) The quadratic element case with Δt_1=1.0E-8 and Δt_2=1.0E-1

mesh	32×32	16×64	8×128	4×256
θ	0.25π	0.077979π	0.019869π	0.004973π
$\kappa(M)$	5.8237	5.7907	5.6770	5.2932
$\kappa(K)$	221.23	4.7039E+3	1.7804E+4	7.1325E+4
$\kappa(M+\Delta t_1 K)$	5.8199	5.7824	5.6257	5.1078
$\kappa(M+\Delta t_2 K)$	146.87	3.1231E+3	1.1827E+4	4.7508E+4

(c) The linear (κ_1) and quadratic (κ_2) element case with different Δt

mesh	32×32	16×64	8×128	4×256
θ_i	0.25π	0.077979π	0.019869π	0.004973π
Δt_i	1.0E-6	4.7059E-7	1.2452E-7	3.1241E-8
$\kappa_1(M+\Delta t_i K)$	1.0082	1.0082	1.0082	1.0082
$\kappa_2(M+\Delta t_i K)$	5.4769	5.4735	5.3177	4.7763

We present condition numbers of both the global stiffness/mass matrix and the coefficient matrix on various meshes for different values of the time step size Δt in Tables 6.9–6.12. These results show that, in general, when the time step size $\Delta t(=\Delta t_1)$ is small, no matter which particular element is being used (either the

TABLE 6.11. The condition number κ on adaptive shape regular mesh

(a) The linear element case with Δt_1=1.0E-6, and Δt_2=1.0E-1

mesh(k)	6	8	10	12	14
$\kappa(M)$	7.0000	28.0000	1.1200E+2	4.4800E+2	1.7920E+3
$\kappa(K)$	14.8891	28.5064	46.6756	69.3849	96.6304
$\kappa(M+\Delta t_1 K)$	6.9977	27.9586	1.1132E+2	4.3741E+2	1.6393E+3
$\kappa(M+\Delta t_2 K)$	10.3706	20.4829	35.0041	54.0979	77.7881

(b) The quadratic element case with Δt_1=1.0E-7, and Δt_2=1.0E-1

mesh(k)	6	8	10	12	14
$\kappa(M)$	31.5721	1.2630E+2	5.0519E+2	2.0208E+3	8.0832E+3
$\kappa(K)$	68.3373	1.3370E+2	2.2075E+2	3.2949E+2	4.5992E+2
$\kappa(M+\Delta t_1 K)$	31.5570	1.2605E+2	5.0125E+2	1.9593E+3	7.1939E+3
$\kappa(M+\Delta t_2 K)$	46.9560	95.8886	1.6584E+2	2.5751E+2	3.7106E+2

TABLE 6.12. The condition number κ on the adaptive shape irregular mesh

(a) The linear element case with Δt_1=1.0E-9, and Δt_2=1.0E-1

mesh	4×6	4×8	4×10	4×12	4×14
$\kappa(M)$	4.0000	16.0000	64.0000	2.5600E+2	1.0240E+3
$\kappa(K)$	12.2011	45.5946	1.8162E+2	7.2630E+2	2.9051E+3
$\kappa(M+\Delta t_1 K)$	4.0000	15.9999	63.9916	2.5546E+2	9.9130E+2
$\kappa(M+\Delta t_2 K)$	8.4323	31.4487	1.2530E+2	5.0110E+2	2.0044E+3

(b) The quadratic element case with Δt_1=1.0E-11, and Δt_2=1.0E-1

mesh	4×6	4×8	4×10	4×12	4×14
$\kappa(M)$	17.5393	70.1564	2.8063E+2	1.1225E+3	4.4900E+3
$\kappa(K)$	77.5100	3.0046E+2	1.2014E+3	4.8057E+3	1.9223E+4
$\kappa(M+\Delta t_1 K)$	17.5393	70.1563	2.8062E+2	1.1221E+3	4.4619E+3
$\kappa(M+\Delta t_2 K)$	53.0174	2.1207E+2	8.4958E+2	3.3988E+3	1.3595E+4

linear element or the quadratic element for example), the condition number $\kappa(M+\Delta t_1 K)$ is nearly the same as the condition number $\kappa(M)$. Meanwhile, the condition number $\kappa(M+\Delta t_2 K)$ is nearly the same as the condition numbers $\kappa(K)$ when the time step size $\Delta t(=\Delta t_2)$ is large. This is consistent with the theoretical prediction in Theorem 5.1.

Table 6.9(a) gives condition numbers, on the uniform shape-regular mesh, for the linear element case while Table 6.9(b) for the quadratic element case. In these two tables, the change ratios of the condition numbers with a small time step size case are also computed by $\log(R(\Delta t_2))$=$\log(\Delta t_2^h/\Delta t_2^{h/2})$. These results show that, when the time step size Δt is small (Δt_1=5.0E-6), the condition numbers $\kappa(M+\Delta t_1 K)$ are nearly the same as the condition numbers $\kappa(M)$, which is nearly a constant on both coarsened meshes and refined meshes. When the time step size is large (Δt_2=1.0E-1), condition numbers $\kappa(M+\Delta t_2 K)$ are getting to be more similar to condition numbers $\kappa(K)$ and the change ratio of the condition number is nearly 2, which implies that the condition number is proportional to the square of the

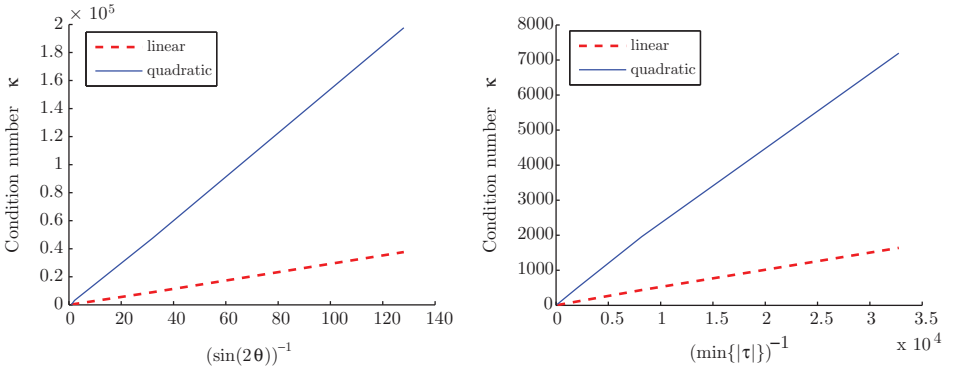


FIGURE 6.4. Plots against $\sin(2\theta)^{-1}$ of the condition number κ for both the linear and quadratic elements with Δt=1.0E-1 on uniform irregular mesh (left); Plots against the inverse of the smallest element area of the condition number κ for both the linear element with Δt=1.0E-6 and the quadratic element with Δt=1.0E-7 on nonuniform regular mesh (right).

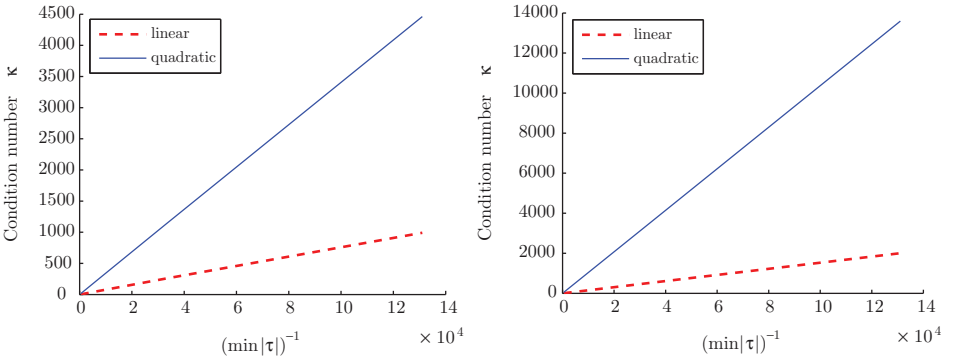


FIGURE 6.5. Plots against the inverse of the smallest element area of the condition number κ for both linear and quadratic elements on the adaptive regular mesh with the large time step size Δt=1.0E-1 (right) and the small time step size Δt=1.0E-9 for the linear element and Δt=1.0E-11 for the quadratic element (left), respectively.

particular edge length of element. This verifies the theoretical discussions in the previous section.

Table 6.10 presents the results on the uniform shape irregular mesh as shown in Figure 6.1(b). In this case, on the 4×256 mesh, the fixed area is $1/2048 \approx$4.8828E-4, and the smallest edge length of the element is $1/256 \approx$ 3.9063E-3. As in the shape regular case, when the time step size is small(Δt=5.0E-7), the condition number is mainly dependent on the ratio of the largest element area to the smallest element area, which is independent of the shape. Then, as shown in Table 6.10, the condition numbers stay nearly some constant because all elements in the mesh are the same, though their shapes are irregular. When the time step size is large (Δt=1.0E-1),

the element shape dominates the condition number as predicted in Theorem 5.1. That is, small angles incur larger condition numbers. More precisely, the condition numbers go inversely with $\sin(2\theta)$ as shown in Figure 6.4 (left). Furthermore, similar to the calculation in (6.3) and (6.4), we have the estimates on the condition number as

$$\kappa(M + \Delta t K) \leq \frac{C_1(1 + \Delta t(|\tau| \sin(2\theta))^{-1})}{1 + C_2 \Delta t}, \tag{6.5}$$

which says that the condition number should stay nearly as a constant when the time step size Δt changes proportionally with $\sin(2\theta)$. The results in Table 6.10(c) verify this prediction. The results in Tables 6.10(a) and 6.10(b) also confirm our estimates.

In Table 6.11, condition numbers on the non-uniform adaptive shape regular mesh Figure 6.2(a) are presented. Here, the mesh "k" represents the mesh on the k-th refinement level, with k being a numeric value. The larger k is, the finer the mesh becomes. On the $k = 14$ mesh, the smallest edge length of element is $1/2^7$=7.8125E-3, and the smallest element area is $\frac{1}{2} \times \frac{1}{2^7} \times \frac{1}{2^7} \approx$ 3.0518E-5. We can see that, when Δt is small, the condition number is large because the smallest element area becomes small but the largest element area does not change. Theorem 5.1 implies that the condition number in this case is decided by the ratio $\max_\tau |\tau| / \min_\tau |\tau|$. That is, the condition number goes inversely with the area of the smallest element. In Figure 6.4 (right), we plot, with respect to $(\min_\tau\{|\tau|\})^{-1}$, the curves of the condition number, respectively, for both the linear element with Δt =1.0E-6 and the quadratic element with Δt =1.0E-7. The perfect linear curves in Figure 6.4 (right) and the results in Table 6.11 confirm the theoretical prediction.

Table 6.12 shows condition numbers on another artificially constructed non-uniform mesh which is shape-irregular as shown in Figure 6.2(b). In this case, on the 4×14 mesh, the smallest edge length of element is $1/2^{14} \approx$ 6.1035E-5, and the smallest element area is $\frac{1}{2} \times \frac{1}{4} \times \frac{1}{2^{14}} \approx$ 7.6294E-6. Similar to the case of adaptive shape-regular case, when Δt is small ($\Delta t = \Delta t_2$), the condition number is large because the smallest element area becomes smaller while the largest element area does not change. Here, in Figure 6.5(left), we also plot with respect to $(\min\{|\tau|\})^{-1}$, the curves of the condition number, respectively, for both the linear element and the quadratic element with $\Delta t = \Delta t_2$. The linear curves in Figure 6.5 (left) show that the condition number goes inversely with the area of the smallest element when the time step size is small, which is again consistent with the prediction of Theorem 5.1.

When Δt is large, the exact condition number estimate on meshes shown in Figure 6.2 is related to the global stiffness matrix K estimate on these adaptive meshes (which will be more carefully analyzed in the future). Generally speaking, in these cases, the condition number increases as the mesh gets refined, as shown in Tables 6.11 and 6.12. In Figure 6.5 (right), we plot the curves of the condition number with respect to $(\min\{|\tau|\})^{-1}$, respectively, for both the linear element and quadratic element with Δt=1.0E-1. The linear curves in Figure 6.5 (right) show that the condition number goes inversely with the area of the smallest element. For these adaptive mesh, as shown in [10], in order to get the optimal condition number estimates for the stiffness matrix, it is advantageous to introduce a non-

uniform density function ρ in the eigenvalue problem (5.3). For example, if we take a density function ρ such that the triangulation $\mathcal{T}$ satisfies approximately that the integral of ρ over τ does not vary significantly over different $\tau \in \mathcal{T}$, then Theorem 5.1 implies that the condition number is dominated by the quantity $Ne\mathcal{Q}(\tau)$, where Ne is the total number of elements, when the time step size Δt is large. This can help to explain why the condition numbers $\kappa(M + \Delta t_2 K)$ in Table 6.11 have little change while the condition numbers $\kappa(M + \Delta t_2 K)$ in Table 6.12 goes inversely with the area of the smallest element. Indeed, this is because the mesh quantity $\mathcal{Q}(\tau)$ stays as a constant on such a non-uniform but shape-regular mesh while $\mathcal{Q}(\tau)$ changes significantly with the mesh refinement for the shape irregular case. We will pursue more careful studies of the related phenomena in the future.

Let us point out that, although our numerical experiments are on some particular meshes shown in Figures 6.1 and 6.2, the results are expected to hold for general simplicial meshes and general finite element spaces as predicted by the theory established here.

7. Conclusion and future work

In this paper, the effects of spatial simplicial meshes on the stability of fully discrete approximations of parabolic equations using general finite element discretization in space and explicit marching in time are explored systematically. More precise mesh dependent permissible time step size estimates are derived. Numerical examples show that the obtained bounds are sharp with respect to mesh geometry. Furthermore, for the implicit time integration scheme, mesh dependent condition number estimates of the resulting linear systems are given and effects of the mesh geometry on the condition number are characterized precisely. Much of our rigorously derived theory is new and the results are applicable to general parabolic equations, general finite element spaces and general simplicial meshes.

We note that exploring the precise characterization of the constants that appeared in various forms of the finite element analysis has been of much interest in the literature (see [15] for an illustration). Obviously, better understanding of the effect of geometry on the permissible time step size can lead to the more efficient time resolution of the problems while maintaining stability. Moreover, the explicit and precise mesh dependent bounds of the condition number for implicit time integration are helpful to not only develop better iterative solvers but also improve mesh generation and optimization strategies on which a better efficiency of solver can be reached. While we focus on simple forward and backward Euler schemes for the time integration, our framework is quite illustrative so that generalization to other high order time integrations schemes can also be made. Related issues for non-linear evolution equations may also be examined along with further studies of sharper estimates on the algebraic constants involved. We will pursue these and other interesting issues in future works.

Acknowledgments

The authors would like to thank the reviewers for their valuable comments and suggestions on the original version of the manuscript.

References

[1] D. Arnold, F. Brezzi and M. Fortin, A stable finite element for the Stokes equations, Calcolo, 12 (1984), 337-344. MR799997 (86m:65136)

[2] O. Axelsson and V. Barker, *Finite Element Solution of Boundary Value Problems*, Academic Press, London, 1983; reprinted as Classics Appl. Math. 35, SIAM, Philadelphia, 2001. MR1856818 (2002g:65001)

[3] R Bank and L. Scott, *On the conditioning of finite element equations with highly refined meshes*, SIAM J. Numer. Anal., 26(1989), 1383-1394. MR1025094 (90m:65192)

[4] M. Batdor, L. Freitag and C. Ollivier-Gooch, *Computational study of the effect of unstructured and mesh quality on solution efficiency*, AIAA, 13th CFD Conf, 1997.

[5] M. Berzins, *Mesh quality: a function of geometry, error estimates or both?* Engineering with Computers, 15 (1999), 236-247.

[6] S. Brenner and L. Scott, *The mathematical theory of finite element Methods*, 2nd edition, Springer-Verlag, New York, 2002. MR1894376 (2003a:65103)

[7] W. Cao, *On the error of linear interpolation and orientation, aspect ratio, and internal angles of a triangle*, SIAM J. Numer. Anal., 43 (2005), 19-40. MR2177954 (2006k:65023)

[8] W. Dörfler, *The conditioning of the stiffness matrix for certain elements approximating the incompressibility condition in fluid dynamics*, Numer. Math., 58(1990), 203-214. MR1069279 (91k:65142)

[9] Q. Du, Z. Huang and D. Wang, *Mesh and Solver Coadaptation in finite element mehods for anisotropic problems, Numerical Methods for Partial Differential Equations*, 21 (2005), 859-874. MR2140812 (2006f:65126a)

[10] 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), 1421-1444. MR2497335 (2010b:65252)

[11] A. Ern, J.L. Guermond, *Evaluation of the condition number in linear systems arising in finite element approximations*, ESAIM: M2AN, 40(2006), 29-48. MR2223503 (2007b:65119)

[12] L. Freitag and C. Ollivier-Gooch, *A cost/benefit analysis of simplicial mesh improvement techniques as measured by solution efficiency*, Int. J. Comp. Geo. Appl., 10 (2000), 361-382. MR1791193

[13] I. Fried, *Bounds on the spectral and maximum norms of the finite element stiffness, flexibility and mass matrices*, Int. J. Solids Structures, 9 (1973), 1013-1034. MR0345400 (49:10136)

[14] I. Fried, *Numerical solution of differential equations*, Academic Press, New York, 1979. MR526039 (80d:65001)

[15] I. Harari and T. Hughes, *What are C and h?: Inequalities for the analysis and design of finite element methods*, Computer Methods in Applied Mechanics and Engineering, 97 (1992), 157-192. MR1167711 (93g:65122)

[16] N. Hu, X.-Z. Guo, I. Katz, *Bounds for eigenvalues and condition numbers in the p-version of the finite element method*, 67(1998), 1423-1450. MR1484898 (99a:65149)

[17] W.Z. Huang and R. D. Russell, *Adaptive Moving Mesh Methods*, (Springer, New York). 2011. MR2722625 (2012a:65243)

[18] Z.-C. Li, H.-T. Huang, *Effective condtion number for the finite element method using local mesh refinements*, Applied Numerical Mathematics, 59(2009), 1779-1795. MR2532444 (2010g:65213)

[19] J. I. Lin, *Bounds on eigenvalues of finite element systems*, International Journal for numerical methods in engineering, 32(1991), 957-967. MR1128904 (92g:80001)

[20] J. I. Lin, *An element eigenvalue theorem and its application for stable time step sizes*, Computer Methods in Applied Mechanics and Engineering, 73 (1989), 283-294. MR1016643 (90k:73006)

[21] A. Ramage and A. Wathen, *On preconditioning for finite element equations on irregular grids*, SIAM J. Matrix Anal. Appl. 15 (1994), 909-921. MR1282702 (95d:65095)

[22] J. Shewchuk, *What is a good linear finite element? Interpolation, conditioning, anisotropy and quality measures*, 2003, Technical report, CS, UC Berkeley.

[23] G. Strang and G. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, NJ, 1973. MR0443377 (56:1747)
[24] V. Thomee, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin, 2nd edition, 2006. MR2249024 (2007b:65003)
[25] A. Wathen, *Realistic eigenvalue bounds for the Galerkin mass matrix*, IMA J. Numer. Anal., 7 (1987), 449-457. MR968517 (90a:65246)
[26] L. Zhu and Q. Du, *Mesh dependent stability for finite element approximations of parabolic problems with mass lumping*, Journal of Computational and Applied Math, 236(2011), 801-811. MR2853505
[27] O. Zienkiewicz, R. Taylor and J. Zhu, *The Finite Element Method, Its Basis and Fundamentals*, Sixth edition, 2005, Elsevier, Oxford.

LMIB and School of Mathematics and Systems Sciences, Beihang University, 100191, Beijing, People's Republic of China
E-mail address: liyongzhu@buaa.edu.cn

Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802
E-mail address: qdu@math.psu.edu