**The Pennsylvania State University**
**The Graduate School**

# ROBUST AND ACCURATE ALGORITHMS FOR SOLVING ANISOTROPIC SINGULARITIES

A Thesis in
Mathematics
by
Long Chen



Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2005

The thesis of Long Chen was reviewed and approved* by the following:

Jinchao Xu
Professor of Mathematics
Thesis Advisor, Chair of Committee

Qiang Du
Professor of Mathematics

Chun Liu
Associate Professor of Mathematics

James Z. Wang
Assistant Professor of Information Sciences and Technology

Ludmil T. Zikatanov
Assistant Professor of Mathematics

Dmitri Burago
Professor of Mathematics,
Director of Graduate Studies in Department of Mathematics

*Signatures are on file in the Graduate School.

# Abstract

This dissertation is focused on the construction of robust and accurate algorithms for mathematical models of physical phenomena that exhibit strong anisotropies, that is, when the quantities have very slow and smooth variations in some directions but have rapid variations in other directions.

Our first result is on the mathematically characterization of optimal or nearly optimal meshes for a general function which could be either isotropic or anisotropic. We give an interpolation error estimate for the continuous and piecewise linear nodal interpolation. Roughly speaking, a nearly optimal mesh is a quasi-uniform triangulation under some new metric defined by the Hessian matrix of the object function. We also prove the error estimate is optimal for strictly convex (or concave) functions.

Based on the interpolation error estimates, we introduce a new concept Optimal Delaunay Triangulation (ODT) and present practical algorithms to construct such nearly optimal meshes. By minimizing the interpolation error globally or locally, we obtain some new functionals for the moving mesh method and several new mesh smoothing schemes.

We then apply our mesh adaptation algorithms to the convection dominated convection-diffusion problems which present anisotropic singularities such as boundary layers. We develop a robust and accurate adaptive finite element method for convection dominated problems by the homotopy of the diffusion parameter.

We give an error analysis of a one dimensional convection dominated convection-diffusion problem that is discretized by the standard finite element method on layer-adapted grids. We find that it is not uniform stable with respect to the perturbation of grid points. We then design a special streamline diffusion finite element method and prove the uniform stability and optimality of our new method.

We also discuss some related concepts and problems on the optimal Delaunay triangulations.

# Table of Contents

**List of Figures** vi

**List of Tables** vii

**Acknowledgments** viii

**Chapter 1**
**Introduction** 1

**Chapter 2**
**Optimal Anisotropic Interpolation Error Estimates** 9
2.1 Optimal Delaunay triangulations . . . . . 11
2.2 Interpolation error estimate: lower bound . . . . . 18
2.3 Interpolation error estimate: upper bound . . . . . 24

**Chapter 3**
**Algorithms to Generate Nearly Optimal Triangulations** 30
3.1 Global algorithms: moving mesh methods . . . . . 32
3.1.1 New optimization problems . . . . . 32
3.1.2 On the optimization problems . . . . . 35
3.2 Local mesh optimizations . . . . . 40
3.2.1 Local refinement and coarsening . . . . . 40
3.2.2 Edge swapping . . . . . 41
3.2.3 Mesh smoothing . . . . . 42
3.3 Mesh smoothing based on optimal Delaunay triangulations . . . . . 43
3.3.1 Mesh quality . . . . . 44
3.3.2 Smoothers based on centroidal patch triangulations . . . . . 47
3.3.3 Smoothers based on optimal Delaunay triangulations . . . . 50

3.3.4 Numerical experiments . . . . . 55

**Chapter 4**
**Adaptive Finite Element Methods for Convection Dominated Problems** **61**
4.1 Standard finite element methods . . . . . 62
4.2 Streamline diffusion finite element methods . . . . . 66
4.3 Post-processing of the Hessian matrix . . . . . 71
4.4 Multilevel homptotic adaptive finite element methods . . . . . 73

**Chapter 5**
**Stability and Accuracy of a New Streamline Diffusion Finite Element Method for One Dimensional Convection Dominated Problems** **77**
5.1 Error analysis of the standard finite element method . . . . . 81
5.1.1 Basic error equation . . . . . 81
5.1.2 Smooth solutions . . . . . 86
5.1.3 Solutions with boundary layers . . . . . 90
5.2 Error analysis of a new streamline diffusion finite element method . 95
5.2.1 Uniform stability . . . . . 95
5.2.2 Uniform convergence . . . . . 99

**Appendix A**
**Related Problems on Optimal Delaunay triangulations** **103**
A.1 Centroidal Voronoi tessellations . . . . . 103
A.2 Sphere covering problem . . . . . 106
A.3 Optimal polytope approximation of convex bodies . . . . . 112

**Bibliography** **117**

# List of Figures


2.1 Empty sphere condition in two dimensions . . . . . 12
2.2 Projection of a lower convex hull . . . . . 14
2.3 Degenerate simplex . . . . . 16

3.1 Mesh distribution function in one dimension . . . . . 36
3.2 Edge-based refinement . . . . . 41
3.3 Coarsening . . . . . 41
3.4 Edge swapping . . . . . 41
3.5 Feasible region in a local patch . . . . . 42
3.6 Moving a grid point in its local patch . . . . . 51
3.7 Weighted average of circumcenters . . . . . 53
3.8 Moving a point to the boundary of an element . . . . . 55
3.9 Comparison of Laplacian smoother, CPT smoother and ODT smoother 56
3.10 Error comparison of Laplacian smoother, CPT smoother and ODT smoother . . . . . 57
3.11 Anisotropic meshes obtained by the ODT smoother . . . . . 59
3.12 Interpolation error of the second example . . . . . 59
3.13 Anisotropic meshes . . . . . 60

4.1 Uniform mesh and its perturbation . . . . . 65
4.2 Bakhvalov mesh . . . . . 67
4.3 Perturbed Bakhvalov mesh . . . . . 68
4.4 Continuation adaptive meshes and corresponding solutions . . . . . 75
4.5 Domain of the fluid well problem . . . . . 75
4.6 Solution and the underlying mesh of the fluid well problem . . . . . 76

A.1 Optimal sphere covering in two dimensions. . . . . 107

# List of Tables

3.1 Computational cost comparison of Laplacian smoother, CPT smoother and ODT smoother . . . . . 57

4.1 Errors of standard FEM on uniform meshes: odd unknowns . . . . 64
4.2 Errors of standard FEM on uniform meshes: even unknowns . . . . 64
4.3 Errors of standard FEM on a special perturbed mesh . . . . . 64
4.4 Errors of standard FEM on Bakhvalov mesh . . . . . 67
4.5 Errors of standard FEM on a perturbed Bakhvalov mesh . . . . . 67
4.6 Errors of SDFEM on uniform meshes: odd unknowns . . . . . 68
4.7 Errors of SDFEM on uniform meshes: even unknowns . . . . . 69
4.8 Errors of SDFEM on a special perturbed mesh . . . . . 69

# Acknowledgments

First and foremost, my greatest thanks go to my supervisor Dr. Jinchao Xu who save my life in mathematics. His guidance made this thesis a reality. I really appreciate his encouragements and deep insights.

I am grateful and indebted to Dr. Ludmil T. Zikatanov and Dr. Qiang Du for inspiration and enlightening discussions on a wide variety of topics. I also wish to thank Dr. Chun Liu and Dr. James Z. Wang for providing valuable comments and suggestions that improved the quality of the thesis. I would also like to thank Dr. Pengtao Sun for his kindness and help during my graduate career.

Finally, I must express my appreciation to my wife, Juan Du. She has been a source of constant support throughout the many long days spent working on this dissertation.

# Pippo

To my family.

Chapter 1

# Introduction

For almost all mathematical models in science and engineering, the exact solution to the underlying equations cannot be written explicitly as analytical formula. Thus, performing any quantitative analysis requires first discretizing the original model and then analyzing the approximate (or discrete) solution.

This dissertation is on the theoretical analysis and the design of robust and accurate algorithms for mathematical models of physical phenomena that exhibit strong anisotropies, that is, when the quantities have very slow and smooth variation in some directions but have rapid variations in other directions.

Many physical phenomena exhibit strong anisotropic behaviors, such as: boundary layer flows in porous media, currents and concentrations in fuel cells, characteristics of semiconductors, stresses and strains in thin plates, shells and anisotropic materials. When an anisotropic phenomenon occurs and an accurate approximation is required, it is natural to use anisotropic meshes stretched according to the solution. That reduces the number of elements needed to partition the domain, better captures the solution behavior, and thus leads to much more efficient algorithms.

The discretization methods that are most commonly used in practice include finite element methods, finite difference methods, finite volume methods, and spectral methods. This work is mainly devoted to the finite element method (FEM) although the techniques can also be in principle applied to other discretizations.

The finite element method firstly uses a family of meshes (also indicated by triangulations, meshes or grids) to construct finite dimensional spaces and then

restrict a weak form of the original equation to those finite dimensional spaces to get a discrete approximation. Two important basic questions arise for such a discretization: (1) How well does the discrete problem approximate the continuous one? (2) Is the method stable? In another word, whether or not a small perturbation of the input data leads to a small change in the solution. The answer to question (1) is related to the approximation property of the finite element space, which crucially depends on the underlying triangulation of the domain. In the classic discretization theory for the FEM which is mostly designed for isotropic problems, a positive answer of (1) usually leads to a positive answer to (2). Namely the FEM will be stable if the mesh is adapted such that the corresponding finite element space provides good approximation property. There are some numerical evidences confirming that it could also be true in the anisotropic case [8, 25, 66, 64, 173, 151, 145, 119, 7, 36, 65, 101, 28, 29, 182]. But the theoretical results, which justify these numerical observations and further give theoretical directions to improve the efficiency of algorithms, are much weaker than for those for isotropic elements.

We will study in depth these two questions, give answers, and develop corresponding algorithms in the anisotropic case.

# Optimal Anisotropic Interpolation Error Estimates

Our first attempt is to mathematically characterize optimal interpolations for a general function. Given a function $u \in \mathcal{C}^2(\bar{\Omega})$ where $\Omega$ is a domain in $\mathbb{R}^n$, we consider the following optimization problem

$$\inf_{T \in \mathcal{P}_N} \|u - u_I\|_{L^p(\Omega)}, \quad 1 \leq p \leq \infty, \tag{1.1}$$

where $u_I$ is the continuous and piecewise linear nodal interpolant of the function $u$ based on the triangulation $\mathcal{T}$ and $\mathcal{P}_N$ denotes the set of all triangulations of $\Omega$ with at most $N$ elements.

This problem is a typical $N$-term approximation problem studied in the nonlinear approximation theory [61]. It can be traced back to de Boor [59, 60] where the equidistribution principle was first introduced and this concept has

been widely used in the literature. In two dimensions optimal meshes for the discontinuous piecewise linear approximation in the sense of minimizing error in $L^2$ norm was studied by Nadler [149]. The $L^\infty$ case was studied by D'Azevedo and Simpson [57, 58]. An optimal mesh obtained in those work is that each triangle under the Hessian metric is equilateral and the error is equidistributed on each triangle. Recently Agouzal, Lipnikov and Vassilevski [3, 2, 137, 136, 138] show the optimality of the anisotropic $L_\infty$ error estimates. We also need to mention that in the studying of the moving mesh method, Huang [104] and Huang and Sun [107] obtained anisotropic interpolation error estimates in $L^2$ and $H^1$ norms. There are other anisotropic interpolation error estimates in the literature [8, 9, 10, 7, 87, 88, 125, 126, 119, 120, 122, 123, 121, 124, 145, 151, 152].

We first study a simplified version of (1.1). We choose a special function $u(\mathbf{x}) = \|\mathbf{x}\|^2$ and assume vertices of triangulations are fixed. We show an optimal triangulation in this simple case is an Delaunay triangulation. As a generalization we introduce the concept Optimal Delaunay Triangulations (ODTs) which are optimizers of (1.1) and prove the existence of ODTs for convex functions [48].

We obtain an asymptotical lower bound for the nonlinear problem (1.1). More specifically, for strictly convex (or concave) functions $u \in C^2(\bar{\Omega})$, we have:

$$\liminf_{N\to\infty} N^{\frac{2}{n}} \|u - u_I\|_{L^p(\Omega),N} \geq C_{n,p} \| \sqrt[n]{|\det \nabla^2 u|} \|_{L^{\frac{pn}{2p+n}}(\Omega)}. \tag{1.2}$$

The equality holds if and only if the triangulation is asymptotic equilateral under the metric

$$H_p = (\det \nabla^2 u)^{-\frac{1}{2p+n}} \nabla^2 u.$$

For the numerical solutions of PDEs, we are more interested in the upper bound. For general functions $u \in \mathcal{C}^2(\bar{\Omega})$ we have:

$$\|u - u_I\|_{L^p(\Omega)} \leq C N^{-2/n} \| \sqrt[n]{\det H} \|_{L^{\frac{pn}{2p+n}}(\Omega)} \tag{1.3}$$

under suitable assumptions of the underlying mesh. Since in general $\nabla^2 u$ is not positive definite, we use a majorant $H$ of the Hessian matrix. (See Chapter 1 for details). One important requirement of the mesh is that it is quasi-uniform under the new metric $H_p$ [46].

# Algorithms to Generate Nearly Optimal Triangulations

With these sharp error estimates, we thought the mesh quality is a function depended concept and define an error-based mesh quality

$$Q(\mathcal{T}, u, p) = \|u - u_{I,\mathcal{T}}\|_{L^p,\Omega}.$$

By improving the error-based mesh quality or equivalently reducing the interpolation error, we develop mesh adaptation techniques which can be applied to the anisotropic case also. We will study both global and local algorithms.

An example of global algorithms is the so-called moving mesh method in the literature [80, 184, 37, 156, 37, 104, 107, 141, 174, 22, 140, 79, 63, 132, 131]. This type of methods can be described by two different sets of domains, $\Omega_c$ and $\Omega$, and corresponding grids, $\mathcal{T}_{N,c}$ and $\mathcal{T}_N$. The adapted mesh $\mathcal{T}_N$ can be viewed as a result of moving the grid $\mathcal{T}_{N,c}$ through the transformation $\mathbf{x} = \mathbf{x}(\xi) : \Omega_c \mapsto \Omega$, which can be obtained by minimizing the following functionals:

$$\min_{\mathbf{x}} \int_{\Omega_c} \Big[ \sum_i (\nabla \mathbf{x}_i)^t G(\mathbf{x}) \nabla \mathbf{x}_i \Big]^q (\xi) d\xi, \quad q > n/2,$$

or

$$\min_{\xi} \int_{\Omega} (\det G)^{1/2} \Big[ \sum_i (\nabla \xi_i)^t G^{-1}(\xi) \nabla \xi_i \Big]^q (\mathbf{x}) d\mathbf{x}, \quad q > n/2.$$

We emphasis that the case that $q = 1$ corresponds to the harmonic mapping but we require $q > n/2$ here. When $n \geq 3$, these minimization problems (which are more or less $p$-Laplacians with $p > n$) is significantly different from the harmonic mapping which has been most commonly used in the literature for the moving mesh method [80, 37, 106, 131]. The metric $G$ used in the above functionals is called monitor function. According to our interpolation error estimate (1.2) and (1.3), $G = (\det H)^{-1/(2p+n)} H$ will be a good choice [46].

If we only move each vertex $\mathbf{x}_i$ in its local patch $\Omega_i$ which consists of all simplices containing $\mathbf{x}_i$ , it is called mesh smoothing [4, 6, 18, 42, 92, 91, 189]. By minimizing the interpolation error in the local patch we derive several, old and new, mesh smoothing schemes in a unified way [42]. Among them, the following result is of

special interest. We prove that if the triangulation $\mathcal{T}$ is optimal in the sense of minimizing $Q(\mathcal{T}, u, 1)$ for a convex function $u \in C^1(\bar{\Omega})$, then for an interior vertex $\mathbf{x}_i$

$$\nabla u(\mathbf{x}_i) = -\frac{1}{|\Omega_i|} \sum_{\tau_j \in \Omega_i} \Big( \nabla |\tau_j|(\mathbf{x}) \sum_{\mathbf{x}_k \in \tau_j, \mathbf{x}_k \neq \mathbf{x}_i} u(\mathbf{x}_k) \Big). \tag{1.4}$$

This means that we can recover the derivative exactly from the nodal values of the function if the triangulation is optimized. If the triangulation is not optimized, we can use the identity (1.4) to compute the new point for mesh smoothing schemes.

For the isotropic function $u(\mathbf{x}) = \|\mathbf{x}\|^2$ which corresponds to the Euclidean metric, Alliez et al. [5] simplify the formula (1.4) to be a weighted average of circumcenters:

$$\mathbf{x}_i = \frac{\sum_{\tau_j \in \Omega_i} |\tau_j| \mathbf{c}_j}{|\Omega_i|},$$

where $\mathbf{c}_j$ is the circumcenter of simplex $\tau_j$.

# Adaptive Finite Element Methods for Convection Dominated Problems

The anisotropic mesh adaptation techniques are successfully applied to convection dominated convection-diffusion equations which contain some core difficulties from more complex models such as Navier-Stokes equations. To be specific, we consider the convection-diffusion equation

$$-\varepsilon \Delta u + \mathbf{b} \cdot \nabla u = f, \tag{1.5}$$

with a proper boundary condition on a bounded domain $\Omega \subset \mathbb{R}^2$. We are interested in the convection-dominated case, namely $\varepsilon$ is sufficiently smaller than $\|\mathbf{b}\|_\infty$.

Due to the small diffusion, the solution to (1.5) presents boundary layers or interior layers. It is well known that the standard FEM approximation on a quasi-uniform grid will yield nonphysical oscillations unless the mesh size compares with $\varepsilon$ [163, 148, 146]. To obtain a reliable numerical approximation, one approach is to use mesh adaptation to capture the layers. Although the mesh adaptation will improve the stability of the finite element method, we find that the accuracy of

the scheme depends on the uniformity of the grid in regions where the solution is smooth. Namely when the grid is only quasi-uniform in the smooth part, in general, we can only expect that a first order convergence. In order to achieve the optimal convergence rate and stability simultaneously, we will combine stabilized methods such as streamline diffusion finite element method (SDFEM) with the mesh adaptation especially the anisotropic mesh adaptation to the layers.

In order to apply the mesh adaptation algorithms, we use post-processing technique developed recently by Bank and Xu [20, 19] to get the Hessian matrix of the unknown solution $u$. Another computational difficulty is the small diffusion parameter $\varepsilon$. If the initial mesh is too coarse compared with $\varepsilon$, the numerical solution would not be able to capture the layers and thus the recovered Hessian is not so accurate. To overcome this difficulty, we will use a homotopic method with respect to the diffusion parameter $\varepsilon$. Namely, we first start our computation for large $\varepsilon$, say $\varepsilon = 1$ and use adaptive grid techniques to obtain a good initial grid. We then start to decrease the value of $\varepsilon$ and use the current grid as an initial grid to obtain a new adaptive grid. We continue in this way until the desired value of $\varepsilon$ is reached. We show the success of this approach by numerical examples [45].

# Stability and Accuracy of a New Streamline Diffusion Finite Element Method for One Dimensional Convection-Dominated Problems

We analyze a specially designed SDFEM for a one dimensional linear convection-dominated convection-diffusion problem:

$$-\varepsilon u'' - bu' = f \ \text{ in } (0,1), \tag{1.6}$$

$$u(0) = u(1) = 0, \tag{1.7}$$

where the positive diffusion constant $\varepsilon$ satisfies $0 < \varepsilon \ll b$. With a special choice of the stabilization bubble function in the SDFEM [49, 50], we are able to prove that our numerical approximation $u_N$ based on a grid $\mathcal{T}_N$ with $N$ elements is nearly

optimal (or so-called quasi-optimal) in the maximum norm, namely

$$\|u - u_N\|_\infty \leq C \inf_{v_N \in V^N} \|u - v_N\|_\infty, \tag{1.8}$$

where $V^N$ is the linear finite element space based on $\mathcal{T}_N$ with appropriate boundary conditions. Here $C$ is a constant that is independent on both $\varepsilon$ and $N$. The estimate (1.8) is the most desirable estimate we may expect to obtain for problem (1.6)-(1.7).

Thanks to (1.8) the convergence of the new SDFEM becomes an approximation problem we studied in the Chapter 1. If, for example, the function $|u''|^{1/2}$ is monotone, there exists a grid such that

$$\|u - u_I\|_\infty \leq C\|u''\|_{1/2} N^{-2}, \tag{1.9}$$

and thus by (1.8)

$$\|u - u_N\|_\infty \leq C\|u - u_I\|_\infty \leq C\|u''\|_{1/2} N^{-2}, \tag{1.10}$$

where $\|u''\|_{1/2} := (\int_0^1 |u''|^{1/2} dx)^2$. Note that $\|u''\|_{1/2}$ is $\varepsilon$-uniformly bounded in many cases, the convergence (1.10) is indeed $\varepsilon$-uniform.

We can construct such an optimal grid by the moving mesh method which has a simple formula in one dimension. We choose grid $\{0 = x_0 < ... < x_{N+1} = 1\}$ such that

$$\int_{x_i}^{x_{i+1}} M(x)dx = \text{constant}, \quad i = 0, 1, 2, ..., N,$$

where $M$ is a monitor function. In the literature, people often choose the arc-length function $M = \sqrt{1 + |u'|^2}$ or its discrete analogue which results only a first order uniform approximation [154, 117, 51, 52]. The optimal choice of the monitor function for a second order uniform convergent scheme remains open. Based on our analysis $M = |u''|^{1/2}$ is evidently a monitor function that leads to the optimal rate of convergence.

# Related Problems on Optimal Delaunay Triangulations

We will discuss several related problems on optimal Delaunay triangulations in the appendix. We first give a dual concept of ODT, an optimal Voronoi tessellation which is known as Centroidial Voronoi Tessellations (CVTs) [70]. Then we apply two special ODTs which minimize $Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty)$ and $Q(\mathcal{T}, \|\mathbf{x}\|^2, 1)$ to the sphere covering problem and the optimal polytope approximation of convex bodies, respectively [43].

Based on this approach, we give a new analysis of those problems and get a new proof of Coxeter-Few-Rogers lower bound $\tau_n$ for the thickness $\theta_n$ in the sphere covering problem. More importantly, we get a new estimate of the constant $\mathrm{del_n}$ in the optimal polytopes approximation to the convex bodies. Namely

$$\frac{n+1}{n+2}\left(\frac{\tau_n}{|B_n|}\right)^{2/n} \leq \mathrm{del_n} \leq \frac{n+1}{n+2}\left(\frac{\theta_n}{|B_n|}\right)^{2/n},$$

where $|B_n|$ is the volume of the unit ball in $\mathbb{R}^n$. Our estimate is asymptotic exact when dimension $n$ goes to infinity, that is

$$\lim_{n\to\infty} \frac{\mathrm{del}_n}{n} = \frac{1}{2\pi e}.$$

The layout of the rest of this dissertation is as follows. In Chapter 2 we give an interpolation error estimate for nodal interpolation and prove its optimality. Based on the interpolation estimates, in Chapter 3, we present practical algorithms to construct nearly optimal meshes. We then apply our mesh adaptation algorithms to the convection dominated problems in Chapter 4. In Chapter 5, we give an error analysis of a one dimensional convection dominated convection-diffusion problem for a special designed SDFEM and prove the uniform stability and convergence of our new method. Appendix is the theoretical complement of Chapter 1. We will study several problems related to optimal Delauany triangulations.

Chapter 2

# Optimal Anisotropic Interpolation Error Estimates

The basic problem we shall study in this chapter is:

$$\inf_{\mathcal{T}\in\mathcal{P}_N} \|u - u_{I,\mathcal{T}}\|_{L^p(\Omega)}, \quad 1 \le p \le \infty, \tag{2.1}$$

where $\Omega$ is a domain in $\mathbb{R}^n$, $\mathcal{P}_N$ denotes all the triangulations of $\Omega$ with at most $N$ elements or $N$ vertices, $u \in \mathcal{C}(\bar{\Omega})$ and $u_{I,\mathcal{T}}$ is the continuous and piecewise nodal linear interpolant of $u$ based on the triangulation $\mathcal{T}$. It is a typical $N$-term approximation problem studied in the nonlinear approximation theory [61].

In this chapter, we shall explore this problem from theoretical aspects. The long term theoretical goal is to characterize the approximation property by the regularity of the object function $u$. It has been done in one dimension [62] and in multi-dimensions for isotropic triangulations [26]. We will present some results without isotropic restriction on the triangulations.

To simplify the problem, we first choose a special function $u(\mathbf{x}) = \|\mathbf{x}\|^2$ and assume vertices of triangulations are fixed. We show optimal triangulations in this simple case are Delaunay triangulations. As a generalization we introduce the concept Optimal Delaunay Triangulations (ODTs) which are optimizers of (2.1) without fixing the vertices. We shall prove the existence of ODTs for convex functions.

Then we present an asymptotic lower bound for strictly convex (or concave )

functions $u \in C^2(\bar{\Omega})$:

$$\liminf_{N\to\infty} N^{\frac{2}{n}} \|u - u_{I,\mathcal{T}_N}\|_{L^p(\Omega)} \geq C_{n,p} \| \sqrt[n]{|\det \nabla^2 u|} \|_{L^{\frac{pn}{2p+n}}(\Omega)}. \tag{2.2}$$

The equality holds if and only if the triangulation is asymptotic equilateral under the metric

$$H_p = (\det \nabla^2 u)^{-\frac{1}{2p+n}} \nabla^2 u.$$

For the numerical solutions of partial differential equations (PDEs), we are more interested in the upper bound. For general functions $u \in \mathcal{C}^2(\bar{\Omega})$, let $H$ be a majorant of the Hessian matrix of $u$, we have:

$$\|u - u_I\|_{L^p(\Omega)} \leq C N^{-2/n} \| \sqrt[n]{\det H} \|_{L^{\frac{pn}{2p+n}}(\Omega)} \tag{2.3}$$

under the following two assumptions about the triangulation:

1. there is no oscillation of $H$ in each element;
2. the mesh to be quasi-uniform under the new metric $(\det H)^{-1/(2p+n)} H$.

The problem (2.1) can be traced back to de Boor [59, 60] where a problem of the best approximation by free knots splines was studied in one spatial dimension. In this work, the equidistribution principle was introduced and this concept has been widely used by other researchers studying grid adaptation. In two dimensions the optimal triangulations for the discontinuous piecewise linear approximation in the sense of minimizing error in $L^2$ norm was studied by Nadler [149]. The $L^\infty$ case was studied by D'Azevedo and Simpson [57, 58]. An optimal mesh obtained in those work is that each triangle under the Hessian metric is equilateral and the error is equidistributed on each triangle. Thereafter anisotropic mesh adaptation which aims to generate equilateral triangles under the metric induced by Hessian matrix was developed in [27, 101, 64] and successfully applied to the computational fluid dynamic problems in two spatial dimensions [101, 65]. Recently Agouzal, Lipnikov and Vassilevski [3, 2, 137, 136, 138] show the optimality of the anisotropic $L_\infty$ error estimates. We also need to mention that in the studying of the moving mesh method, Huang [104] and Huang and Sun [107], by formulating the conditions in terms of the isotropy and the equidistribution, the authors obtained interpolation

error estimates in $L^2$ and $H^1$ norms.

There are other anisotropic interpolation error estimates in the literature [8, 9, 10, 7, 87, 88, 125, 126, 119, 120, 122, 123, 121, 124, 145, 151, 152]. Apel [7] obtained some estimates under a condition on the coordinate orientation and on the maximal allowable mesh angle. Formaggia and Perotto [87] exploited the spectral properties of the affine map from the reference triangle to the general triangle to get anisotropic estimates for the $L^2$ and $H^1$ interpolation error on linear finite elements in two dimensions. Kunert [119] introduced the matching function to measure the alignment of an anisotropic function and an anisotropic mesh and presented error estimates using the matching function. Yet the overall optimal convergent rate in terms of the number of degree of freedom is not easy to get from those approaches.

## 2.1 Optimal Delaunay triangulations

In this section, we first study a simplified version of (2.1). We choose a special function $u(\mathbf{x}) = \|\mathbf{x}\|^2$ and assume vertices of triangulations are fixed. We shall show an optimal triangulation in this simple case is an Delaunay triangulation. As a generalization we introduce the concept Optimal Delaunay Triangulations (ODTs) which are optimizers of (2.1) and prove its existence for convex functions.

### Delaunay Triangulation

Delaunay triangulation (DT) is the most commonly used unstructured triangulation in many applications. It is often defined as the dual of Voronoi diagram [81]. We will use an equivalent definition [23, 89] which only involves the triangulation itself. The duality between Delaunay triangulations and Voronoi diagrams will be discussed in the Appendex.

**Definition 2.1.** *Let $V$ be a finite set of points in $\mathbb{R}^n$, the* convex hull *of $V$ is the smallest convex set which contains these points and is denoted by $CH(V)$. A* Delaunay triangulation *of $V$ is a triangulation of $CH(V)$ so that it satisfies* empty sphere condition *for any simplex: there are no points in $V$ inside the circumsphere of any simplex. See Figure 2.1 for an illustration in two dimensions.*

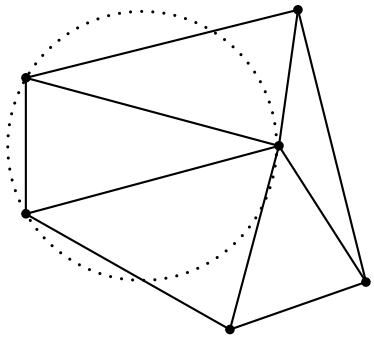


**Figure 2.1.** Empty sphere condition in two dimensions

There are many characterizations for Delaunay triangulations in two dimensions. Lawson [128] and Sibson [172] observed that Delaunay triangulations maximize the minimum angle of any triangle. Lambert [127] showed that Delaunay triangulations maximize the arithmetic mean of the radius of inscribed circles of the triangles. Rippa [157] showed that Delaunay triangulations minimize the integral of the squared gradients.

We shall characterize Delaunay triangulations from a function approximation point of view. Let us begin with the following definition.

**Definition 2.2.** *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain, $\mathcal{T}$ a triangulation of $\Omega$, and $u_{I,\mathcal{T}}$ be the piecewise linear and globaly continuous interpolation of a given function $u \in \mathcal{C}(\bar{\Omega})$ based on the triangulation $\mathcal{T}$. We define an* error-based mesh quality *$Q(\mathcal{T}, u, p)$ as*

$$Q(\mathcal{T}, u, p) = \|u - u_{I,\mathcal{T}}\|_{L^p(\Omega)} = \left( \int_\Omega |u(\mathbf{x}) - u_{I,\mathcal{T}}(\mathbf{x})|^p d\mathbf{x} \right)^{1/p}.$$

By choosing a special function $u(\mathbf{x}) = \|\mathbf{x}\|^2$, we can characterize the Delaunay triangulation as an optimal triangulation which achieve the best error-based mesh quality.

**Theorem 2.3.** *For a finite point set $V$, we let $\Omega = CH(V)$ and denote $\mathcal{P}_V$ all possible triangulations of $\Omega$ by using the points in $V$. We have*

$$Q(DT, \|\mathbf{x}\|^2, p) = \min_{\mathcal{T} \in \mathcal{P}_V} Q(\mathcal{T}, \|\mathbf{x}\|^2, p), \quad \forall\, 1 \le p \le \infty.$$

This type of result was first proved in $\mathbb{R}^2$ by D'Azevedo and Simpson [58] for $p = \infty$ and Rippa [158] for $1 \le p < \infty$. Rajan proved the case $p = \infty$ in multiple dimensions [155]. Theorem 2.3 is a generalization of their work to $L^p$ norm in multiple dimensions.

Actually the linear interpolant $u_{I,DT}$ based on an Delaunay triangulation is pointwise optimal.

**Lemma 2.4.** *Let $u(\mathbf{x}) = \|\mathbf{x}\|^2$ and $\mathcal{P}_V$ the set defined in Theorem 2.3, we have*

$$u_{I,DT}(\mathbf{x}) \le u_{I,\mathcal{T}}(\mathbf{x}), \quad \textit{for any } \mathcal{T} \in \mathcal{P}_V.$$

*Proof.* Let us introduce some notation first. We will identify $\mathbb{R}^{n+1}$ as $\mathbb{R}^n \times \mathbb{R}$ and write a point in $\mathbb{R}^{n+1}$ as $(\mathbf{x}, x_{n+1})$, where $\mathbf{x} \in \mathbb{R}^n$ and $x_{n+1} \in \mathbb{R}$. For a point $\mathbf{x}$ in $\mathbb{R}^n$, we can lift it to the paraboloid $(\mathbf{x}, \|\mathbf{x}\|^2)$ living in $\mathbb{R}^{n+1}$ and denote this lifting operation by $'$, namely $\mathbf{x}' = (\mathbf{x}, \|\mathbf{x}\|^2)$.

Let $CS_\tau$ be the circumsphere of $\tau$ and let $\mathbf{x}_o$, $R_\tau$ be the its center and radius respectively. For a simplex $\tau$, we consider the hyperplane in $\mathbb{R}^{n+1}$ given by $F_\tau = 0$ where

$$F_\tau := x_{n+1} - 2\mathbf{x} \cdot \mathbf{x}_o + \|\mathbf{x}_o\|^2 - R_\tau^2.$$

Noting that for a lifting point $\mathbf{x}'$, $x_{n+1} = \|\mathbf{x}\|^2$ and thus

$$F_\tau(\mathbf{x}') = \|\mathbf{x} - \mathbf{x}_o\|^2 - R_\tau^2. \tag{2.4}$$

Based on (2.4) it is easy to check that: the lifting of vertices of $\tau$ lie on $F_\tau$ and they form an $n$-simplex $\tau'$, and a point $\mathbf{x}$ in $\mathbb{R}^n$ is out of $CS_\tau$ if and only if its lifting point $\mathbf{x}'$ in $\mathbb{R}^{n+1}$ lies above the hyperplane $F_\tau = 0$. See Figure 2.2 for an illustration in two dimensions.

Let us fix a triangulation $\mathcal{T} \in \mathcal{P}_V$. For any $\mathbf{x} \in \Omega$, let $\tau_1, \tau_2$ be the simplices containing $\mathbf{x}$ in DT and $\mathcal{T}$ respectively. By our definition of DT, vertices of $\tau_2$ cannot be enclosed by the circumsphere of $\tau_1$. It means $\tau_2'$ lie above $\tau_1'$. Note that

$$(\tau_1, u_{I,DT}(\tau_1)) = \tau_1', \quad (\tau_2, u_{I,T}(\tau_2)) = \tau_2'$$

and $u(\mathbf{x})$ is convex, we get $u(\mathbf{x}) \le u_{I,DT}(\mathbf{x}) \le u_{I,\mathcal{T}}(\mathbf{x})$. ∎

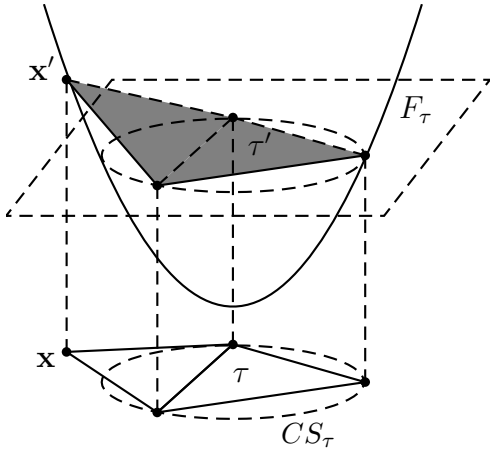



**Figure 2.2.** Projection of a lower convex hull

We now give a geometry explanation of Lemma 2.4. For a given point set $V$ in $\mathbb{R}^n$, we have a set of points $V'$ in $\mathbb{R}^{n+1}$ by lifting point in $V$ to the paraboloid. The convex hull $CH(V')$ can be divided into *lower* and *upper* parts; a facet belongs to the lower convex hull if it is supported by a hyperplane that separates $V'$ from $(\mathbf{0}, -\infty)$. We may assume the facets of the lower convex hull are simplices since if $n+2$ more vertices forms a facet, we can choose any triangulation of this facet. Brown [35] discovered that the projection of a lower convex hull of $V'$ in $\mathbb{R}^{n+1}$ is a DT of $V$ in $\mathbb{R}^n$. It is known as the lifting method in the mesh generation community [82]. Figure 2.2 is an illustration for a two dimensional Delaunay triangulation of four vertices obtained by the projection. Note that the lower convex hull of $V'$ is exactly the graph of the linear interpolant $u_{I,DT}$.

Theorem 2.3 also implies the existence of a Delaunay triangulation for a given points set. In general Delaunay triangulation is not unique since $n+2$ (or more) points may lie on a common sphere and any triangulation of those points will be a Delaunay triangulation. Fortunately this is the only possibility [81]. If we assume that points are affine independent and no $n+2$ points lie on a common sphere, the Delaunay triangulation is then uniquely determined by these points.

## Optimal Delaunay Triangulations

We have shown that when grid points are fixed, Delaunay triangulations optimize the connectivity when the vertices of triangulations are fixed. Now we free the

vertices to further optimize the triangulation.

**Definition 2.5.** *Let $\mathcal{P}_N$ stand for the set of all triangulations with at most $N$ vertices. Given a continuous function $u$ on $\bar{\Omega}$ and $1 \leq p \leq \infty$, a triangulation $\mathcal{T}_N^* \in \mathcal{P}_N$ is* optimal *if*

$$Q(\mathcal{T}_N^*, u, p) = \inf_{\mathcal{T} \in \mathcal{P}_N} Q(\mathcal{T}, u, p).$$

*We call it an Optimal Delaunay Triangluation (ODT) with respect to $u$ and $p$.*

The following theorem concerns the existence of optimal Delaunay triangulations.

**Theorem 2.6.** *Given $1 \leq p \leq \infty$, an integer $N$ and a convex function $u$, there exists an optimal Delaunay triangulation $\mathcal{T}_N^* \in \mathcal{P}_N$ with respect to $u$ and $p$.*

*Proof.* By the lifting method, we note that adding a new point into a simplex $\tau$ the new triangulation obtained by connecting it to the vertices of $\tau$ is not worse than the original one since $u$ is convex. Therefore, we may prove the result for the triangulation with exactly $N$ vertices.

Let us take a sequence of triangulations $\{\mathcal{T}^k\}_{k=1}^{\infty} \subset \mathcal{P}_N$ with vertices $V^k = (\mathbf{x}_1, \cdots, \mathbf{x}_N) \in \Omega^N$ such that

$$\lim_{k \to \infty} Q(\mathcal{T}^k, u, p) = \inf_{\mathcal{T} \in \mathcal{P}_N} Q(\mathcal{T}, u, p).$$

By the compactness of $\bar{\Omega}$, we may suppose that there exists $V^* \in \bar{\Omega}^N$ such that $\lim_{k \to \infty} V^k = V^*$, namely

$$\lim_{k \to \infty} \mathbf{x}_i^k = \mathbf{x}_i^*, \quad \forall\, \mathbf{x}_i^k \in V^k, i = 1, \cdots, N.$$

Because of the finite possible connectivity for $N$ vertices, again by the compactness argument, we may also assume $\{\mathcal{T}^k\}_{k=1}^{\infty}$ yield the same connectivity and index the simplices as $\tau_1^k, \cdots, \tau_{N_T}^k$. $V^*$ with the same connectivity (i.e. with simplices $\{\tau_i^{\infty}\}$) yields a triangulation $\mathcal{T}^*$. It might be not a valid triangulation if some $\tau_i^{\infty}$ is degenerate (see Figure 2.3). More precisely, let us define the signed volume $\mathrm{vol}(\tau)$

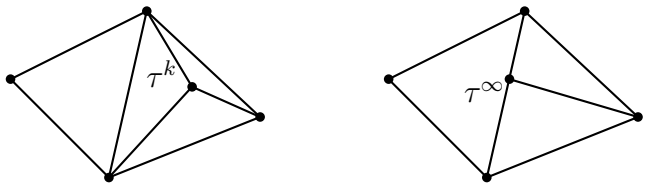



**Figure 2.3.** Degenerate simplex

of a simplex $\tau$ with vertices $\mathbf{x}_1, \cdots, \mathbf{x}_{n+1}$ as

$$\mathrm{vol}(\tau) := \det |\mathbf{x}_2 - \mathbf{x}_1 \ \mathbf{x}_3 - \mathbf{x}_1 \ \cdots \ \mathbf{x}_{n+1} - \mathbf{x}_1|,$$

which is obviously a continuous function with respect to its vertices. Since $\mathrm{vol}(\tau_j^k) > 0$, for $k = 1, \cdots, n, \cdots$, and $j = 1, \cdots, N_T$, we conclude that $\mathrm{vol}(\tau_j^\infty) \geq 0, j = 1, \cdots, N_T$. The linear interpolation is well defined except on the boundary of the degenerate simplex which is a measure zero set. Thus the interpolation error $Q(\mathcal{T}^*, u, p)$ is still well defined. To obtain a valid triangulation, we use $V^*$ to construct a Delaunay triangulation $DT^*$ with respect to $u$. A careful check of the proof of Lemma 2.4 tells us that $Q(DT^*, u, p) \leq Q(\mathcal{T}^*, u, p)$.

Since $Q(\mathcal{T}^k, u, p) = \sum_j Q(\tau_j^k, u, p)$, the quality $Q(\mathcal{T}^k, u, p)$ is continuous with respect to $k$, and thus

$$Q(DT^*, u, p) \leq Q(\mathcal{T}^*, u, p) = \lim_{k \to \infty} Q(\mathcal{T}^k, u, p) = \inf_{\mathcal{T} \in \mathcal{P}_N} Q(\mathcal{T}, u, p).$$

□

We can also define the optimal triangulation in the set of triangulations with at most $N$ simplices. It is slightly different with the optimal triangulation in $\mathcal{P}_N$ since with the same vertices, we can have two triangulations with different number of simplices. The existence of such optimal triangulation can be proved in a similar way.

## Non-uniform density

A density $\rho(\mathbf{x})$ on $\Omega$ is a integrable function such that $\rho(\mathbf{x}) > 0$ and $\int_\Omega \rho(\mathbf{x}) d\mathbf{x} < \infty$. Without loss of generality, we always assume $\int_\Omega \rho(\mathbf{x}) d\mathbf{x} = 1$. Given a density

function $\rho$, we can define a new measure $d\rho = \rho(\mathbf{x})dx$ and the corresponding weighted norm

$$\|f\|_{L^p,\rho} := (\int_\Omega |f|^p d\rho)^{1/p}.$$

We can generalize our error-based mesh quality to be

$$Q(\mathcal{T}, u, \rho, p) = \|u - u_I\|_{L^p,\rho}$$

and define optimal Delaunay triangulation with respect to the density $\rho$ in the same way. In the isotropic case we can use a non-uniform density to control the mesh size and simply choose $u(\mathbf{x}) = \|\mathbf{x}\|^2$ to control the shape of elements.

**Definition 2.7.** *For a density function $\rho$, we define the density-based mesh quality*

$$Q(\mathcal{T}, \rho, p) = Q(\mathcal{T}, \|\mathbf{x}\|^2, \rho, p).$$

*An optimal Delaunay triangulation $\mathcal{T}_N^*$ with respect to $\rho$ and $p$ is that*

$$Q(\mathcal{T}_N^*, \rho, p) = \inf_{\mathcal{T}\in\mathcal{P}_N} Q(\mathcal{T}, \rho, p).$$

We would like to point out that we only change the measure of the integration, while Lemma 2.4 is a result on the pointwise value of functions, the optimality of Delaunay triangulation still holds. The existence of optimal Delaunay triangulation with respect to $\rho$ can be obtained similarly. We now list the results below and skip the proof.

**Theorem 2.8.** *For a finite point set $V$, we let $\Omega = CH(V)$ and denote $\mathcal{P}_V$ all possible triangulations of $\Omega$ by using the points in $V$. We have*

$$Q(DT, \rho, p) = \min_{\mathcal{T}\in\mathcal{P}_V} Q(\mathcal{T}, \rho, p), \quad \forall\, 1 \le p \le \infty.$$

**Theorem 2.9.** *Given $1 \le p \le \infty$, an integer $N$ and a density function $\rho$, there exists an optimal Delaunay triangulation $\mathcal{T}_N^* \in \mathcal{P}_N$ with respect to $\rho$ and $p$.*

## 2.2 Interpolation error estimate: lower bound

In this section, we shall present an asymptotic lower bound for strictly convex (or concave) functions $u \in C^2(\bar{\Omega})$:

$$\liminf_{N\to\infty} N^{\frac{2}{n}} \|u - u_{I,\mathcal{T}_N}\|_{L^p(\Omega)} \geq C_{n,p} \| \sqrt[n]{|\det \nabla^2 u|} \|_{L^{\frac{pn}{2p+n}}(\Omega)}.$$

### Notation

Let us first introduce some standard short-hand notation for multiple indices. A multi-index $\alpha$ is an $m$-tuple of non-negative integers $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)$. The length of $\alpha$ is defined by $|\alpha| = \sum_{i=1}^{m} \alpha_i$. For a given vector $\mathbf{x} = (x_1, x_2, \cdots, x_m)$, we define $\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_m^{\alpha_m}$.

For a given multi-index $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_{n(n+1)/2})$, we define an $(n+1)\times(n+1)$ symmetric matrix $B = (b_{ij})$ by $b_{ii} = 0$, $b_{ij} = \alpha_{(i-1)\times(n+1)+j}$, for $i < j$ and a new multi-index $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \cdots, \tilde{\alpha}_{n+1})$ by $\tilde{\alpha}_i = \sum_{j=1}^{n+1} b_{ij}$.

Assume that we are given a strictly convex (or concave) function $u \in \mathcal{C}^2(\bar{\Omega})$, namely $\nabla^2 u$ is positive (or negative) definite and continuous on $\bar{\Omega}$. We can define a metric $H = \nabla^2 u$ and its average over a simplex $\tau$:

$$H_\tau := \frac{1}{|\tau|} \int_\tau H(\mathbf{x}) d\mathbf{x}.$$

Given $1 \leq p \leq \infty$, we introduce a scaled Hessian matrix as follows

$$H_p = \mu_p H, \quad \mu_p = (\det H)^{-\frac{1}{2p+n}}, \tag{2.5}$$

and its discrete version on a triangulation $\mathcal{T}_N$

$$H_{\tau,p} = \mu_{\tau,p} H_\tau, \quad \mu_{\tau,p} = [\det(H_\tau)]^{-\frac{1}{2p+n}}, \quad \forall \tau \in \mathcal{T}_N. \tag{2.6}$$

For a simplex $\tau$ with vertices $\{\mathbf{a}_i\}_{i=1}^{n+1}$, we define $d_{ij}^2 = (\mathbf{a}_i - \mathbf{a}_j) H_\tau (\mathbf{a}_i - \mathbf{a}_j)$, for $i, j = 1, 2, \cdots, n+1$. It is more convenient to order them with one single index, say lexicographically, i.e. $d_k, k = 1, 2, \cdots, m$ for $m = n(n+1)/2$. We use

$\mathbf{d} = (d_1, d_2, \cdots d_m)$ to denote the vector composed by edge lengths. We use $|\Omega|_{H_p}$ for the volume of $\Omega$ under the metric $H_p$ i.e.

$$|\Omega|_{H_p} = \int_\Omega (\det H_p)^{1/2}(\mathbf{x})d\mathbf{x} = \int_\Omega (\det H)^{\frac{p}{2p+n}}(\mathbf{x})d\mathbf{x}.$$

Its discrete version is

$$|\tau|_{H_{\tau,p}} = (\det H_{\tau,p})^{1/2}|\tau|, \quad \text{and} \quad |\Omega|_{H_{\tau,p}} = \sum_{\tau\in\mathcal{T}_N} |\tau|_{H_{\tau,p}}. \tag{2.7}$$

Furthermore we will define

$$\kappa_0(\tau) = \min_{\xi\in\mathbb{R}^n, x\in\tau} \frac{\xi^t H(\mathbf{x})\xi}{\xi^t H_\tau \xi}, \text{ and } \kappa_1(\tau) = \min_{x\in\tau} \left[\frac{\det H_\tau}{(\det H)(\mathbf{x})}\right]^{1/n}. \tag{2.8}$$

Since $u$ is strictly convex (or concave), the above functions are well defined.

In the statement of our main theorems, we need to use a $L^r$-metric for all $r > 0$, namely

$$\|u\|_{L^r(\Omega)} := \Big\{ \int_\Omega |u|^r(\mathbf{x})d\mathbf{x}\Big\}^{\frac{1}{r}}.$$

We note that $\|\cdot\|_{L^r}$ is not a norm when $0 < r < 1$.

## Main Result

**Lemma 2.10.** *Let $\lambda := (\lambda_1, \lambda_2, \cdots, \lambda_{n+1})$ be the barycentric coordinates of $\tau$ and $\mathbf{a}_k(\mathbf{x}, t) = \mathbf{a}_k + t(\mathbf{x} - \mathbf{a}_k)$. For the nodal interpolant $u_I$ of a function $u \in \mathcal{C}^2(\bar{\tau})$, we have*

$$(u_I - u)(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n+1} \lambda_i(\mathbf{x})(\mathbf{x} - \mathbf{a}_i)^t \left[2\int_0^1 t\nabla^2 u(\mathbf{a}_i(\mathbf{x}, t))\right](\mathbf{x} - \mathbf{a}_i). \tag{2.9}$$

*Proof.* By Taylor expansion,

$$u(\mathbf{a}_i) = u(\mathbf{x}) + \nabla u(\mathbf{x})(\mathbf{a}_i - \mathbf{x}) + \frac{1}{2}(\mathbf{x} - \mathbf{a}_i)^t \left[2\int_0^1 t\nabla^2 u(\mathbf{a}_i(\mathbf{x}, t))\right](\mathbf{x} - \mathbf{a}_i).$$

Multiplying both sides by $\lambda_i$ and using the fact

$$\sum_{i=1}^{n+1} \lambda_i(\mathbf{x})u(\mathbf{a}_i) = u_I(\mathbf{x}), \ \sum_{i=1}^{n+1} \lambda_i(\mathbf{x})u(\mathbf{x}) = u(\mathbf{x}), \ \text{ and } \ \sum_{i=1}^{n+1} \lambda_i(\mathbf{x})(\mathbf{a}_i - \mathbf{x}) = 0,$$

we obtain the desired result. □

**Lemma 2.11.** *For a strictly convex function $u \in \mathcal{C}^2(\bar{\tau})$, we have*

$$\|u - u_I\|^p_{L^p(\tau)} \geq \kappa_0(\tau)\frac{p!n!}{2^p(2p+n)!}|\tau| \sum_{\alpha,|\alpha|=p} \frac{\tilde{\alpha}!}{\alpha!}\mathbf{d}^{2\alpha}, \ \textit{when } 1 \leq p < \infty \quad (2.10)$$

$$\|u - u_I\|_{L^\infty(\tau)} \geq \frac{\kappa_0(\tau)}{2(n+1)^2} \sum_{i=1}^{m} d_i^2, \ \textit{when } p = \infty. \quad (2.11)$$

*Proof.* By (2.9) and (2.8)

$$\begin{aligned} |(u - u_I)(\mathbf{x})| &\geq \frac{\kappa_0(\tau)}{2} \sum_{j=1}^{n+1} \lambda_j(\mathbf{x})(\mathbf{x} - \mathbf{a}_j)^t H_\tau(\mathbf{x} - \mathbf{a}_j) \\ &= \frac{\kappa_0(\tau)}{2} \sum_{i,j=1}^{n+1} (\lambda_i\lambda_j)(\mathbf{x})(\mathbf{a}_i - \mathbf{a}_j)H_\tau(\mathbf{x} - \mathbf{a}_j). \end{aligned}$$

Using the symmetry of the index $i$ and $j$, we can also write this inequality as

$$|(u - u_I)(\mathbf{x})| \geq \frac{\kappa_0(\tau)}{2} \sum_{i,j=1}^{n+1} (\lambda_i\lambda_j)(\mathbf{x})(\mathbf{a}_j - \mathbf{a}_i)H_\tau(\mathbf{x} - \mathbf{a}_i).$$

Summing the above two inequalities gives that

$$|(u - u_I)(\mathbf{x})| \geq \frac{\kappa_0(\tau)}{2} \sum_{i,j=1, i<j}^{d+1} (\lambda_i\lambda_j)(\mathbf{x})d_{ij}^2. \quad (2.12)$$

The inequality (2.10) can then be obtained by using two elementary identities as follows:

$$(\sum_{i=1}^{n} x_i)^p = \sum_{\alpha,|\alpha|=p} \frac{p!}{\alpha!}x^\alpha,$$

and

$$\int_\tau \lambda^\alpha(\mathbf{x})d\mathbf{x} = \frac{\alpha!n!}{(|\alpha|+n)!}|\tau|.$$

When $p = \infty$, we choose $\mathbf{x}_\tau$, the barycenter of $\tau$, in (2.12). Since $\lambda_i(\mathbf{x}_\tau) = 1/(n+1)$ for $i = 1, .., n$, we get (2.11). □

**Lemma 2.12.** *Let*

$$E_p(\mathbf{x}) = \sum_{\alpha,|\alpha|=p} \frac{\tilde{\alpha}!}{\alpha!}\mathbf{x}^\alpha, \quad 1 \le p < \infty,$$

*then for any* $\mathbf{x} = (x_1, x_2, \cdots x_m)$ *with* $x_i > 0$, $i = 1, 2, \cdots, m$,

$$E_p(\mathbf{x}) \ge E_p(\mathbf{x}^*) \left(\sum_{i=1}^m x_i\right)^p.$$

*The equality holds if and only if* $\mathbf{x} = \mathbf{x}^*$ *with* $\mathbf{x}^* = (1/m, 1/m, \cdots, 1/m)$.

*Proof.* Note that $E_1(\mathbf{x}) = \sum_{i=1}^m x_i$, we will prove the result for $p > 1$. We first compute the minimum of $E_p(\mathbf{x})$ under the constrain $\sum_{i=1}^m x_i = 1$. Let

$$F(\mathbf{x}, \lambda) = E_p(\mathbf{x}) - \lambda(\sum_{i=1}^m x_i - 1).$$

The critical point of $F$ satisfies

$$\lambda = \partial_1 E_p(\mathbf{x}) = \partial_2 E_p(\mathbf{x}) = \cdots = \partial_m E_p(\mathbf{x}).$$

Here for convenience in notation, we let $\partial_i E_p$ denote the partial derivative $\dfrac{\partial E_p}{\partial x_i}$. Since $E_p$ is symmetric with respect to $x_i$, namely

$$E_p(\sigma(\mathbf{x})) = E_p(\sigma^2(\mathbf{x})) = \cdots = E_p(\sigma^m(\mathbf{x})),$$

where $\sigma$ is the cyclic permutation: $\sigma((x_1, x_2, \cdots x_m)) = (x_2, \cdots x_m, x_1)$. We then have

$$\partial_1 E_p(\sigma(\mathbf{x})) = \partial_2 E_p(\sigma^2(\mathbf{x})) = \cdots = \partial_m E_p(\sigma^m(\mathbf{x})).$$

We conclude that the point $(\mathbf{x}^*, \partial_1 E_p(\mathbf{x}^*))$ is a critical point of $F$.

Note $E_p(\mathbf{x})$ is $p$-th homogeneous, by differentiating the Euler formula $\mathbf{x}\cdot\nabla E_p(\mathbf{x}) = pE_p(\mathbf{x})$, we get $\mathbf{x}\cdot(\nabla^2 E_p) = (p-1)\nabla E_p(\mathbf{x})$ and thus

$$\mathbf{x}\cdot\nabla^2 E_p(\mathbf{x})\mathbf{x}^t = (p-1)\mathbf{x}\cdot\nabla E_p = (p-1)pE_p(\mathbf{x}) > 0.$$

Since $F$ has the same quadratic part as $E_p$ as a function of $\mathbf{x}$, we conclude that $F$ achieves the minimum at $\mathbf{x}^*$ under the constraint $\sum_{i=1}^m x_i = 1$.

By combining this result with the following obvious identity

$$E_p\big(\frac{\mathbf{x}}{\sum_{i=1}^m x_i}\big) = \frac{E_p(\mathbf{x})}{\big(\sum_{i=1}^m x_i\big)^p},$$

we complete the proof. □

The following lemma is a well known geometry inequality between the total edge length and the volume of a simplex, for example, see [147] (p.517).

**Lemma 2.13.**

$$\sum_{i=1}^m d_i^2 \geq \frac{n(n+1)n!^{2/n}}{(n+1)^{1/n}}|\tau|_{H_{\tau,p}}^{2/n}.$$

*The equality holds if and only if $\tau$ is equilateral under the metric $H_{\tau,p}$.*

For a family of triangulations $\{\mathcal{T}_N\}$ of $\Omega$ we define

$$h_N = \max_{\tau\in T_N}\operatorname{diam}(\tau),$$

$$\kappa_0^N(\mathbf{x}) = \kappa_0(\tau),\ \kappa_1^N(\mathbf{x}) = \kappa_1(\tau), \quad x \in \tau.$$

For an edge $E \in \mathcal{T}$ we use $d_E$ to denote its length under the metric (2.6)

$$\rho_N = \frac{\max_{E\in\mathcal{T}_N} d_E}{\min_{E\in\mathcal{T}_N} d_E}.$$

We say all edges are *asymptotically equal* if $\lim_{N\to\infty}\rho_N = 1$.

**Theorem 2.14.** *Assume that $u \in C^2(\bar{\Omega})$ is a strictly convex (or concave) function*

*and* $\{\mathcal{T}_N\}$ *are a family of triangulations of* $\Omega$ *satisfying* $\lim_{N\to\infty} h_N = 0$. *Then*

$$\liminf_{N\to\infty} N^{\frac{2}{n}} \|u - u_{I,\mathcal{T}_N}\|_{L^p(\Omega)} \geq C_{n,p} \| \sqrt[n]{|\det \nabla^2 u|} \|_{L^{\frac{pn}{2p+n}}(\Omega)}, \tag{2.13}$$

*where*

$$\begin{aligned} C_{n,p} &= \Big( \sum_{\alpha,|\alpha|=p} \frac{\tilde{\alpha}!}{\alpha!} \Big) \frac{p!n!}{(2p+n)!} \frac{n!^{2p/n}}{(n+1)^{p/n}}, \quad 1 \leq p < \infty, \\ C_{n,\infty} &= \frac{n}{2(n+1)} \frac{n!^{2/n}}{(n+1)^{1/n}}. \end{aligned}$$

*Furthermore the equality holds if all edges in* $\{\mathcal{T}_N\}$ *are asymptotically equal under the new metric (2.6), namely* $\lim_{N\to\infty} \rho_N = 1$.

*Proof.* We first prove the case $1 \leq p < \infty$. By (2.10) in Lemma 2.12 and Lemma 2.13 we have

$$|\tau| E_p(\mathbf{d}^2) = \mu_{\tau,p}^{-p} |\tau| E_p(\mathbf{d}^2) \geq C \mu_{\tau,p}^{-p} |\tau| \left( \sum_{i=1}^m d_i^2 \right)^p \geq C \mu_{\tau,p}^{-p} |\tau| |\tau|_{H_{\tau,p}}^{2p/n},$$

where $\mu_{\tau,p}$ and $|\tau|_{H_{\tau,p}}$ are defined by (2.6) and (2.7). The constant $C$ only depends on $p$ and $n$ but not $\tau$ and the equality holds if and only if simplex $\tau$ is equilateral under the new metric $H_{\tau,p}$ i.e. all edges of $\tau$ are equal under the new metric $H_{\tau,p}$.

By (2.20) and the definition of $\mu_{\tau,p}$, we know

$$\mu_\tau^{-p} |\tau| \| |\tau|_{H_{\tau,p}}^{2p/n} = |\tau|_{H_{\tau,p}}^{\frac{2p+n}{n}}.$$

Consequently

$$\begin{aligned} \|u - u_I\|_{L^p(\Omega)}^p &\geq C \sum_{\tau\in T_N} \kappa_0(\tau) |\tau|_{H_{\tau,p}}^{\frac{2p+n}{n}} \\ &\geq C \kappa_0^N(\mathbf{x}) \kappa_1^N(\mathbf{x}) \sum_{\tau\in\mathcal{T}_N} \left( \int_\tau \left[ \sqrt[n]{|\det \nabla^2 u|} \right]^{\frac{np}{2p+n}} (\mathbf{x}) d\mathbf{x} \right)^{\frac{2p+n}{n}} \\ &\geq C N^{-2p/n} \kappa_0^N(\mathbf{x}) \kappa_1^N(\mathbf{x}) \left( \int_\Omega \left[ \sqrt[n]{|\det \nabla^2 u|} \right]^{\frac{np}{2p+n}} (\mathbf{x}) d\mathbf{x} \right)^{\frac{2p+n}{n}}. \end{aligned}$$

The last inequality becomes equality if and only if all the volumes of simplices

under the new metric are equal.

Since $\nabla^2 u$ is continuous, $\kappa_0^N(\mathbf{x})\kappa_1^N(\mathbf{x})$ are uniform bounded and it approaches to 1 when $N$ goes to $\infty$. The desired inequality of the theorem for $1 \le p < \infty$ then follows. Furthermore, if all edges are asymptotically equal under the new metric, we can achieve all the equalities when $N \to \infty$.

For $p = \infty$, we make use of (2.11) in Lemma 2.12 and follow the same procedure to get the desired inequality.

The formula for $C_{n,p}$ and $C_{n,\infty}$ follow from careful computation. □

The strictly convexity assumption of the function is to ensure $\nabla^2 u$ is a Riemannian metric so that locally we can apply Lemma 2.13. For $p = \infty$ the strictly convexity assumption can be relaxed to be nonsingular, i.e. the indefinite metric is allowed in this case; see [153, 138].

## 2.3 Interpolation error estimate: upper bound

In general, it is difficult to find useful sufficient conditions for optimal Delaunay triangulations. Instead we try to find a triangulation $\mathcal{T}_N$ such that

$$Q(\mathcal{T}_N, u, p) \le c\, Q(\mathcal{T}_N^*, u, p),$$

where $c$ is a constant independent of $N$ and $\mathcal{T}_N^*$ is an optimal Delaunay triangulation in $\mathcal{P}_N$. We call such a $\mathcal{T}_N$ a *nearly optimal triangulation* or *quasi-optimal triangulation*.

In this section, we will give two practical sufficient conditions for a triangulation to be nearly optimal. Under these conditions, we are able to prove a upper bound for the interpolation error

$$\|u - u_I\|_{L^p(\Omega)} \le CN^{-2/n} \| \sqrt[n]{\det H} \|_{L^{\frac{pn}{2p+n}}(\Omega)}, \ 1 \le p \le \infty,$$

where $H$ is a majorant of $\nabla^2 u$.

## Assumptions about the mesh

For a general function $u \in \mathcal{C}^2(\bar{\Omega})$, $\nabla^2 u$ may not be positive definite. We will use a majorant of $\nabla^2 u$ to define a new Riemannian metric.

**Definition 2.15.** *Given a function $u \in \mathcal{C}^2(\bar{\Omega})$ where $\Omega$ is an open set of $\mathbb{R}^n$, a symmetric positive definite matrix function $H \in (C(\bar{\Omega}))^{n\times n}$ is called a* majorant *of the Hessian matrix for $u$ if it satisfies*

$$|\xi^t(\nabla^2 u)(\mathbf{x})\xi| \leq c_0 \xi^t H(\mathbf{x})\xi, \quad \xi \in \mathbb{R}^n, \mathbf{x} \in \Omega \tag{2.14}$$

*for some positive constant $c_0$.*

One example of $H$ can be constructed as follows. First, we diagonalize the Hessian: $\nabla^2 u = Q^t \mathrm{diag}(\sigma_i) Q$ and then define

$$H = Q^t \mathrm{diag}(|\sigma_i|) Q + \delta I, \quad \delta \geq 0. \tag{2.15}$$

It is easy to see that this matrix $H$ is a majorant of the Hessian matrix of $u$ and it satisfies (2.14) for any $\delta > 0$ with $c_0 = 1$. When $\nabla^2 u$ is singular, a positive parameter $\delta$ is critical to control the variation of $H$. A careful analysis in Huang [105] shows that $\delta$ can be used to control the ratio of mesh points in the singular region and smooth region of the function.

Let $u \in \mathcal{C}^2(\bar{\Omega})$ and $H$ be a majorant of the Hessian of $u$. Associated with $u$, we consider a triangulation $\mathcal{T}_N$ of $\Omega$ that is adaptively obtained according to the properties of $u$. We will now make two major assumptions on $\mathcal{T}_N$ in its relation with $u$.

Our first assumption on $\mathcal{T}_N$ is a local one. We assume that the there is no oscillation of $H$ in each element.

**(A1)** There exist two positive constants $\alpha_0$ and $\alpha_1$ such that

$$\alpha_0 \xi^t H_\tau \xi \leq \xi^t H(\mathbf{x}) \xi \leq \alpha_1 \xi^t H_\tau \xi, \quad \xi \in \mathbb{R}^n. \tag{2.16}$$

The above assumption is hard to satisfy where $\nabla^2 u$ is nearly singular. In this event, the introduction of the relaxation parameter $\delta$ in (2.15) becomes critical for this assumption.

Our second assumption is both local and global. Let

$$\mathrm{diam}_{H_{\tau,p}}(\tau) = \max_{\mathbf{x},\mathbf{y}\in\tau}(\mathbf{x}-\mathbf{y})^t H_{\tau,p}(\mathbf{x}-\mathbf{y})$$

be the diameter of $\tau$ under the new metric $H_{\tau,p}$. The following assumption means that $\mathcal{T}_N$ is quasi-uniform under the new metric induced by $H_{\tau,p}$.
**(A2)** There exists two positive constants $\beta_0$ and $\beta_1$ such that

$$\frac{\mathrm{diam}_{H_{\tau,p}}(\tau)}{|\tau|^{1/n}_{H_{\tau,p}}} \le \beta_0, \quad \forall \tau \in T_N, \tag{2.17}$$

and

$$\frac{\max_{\tau\in T_N}|\tau|_{H_{\tau,p}}}{\min_{\tau\in\mathcal{T}_N}|\tau|_{H_{\tau,p}}} \le \beta_1. \tag{2.18}$$

The inequality (2.17) means that each $\tau$ is shape-regular under the metric $H_{\tau,p}$, namely all edges of $\tau$ are of comparable size. This is related to the so-called *isotropy* criterion considered in [105]. The inequality (2.18) means that all elements $\tau$ are of comparable size (under the new metric). This is related to the so-called *equidistribution* criterion, considered in [57] and [105].

It is easy to see that **(A2)** implies that

$$\sum_{i=1}^{n(n+1)/2} d^2_{\tau,i} \le \frac{n(n+1)}{2}\beta_0|\tau|^{2/n}_{H_{\tau,p}}, \quad \forall \tau \in T_N,$$

and

$$|\tau|_{H_{\tau,p}} \le \beta_1 N^{-1}|\Omega|_{H_{\tau,p}}, \quad \forall \tau \in T_N.$$

## Main Result

Given a simplex $\tau$ with vertices $\{\mathbf{a}_k\}_{k=1}^{n+1}$, we first derive a linear interpolant error estimate in terms of the majorant of the Hessian matrix.

**Lemma 2.16.** *For a given $u \in C^2(\bar{\tau})$, let $H$ be a majorant of the Hessian matrix of $u$, we then have*

$$|(u-u_I)(\mathbf{x})| \le c_0 \sum_{j<k}(\mathbf{a}_j-\mathbf{a}_k)^t H_k(\mathbf{x})(\mathbf{a}_j-\mathbf{a}_k), \quad x \in \tau$$

*where*

$$H_k(\mathbf{x}) = 2\int_0^1 tH(\mathbf{a}_k(t)), \quad \mathbf{a}_k(t) = \mathbf{a}_k + t(\mathbf{x} - \mathbf{a}_k).$$

*Proof.* Let $\{\lambda_k(\mathbf{x})\}$ be the barycentric coordinates of $\tau$. By Taylor expansion (2.9),

$$u_I - u = \frac{1}{2}\sum_{k=1}^{n+1} \lambda_k(\mathbf{x})(\mathbf{x} - \mathbf{a}_k)^t \left[2\int_0^1 t\nabla^2 u(\mathbf{a}_k(t))\right](\mathbf{x} - \mathbf{a}_k).$$

Note that $\mathbf{x} - \mathbf{a}_k = \sum_i \lambda_i(\mathbf{a}_i - \mathbf{a}_k)$ we have

$$|(u_I - u)(\mathbf{x})| \leq \frac{c_0}{2}\sum_{i,j,k=1}^{n+1} \lambda_i\lambda_j\lambda_k(\mathbf{a}_j - \mathbf{a}_k)^t H_k(\mathbf{a}_j - \mathbf{a}_k).$$

By Cauchy-Schwarz inequality:

$$\begin{aligned}
|(u_I - u)(\mathbf{x})| &\leq \frac{c_0}{2}\sum_{i,j,k=1}^{n+1} \lambda_i\lambda_j\lambda_k\sqrt{(\mathbf{a}_i - \mathbf{a}_k)^t H_k(\mathbf{a}_i - \mathbf{a}_k)}\cdot\sqrt{(\mathbf{a}_j - \mathbf{a}_k)^t H_k(\mathbf{a}_j - \mathbf{a}_k)} \\
&\leq \frac{c_0}{2}\left(\sum_{i,j,k=1}^{n+1} \lambda_i\lambda_j\lambda_k(\mathbf{a}_i - \mathbf{a}_k)^t H_k(\mathbf{a}_i - \mathbf{a}_k)\right)^{1/2} \\
&\quad \left(\sum_{i,j,k=1}^{n+1} \lambda_i\lambda_j\lambda_k(\mathbf{a}_j - \mathbf{a}_k)^t H_k(\mathbf{a}_j - \mathbf{a}_k)\right)^{1/2} \\
&= \frac{c_0}{2}\sum_{j,k=1}^{n+1} \lambda_j\lambda_k(\mathbf{a}_j - \mathbf{a}_k)^t H_k(\mathbf{a}_j - \mathbf{a}_k) \\
&\leq c_0 \sum_{j,k=1,j<k}^{n+1} (\mathbf{a}_j - \mathbf{a}_k)^t H_k(\mathbf{a}_j - \mathbf{a}_k).
\end{aligned}$$

In the third step, we use the fact $\sum_{i=1}^{n+1} \lambda_i = 1$. □

**Theorem 2.17.** *Let $u \in C^2(\bar{\Omega})$ and the triangulation $\mathcal{T}_N$ satisfy assumptions (A1) and (A2), the following error estimate holds:*

$$\|u - u_I\|_{L^p(\Omega)} \leq CN^{-2/n}\|\sqrt[n]{\det H}\|_{L^{\frac{pn}{2p+n}}(\Omega)}, \ 1 \leq p \leq \infty, \tag{2.19}$$

*for some constant $C = C(n, p, c_0, \alpha_0, \alpha_1, \beta_0, \beta_1)$.*

*Proof.* We first deal with $p = \infty$ case. Let us assume $u - u_I$ attains the maximum at point $\mathbf{x}^*$ and and let $\tau^*$ be a simplex containing $\mathbf{x}^*$. By Lemma 2.16 and assumptions (A1) and (A2)

$$|(u - u_I)(x^*)| \leq C \sum_{i=1}^{m} d_i^2 \leq C|\tau^*|_{H_{\tau,p}}^{2/n} \leq CN^{-2/n}|\Omega|_{H_\tau}^{2/n} \leq CN^{-2/n}|\Omega|_H^{2/n}.$$

The desired result then follows.

For $1 \leq p < \infty$,

$$|(u - u_I)(\mathbf{x})| \leq C \sum_{j<k} (\mathbf{a}_j - \mathbf{a}_k)^t H_\tau (\mathbf{a}_j - \mathbf{a}_k) = C\mu_{\tau,p}^{-1} \sum_i d_{\tau,i}^2 \leq C\mu_{\tau,p}^{-1}|\tau|_{H_{\tau,p}}^{2/n}.$$

Thus

$$\int_\tau |(u - u_I)(\mathbf{x})|^p d\mathbf{x} \leq C\mu_{\tau,p}^{-p}|\tau||\tau|_{H_{\tau,p}}^{2p/n}.$$

Note that

$$|\tau| = (\det H_{\tau,p})^{-1/2}|\tau|_{H_{\tau,p}} = \mu_{\tau,p}^{-n/2}(\det H_\tau)^{-1/2}|\tau|_{H_{\tau,p}}, \tag{2.20}$$

thus

$$\int_\tau |(u - u_I)(\mathbf{x})|^p d\mathbf{x} \leq C\mu_{\tau,p}^{-(p+n/2)}(\det H_\tau)^{-1/2}|\tau|_{H_{\tau,p}}^{\frac{2p+n}{n}} = C|\tau|_{H_{\tau,p}}^{\frac{2p+n}{n}},$$

since by definition (2.6) of $\mu_{\tau,p}$,

$$\mu_{\tau,p}^{-(p+n/2)}(\det H_\tau)^{-1/2} = 1.$$

By assumptions (A1) and (A2)

$$\begin{aligned} \int_\Omega |(u - u_I)(\mathbf{x})|^p d\mathbf{x} &= \sum_{\tau \in T_N} \int_\tau |(u - u_I)(\mathbf{x})|^p d\mathbf{x} \leq C \sum_{\tau \in T_N} |\tau|_{H_{\tau,p}}^{\frac{2p+n}{n}} \\ &\leq CN^{-\frac{2p}{n}}|\Omega|_{H_{\tau,p}}^{\frac{2p+n}{n}} \leq CN^{-\frac{2p}{n}}|\Omega|_{H_p}^{\frac{2p+n}{n}} \end{aligned}$$

which yields the desired result. □

We would like to point out that the above theorem can be improved and generalized in many ways. For example, for $p \neq \infty$, the assumption (A2) can be slightly relaxed. More precisely, the assumption (A2) can be replaced by

**(A2')** There exists two piecewise constant functions $\beta_0$ and $\beta_1$ such that

$$\sum_{i=1}^{n(n+1)/2} d_{\tau,i}^2 \leq \beta_0(\tau)|\tau|_{H_{\tau,p}}^{2/n}, \quad \forall \tau \in T_N,$$

$$|\tau|_{H_{\tau,p}} \leq \beta_1(\tau) N^{-1} |\Omega|_{H_{\tau,p}}, \quad \forall \tau \in T_N.$$

and

$$\sum_{\tau} \beta_0(\tau) \beta_1^{\frac{2p+n}{n}}(\tau) \leq \beta_2,$$

where $\beta_2$ is constant. Such a relaxation is of practical significance. This means that optimal error estimates are still valid on a mesh that has a few exceptional elements that do not satisfy the isotropic or equidistribution assumptions.

**Remark: suboptimal metric.** In some literature (see [149, 101, 64]) only the metric induced by majorant of the Hessian matrix (without using the scaling $\mu_{\tau,p}$ for different $p$) is used to get an anisotropic mesh optimization methodology. To see why this would work to some degree, let us assume the triangulation is quasi-uniform under the metric $H_\tau$. Then

$$\begin{aligned} \|u - u_I\|_{L^p(\Omega)}^p &\leq C \sum_{\tau} |\tau| |\tau|_{H_\tau}^{2p/n} \\ &\leq C N^{-2p/n} |\Omega| |\Omega|_{H_\tau}^{2p/n} \\ &= C N^{-2p/n} \left( \int_\Omega 1 d\mathbf{x} \right) \left( \int_\Omega (\det H)^{1/2} d\mathbf{x} \right)^{\frac{2p}{n}}. \end{aligned}$$

It implies that we can still get a reasonable convergent rate if we only use the majorant of the Hessian metric without scaling. But this is not optimal since, comparing Theorem 2.17, we have

$$\left( \int_\Omega 1 d\mathbf{x} \right) \left( \int_\Omega (\det H)^{1/2} d\mathbf{x} \right)^{\frac{2p}{n}} \geq \left( \int_\Omega (\det H)^{\frac{p}{2p+n}} d\mathbf{x} \right)^{\frac{2p+n}{pn}}.$$

Chapter 3

# Algorithms to Generate Nearly Optimal Triangulations

In this chapter, we shall discuss practical algorithms to generate nearly optimal triangulations. There are mainly two types of algorithms: one is global and another is local.

The global algorithm we shall consider is known as the moving mesh method in the literature [80, 184, 37, 156, 37, 104, 107, 141, 174, 22, 140, 79, 63, 132, 131]. It can be described by two different sets of domains, $\Omega_c$ and $\Omega$, and corresponding grids, $\mathcal{T}_{N,c}$ and $\mathcal{T}_N$. The adapted mesh $\mathcal{T}_N$ can be viewed as a result of moving the grid $\mathcal{T}_{N,c}$ through the transformation $\mathbf{x} = \mathbf{x}(\xi) : \Omega_c \mapsto \Omega$.

Let $G$ be a metric on $\Omega$, $E$ be the standard Euclidean metric on $\Omega_c$ and $\mathcal{T}_{N,c}$ a quasi-uniform mesh under $E$. $\mathcal{T}_N$ is quasi-uniform under $G$ is more or less equivalent to that the transformation is quasi-conform and quasi-isometric. Thus the transformation can be obtained by minimizing the following functional:

$$\min_{\mathbf{x}} \int_{\Omega_c} \Big[ \sum_{i=1}^{n} (\nabla x_i)^t G(\mathbf{x}) \nabla x_i \Big]^q (\xi) d\xi, \quad q > n/2.$$

It is often more convenient to study the moving meshes in terms of $\xi = \xi(\mathbf{x})$. In this case, it needs to solve the following optimization problem:

$$\min_{\xi} \int_{\Omega} (\det G)^{1/2} \Big[ \sum_{i=1}^{n} (\nabla \xi_i)^t G^{-1}(\xi) \nabla \xi_i \Big]^q (\mathbf{x}) d\mathbf{x}, \quad q > n/2.$$

We note that the case that $q = 1$ corresponds to the harmonic mapping but we require $q > n/2$ here. When $n \geq 3$, these minimization problems (which are more or less $p$-Laplacians with $p > n$) is significantly different from the harmonic mapping which has been most commonly used in the literature for the moving mesh method [80, 37, 106, 131]. The metric $G$ used in the above functionals is called monitor function. According to the interpolation error estimate in Chapter 1, $G = (\det H)^{-1/(2p+n)} H$ will be a good choice.

The main constrain of the moving mesh method is that the mapping needs to preserve the topological structure of $\mathcal{T}_{N,c}$ and thus may not be able to produce a nearly optimal mesh. To further improve the quality of the mesh, we need local mesh optimization methods.

We will mainly discuss three types of local mesh optimization methods: local refinement or coarsening [17, 159, 118], edge swapping [128], and mesh smoothing [18, 92]. According to our understanding of the mesh quality, refinement and coarsening mainly aims to optimize the mesh density, while edge swapping and mesh smoothing mainly aim to optimize the shape regularity. Our main contribution is the introduction of several new mesh smoothing schemes based on optimal Delaunay triangulations.

In the existing literature, there are mainly two types of smoothing methods, namely Laplacian smoothing and optimization-based smoothing. Laplacian smoothing [85], in its simplest form, is to move each vertex to the arithmetic average of the neighboring points. It is easy to implement and require a very low computational cost, but it operates heuristically and does not guarantee an improvement in the geometric mesh qualities. Thus an optimization-based smoothing has been proposed: the vertex is moved so as to optimize some mesh quality [18, 170, 92]. The price for the guaranteed quality improvement is that the computational time involved is much higher than that of Laplacian smoothing.

Our mesh smoothing schemes essentially belong to the optimization-based smoothing. Instead of geometric mesh qualities, we try to minimize the interpolation error in the local patch. With several formulas of the interpolation error, in isotropic case, we can solve the optimization problem exactly and thus the computational cost is as low as that of Laplacian smoothing, while the error-based mesh quality is guaranteed to be improved. The schemes can be applied to anisotropic

case easily.

## 3.1 Global algorithms: moving mesh methods

In this section, we shall apply the error estimate obtained in Chapter 1 to derive some new optimization problem that can be used in the moving mesh method. The basic idea in our argument below is not new (see for example, Huang [105]), but the optimization problems we will obtain appear to be new and especially appear to be more appropriate than those in the literature. The main new feature of our new optimization problems is that it addresses the isotropic and equidistribution criteria simultaneously.

We first give a motivation of our approach. Let us assume $\mathbf{x} : \Omega_c \mapsto \Omega$ is a differentiable homeomorphism and denote the Jacobian of the transformation $\mathbf{x} = \mathbf{x}(\xi) : \Omega_c \mapsto \Omega$ by

$$J = \frac{\partial \mathbf{x}}{\partial \xi} = \left( \frac{\partial x_i}{\partial \xi_j} \right).$$

Let $G$ be a metric on $\Omega$, then the transformation induces a new metric on $\Omega_c$, that is $G_c = J^t G J$. Now we have three manifolds $(\Omega, G), (\Omega_c, E)$, and $(\Omega_c, G_c)$, where $E$ is the standard Euclidean metric. $\mathcal{T}_N$ is quasi-uniform in $(\Omega, G)$ if $\mathcal{T}_{N,c}$ is quasi-uniform in $(\Omega_c, G_c)$. But $\mathcal{T}_{N,c}$ is chosen as a quasi-uniform triangulation in $(\Omega_c, E)$. Thus it is equivalent to asking that

$$J^T G J = cE \quad \text{or} \quad J^{-T} G^{-1} J^{-1} = cE. \tag{3.1}$$

In general we may not get the identity since it is equivalent to asking the manifold $(\Omega_c, G_c)$ be flat. We shall try to approximate this identity as well as possible.

### 3.1.1 New optimization problems

Let us first obtain a new optimization problem in terms of $\mathbf{x} = \mathbf{x}(\xi)$. Let $\{\mu_i\}_{i=1}^n$ be the eigenvalues of $G_c$. Note that $\mu_i > 0$ since $G$ is a metric and $J$ is non-signular.

Following [105], we begin our derivation with the following elementary inequality

$$\frac{1}{n}\sum_{i=1}^{n}\mu_i \geq \Big(\prod_{i=1}^{n}\mu_i\Big)^{1/n}. \tag{3.2}$$

Therefore,

$$\frac{1}{n}\text{trace}(G_c) \geq (\det G_c)^{1/n}.$$

The above equality holds if and only if the matrix $G_c$ is a scalar matrix.

But

$$\text{trace}(G_c) = \sum_{i=1}^{n}\big[(\nabla x_i)^t G \nabla x_i\big], \quad \text{and} \quad (\det G_c) = (\det J)^2(\det G),$$

thus

$$\sum_{i=1}^{n}(\nabla x_i)^t G \nabla x_i \geq n\big[(\det J)^2(\det G)\big]^{1/n}.$$

Raising this inequality with power $q > n/2$ and integrating the result on $\Omega_c$ gives:

$$\int_{\Omega_c}\Big(\sum_{i=1}^{n}(\nabla x_i)^t G \nabla x_i\Big)^q(\xi)d\xi \geq n^q\int_{\Omega_c}\big[(\det J)^2(\det G)\big]^{q/n}(\xi)d\xi.$$

By the Hölder inequality,

$$|\Omega|^{1-n/2q}\left\{\int_{\Omega_c}\big[(\det J)^2(\det G)\big]^{q/n}(\xi)d\xi\right\}^{n/2q} \geq \int_{\Omega_c}(\det J)(\det G)^{1/2}(\xi)d\xi.$$

Therefore

$$\begin{aligned}
\int_{\Omega_c}\big[(\det J)^2(\det G)\big]^{q/n}(\xi)d\xi &\geq |\Omega_c|^{1-2q/n}\left[\int_{\Omega_c}(\det J)(\det G)^{1/2}(\xi)d\xi\right]^{2q/n} \\
&= |\Omega_c|^{1-2q/n}\left[\int_{\Omega}(\det G)^{1/2}(\mathbf{x})d\mathbf{x}\right]^{2q/n} \\
&= |\Omega_c|^{1-2q/n}|\Omega|_G^{2q/n}.
\end{aligned}$$

In summary, we have

$$\begin{aligned}\int_{\Omega_c}\Big(\sum_{i=1}^{n}(\nabla x_i)^t G\nabla x_i\Big)^q(\xi)d\xi &\geq n^q\int_{\Omega_c}\big[(\det J)^2(\det G)\big]^{q/n}(\xi)d\xi\\ &\geq n^q|\Omega_c|^{1-2q/n}|\Omega|_G^{2q/n}.\end{aligned}$$

By the well-known properties of two inequalities that we used above, we note that the first inequality holds with equality if and only if $G_c$ has all equal eigenvalues (*isotropic*), namely $G_c$ is a scalar matrix, and the second inequality holds with equality if and only if $(\det G_c)^{1/2} = (\det J)(\det G)^{1/2}$ =constant (*equidistribution*). Thus both inequalities hold with equalities if and only if both isotropy and equidistribution conditions are satisfied. This argument leads to the optimization problem:

$$\min_x \int_{\Omega_c}\Big[\sum_{i=1}^{n}(\nabla \mathbf{x}_i)^t G(\mathbf{x})\nabla \mathbf{x}_i\Big]^q(\xi)d\xi, \quad q > n/2. \tag{3.3}$$

For a given $q > n/2$, the minimizer of the above functionals is *expected* to satisfy both equidistribution and isotropy conditions simultaneously. Since the lower bound in the derivation may not be attainable, minimizers for different $q$ may be different. The index $q$ will control the mesh density. A larger $q$ will lead to a more equidistributed grid.

A similar optimization problem can also be obtained in terms of $\xi = \xi(\mathbf{x})$. From a computational point of view, $\mathbf{x} = \mathbf{x}(\xi)$ is the transformation that we need to use, but, as we shall explain later, there are some advantages of using its inverse transformation to define the optimization problem.

We can apply similar arguments for the map $\xi = \xi(\mathbf{x})$. Note that $(G_c)^{-1} = J^{-t}G^{-1}J^{-1}$. Similar to (3.2), we have

$$\Big(\sum_{i=1}^{n}\nabla\xi_i^t G^{-1}\nabla\xi_i\Big) \geq n\Big[(\det J)^2(\det G)\Big]^{-1/n},$$

and for $q > n/2$

$$\int_{\Omega}(\det G)^{1/2}\Big(\sum_{i=1}^{n}\nabla\xi_i^t G\nabla\xi_i\Big)^q(\mathbf{x})d\mathbf{x} \geq n^q\int_{\Omega}\Big[(\det J)^2(\det G)\Big]^{-q/n}(\mathbf{x})d\mathbf{x}.$$

Let $r = 2q/n$ and $r'$ be its congruent index, i.e. $1/r + 1/r' = 1$. By the Hölder inequality,

$$\Big\| \big[(\det J)(\det G)\big]^{-\frac{1}{2r'}} \Big\|_{L^r} \Big\| (\det G)^{\frac{1}{2r'}} \Big\|_{L^{r'}} \geq \int_\Omega (\det J)^{-1}(\mathbf{x}) d\mathbf{x} = \int_{\Omega_c} 1 d\xi = |\Omega_c|.$$

The equality holds if and only if $(\det J)(\det G)^{1/2}$ =constant which is equivalent to the equidistribution condition.

Since $\big\|(\det G)^{\frac{1}{2r'}}\big\|_{L^{r'}} = \int_\Omega (\det G)^{1/2} d\mathbf{x} =$ constant, we then end up with the following optimization problem:

$$\min_\xi \int_\Omega \Big[(\det G)(\mathbf{x})\Big]^{1/2} \Big[\sum_{i=1}^n (\nabla \xi_i)^t G^{-1}(\mathbf{x}) \nabla \xi_i\Big]^q d\mathbf{x}, \quad q > n/2. \tag{3.4}$$

### 3.1.2 On the optimization problems

We shall now give some brief discussions on the minimization problems (3.3) and (3.4). The discussion for one dimensional case is easy. The situation for multiple dimensions is more complicated. One main conclusion for multiple dimensions is that our optimization problem is significantly different from the harmonic mapping approach which has been used mostly in the literature [80, 37, 106, 131].

**One dimension** $n = 1$

In one dimension, namely $n = 1$, as we shall see now, both optimization problems can be solved exactly and in fact the solutions to both problems are identical to each other. The result we obtain in one dimension is not new and it coincide, for example, with the so-called grading function obtained by Carey and Dinh [39].

Without loss of generality, we may suppose $\Omega = (0, 1)$. Solving functional (3.3) gives us

$$(G^q x'^{2q})' = 0 \Rightarrow G^{1/2}(x) x'(\xi) = c \Rightarrow \xi'(x) = c^{-1} G^{1/2}(x). \tag{3.5}$$

We obtain

$$\xi(x) = \frac{\int_0^x G^{1/2}(t) dt}{\int_0^1 G^{1/2}(t) dt}. \tag{3.6}$$

which will be called the mesh distribution function. The asymptotic optimal mesh can be obtained by the inverse of the mesh distribution function, more precisely

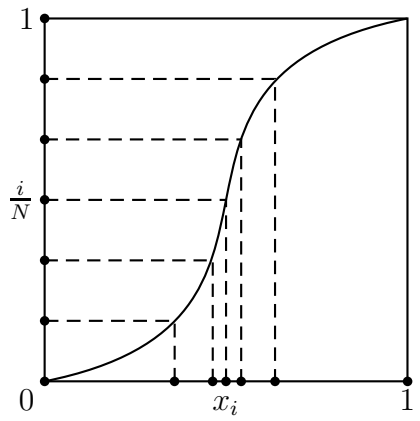



**Figure 3.1.** Mesh distribution function in one dimension

$x_i = \xi^{-1}(\xi_i)$, $i = 0, 1, ..., N$, $\xi_i = i/N$ (see Figure 3.1.2), which may be obtained by the Newton method or a discrete version of the mesh distribution function. It can be used to get a fast curve simplification algorithm [47].

Now, we consider the optimization problem (3.4). In this case, the Euler-Lagrangian equation becomes:

$$(G^{1/2-q}\xi'(x)^{2q-1})' = 0$$

which gives

$$\xi'(x) = cG^{1/2}(x).$$

Thus we get the same mesh distribution function from (3.4).

We summarize the above discussions as follows:

**Proposition 3.1.** *In one dimension, both optimization problems (3.3) and (3.4) are equivalent and their minimizer, known as the mesh distribution function, is given by (3.6), which is independent of the parameter q.*

### Multiple dimensions

The situation in multiple dimensionss is much more complex. Let us first point out that the optimization problem (3.4) appears to have more desirable properties than (3.3). For example, the functional in (3.4) is strictly convex while the functional in (3.3) is more complex. Especially, for $q = 1$, the Euler-Lagrange equation for

(3.4) is linear while the Euler-Lagrange equation for (3.3) is always nonlinear.

Since the problem (3.4) is strictly convex, the existence and uniqueness of the minimizer is then obvious for this problem. A less obvious question is if the minimizer $\xi = \xi(\mathbf{x})$ is actually a homeomorphism between $\Omega$ and $\Omega_c$ and furthermore if its Jacobian is nonsingular. These two questions is already well-studied for the special case that $n = 2$ and $q = 1$ in the literature and the answers to both questions are affirmative under the assumption that the computational domain $\Omega_c$ is convex (see [102, 169]). One natural question to ask is if these results can also be extended to the case that $q > 1$. This question is significant especially for $n \geq 3$ and it will be a subject of further study.

The optimization problem (3.3), on the other hand, is more complex to analyze than (3.4). Let us now only discuss the existence for this problem since the answers to other relevant questions are not so obvious.

**Theorem 3.2.** *Given a homeomorphism $g : \partial\Omega_c \mapsto \partial\Omega$, there is a transformation $\mathbf{x} : \Omega_c \to \Omega$ such that it is a minimizer of the minimization problem (3.3) subject to*

$$\mathbf{x} \in (W^{1,2q}(\Omega_c))^n, \quad \mathbf{x}(\xi) = g(\xi), \quad \xi \in \partial\Omega_c.$$

This theorem can be established by using the following simple lemma.

**Lemma 3.3.** *Assume that $G \in (L^\infty(\Omega))^{n\times n}$ and it is continuous almost everywhere in $\Omega^{n\times n}$ and uniformly symmetric positive definite in $\Omega$. Then*

$$\int_{\Omega_c} \Big[\sum_{i=1}^{n} (\nabla x_i)^t G(x) \nabla x_i\Big]^q (\xi) d\xi$$

*is lower semi-continuous in $(W^{1,2q}(\Omega_c))^n$ for a given $q > n/2$.*

*Proof.* Let us first assume that $G$ is continuous everywhere in $\Omega^{n\times n}$, namely $G \in (C_B(\Omega))^{n\times n}$. When $q > n/2$, $W^{1,2q}((\Omega_c)^n)$ is compactly embedded in $(L^\infty(\Omega_c))^n$. Thus if $(\mathbf{x}^m)$ is a bounded sequence in $(W^{1,2q}(\Omega_c))^n$ that is weakly convergent to $\mathbf{x}^*$, we have $\mathbf{x}^m \to \mathbf{x}^*$ strongly in $(C(\Omega_c))^n$. By a simple mean value theorem:

$$\left| \int_{\Omega_c} \Big[\sum_{i=1}^{n} (\nabla x_i^m)^t G(\mathbf{x}^m) \nabla x_i^m\Big]^q (\xi) d\xi - \int_{\Omega_c} \Big[\sum_{i=1}^{n} (\nabla x_i^m)^t G(\mathbf{x}^*) \nabla x_i^m\Big]^q (\xi) d\xi \right|$$

$$
\begin{aligned}
&= q\int_{\Omega_c}\left\{\sum_{i=1}^{n}(\nabla x_i^m)^t\Big[\theta G(\mathbf{x}^m)+(1-\theta)G(\mathbf{x}^m)\Big]\nabla x_i^m\right\}^{q-1}\\
&\quad\cdot\left\{\sum_{i=1}^{n}(\nabla x_i^m)^t\Big[G(\mathbf{x}^m)-G(\mathbf{x}^*)\Big]\nabla x_i^m\right\}(\xi)d\xi\\
&\le q\|G\|_{L^\infty}\|G(\mathbf{x}^m)-G(\mathbf{x}^*)\|_{L^\infty}\int_{\Omega_c}\sum_{i=1}^{n}|\nabla x_i^m|^{2q}(\xi)d\xi.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\liminf_{m\to\infty}\int_{\Omega_c}\Big[\sum_{i=1}^{n}(\nabla x_i^m)^tG(\mathbf{x}^m)\nabla x_i^m\Big]^q(\xi)d\xi\\
=\;&\liminf_{m\to\infty}\int_{\Omega_c}\Big[\sum_{i=1}^{n}(\nabla x_i^m)^tG(\mathbf{x}^*)\nabla x_i^m\Big]^q(\xi)d\xi\\
\ge\;&\int_{\Omega_c}\Big[\sum_{i=1}^{n}(\nabla x_i^*)^tG(\mathbf{x}^*)\nabla x_i^*\Big]^q(\xi)d\xi
\end{aligned}
$$

where in the last step, we have used the fact that

$$
\int_{\Omega_c}\Big[\sum_{i=1}^{n}(\nabla x_i)^tG(\mathbf{x}^*)\nabla x_i\Big]^q(\xi)d\xi
$$

is semicontinuous as a convex nonlinear functional of $x$. This completes the Lemma when $G\in(C_B(\Omega))^{n\times n}$. For general $G\in(L^\infty(\Omega))^{n\times n}$, the result follows from the standard density argument. □

**Comparisons with harmonic mapping**

Theories based on harmonic mapping have been used extensively in formulating variational mesh generation techniques [80, 37, 106, 131]. The advantages of this approach include

- It is easy to get the existence and uniqueness of the minimizer.
- The Euler-Lagrange equation of this functional is a linear elliptic equation which is easy to solve.

For a comprehensive treatment, we refer to the book [139] and the references therein. Here we only point out that it corresponds to $q = 1$ case in (3.4). Based on the requirement $q > n/2$, we see for one dimension, the harmonic mapping approach is justified by our theory. But for multiple dimensions, it seems that the harmonic mapping only addresses the isotropy property not the equidistribution property. We also note that $n = 2$ is the board line case for $q > n/2$ when harmonic mapping ($q = 1$) is used. Hence the harmonic mapping may still be a reasonable approach in two dimensions but it may not be the case for $n \geq 3$.

**Remarks**

The matrix $G(\mathbf{x})$ used in (3.3) and (3.4) is called *monitor function* in the literature. It is widely used in the existing variational mesh adaptation methods, e.g. see [115, 80, 37, 106, 30, 185]. Based on our interpolation error estimate in Chapter 1, the following choice of monitor function is evident:

$$G(\mathbf{x}) = (\det H)(\mathbf{x})^{-\frac{1}{2p+n}} H(\mathbf{x}). \tag{3.7}$$

The optimization problems have been obtained based on the assumption that the computational grid is quasi-uniform (with respect to the usual Euclidean metric). Now we shall propose a slight modification of the monitor function to relax the quasi-uniform assumption on the computational domain. This modification is useful when the moving mesh method is combined with the local grid refinement method.

Given a computational grid $\mathcal{T}_{N,c}$ which may not be quasi-uniform but is assumed to be shape-regular on the Euclidean metric (namely the all the elements are locally isotropic). For this grid, we introduce a size function $s(\xi) = |\tau_c(\xi)|^{2/n}$, where $\tau_c(\xi) \in \mathcal{T}_{N,c}$ and $\xi \in \tau_c(\xi)$. Noting that $s(\xi)$ is a scalar function, if we use

$$G(\mathbf{x}) = s(\xi(\mathbf{x}))\,[(\det H)(\mathbf{x})]^{-\frac{1}{2p+n}} H(\mathbf{x})$$

in (3.3) or (3.4), the isotropy condition is still satisfied for the minimizer of (3.3) or (3.4). Meanwhile the minimizer is aimed to satisfy

$$s^n(\det J)^2(\det G) = (\det G_c)|\tau_c|^2 = |\tau_c|^2_{G_c} = \text{constant},$$

which means the triangulation satisfies the equidistribution condition.

By using this modified metric, we can use a shape-regular mesh in the computational domain. Conventional isotropic refinement strategies, namely regular division [16, 14] and bisection [159] will result in a regular mesh. Thus it is possible to combine moving mesh ($r$-type adaptive method) and local refinement ($h$-type adaptive method) together to get a more efficient algorithm to solve equations. A combination of those two methods has been studied in [1, 11, 97, 38].

## 3.2 Local mesh optimizations

In this section, we shall discuss local mesh optimization techniques which aim to improve the mesh quality. Here we use the error-based mesh quality for a triangulation $\mathcal{T}$:

$$Q(\mathcal{T}, u, p) = \|u - u_{I,\mathcal{T}}\|_{L^p(\Omega)}, 1 \leq p \leq \infty.$$

We will briefly discuss three types mesh improvements: local refinement or coarsening, edge swapping, and mesh smoothing. We will derive those techniques by minimizing the interpolation error. Our novel mesh smoothing schemes will be presented in the next section.

### 3.2.1 Local refinement and coarsening

We compute edge lengths under the new metric $H_p$ and mark edges whose lengths are greater than $d$, where $d$ is a user defined edge length. We connect marked edges element-wise according to there different situations; see Figure 3.2. Our edge-based refinement will automatically result in a conform triangulation and thus save a lot of work of programming.

The coarsening operates like an inverse procedure of refinement. It marks the one whose length is less than $d$. We then shrink this edge to a point and connect to the vertices of the patch of the edge (see Figure 3.2).

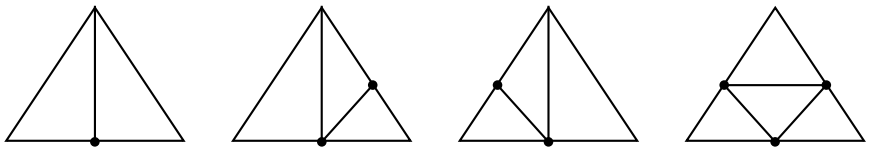


**Figure 3.2.** Edge-based refinement

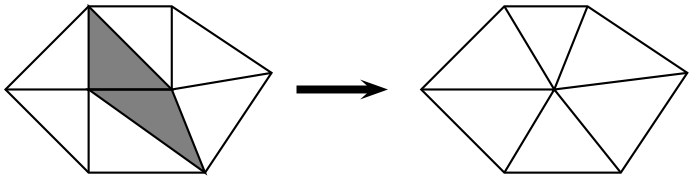


**Figure 3.3.** Coarsening

### 3.2.2 Edge swapping

Now we consider the edge swapping for four points $\{\mathbf{x}_i\}_{i=1}^4$ which form two adjacent triangles and a convex quadrilateral. Let $\mathcal{T}_1 = \triangle_{123} \cup \triangle_{134}$ and $\mathcal{T}_2 = \triangle_{124} \cup \triangle_{234}$, where $\triangle_{ijk}$ stands for the triangle made up by $\mathbf{x}_i, \mathbf{x}_j$, and $\mathbf{x}_k$. We choose triangulation $\mathcal{T}_1$ if and only if $Q(\mathcal{T}_1, u, p) \leq Q(\mathcal{T}_2, u, p)$, for some $1 \leq p \leq \infty$. In [48], we show this criteria is equivalent to the empty circle criteria when $u(\mathbf{x}) = \|\mathbf{x}\|^2$. Thus it is an appropriate generalization of the edge swapping used in the isotropic case to the anisotropic case.

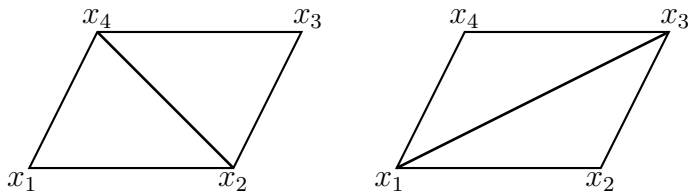



**Figure 3.4.** Edge swapping

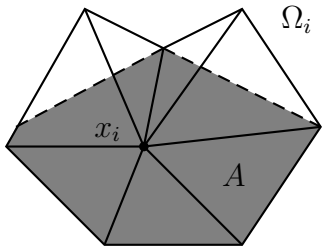



**Figure 3.5.** Feasible region in a local patch

### 3.2.3 Mesh smoothing

Mesh smoothing is a local algorithm which aims to improve the shape regularity, by adjusting the location of a vertex $\mathbf{x}_i$ in its local patch $\Omega_i$, which consists of all simplexes containing $\mathbf{x}_i$, without changing the connectivity. To ensure that the moving will not destroy a valid triangulation, namely non-overlapping or inverted simplexes generated, we restrict to the feasible region $A$, which is the biggest convex set contained in $\Omega_i$ such that $\mathbf{x} \in A$ will not result in overlapping simplexes; see Figure 3.5. Several sweeps through the mesh can be performed to improve the overall mesh quality. A general mesh smoothing algorithm is listed below:

**General mesh smoothing algorithm**

For k=1:step

For i=1:N

$\mathbf{x}^* = smoother(\mathbf{x}_i, \Omega_i)$

If $\mathbf{x}^*$ is acceptable then $\mathbf{x}_i = \mathbf{x}^*$

End

End

The key in the mesh smoothing is the smoother. Namely how to compute the new location by using the information in the local patch. Because the mesh may contain millions of vertices, it is critical that the smoother is computationally inexpensive. Laplacian smoothing, the simplest inexpensive smoother, is to move each vertex to the arithmetic average of the neighboring points.

**Laplacian smoother**

$$\mathbf{x}^* = \frac{1}{k} \sum_{\mathbf{x}_j \in \Omega_i, \mathbf{x}_j \neq \mathbf{x}_i} \mathbf{x}_j, \tag{3.8}$$

where $k$ is the number of vertices of $\Omega_i$. It is low-cost and works in some heuristic way since it is not directly related to most geometrical mesh qualities. Later we will derive Laplacian smoother by minimizing our error-based mesh quality.

An optimization-based smoothing has been proposed in [18, 170, 93]. An objected function $\phi(\mathbf{x})$ is composed by combining the element qualities in the patch. A typical choice [92] is $\phi(\mathbf{x}) = \min_{1 \leq j \leq k} q_j(\mathbf{x})$, where $q_j(\mathbf{x})$ is the quality for simplex $\tau_j \in \Omega_i$. Then one uses the steepest descent optimization or GLP (generalized linear program) [6] to find the optimal point $x^*$.

**Optimization-based smoother**

$$\mathbf{x}^* = \text{argmax}_{\mathbf{x} \in A} \phi(\mathbf{x}). \tag{3.9}$$

The optimization-based smoother is designed to improve the mesh quality and the theoretical results developed for GLP ensure that the expected time for one sweep is a linear function of the problem size [6]. But it is often expensive than Laplacian smoothing. Numerical comparison can be found at [91]. It is worthy noting that in two dimensions Zhou and Shimada [189] proposed an angle-based approach mesh smoothing that strikes a balance between geometric mesh quality and computational cost.

All the mesh smoothing schemes we discussed above are designed for isotropic mesh adaptation. In next section we shall develop several mesh smoothers by minimizing the error-based or the metric-based mesh quality locally, which will be a unified way to derive isotropic and anisotropic mesh smoothing scheme.

## 3.3 Mesh smoothing based on optimal Delaunay triangulations

In this section, we will derive several, old and new, mesh smoothing schemes based on optimal Delaunay triangulations. The main idea is to consider the local opti-

mization problem

$$\min_{\mathbf{x}\in A} Q(\Omega_i, u, 1).$$

Several mesh smoothing schemes are obtained by various formulation of the interpolation error $Q(\Omega_i, u, 1)$.

### 3.3.1 Mesh quality

In the finite element research group, it is believed that the angles of triangles should remain bounded away from 0 and $\pi$. Probably it is based on the following two works. Babuska and Aziz [13] show that as the maximum angle approaches $\pi$, the interpolation error in $H^1$ norm grows. Fried [94] shows that the condition number of the local stiffness matrix grows like $1/\sin\theta$ where $\theta$ is the minimum angle. Hence certain geometric quality measures, for which the equilateral triangle has the best one, can be defined to evaluate the shape of an element to exclude the large and small angle in triangulations. For survey and comparison of a wider range of shape measure in general use, we refer to [67, 86].

On the other hand in order to approximate anisotropic functions, (with sharp boundary layers or internal layers) long thin elements can be good for linear approximation if we measure the error in $L^p$ norm rather than in $H^1$ norm. Based on our interpolation error estimates obtained in Chapter 1, we are going to use error-based mesh quality

$$Q(\mathcal{T}, u, 1) = \int_\Omega |u - u_I|(\mathbf{x}) d\mathbf{x}$$

and density-based mesh quality

$$Q(\mathcal{T}, \rho, 1) = \int_{\Omega_i} |u - u_I|(\mathbf{x})\rho(\mathbf{x}) d\mathbf{x}, \quad \text{with } u(\mathbf{x}) = \|\mathbf{x}\|^2.$$

The function $u$ or density $\rho$ used in the above qualities could be: 1) the solution of some equations when we are solving PDEs; 2) or if we just want to generate a mesh for some domain $\Omega$, $u$ or $\rho$ is related to the curvature of the boundary or local feature of the domain [178].

We now discuss the relation of the error-based mesh quality with commonly

used distortion metric. The following lemma is a special case of Lemma 2.10 in Chapter 1.

**Lemma 3.4.** *For a quadratic function $u$ and a simplex with vertices $\{\mathbf{x}_i\}_{i=1}^n$, we have*

$$Q(\tau, u, 1) = \frac{|\tau|}{2(n+2)(n+1)} \sum_{i,j}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^2_{\nabla^2 u},$$

*where* $\|\mathbf{v}\|_{\nabla^2 u} = \mathbf{v}^T \nabla^2 u \mathbf{v}$.

In particular for $u(\mathbf{x}) = \|\mathbf{x}\|^2$, we have

$$Q(\tau, \|\mathbf{x}\|^2, 1) = \frac{|\tau|}{2(n+2)(n+1)} \sum_{k=1}^{n(n+1)/2} d_k^2 \geq C_n |\tau|^{1+2/n}, \tag{3.10}$$

where $d_k$ is the edge length of the simplex and $C_n = \dfrac{n}{2(n+2)} \dfrac{n!^{2/n}}{(n+1)^{1/n}}$. The equality holds if and only if $\tau$ is equilateral.

The *distortion metric* [18, 24] for a simplex can be defined by the following ratio (or its reciprocal)

$$\frac{\sum_{k=1}^{n(n+1)/2} d_k^2}{\tilde{C}_n |\tau|^{2/n}} = \frac{Q(\tau, \|\mathbf{x}\|^2, 1)}{C_n |\tau|^{1+2/n}},$$

where $\tilde{C}_n$ is used to normalize the quality. The optimization of the distortion metric will lead to equilateral simplexes. The inequality (3.10) also implies that if we minimize the interpolation error directly, the optimal one (when $|\tau|$ is fixed) is an equilateral one. We are going to present new formulas of the interpolation error which help us to develop or improve mesh smoothing schemes.

**Theorem 3.5.** *For a quadratic function $u$, and a triangulation $\mathcal{T}_N$ with $N$ vertices, we have*

$$Q(\mathcal{T}_N, u, 1) = \frac{1}{n+1} \sum_{i=1}^{N} \int_{\Omega_i} \|\mathbf{x} - \mathbf{x}_i\|^2_{\nabla^2 u} d\mathbf{x}. \tag{3.11}$$

*For a density function $\rho$, we denote $\rho_i = \rho\varphi_i$ where $\varphi_i$ is the nodal basis function at $\mathbf{x}_i$. We have*

$$Q(\mathcal{T}_N, \rho, 1) = \sum_{i=1}^{N} \int_{\Omega_i} \|\mathbf{x} - \mathbf{x}_i\|^2 \rho_i(\mathbf{x}) d\mathbf{x}. \tag{3.12}$$

*Proof.* Let $\{\lambda_i(\mathbf{x})\}_{i=1}^{n+1}$ be the barycenter coordinate of $\mathbf{x}$ in the simplex $\tau$. Then $\mathbf{x} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i$ and

$$\begin{aligned}
\sum_{k=1}^{n+1} \int_\tau \|\mathbf{x} - \mathbf{x}_k\|^2_{\nabla^2 u} &= \sum_{i,j,k=1}^{n+1} \int_\tau \lambda_i \lambda_j (\mathbf{x}_i - \mathbf{x}_k)^T \nabla^2 u (\mathbf{x}_j - \mathbf{x}_k) \\
&= \frac{|\tau|}{(n+2)(n+1)} \sum_{i,j,k=1}^{n+1} (\mathbf{x}_i - \mathbf{x}_k)^T \nabla^2 u (\mathbf{x}_j - \mathbf{x}_k) \\
&= \frac{n+1}{2} \frac{|\tau|}{(n+2)(n+1)} \sum_{i,j}^{n+1} \|\mathbf{x}_i - \mathbf{x}_j\|^2_{\nabla^2 u} \\
&= (n+1) Q(\mathcal{T}_N, u, 1).
\end{aligned}$$

The last equality follows from Lemma 3.4. The third one is obtained by summing up the following basic identity with $i, j, k = 1, \cdots, n+1$:

$$\|\mathbf{x}_i - \mathbf{x}_j\|^2_{\nabla^2 u} = \|\mathbf{x}_i - \mathbf{x}_k\|^2_{\nabla^2 u} + \|\mathbf{x}_j - \mathbf{x}_k\|^2_{\nabla^2 u} - 2(\mathbf{x}_i - \mathbf{x}_k)^T \nabla^2 u (\mathbf{x}_j - \mathbf{x}_k).$$

Noting that

$$\sum_{\tau \in \mathcal{T}_N} \sum_{k=1}^{n+1} \int_\tau \|\mathbf{x} - \mathbf{x}_{\tau,k}\|^2_{\nabla^2 u} = \sum_{i=1}^{N} \int_{\Omega_i} \|\mathbf{x} - \mathbf{x}_i\|^2_{\nabla^2 u},$$

we get the identity (3.11).

Recall that for the density-based quality, $u(\mathbf{x}) = \|\mathbf{x}\|^2$. On a simplex $\tau$, we have

$$(u_I - u)(\mathbf{x}) = \sum_{i=1}^{n+1} \lambda_i(\mathbf{x}) \mathbf{x}_i^2 - \mathbf{x}^2 = \sum_{i=1}^{n+1} \lambda_i(\mathbf{x}) \|\mathbf{x}_i - \mathbf{x}\|^2.$$

The desired result (3.12) then follows from the integration using the measure $d\rho$ and the summation. □

It is interesting to note that $\rho = \sum_{i=1}^{N} \rho_i$ since $\sum_{i=1}^{N} \varphi_i = 1$ is a natural partition of unity using the triangulation $\mathcal{T}_N$. Another possible partition of unity is the piecewise constant decomposition corresponding to the dual mesh i.e. Voronoi tessellation; see Appendix A for discussion of Voronoi tessellation.

We now present another formula of $Q(\mathcal{T}_N, u, 1)$ for a convex function $u$ and $Q(\mathcal{T}_N, \rho, 1)$ for a given density.

**Theorem 3.6.** *For a convex function $u$ and a triangulation $\mathcal{T}_N$ with $N$ vertices*

$$Q(\mathcal{T}_N, u, 1) = \frac{1}{n+1}\sum_{i=1}^{N} u(\mathbf{x}_i)|\Omega_i| - \int_\Omega u(\mathbf{x})d\mathbf{x}. \tag{3.13}$$

*For a density function $\rho$ and a triangulation $\mathcal{T}_N$ with $N$ vertices*

$$Q(\mathcal{T}_N, \rho, 1) = \sum_{i=1}^{N} \|\mathbf{x}_i\|^2|\Omega_i|_{\rho_i} - \int_\Omega \|\mathbf{x}\|^2\rho(\mathbf{x})d\mathbf{x}, \tag{3.14}$$

*where* $|\Omega_i|_{\rho_i} = \int_{\Omega_i} \rho_i(\mathbf{x})d\mathbf{x} = \int_{\Omega_i} \varphi_i(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}$.

*Proof.* Note that $u_I(\mathbf{x}) = \sum_{\mathbf{x}_i\in\mathcal{T}} u(\mathbf{x}_i)\varphi_i(\mathbf{x})$ and $u_I(\mathbf{x}) \geq u(\mathbf{x})$ in $\Omega$. We then have

$$\begin{aligned}\int_\Omega |u_I - u|\rho(\mathbf{x})d\mathbf{x} &= \int_\Omega u_I(\mathbf{x})\rho(\mathbf{x})d\mathbf{x} - \int_\Omega u(\mathbf{x})\rho(\mathbf{x})d\mathbf{x} \\ &= \sum_{\mathbf{x}_i\in\mathcal{T}} u(\mathbf{x}_i)\int_\Omega (\varphi_i\rho)(\mathbf{x})d\mathbf{x} - \int_\Omega u(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}.\end{aligned}$$

Choosing $\rho(\mathbf{x}) = 1$ or $u(\mathbf{x}) = \|\mathbf{x}\|^2$, we get (3.13) and (3.14), respectively. □

### 3.3.2 Smoothers based on centroidal patch triangulations

In this subsection we develop mesh smoothing scheme by considering the metric-based mesh quality

$$Q(\mathcal{T}, \rho, 1) = \sum_{i=1}^{N}\int_{\Omega_i} \|\mathbf{x} - \mathbf{x}_i\|^2\rho_i(\mathbf{x})d\mathbf{x}.$$

The nonuniform function $\rho(\mathbf{x})$ is to control the mesh density, which aims to equidistribute the error or the volume of element under this metric, while the isotropic function $u(\mathbf{x}) = \|\mathbf{x}\|^2$ is to improve the shape regularity of elements under the Euclidean metric.

To improve the quality $Q(\mathcal{T}, \rho, 1)$, let us consider the following optimization problem:

$$\min_{\mathbf{z}\in\Omega_i}\int_{\Omega_i} \|\mathbf{x} - \mathbf{z}\|^2\rho_i(\mathbf{x})d\mathbf{x}. \tag{3.15}$$

**Lemma 3.7.** *The minimizer of (3.15) is the centroid of $\Omega_i$ with respect to the density $\rho_i(\mathbf{x})$, namely*

$$\mathbf{x}^* = \frac{\int_{\Omega_i} \mathbf{x}\rho_i(\mathbf{x})d\mathbf{x}}{\int_{\Omega_i} \rho_i(\mathbf{x})d\mathbf{x}}. \tag{3.16}$$

*Proof.* Let $\mathbf{x}^*$ be the centroid of $\Omega_i$ with respect to the density $\rho_i$. For any point $\mathbf{z} \in \Omega_i$, we have

$$\begin{aligned}\int_{\Omega_i} ||\mathbf{x}-\mathbf{z}||^2\rho_i(\mathbf{x})d\mathbf{x} &= \int_{\Omega_i} (\mathbf{x}-\mathbf{x}^*)\cdot(\mathbf{x}-\mathbf{z})\rho_i(\mathbf{x})d\mathbf{x} \\ \leq \Big(\int_{\Omega_i} ||\mathbf{x}-\mathbf{x}^*||^2\rho_i(\mathbf{x})d\mathbf{x}\Big)^{1/2} &\Big(\int_{\Omega_i} ||\mathbf{x}-\mathbf{z}||^2\rho_i(\mathbf{x})d\mathbf{x}\Big)^{1/2}.\end{aligned}$$

Thus

$$\int_{\Omega_i} ||\mathbf{x}-\mathbf{x}^*||^2\rho_i(\mathbf{x})d\mathbf{x} \leq \int_{\Omega_i} ||\mathbf{x}-\mathbf{z}||^2\rho_i(\mathbf{x})d\mathbf{x},$$

which shows $\mathbf{x}^*$ is the minimal point. □

**Definition 3.8.** *For a triangulation $\mathcal{T}$ if for any vertex $\mathbf{x}_i \in \mathcal{T}$, $\mathbf{x}_i$ is also the centroid of its patch $\Omega_i$ with respect to the density $\rho_i$, we call it Centroidal Patch Triangulation (CPT) with respect to the density $\rho$.*

This is a mimic definition of Centroidal Voronoi Tessellations [70] for which the generator and centroid of each Voronoi region coincide. For various application of CVT to the mesh generation and numerical solution of PDEs, we refer to [70, 73, 71, 76, 77, 78, 72, 74, 75]. Especially the smoother based on the CVT in Du and Gunzburger [71] is mostly related work.

Formula 3.16 can be severed as smoother. To further save the computational cost without drastic change in the result, we use one point numerical quadrature to evaluate the integral. Let $\mathbf{x}_\tau$ be the centroid of $\tau$, i.e. $\mathbf{x}_\tau = \sum_{\mathbf{x}_k\in\tau} \mathbf{x}_k/(n+1)$ and $\rho_\tau = \rho_i(\mathbf{x}_\tau) = \rho(\mathbf{x}_\tau)/(n+1)$. We use the numerical quadrature $\int_\tau \mathbf{x}\rho_i(\mathbf{x})d\mathbf{x} \approx \mathbf{x}_\tau\rho_\tau|\tau|$ to get our CPT mesh smoother.

**CPT smoother**

$$\mathbf{x}^* = \frac{\sum_{\tau\in\Omega_i} \mathbf{x}_\tau\rho_\tau|\tau|}{\sum_{\tau\in\Omega_i} \rho_\tau|\tau|}. \tag{3.17}$$

What is the right choice of the density function $\rho_\tau$? It could be *a priori* one. Namely the density is given by the user. For example the mesh around the transi-

tion layer will quickly change from a small size to a much larger size. In practice, especially when solving partial differential equations, the density is given by *a posteriori* error estimate.

The uniform density $\rho = 1$ corresponds to the uniform mesh.

**CPT smoother: uniform density**

$$\mathbf{x}^* = \frac{\sum_{\tau \in \Omega_i} \mathbf{x}_\tau |\tau|}{|\Omega_i|}, \tag{3.18}$$

Another choice of the density function is related to the volume of the element. Recall that the mesh smoothing mainly takes care of the isotropic property of the mesh. It is reasonable to assume that after refinement and coarsening the mesh size is almost equidistributed according to the appropriate mesh density function. Namely $|\tau|_{\rho_\tau E} = \rho_\tau^{n/2}|\tau|$ =constant. Therefore we may choose $\rho_\tau = |\tau|^{-2/n}$. With this choice, the mesh smoother (3.17) is a generalized Laplacian smoother.

**CPT smoother: Laplacian**

$$\mathbf{x}^* = \frac{\sum_{\tau \in \Omega_i} \mathbf{x}_\tau |\tau|^{1-2/n}}{\sum_{\tau \in \Omega_i} |\tau|^{1-2/n}}. \tag{3.19}$$

When $n = 2$, the formula (3.19) is

$$\mathbf{x}^* = \frac{2}{3}\frac{\sum_{j=1}^{k} \mathbf{x}_j}{k} + \frac{1}{3}\mathbf{x}_i,$$

where $k$ is the number of neighbors of vertex $x_i$. It is a lumped Laplacian smoothing. This relation shows that why Laplacian works in some sense.

Since $\cup_i \Omega_i$ is an overlapping decomposition of $\Omega$, the change of $\Omega_i$ will affect other patches and thus the overall metric-based energy $Q(\mathcal{T}, \rho, 1)$ will not necessarily be reduced. In the next subsection we shall use Theorem 3.6 to get better mesh smoothers.

### 3.3.3 Smoothers based on optimal Delaunay triangulations

We will develop mesh smoothers by considering the optimization problem

$$\min_{\mathbf{x}_i \in \Omega_i} Q(\Omega_i, u, 1) \quad \text{or} \quad \min_{\mathbf{x}_i \in \Omega_i} Q(\Omega_i, \rho, 1)$$

We treat $Q(\mathcal{T}, u, 1)$ as a function of the vertices. Let us first compute the gradient $\nabla_{\mathbf{x}_i} Q(\mathcal{T}, u, 1)$.

**Lemma 3.9.** *For a convex function $u$*

$$\nabla_{\mathbf{x}_i} Q(\mathcal{T}, u, 1) = \frac{1}{n+1} \sum_{\tau_j \in \Omega_i} \Big( \nabla_{\mathbf{x}_i} |\tau_j|(\mathbf{x}_i) \sum_{\mathbf{x}_k \in \tau_j, \mathbf{x}_k \neq \mathbf{x}_i} u(\mathbf{x}_k) \Big) + \frac{|\Omega_i|}{n+1} \nabla_{\mathbf{x}_i} u(\mathbf{x}_i).$$

*Proof.* First note that $Q(\mathcal{T}, u, 1) = Q(\Omega_i, u, 1) + Q(\mathcal{T} \backslash \Omega_i, u, 1)$ and the second term does not depends on the position of $\mathbf{x}_i$. Therefore $\nabla_{\mathbf{x}_i} Q(\mathcal{T}, u, 1) = \nabla_{\mathbf{x}_i} Q(\Omega_i, u, 1)$.

Then by Theorem 3.6,

$$Q(\Omega_i, u, 1) = \frac{1}{n+1} \sum_{\tau_j \in \Omega_i} \Big( |\tau_j(\mathbf{x}_i)| \sum_{\mathbf{x}_k \in \tau_j, \mathbf{x}_k \neq \mathbf{x}_i} u(\mathbf{x}_k) \Big) + \frac{|\Omega_i|}{n+1} u(\mathbf{x}_i) - \int_{\Omega_i} u(\mathbf{x}) d\mathbf{x}.$$

Since we only adjust the location of $\mathbf{x}_i$, $\Omega_i$ is fixed and $\int_{\Omega_i} u(\mathbf{x}) d\mathbf{x}$ is a constant. We then get the desired result. □

If the triangulation is optimal, for an interior point $\mathbf{x}_i$, it should be a critical point. Therefore $\nabla_{\mathbf{x}_i} Q(\mathcal{T}, u, 1) = 0$ and we obtain the following theorem.

**Theorem 3.10.** *If the triangulation $\mathcal{T}$ is optimal in the sense of minimizing $Q(\mathcal{T}, u, 1)$ for a convex function $u \in \mathcal{C}^1(\Omega)$, then for an interior vertex $\mathbf{x}_i$, we have*

$$\nabla u(\mathbf{x}_i) = -\frac{1}{|\Omega_i|} \sum_{\tau_j \in \Omega_i} \Big( \nabla_{\mathbf{x}_i} |\tau_j| \sum_{\mathbf{x}_k \in \tau_j, \mathbf{x}_k \neq \mathbf{x}_i} u(\mathbf{x}_k) \Big).$$

When $n = 1$, Theorem 3.10 says that if the grid optimize the interpolation error in $L^1$ norm, it should satisfy

$$u'(x_i) = \frac{u(x_{i+1}) - u(x_{i-1})}{x_{i+1} - x_{i-1}}. \tag{3.20}$$

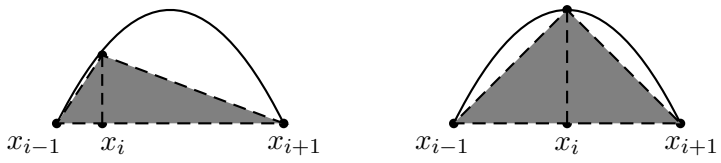



**Figure 3.6.** Moving a grid point in its local patch

We use Figure 3.6 to illustrate (3.20). We move the grid point $x_i$ in its local patch $[x_{i-1}, x_{i+1}]$. It is easy to see that minimizing $Q(\Omega_i, u, 1)$ is equivalent to maximizing the area of the shadowed triangle. Since the base edge is fixed, it is equivalent to maximizing the height. Thus (3.20) holds.

In two dimensions, since

$$|\tau_j|(x,y) = \begin{vmatrix} x_{j+1} - x_j & x - x_j \\ y_{j+1} - y_j & y - y_j \end{vmatrix},$$

we can get a similar formula

$$\begin{aligned} u_x(x_i, y_i) &= \sum_j \omega_j^x u(x_j, y_j), \\ u_y(x_i, y_i) &= \sum_j \omega_j^y u(x_j, y_j), \end{aligned}$$

where

$$\omega_j^x = \frac{y_{j+1} - y_{j-1}}{|\Omega_i|}, \text{ and } \omega_j^y = \frac{x_{j-1} - x_{j+1}}{|\Omega_i|}.$$

The significance of Theorem 3.10 is that we can recover the derivative exactly from the nodal values of the function if the triangulation is optimized. With the gradient information, we can approximate $u$ by higher degree polynomials or construct *a posteriori* error indicator.

If the triangulation is not optimized, Theorem 3.10 can be used to solve the critical point. And the critical point can be used as the new location for the mesh smoother. When $u(\mathbf{x}) = u_H(\mathbf{x}) := \mathbf{x}^T H \mathbf{x}$ is a non-degenerate quadratic function, i.e. $H$ is a $n \times n$ nonsingular matrix. We can solve the critical point exactly and get a mesh smoother based on ODTs.

**ODT smoother: quadratic function**

$$\mathbf{x}^* = -\frac{H^{-1}}{|\Omega_i|} \sum_{\tau_j \in \Omega_i} \Big( \nabla |\tau_j| \sum_{\mathbf{x}_k \in \tau_j, \mathbf{x}_k \neq \mathbf{x}_i} \|\mathbf{x}_k\|_H^2 \Big). \tag{3.21}$$

Since the approximation error only depends on the second derivative,

$$Q(\Omega_i, u_H(\mathbf{x}), p) = Q(\Omega_i, u_H(\mathbf{x} - \mathbf{x}_0), p)$$

for any fixed point $\mathbf{x}_0$. (3.21) can be written as

$$\mathbf{x}^* = \mathbf{x}_0 - \frac{H^{-1}}{|\Omega_i|} \sum_{\tau_j \in \Omega_i} \Big( \nabla |\tau_j| \sum_{\mathbf{x}_k \in \tau_j, \mathbf{x}_k \neq \mathbf{x}_i} \|\mathbf{x}_k - \mathbf{x}_0\|_H^2 \Big). \tag{3.22}$$

When the goal of the mesh adaptation is to get a uniform and shape regular mesh, we choose $u(\mathbf{x}) = \|\mathbf{x}\|^2$ and get the smoother for uniform density.

**ODT smoother: uniform density**

$$\mathbf{x}^* = -\frac{1}{2|\Omega_i|} \sum_{\tau_j \in \Omega_i} \Big( \nabla_{\mathbf{x}_i} |\tau_j| \sum_{\mathbf{x}_k \in \tau_j, \mathbf{x}_k \neq \mathbf{x}_i} \|\mathbf{x}_k\|^2 \Big). \tag{3.23}$$

Recently Alliez et al. [5] express the optimal location as weighed average of circumcenters; see Figure 3.7 for an illustration in two dimensions. We will present the formula after the following two special cases of (3.23).

**Lemma 3.11.** *If vertices of the patch $\Omega_i$ lie on a common circum-sphere, then the optimal location is the center of circum-sphere.*

*Proof.* Let $\mathbf{c}_o$ be the center of the circum-sphere and $R$ its radius. We choose $\mathbf{x}_0 = \mathbf{c}_o$ in (3.22) to get

$$\mathbf{x}^* = \mathbf{c}_o - \frac{nR^2}{|\Omega_i|} \sum_{\tau_j \in \Omega_i} \nabla_{\mathbf{x}_i} |\tau_j| = \mathbf{c}_o - \frac{nR^2}{|\Omega_i|} \nabla_{\mathbf{x}_i} |\Omega_i| = \mathbf{c}_o.$$

The last equality comes from the fact that $|\Omega_i|$ does not depends on $\mathbf{x}_i$. □

As a consequence, (3.23) gives a simple formula to compute the circumcenter of a simplex if we choose $\Omega_i$ as a simplex.

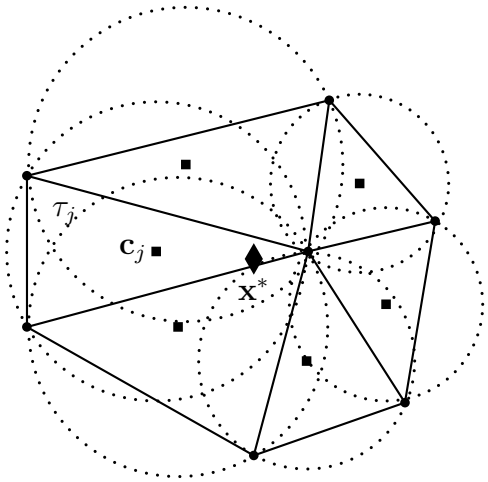



**Figure 3.7.** Weighted average of circumcenters

**Lemma 3.12.** *For a simplex $\tau$ with vertices $\mathbf{x}_i, i = 1, ..., n+1$, the center of the circum-sphere $\mathbf{c}$ is:*

$$\mathbf{c} = -\frac{1}{2|\tau|}\sum_{i=1}^{n+1}\Big(\nabla_{\mathbf{x}_i}|\tau|\sum_{j\neq i, j=1}^{n+1}\|\mathbf{x}_j\|^2\Big). \tag{3.24}$$

**Theorem 3.13.** *Let $\mathbf{c}_j$ be the center of the circum-sphere of $\tau_j$. Then the optimal location $\mathbf{x}^*$ in (3.23) can be written as*

$$\mathbf{x}^* = \frac{\sum_{\tau_j\in\Omega_i}|\tau_j|\mathbf{c}_j}{\Omega_i}. \tag{3.25}$$

*Proof.* Based on the formula of the circum-center (3.24), we get

$$\nabla_{\mathbf{x}_i}|\tau_j|\sum_{\mathbf{x}_k\in\tau_j,\mathbf{x}_k\neq\mathbf{x}_i}\|\mathbf{x}_k\|^2 = -2|\tau_j|\mathbf{c}_j - \sum_{\mathbf{x}_k\in\tau_j,\mathbf{x}_k\neq x_i}\Big(\nabla_{\mathbf{x}_k}|\tau_j|\sum_{\mathbf{x}_j\in\tau_j,\mathbf{x}_j\neq\mathbf{x}_k}\|\mathbf{x}_j\|^2\Big).$$

When we sum the above equality in the patch, the second term will cancel each other. We then get the desired result. □

It is interesting to compare the CPT smoother (3.18) and ODT smoother (3.25) for the uniform density. For an ODT, the node $\mathbf{x}_i$ is a weighted average of circum-

centers while for an CPT it is the average of centroids. This is the main difference of the ODT smoother with the CPT smoothers including Laplacian smoother. Note that the circum-centers are vertices of Voronoi diagram, similar smoother can be obtained based on CVTs. The difference between CVT and ODT smoother will be the weight used in the average.

For non-uniform density $\rho(\mathbf{x})$, if we use piecewise constant approximation of $\rho$ on each simplex, we then formally get the ODT smoother for non-uniform density.

**ODT smoother: non-uniform density**

$$\mathbf{x}^* = \frac{\sum_{\tau\in\Omega_i} \rho_\tau |\tau| \mathbf{c}_\tau}{\sum_{\tau\in\Omega_i} \rho_\tau |\tau|}. \tag{3.26}$$

Again the difference between (3.17) and (3.26) is the centers used in the formula.

When $u$ is a convex quadratic function, the optimization of the interpolation error is a quadratic optimization problem. But the critical point $\mathbf{x}^*$ may be out of the acceptable region $A$. After we compute the critical point $\mathbf{x}^*$ we can further simplify our optimization problem to be

$$\min_{\mathbf{x}\in A} \|\mathbf{x} - \mathbf{x}^*\|^2_{\nabla^2 u}. \tag{3.27}$$

The problem (3.27) is to find the projection (under the metric $\nabla^2 u$) of $\mathbf{x}^*$ to the convex set $A$. Then the projection of $\mathbf{x}^*$ is on the boundary of the patch; see Figure 3.8. For the sake of conformity we need to connect this hanging point to the related points which will reduced the error since $\|\mathbf{x}\|^2$ is convex. For two dimensional triangulations, it looks like we perform an edge swapping before a local smoothing. If the point $\mathbf{x}^*$ is on the boundary of $\Omega$, we will eliminate an element by moving an interior point to the boundary. Conversely a point on the boundary can be moved into the interior. Some boundary points, which are called corner points, are fixed to preserve the geometric shape of the domain. But we free other boundary points. This freedom can change the density of points near the boundary and yield a better mesh since the interpolation error is reduced after each local adjustment.

For the efficiency of algorithm, we only compute the projection when the critical point $\mathbf{x}^*$ is not acceptable. The cost of this algorithm is a little bit higher if we

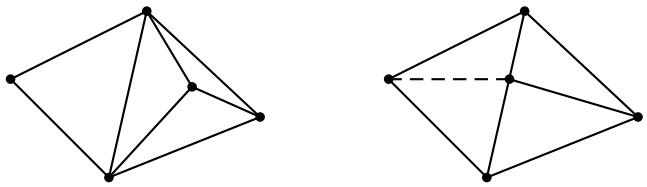


**Figure 3.8.** Moving a point to the boundary of an element

need to compute the projection and change the topological structure of the mesh. But the overall cost for one sweep will not increase too much since it operates like a smart-Laplacian smoothing [170].

For a general function $u$, we can use line search to solve the following optimization problem.

**ODT smoother: general case**

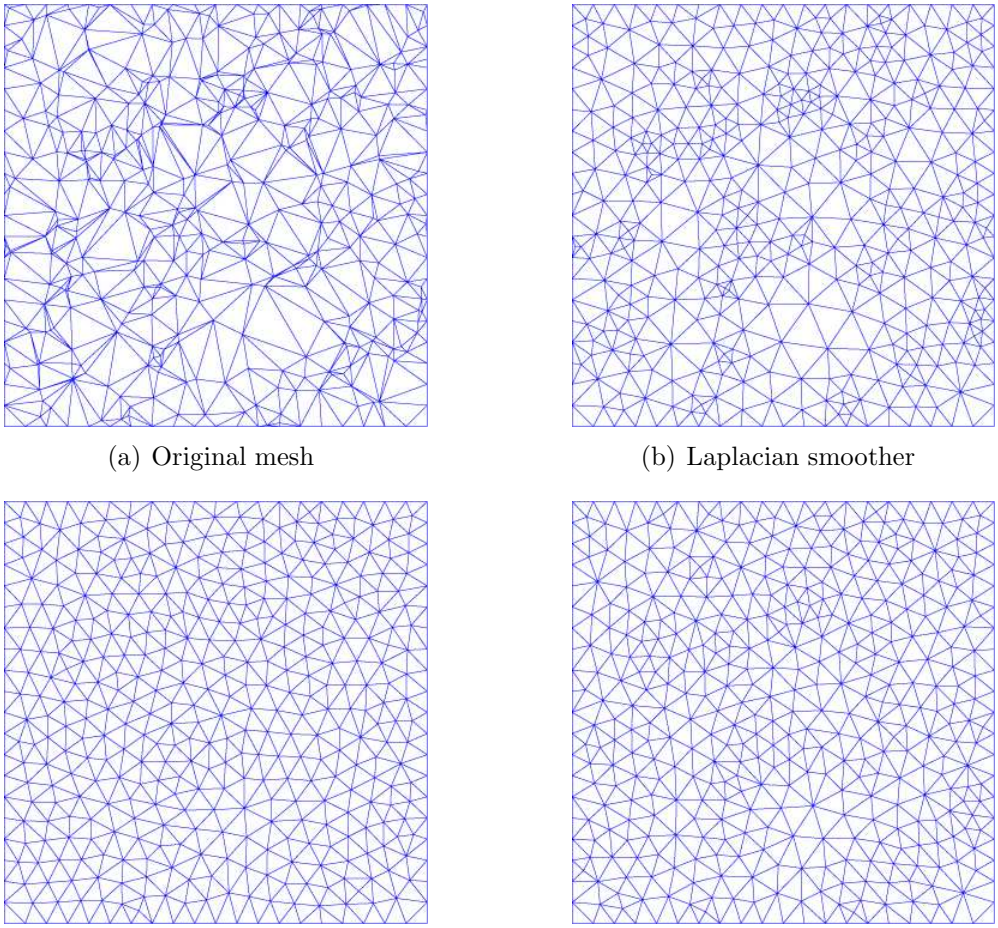
$$x^* = \text{argmin}_{\mathbf{x}\in A} E(\mathbf{x}). \tag{3.28}$$

An alternative approach to solve (3.28) approximately is to compute an average Hessian matrix $H_{\Omega_i}$ in the local patch, and using ODT smoother for the quadratic function $u_q(\mathbf{x}) := \mathbf{x}^T H_{\Omega_i} \mathbf{x}$. This approach is successfully applied in the construction of optimal meshes in [46].

### 3.3.4 Numerical experiments

We shall present several examples in this section to show the efficiency of our new smoothers in the isotropic grid adaptation as well as the anisotropic case.

The first example is to compare our new smoothers with Laplacian smoother for the isotropic grid adaptation and to show the reduction of the interpolation error for those smoothers. We place 20 equally spaced nodes on each edge of the boundary of square $[0, 1] \times [0, 1]$ and 361 nodes in the square. The nodes in the domain are placed randomly while the nodes on the boundary is equally spaced since in this example we only move the interior nodes. We use 'delaunay' command of the Matlab 6.1 to generate the original mesh; see Fig 3.9(a). In this example, the goal of the mesh smoothing is to get an equilateral mesh. Namely triangles are almost equilateral and the density is uniform. We implemented Laplacian smoothing, CPT smoothing (3.18) and ODT smoothing (3.23). In one iteration



(a) Original mesh (b) Laplacian smoother

(c) ODT smoother (d) CPT smoother

**Figure 3.9.** Comparison of Laplacian smoother, CPT smoother and ODT smoother

we apply the mesh smoothing for each node and then do the edge swapping once. We incorporate the edge swapping in our mesh smoothing since it can change the topological structure of the mesh. In practice, the edge swapping always come with the mesh smoothing. We perform 10 iterations and present meshes obtained by different smoothers in Figure 3.9. According to our theory, Laplaican smoothing is not good for the uniform density. Figure 3.9(b) shows that the triangle size is not uniform. We also test CPT smoother (3.18) and ODT smoother (3.23) which are designed for the uniform density. Both of them get better meshes than Laplacian smoothing; see Figure 3.9(c) and 3.9(d).

In Figure 3.10, we plot the interpolation error of each mesh smoother. In this example, $u(\mathbf{x}) = \|\mathbf{x}\|^2$. Therefore we only need to compare $\int_\Omega u_I(\mathbf{x})d\mathbf{x}$ which can

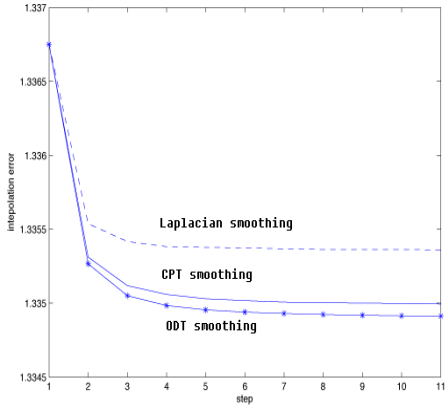



**Figure 3.10.** Error comparison of Laplacian smoother, CPT smoother and ODT smoother

| Step | Laplacian | CPT | ODT |
|---|---|---|---|
| 1 | 0.19 | 0.18 | 0.20 |
| 2 | 0.15 | 0.16 | 0.16 |
| 3 | 0.15 | 0.16 | 0.16 |
| 4 | 0.15 | 0.16 | 0.15 |
| 5 | 0.15 | 0.16 | 0.17 |
| 6 | 0.15 | 0.15 | 0.16 |
| 7 | 0.15 | 0.15 | 0.16 |
| 8 | 0.16 | 0.16 | 0.15 |
| 9 | 0.15 | 0.17 | 0.16 |
| 10 | 0.13 | 0.16 | 0.17 |

**Table 3.1.** Computational cost comparison of Laplacian smoother, CPT smoother and ODT smoother

be evaluated exactly. The initial interpolation error is plotted in the location 'step 1'. Figure 3.10 clearly shows the reduction of the interpolation after each iteration. The ODT smoother is better than the others since it has a provably error reduction property. The numeric convergence of the interpolation errors for those smoothers is very clear from this picture.

The computational cost of those smoothing schemes in each iteration is listed in the Table 3.1. In order to compare the efficiency of the smoothing schemes, we do not include the computational time for the edge swapping in each iteration. Table 3.1 clearly shows that all of those three mesh smoothing schemes have similar computational cost. Thus it is fair to say that ODT smoother is very desirable for

isotropic uniform mesh generation.

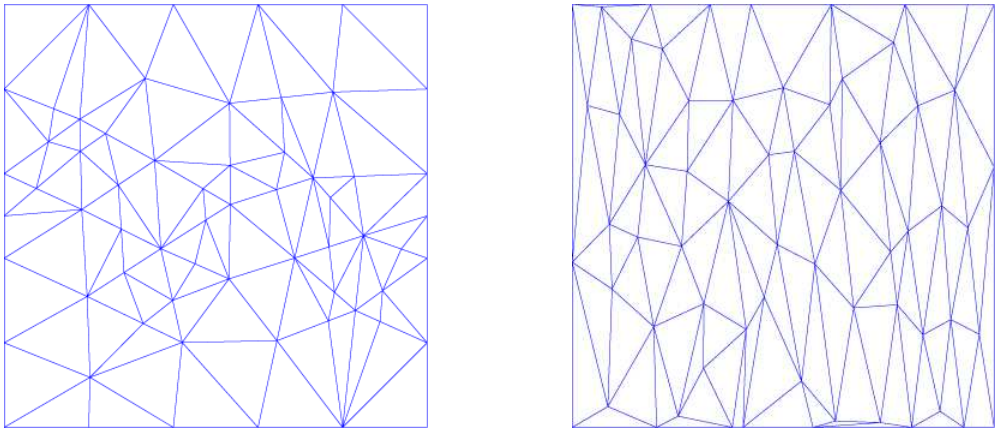
Our second example is to use ODT smoothers to generate an anisotropic mesh. We set $u(x, y) = 10x^2 + y^2$ to be an anisotropic function. The optimal mesh under the Hessian matrix of $u$ should be long and thin vertically. We also include the edge swapping in each iteration. In Figure 3.11 we list several meshes after different iterations. Since the desirable mesh is anisotropic, the number of boundary points on the vertical edges should be much less than that of points on the horizontal edges. Therefore we free the boundary points except four corner points. From those pictures, it is clear that some points are projected to the boundary and some are moved into the square. We also plot the interpolation error in the Figure 3.12. Since the local mesh smoothing is a Gauss-Seidel like algorithm, we see the Gauss-Seidel type convergence result for those mesh smoothing schemes; see Figure 3.10 and 3.12. An ongoing project is to develop a multigrid-like mesh smoothing schemes. It is essentially a multilevel constraint nonlinear optimization problem which is well studied in the literature( see, for example, Tai and Xu [177]). Recent works about multilevel nonlinear optimization problems can be found at [69, 68, 83].

The third example is to show a successful application of the ODT smoother in the anisotropic mesh generation. The function we approximate is

$$f(x, y) = e^{-(\frac{r-0.5}{\epsilon})^2} + 0.5r^2$$

where $r^2 = (x+0.1)^2+(y+0.1)^2$ and $\epsilon = 10^{-3}$. This function changes dramatically at the $\epsilon$ neighborhood of $r = 0.5$. We use offset $(x + 0.1, y + 0.1)$ to avoid the non-smoothing Hessian matrix at $(0, 0)$ and quadratic function $0.5r^2$ to ensure that Hessian matrix is not singular when $r$ is far away from the circle so that we can focus our attention on the interior layer. We use our local refinement, edge swapping and ODT smoother to improve the mesh. Here we present several pictures of our meshes. For the optimality of the $L^p$ norm of the interpolation error, see [46] for details.

We would like to mention a recent work [5] in which ODT smoothing is successfully applied to three dimensional meshing. It has the advantage to produce much less slivers, which are a certain type of badly-shaped tetrahedra, than standard



(a) Original mesh

(b) Mesh after 1 iteration

(c) Mesh after 5 iterations

(d) Mesh after 20 iterations

**Figure 3.11.** Anisotropic meshes obtained by the ODT smoother

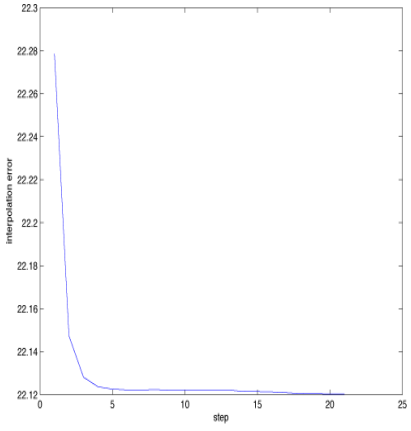
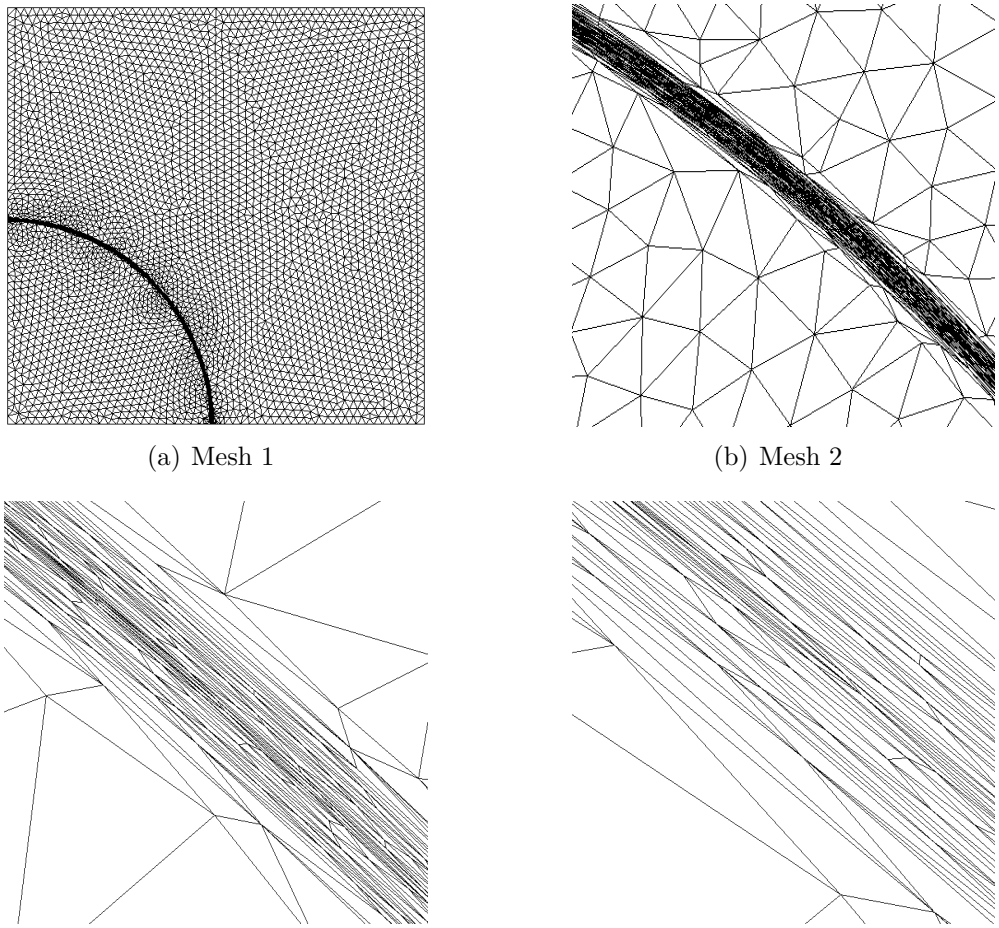



**Figure 3.12.** Interpolation error of the second example



(a) Mesh 1

(b) Mesh 2

(c) Mesh 3

(d) Mesh 4

**Figure 3.13.** Anisotropic meshes

Delaunay refinement. For numerical examples, we refer to [5].

Chapter 4

# Adaptive Finite Element Methods for Convection Dominated Problems

In this Chapter we shall develop a class of adaptive finite element methods (FEMs) for the convection-dominated problems. One simple model is the following convection-diffusion problem:

$$-\varepsilon\Delta u + \mathbf{b}\cdot\nabla u = f, \tag{4.1}$$

which is posed on a bounded domain $\Omega \subset \mathbb{R}^2$ with a proper boundary condition. We are interested in the convection-dominated case, namely $\varepsilon \ll \|\mathbf{b}\|_\infty$.

Due to the small diffusion, the solution to (4.1) has boundary layers or interior layers. It is well known that the standard FEM approximation on quasi-uniform meshes will yield nonphysical oscillations unless the mesh size compares with $\varepsilon$ [163, 148, 146]. To obtain a robust numerical approximation, one approach is to use adapted mesh to capture the layers. Although the right mesh adaptation will improve the stability of the finite element method, we find that the accuracy of the standard FEM depends on the uniformity of the grid in regions where the solution is smooth. Namely when the grid is only quasi-uniform in the smooth part, in general, we can only expect that a first order convergence. In order to achieve the optimal convergence rate and stability simultaneously, we will combine stabilized methods such as streamline diffusion finite element method (SDFEM) [109] with the mesh adaptation especially the anisotropic mesh adaptation to the layers.

In order to apply the mesh adaptation algorithms developed in Chapter 3,

we will use post-processing technique to get the Hessian matrix of the unknown solution $u$ since taking piecewise second derivatives to piecewise linear functions will given no approximation to Hessian matrix. The Hessian recovery algorithm used in the numerical examples is based on a new approach due to Bank and Xu [20, 19] where they use the smoothing iteration of the multigrid method. This scheme proves to be very efficient for recovering Hessian matrix in the isotropic case. It can be extended to anisotropic meshes with some proper modifications and numerical experiments have given satisfactory results.

Another computational difficulty is on the small diffusion $\varepsilon$. If the initial mesh is too coarse compared with $\varepsilon$, the numerical solution would not be able to capture the layers and thus the recovered Hessian is not so accurate. To overcome it, we will use a homotopic method with respect to the diffusion parameter $\varepsilon$. Namely, we first start our computation for large $\varepsilon$, say $\varepsilon = 1$ and use adaptive grid techniques to obtain a good initial grid. We then start to decrease the value of $\varepsilon$ and use the current grid as an initial grid to obtain a new adaptive grid. We continue in this way until the desired value of $\varepsilon$ is reached. We show the success of this approach by numerical examples.

## 4.1 Standard finite element methods

In this section, we shall show the accuracy of the standard FEM highly depends on the uniformity of the mesh in regions where the solution is smooth. Since the equation (4.1) is linear, we first illustrate this point by looking at the problem without boundary layer.

Let us consider the following equation posed on a domain $\Omega \subset \mathbb{R}^2$:

$$\begin{cases} -\varepsilon \Delta u + \mathbf{b} \cdot \nabla u = & f, \quad x \in \Omega, \\ u = & g, \quad x \in \partial\Omega. \end{cases} \tag{4.2}$$

The weak formulation of (4.2) is to find $u \in H^1(\Omega)$ such that $u = g$ on $\partial\Omega$ and

$$a(u, v) = (f, v), \quad \forall\, v \in H_0^1(\Omega),$$

where

$$a(u,v) = \varepsilon \int_\Omega \nabla u \nabla v + \int_\Omega (\mathbf{b}\cdot\nabla u)v, \text{ and } (f,v) = \int_\Omega fv.$$

Given a triangulation $\mathcal{T}$ of the domain $\Omega$, let $V^h$ be the piecewise linear and continuous finite element space on $\mathcal{T}$. The standard FEM is to find $u_h \in V^h$ such that

$$a(u_h, v_h) = (f, v_h), \quad \forall\, v_h \in V_h \cap H_0^1.$$

## Numerical examples

In the following numerical examples, we fix $\varepsilon = 10^{-8}$ which is small enough to observe many interesting results. We choose

$$\mathbf{b} = (1,0), f = -6\varepsilon(x+y) + \mathbf{b}\cdot(3x^2, 3y^2), \text{ and } g = x^3 + y^3,$$

such that the real solution of (4.2) is $u = x^3 + y^3$. The underlying triangulation of $\Omega$ is obtained in the following way. We first decompose the domain into uniform rectangles and then divide each rectangle into triangles by its diagonal. All the diagonals used are in the same direction.

We apply standard FEM on uniform meshes to obtain the numerical approximation $u_h$. The error will be measured in $L^\infty$ and $L^2$ norms. We use $N$ to denote the number of unknowns, i.e. the number of interior grid points. The convergence rate is computed by the formula $\ln \|u - u_h\| / \ln N$. Since the real solution of (4.2) is smooth, by classic interpolation error estimates, for nodal interpolation $u_I$, $\|u - u_I\|_\infty \le CN^{-1}$. We may expect that the standard FEM approximation $u_h$ will also give the same order approximation, i.e. $\|u - u_h\|_\infty \le CN^{-1}$. However our computation will show for the convection dominated problems, the stability and accuracy of the standard FEM is subtle.

In the classic FEM error estimate, we use mesh size $h$ to measure the convergence rate. For uniform meshes, $h^2 = N^{-1}$ and $\|u - u_I\| \le Ch^2$. For this reason, we say the optimal convergence rate is second order.

| $N$ | $\|u-u_h\|$ | Rate | $\|u-u_h\|_\infty$ | Rate |
|---|---|---|---|---|
| 441 | 0.26478E+02 | 0.53807 | 0.52990E+02 | 0.65201 |
| 961 | 0.59294E+01 | 0.25916 | 0.11893E+02 | 0.36051 |
| 2025 | 0.13909E+01 | 0.04334 | 0.28173E+01 | 0.13605 |
| 3969 | 0.37183E+00 | 0.11939 | 0.77940E+00 | 0.03008 |
| 7921 | 0.96102E-01 | 0.26092 | 0.22641E+00 | 0.16547 |
| 16129 | 0.25373E-01 | 0.37922 | 0.71837E-01 | 0.27181 |

**Table 4.1.** Errors of standard FEM on uniform meshes: odd unknowns

| $N$ | $\|u-u_h\|$ | Rate | $\|u-u_h\|_\infty$ | Rate |
|---|---|---|---|---|
| 484 | 0.97785E-03 | 1.12101 | 0.29258E-02 | 0.94373 |
| 1024 | 0.20663E-02 | 0.89188 | 0.66374E-02 | 0.72352 |
| 2116 | 0.23770E-03 | 1.08974 | 0.76847E-03 | 0.93651 |
| 4096 | 0.13634E-03 | 1.07004 | 0.49990E-03 | 0.91384 |
| 8100 | 0.65584E-04 | 1.07029 | 0.23122E-03 | 0.93028 |
| 16384 | 0.39648E-04 | 1.04446 | 0.16033E-03 | 0.90048 |

**Table 4.2.** Errors of standard FEM on uniform meshes: even unknowns

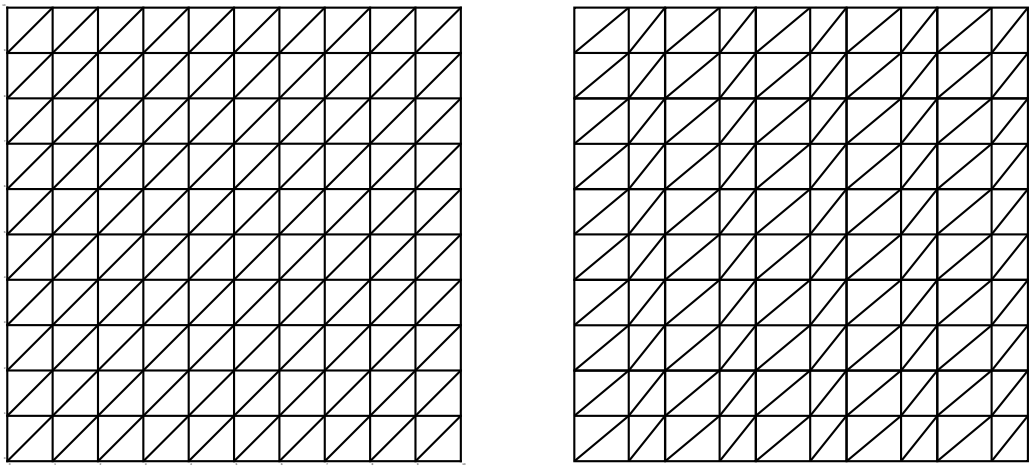
## Even and Odd unknowns.

We first apply standard FEM on uniform meshes. Table 4.1 and 4.2 contain the result for even unknowns and odd unknowns. It is clear that when the number of unknowns is odd, the standard FEM is not stable. Whe the number of unknowns is even, it is of order $N^{-1}$ as expected.

This example shows that for a smooth solution and uniform meshes, when $\varepsilon$ is small, the stability of the scheme depends on the parity of the grid. An intuitive way to understand this interesting phenomenon is to consider the limiting case as

| $N$ | $\|u-u_h\|$ | Rate | $\|u-u_h\|_\infty$ | Rate |
|---|---|---|---|---|
| 484 | 0.90014E-02 | 0.76194 | 0.23730E-01 | 0.60514 |
| 1024 | 0.72538E-02 | 0.71070 | 0.22888E-01 | 0.54493 |
| 2116 | 0.44563E-02 | 0.70697 | 0.11733E-01 | 0.58054 |
| 4096 | 0.32361E-02 | 0.68929 | 0.87461E-02 | 0.56976 |
| 8100 | 0.23147E-02 | 0.67430 | 0.60969E-02 | 0.56669 |
| 16384 | 0.16363E-02 | 0.66109 | 0.44221E-02 | 0.55865 |

**Table 4.3.** Errors of standard FEM on a special perturbed mesh



(a) Uniform mesh (b) Perturbed mesh

**Figure 4.1.** Uniform mesh and its perturbation

$\varepsilon$ goes to zero, that is $-\mathbf{b}\cdot\nabla u = f$. Both the PDE operator and the corresponding discritization matrix are skew symmetric. Thus if the dimension of the matrix is odd, it has a zero eigenvalue. The corresponding eigenvector spans the kernel of the discrete problem which makes the scheme unstable. This is also true for high order elements.

## Perturbation of the uniform mesh

Let us take the even unknowns such that the scheme is stable. But we perturb our mesh in the following way. We make a small rightward perturbation on those odd grid points on the $x$-direction (convectional direction), say $(4N)^{-1}$. Figure 4.1(a) is the uniform mesh and Figure 4.1(b) is its perturbation. The approximation error on the perturbed mesh is listed in Table 4.3.

This example shows that the accuracy of the scheme depends on the uniformity of the grid in the smooth part of the solution. When the grid is only quasi-uniform in the smooth part, in general, we can only expect that $\|u - u_h\| \leq CN^{-1/2}$, while the interpolation error is $\|u - u_I\| \leq CN^{-1}$.

### Real solution with boundary layers

Let us take $u = x^2y^2(1-e^{-\frac{1-x}{\varepsilon}})(1-e^{-\frac{1-y}{\varepsilon}})$ as the real solution of (4.2). By plugging this real solution back into (4.2), we then get the right-hand side $f$ and Dirichlet boundary condition $g$ as follows:

$$\begin{aligned} f &= 2xy^2e^{\frac{-1+x}{\varepsilon}} - 2xy^2e^{\frac{-2+x+y}{\varepsilon}} + 2x^2ye^{\frac{-1+y}{\varepsilon}} - 2x^2ye^{\frac{-2+x+y}{\varepsilon}} + 2y^2\varepsilon e^{\frac{-1+y}{\varepsilon}} \\ &\quad +2y^2\varepsilon e^{\frac{-1+x}{\varepsilon}} + 2x^2\varepsilon e^{\frac{-1+y}{\varepsilon}} + 2x^2\varepsilon e^{\frac{-1+x}{\varepsilon}} + 2x^2y - 2y^2\varepsilon e^{\frac{-2+x+y}{\varepsilon}} - 2y^2\varepsilon \\ &\quad -2x^2\varepsilon e^{\frac{-2+x+y}{\varepsilon}} - 2x^2\varepsilon - 2xy^2e^{\frac{-1+y}{\varepsilon}} - 2x^2ye^{\frac{-1+x}{\varepsilon}} + 2xy^2 \\ g &= x^2y^2(1-e^{-\frac{1-x}{\varepsilon}})(1-e^{-\frac{1-y}{\varepsilon}}) \end{aligned}$$

where we take $\varepsilon = 10^{-4}, b_1 = 1, b_2 = 1$. Again $\Omega = (0,1)^2$.

To solve the boundary layers, we will use Bakhvalov mesh. Given an integer $M$, let $\tau = 1 - 2\varepsilon \ln M$ be the transition point. When $x \in (\tau, 1)$ the $x$-coordinates of grid points is given by

$$x_i = 1 - \tau\varepsilon \ln(1 - (1-\varepsilon)\frac{M+1-i}{M}).$$

The same rule is applied for $y$ direction. All the other parts of mesh are uniform meshes.

By solving this convection dominated problem with the standard FEM on a series of refined Bakhvalov triangle meshes, we get the numerical results listed in Table 4.1.

We then perturb the mesh by slightly rightward and upward perturbing on those odd grid points along the $x, y$-direction, say $h/4$, the convergence rate is decrease.

## 4.2 Streamline diffusion finite element methods

Based on examples in the previous section, it is necessary to invoke some form of upwinding to stabilize the standard FEM. In this section we will give a brief introduction of the streamline diffusion finite element method (SDFEM) introduced first by Hughes and Brooks in [109]. We then derive it from residual bubble free approach following Brezzi et. al. [31].

| $N$ | $\|u_I - u_h\|$ | Rate |
|---|---|---|
| 841 | 1.39505E-03 | 0.97627 |
| 1521 | 7.54577E-04 | 0.98119 |
| 2401 | 4.67480E-04 | 0.98516 |
| 3481 | 3.33972E-04 | 0.98153 |
| 4761 | 2.33654E-04 | 0.98741 |
| 6241 | 1.81715E-04 | 0.98560 |
| 7921 | 2.03149E-04 | 0.94700 |
| 9801 | 1.43943E-04 | 0.96255 |

**Table 4.4.** Errors of standard FEM on Bakhvalov mesh

| $N$ | $\|u_I - u_h\|$ | Rate |
|---|---|---|
| 841 | 1.93581E-02 | 0.58572 |
| 1521 | 8.97218E-03 | 0.64331 |
| 2401 | 8.90839E-03 | 0.60649 |
| 3481 | 4.85244E-03 | 0.65336 |
| 4761 | 4.79265E-03 | 0.63067 |
| 6241 | 2.7745E-03 | 0.67368 |
| 7921 | 2.44691E-03 | 0.66979 |
| 9801 | 1.91189E-03 | 0.68112 |

**Table 4.5.** Errors of standard FEM on a perturbed Bakhvalov mesh

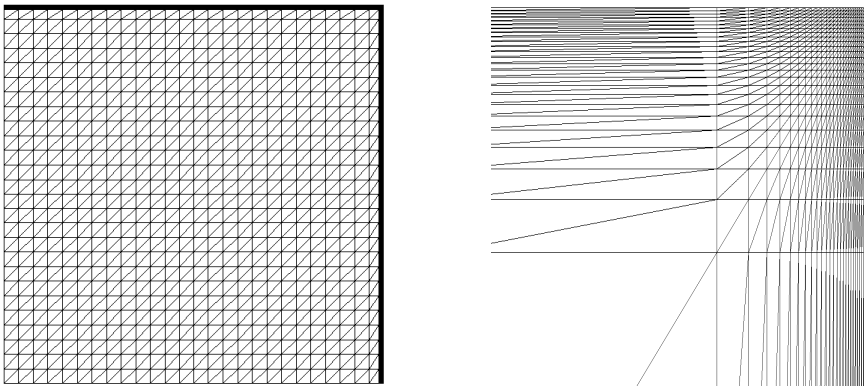


**Figure 4.2.** Bakhvalov mesh

The bilinear form of the standard FEM $a(\cdot,\cdot)$ is weak positive due to the small diffusion $\varepsilon$ and thus suffers from a lack of stability. The SDFEM is to add diffusion in the convection direction to enhance the stability. It reads as: find $u_h \in V^h$ such

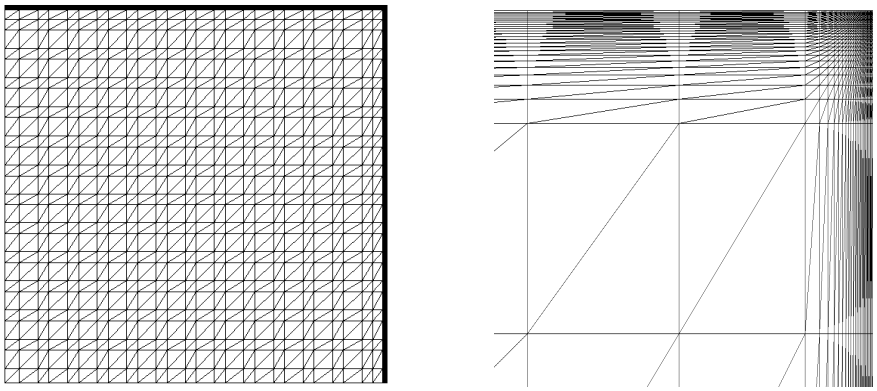


**Figure 4.3.** Perturbed Bakhvalov mesh

| $N$ | $\|u-u_h\|$ | Rate | $\|u-u_h\|_\infty$ | Rate |
|---|---|---|---|---|
| 441 | 0.51185E-03 | 1.24445 | 0.93345E-03 | 1.14577 |
| 961 | 0.25458E-03 | 1.20500 | 0.48249E-03 | 1.11191 |
| 2025 | 0.12726E-03 | 1.17810 | 0.24844E-03 | 1.09024 |
| 3969 | 0.67087E-04 | 1.15969 | 0.13402E-03 | 1.07618 |
| 7921 | 0.34423E-04 | 1.14476 | 0.70148E-04 | 1.06546 |
| 16129 | 0.17197E-04 | 1.13236 | 0.35612E-04 | 1.05723 |

**Table 4.6.** Errors of SDFEM on uniform meshes: odd unknowns

that

$$a(u_h, v_h) + \sum_{\tau\in\mathcal{T}} \delta_\tau \int_\tau (\mathbf{b}\cdot\nabla u_h - f)(\mathbf{b}\cdot\nabla v_h) = (f, v_h), \quad \forall v_h \in V_h \cap H_0^1.$$

Here $\delta_\tau$ is a stabilization parameter and we will discuss its choice in next subsection. Nothing that we can rewrite it as: to find $u_h \in V^h$ such that

$$a(u_h, v_h + \delta_\tau \mathbf{b}\cdot\nabla v_h) = (f, v_h + \delta_\tau \mathbf{b}\cdot\nabla v_h), \quad \forall v_h \in V_h \cap H_0^1.$$

It is the standard FEM applied to an augment test space and thus this method was also referred to as streamline upwind Petrov-Galerkin (SUPG) method. We compute examples in previous sections using SDFEM with $\delta_\tau = h_\tau$ the diameter of $\tau$. We list errors in the following tables. It is clear that SDFEM approximation is stable and accurate for the smooth solution.

| $N$ | $\|u-u_h\|$ | Rate | $\|u-u_h\|_\infty$ | Rate |
|---|---|---|---|---|
| 484 | 0.47172E-03 | 1.23892 | 0.86349E-03 | 1.14113 |
| 1024 | 0.24017E-03 | 1.20237 | 0.45645E-03 | 1.10973 |
| 2116 | 0.12210E-03 | 1.17675 | 0.23877E-03 | 1.08916 |
| 4096 | 0.65089E-04 | 1.15893 | 0.13014E-03 | 1.07564 |
| 8100 | 0.33683E-04 | 1.14433 | 0.68680E-04 | 1.06516 |
| 16384 | 0.16935E-04 | 1.13212 | 0.35081E-04 | 1.05707 |

**Table 4.7.** Errors of SDFEM on uniform meshes: even unknowns

| $N$ | $\|u-u_h\|$ | Rate | $\|u-u_h\|_\infty$ | Rate |
|---|---|---|---|---|
| 484 | 0.52852E-03 | 1.22053 | 0.11069E-02 | 1.10096 |
| 1024 | 0.27048E-03 | 1.18522 | 0.57888E-03 | 1.07544 |
| 2116 | 0.13795E-03 | 1.16081 | 0.30519E-03 | 1.05711 |
| 4096 | 0.73689E-04 | 1.14402 | 0.16633E-03 | 1.04614 |
| 8100 | 0.38189E-04 | 1.13038 | 0.87471E-04 | 1.03829 |
| 16384 | 0.19220E-04 | 1.11907 | 0.44782E-04 | 1.03191 |

**Table 4.8.** Errors of SDFEM on a special perturbed mesh

## Residual free bubbles approach

We begin with the augment of the finite element space. Let $V^b := \oplus_{\tau\in\mathcal{T}} H_0^1(\tau)$. $V^b$ will take care of the fine scale in each element $\tau$ and thus improve the approximation. The standard FEM applied to the space $V^h \bigoplus V^b$ is to find $u_h+u_b \in V^h \bigoplus V^b$ such that

$$a(u_h+u_b, v_h) = (f, v_h), \quad \forall v_h \in V^h \cap H_0^1 \tag{4.3}$$

$$a(u_h+u_b, v_b) = (f, v_b), \quad \forall v_b \in V^b. \tag{4.4}$$

The space $V^h \bigoplus V^b$ is not finite dimensional. But we do not need to compute the fine scale part $u_b$. Instead we will approximate the affection $a(u_b, v_h)$ using the equation (4.4) and substitute it into (4.3).

Let us define $\mathcal{L}u := -\varepsilon u + \mathbf{b}\cdot\nabla u$. The operator should be understood in the weak sense, i.e. $(\mathcal{L}u, v) = a(u, v)$. The discrete operator is defined similarly as $(\mathcal{L}_h u_h, v_h) = a(u_h, v_h)$. The dual of $\mathcal{L}$ is defined by $(\mathcal{L}u, v) = (u, \mathcal{L}^* v)$. Formally $\mathcal{L}^* v = -\varepsilon v - \mathbf{b}\cdot\nabla v$ if $\mathbf{b}$ is constant. Using those notations, we can rewrite (4.3)

as

$$a(u_h, v_h) + (u_b, \mathcal{L}^* v_h) = (f, v_h). \tag{4.5}$$

and (4.4) as

$$a(u_b, v_b) = -(\mathcal{L}u_h - f, v_b),\ \forall v_b \in V^b. \tag{4.6}$$

For further simplification, we assume both $\mathbf{b}$ and $f$ are piecewise constant on each element $\tau$. Since we use piecewise linear element, $\mathcal{L}u_h$ is also piecewise constant. The solution to the equation (4.6) can be written as $u_b = \sum_\tau C_\tau u_\tau$, with $C_\tau = -(\mathcal{L}u_h - f)_{|\tau}$ and $u_\tau$ satisfies

$$\begin{cases} -\varepsilon \Delta u_\tau + \mathbf{b}_{|\tau} \cdot \nabla u_\tau = & 1 \quad \text{in } \tau, \\ u_\tau = & 0 \quad \text{on } \partial\tau. \end{cases} \tag{4.7}$$

The effect

$$\begin{aligned} (u_b, \mathcal{L}^* v_h) &= \sum_\tau (C_\tau u_\tau, \mathcal{L}^* v_h)_{|\tau} = \sum_\tau C_\tau \mathcal{L}^* v_h (u_\tau, 1)_{|\tau} \\ &= \sum_\tau \frac{1}{|\tau|} \int_\tau u_\tau dx (C_\tau, \mathcal{L}^* v_h)_{|\tau} = \sum_\tau \bar{u}_\tau (\mathcal{L}u_h - f, \mathbf{b} \cdot \nabla v_h)_{|\tau}, \end{aligned}$$

where

$$\bar{u}_\tau = \frac{1}{|\tau|} \int_\tau u_\tau dx. \tag{4.8}$$

Thus (4.5) becomes

$$a(u_h, v_h) + \sum_{\tau \in \mathcal{T}} \bar{u}_\tau \int_\tau (\mathbf{b} \cdot \nabla u_h - f)(\mathbf{b} \cdot \nabla v_h) = (f, v_h),\ \forall v_h \in V_h \cap H_0^1,$$

which is the SDFEM with $\delta_\tau = \bar{u}_\tau$.

A priori error estimate of residual free bubble finite element on general quasi-uniform meshs can be found at [32, 33]. Since quasi-uniform meshes cannot solve the layers, we shall combine anisotropic mesh adaptation to further improve the accuracy of the SDFEM.

## 4.3 Post-processing of the Hessian matrix

In this section, we will discuss how to obtain the Hessian matrix of the solution when linear finite element approximation is used for the discretization of partial differential equations. Since taking piecewise second derivatives to piecewise linear functions will given no approximation to Hessian matrix, special post-processing techniques need to be used to obtain reasonable Hessian matrix approximation from linear finite elements.

One most popular technique is the ZZ patch recovery scheme proposed by Zikienwicz-Zhu [191, 190]. We will present a simple example. Noting that to define an element in $V^h$, it suffices to determine nodal values. We will define an averaging operator on the local patch $\Omega_i$

$$A_h(\nabla u_h)(x_i) = \sum_{\tau \in \Omega_i} \omega_\tau \nabla u_h|_\tau, \quad \text{and} \quad A_h(\nabla u_h) = \sum A_h(\nabla u_h)(x_i)\phi_i,$$

where $\phi_i$ is the nodal basis function and $\omega_\tau$ is the weight. A natural choice is the normalized volume.

$$A_h(\nabla_h)(x_i) = \sum_{\tau \in \Omega_i} \frac{|\tau|}{|\Omega_i|} \nabla u_h|_\tau \tag{4.9}$$

(4.9) is the simplest ZZ recovery scheme. It can be thought as the best constant approximation in $\Omega_i$ for the piecewise constant function $\nabla u_h$. General ZZ type recovery scheme is to take the least square fitting of $\nabla u_h$ in its local patch.

We can also use values of $\nabla u_h$ to get a linear least square fitting of $u$ itself and then take the gradient to approximate $\nabla u$. This leads to the so-called polynomial preserving recovery studied by Zhang and Naga [187].

A global $L^2$ projection $Q_h$ can also serve as a recovery operator. Namely

$$(Q_h(\nabla u_h), v_h) = (\nabla u_h, v_h), \quad \forall v_h \in (V^h)^2.$$

Although the $L^2$ projection operator is global, the overall work estimate is still $O(N)$ for a mesh with $N$ vertices.

The theoretical reason for ZZ method to work is largely understood to be related to the superconvergence phenomenon for second order elliptic boundary value problems discretized on a finite element grid that has certain local symmetry, see

Walhbin [180], Chen and Huang [41], Babuska and Strouboulis [14]. These classic superconvergence results can be used to justify the effectiveness of Zikienwicz-Zhu scheme, see [186, 130] for nearly structured grids. A significant improvement of this type of analysis was given recently by Bank and Xu [20, 19]. In [19] they gave superconvergence estimates for piecewise linear finite element approximation on quasi-uniform triangular meshes where most pairs of triangles sharing a common edge form $O(h^2)$ approximate parallelograms except a region with $O(h^\sigma)$ area. Based on this superconvergence result, they show that $Q_h\nabla u_h$ is a superconvergent approximation to $\nabla u$:

$$\|\nabla u - Q_h\nabla u_h\|_{0,\Omega} \lesssim h^{1+\min(1,\sigma)}|\log h|^{1/2}\|u\|_{3,\infty,\Omega}. \tag{4.10}$$

This result leads to a theoretical justification of ZZ method for such type of grids, see Xu and Zhang [183]. Recently Chen [44] generalize the superconvergence result to three dimensions tetrahedral linear finite elements. Recently, Carstensen and Bartels [21, 40] also gave theoretical and numerical support for the robust reliability of all averaging techniques on unstructured grids.

For totally unstructured meshes, it is indeed possible to devise certain averaging process to obtain superconvergence results. Here is an illustration of the main idea. Let $u_h$ be an approximation of $u$(for any problems) satisfy

$$(u-u_h)(x) = O(h^2) \text{ and } \nabla(u-u_h)(x) = O(h).$$

We have

$$\begin{aligned}
\partial_i u(x) &= \frac{1}{2H}\Big(u(x+He_i) - u(x-He_i)\Big) + O(H^2) \\
&= \frac{1}{2H}\Big(u_h(x+He_i) - u_h(x-He_i)\Big) + O(H^2 + \frac{h^2}{H}) \\
&= \frac{1}{2H}\int_{[x-He_i,x+He_i]} \partial_i u_h + O(H^2 + \frac{h^2}{H}) \\
&= \frac{1}{2H}\int_{[x-He_i,x+He_i]} \partial_i u_h + O(h^{4/3}) \quad \text{if } H = O(h^{2/3}).
\end{aligned}$$

It implies that if we average $u_h$ to a larger neighborhood of $\tau$, we may get

a superconvergent approximation of $\nabla u$. Based on this observation Hoffmann, Schatz and Wahlbin give a class of recovery schemes [103, 168].

Another averaging approach is taken by Bank and Xu [20, 19]. When the mesh does not satisfy the $O(h^2)$ parallelogram property or $\sigma$ becomes very close to zero, then the superconvergence demonstrated in (4.10) will be diminished. Intuitively, it appears that this is due mainly to high frequency errors introduced by the small nonuniformities of the mesh. Preferentially attenuating high frequency errors in mesh functions is of course a widely studied problem in multilevel iterative methods. Thus, to enhance the superconvergence effect on general shape regular meshes, we compute $S^m Q_h u_h$, where $S$ is an appropriate multigrid-like smoothing operator. In [20] Bank and Xu shows that

$$\|\nabla u - S^m Q_h \nabla u_h\| \leq C(u) h \Big( m h^{1/2} + \varepsilon_m \Big),$$

where $1/2 < \alpha \leq 1$ and

$$\varepsilon_m = [\kappa/(\kappa+1]^m \ \ (\text{for } m \leq \tfrac{\kappa\alpha}{2}) \quad \text{and } (2m+1)^{-\alpha/2} \ (\text{for } m > \tfrac{\kappa\alpha}{2}).$$

Furthermore we can obtain reasonable Hessian approximation by taking derivatives again:

$$\|\partial_i(\partial_k u - S^m Q_h \partial_k u_h))\| \leq C(u) \Big( m h^{1/2} + \varepsilon_m \Big).$$

The gradient recovery algorithm used in the numerical examples of this Chapter is based on this approach. It can be extended to anisotropic grids with some proper modifications, but a theoretical justification of such extensions is still lacking. Nevertheless, numerical experiments have given satisfactory results.

## 4.4 Multilevel homptotic adaptive finite element methods

Our multilevel homotopic adaptive FEM for the convection dominated problem $\mathcal{L}^\varepsilon u := -\varepsilon \Delta u + \mathbf{b} \cdot \nabla u = f$ can be roughly described as follows.

Given $\rho = \varepsilon_0 \gg \varepsilon$ and $h = h_0$, we generate an initial quasi-uniform mesh $\mathcal{T}_h$.

1. Discretize the equation $\mathcal{L}^{\rho}u = f$ on the mesh $\mathcal{T}_h$ using SDFEM and solve it to get the solution $u_h$.

2. Post-processing $u_h$ to get a recovered Hessian matrix $\partial_i S^m Q_h \partial_k u_h$.

3. Global or local move $\mathcal{T}_h$ using the recovered Hessian matrix.

4. If $\rho = \varepsilon$, locally improve the grid using the recovered Hessian matrix several times and then stop. Otherwise go to (5).

5. Global refine $\mathcal{T}_h$, and set $\rho = \rho/2$ and $h \leftarrow h/2$. Go to (1).

We will list two examples below to show the success of this approach.

**A simple example**

Let us consider the following convection dominated model problem:

$$\begin{cases} -\varepsilon\Delta u + u_x = 1 \quad , & x \in \Omega, \\ u = 0 \quad , & x \in \partial\Omega. \end{cases} \tag{4.11}$$

on the unit square domain $\Omega = (0,1)^2$ with $\varepsilon = 0.001$. By applying above algorithm we solve this convection-diffusion problem and try to catch out layers of the solution by means of moving mesh and local optimized mesh. The following pictures in Fig. 4.4 describe that how the adaptive mesh and numerical solution change in this multilevel homotopic adaptive process step by step.

**Fluid well example**

We consider a convection- iffusion problem on a rectangular domain with an inner disk excluded. The equation is listed below and the domain is given in Figure 4.4.

$$\begin{aligned} -\varepsilon\Delta u + u_x &= 0 \text{ in } \Omega, \\ \frac{\partial u}{\partial n} &= 0 \text{ on the right boundary }, \\ u &= 1 \text{ on the inner circle}, \\ u &= 0 \text{ the rest of the boundary}. \end{aligned}$$

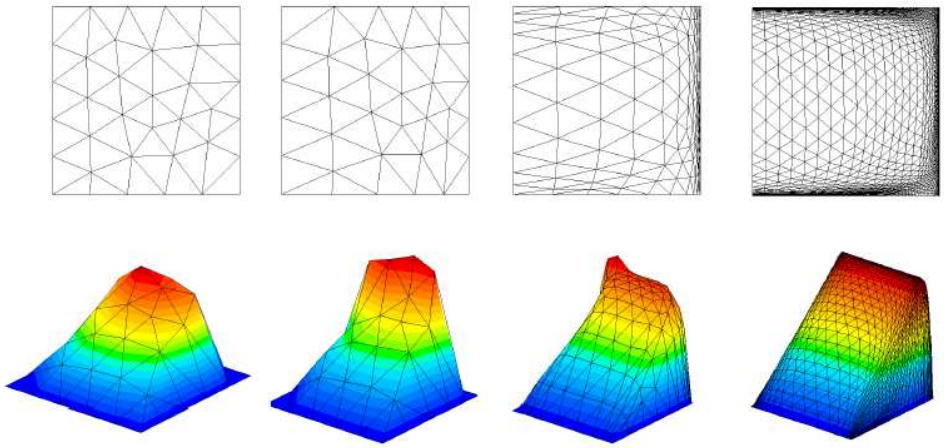


**Figure 4.4.** Continuation adaptive meshes and corresponding solutions

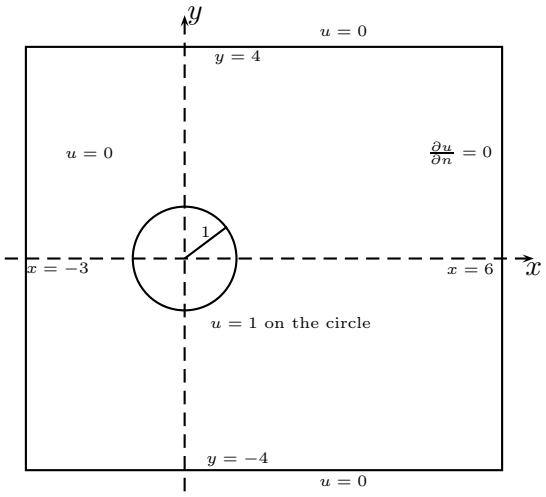
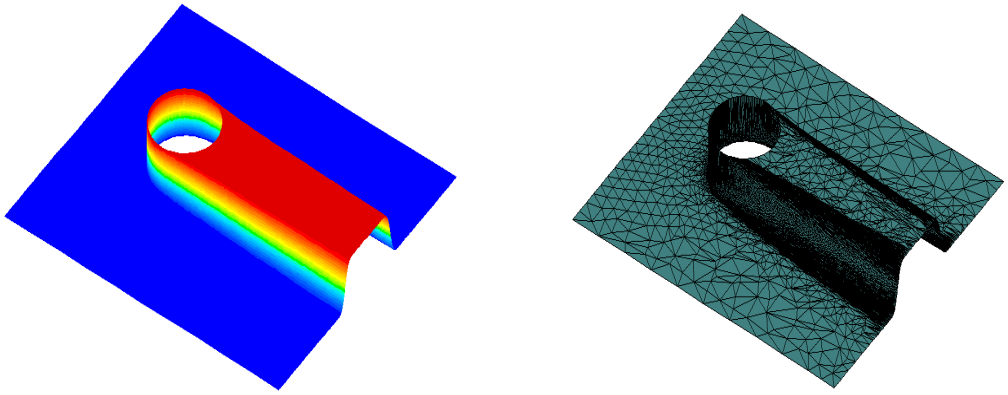



**Figure 4.5.** Domain of the fluid well problem

The fluid flow will spray out from the inner well and flow rightward to the right boundary very quickly, then there are two narrow sections with singularity. We will present the solution and the underlying mesh we obtained by our adaptive method.



(a) Fluid well (b) Underlying mesh

**Figure 4.6.** Solution and the underlying mesh of the fluid well problem

Chapter 5

# Stability and Accuracy of a New Streamline Diffusion Finite Element Method for One Dimensional Convection Dominated Problems

The model problem we will study in this chapter is a linear convection dominated stationary convection diffusion problem:

$$-\varepsilon u'' - bu' = f \ \text{ in } (0,1), \tag{5.1}$$

$$u(0) = g_0, u(1) = g_1, \tag{5.2}$$

where the diffusion constant $\varepsilon$ satisfies $0 < \varepsilon \ll b$. For the simplicity of analysis, we assume the convection coefficient $b$ is constant and positive. Most results in this chapter remain true if $b$ is a smooth enough function with a positive lower bound [49].

The solution to (5.1)-(5.2) typically has a boundary layer at $x = 0$ and thus the standard FEM approximation on a uniform grid will yield nonphysical oscillation unless the mesh size compares with $\varepsilon$ (see, e.g. [163, 148, 146]). To obtain a reliable numerical approximation, one approach is to use highly non-uniform mesh which is adapted to capture the boundary layer. Examples of this approach are layer-adapted grids [15, 171, 134, 162] or grids by the equidistribution of monitor

functions [60, 59, 181, 39]. Another approach is to invoke some form of upwinding to stabilize the scheme. The SDFEM, introduced first by Hughes and Brooks in [109], is one of such stabilized methods and it can also be derived from more general approaches based on, for example, residual-free bubble finite element method [34, 90] and multiscale variational methods [108, 110].

The above approaches and their combinations have been observed to work well in practice. Their error analysis, however, is not so easy. The classic convergence result for the standard FEM is not appropriate in the sense that the accuracy depends not only on the number of the grid nodes $N$ but also on the parameter $\varepsilon$. When $\varepsilon$ is small, the error bound becomes prohibitively large. This chapter is devoted to the $\varepsilon$-uniform convergence. All the error bounds in this chapter are $\varepsilon$-uniform unless it is explicitly expressed otherwise.

## Smooth solutions

To isolate the stability issue, the first question we would like to ask is about a smooth solution to this equation. Namely, if the solution to (5.1)-(5.2) happens to be smooth (say, $\|u^{(3)}\|_\infty$ is uniformly bounded), does the standard FEM on a quasiuniform grid $\mathcal{T}_N$ give an accurate approximation to the solution?

Let $u_N$ be the standard finite element approximation of the exact solution $u$ based on the grid $\mathcal{T}_N$. Here are the answers to the above question.

1. The stability of the standard FEM depends on the parity of the number of unknowns for uniform grids. Namely $\|u - u_N\|_\infty \leq CN^{-2}$ if the number of unknowns is even, while the method is not $\varepsilon$-uniformly stable if the number of unknowns is odd.

2. When the number of unknowns is even, the method can be stabilized if we only move any one grid point within $O(\varepsilon)$ to one of its neighbor.

3. For quasi-uniform grids, we show that $\|u - u_N\|_\infty = O(N^{-1})$ at best in general. Although for the nodal interpolation $u_I$ the error is of second order, namely $\|u - u_I\|_\infty \leq CN^{-2}$.

## Solutions with boundary layers

In most cases of interest, such as the homogeneous boundary condition with a smooth source term, the solution to (5.1)-(5.2) has a boundary layer at $x = 0$. In order to capture the boundary layer highly nonuniform layer-adapted grids need to be adopted. Among them, Bakhvalov grid [15] and Shishkin grid [171] are two commonly used grids. For these two grids, the uniform convergence of the standard FEM is well known [146, 116, 134]: the following two error estimates

$$\|u - u_N\|_\infty \leq CN^{-2}, \text{ and } \|u - u_N\|_\infty \leq CN^{-2} \ln^2 N, \tag{5.3}$$

are valid for Bakhvalov grid and Shishkin grid respectively.

A careful analysis in this chapter will provide some further insights to this type of results. Namely the optimality of the convergence rate in (5.3) depends crucially on the uniformity of the grid in the smooth part. If the grid is only quasiuniform outside of the boundary layer, we can only expect in general

$$\|u - u_N\|_\infty = O(N^{-1}).$$

But if the grid is indeed uniform away from the boundary layer, the estimate (5.3) remains valid even if the grid is locally perturbed (in a locally quasi-uniform manner) within the boundary layer. From both theoretical and practical points of view, we find this is a rather significant phenomenon for singularly perturbed problems.

## Uniform stability of a new SDFEM

For singularly perturbed problems, special stablized methods such as the streamline diffusion finite element method (SDFEM) are more often used than the standard finite element method. Many convergence estimates of the SDFEM [111, 112, 150, 188, 32, 33] have been done for quasiuniform meshes which show that the SDFEM is able to capture the main feature of the solution without using layer-adapted meshes to resolve the boundary layer. Nevertheless, very few $\varepsilon$-uniformly pointwise convergence results are obtained inside the boundary layer [175, 135, 176].

We will propose a new SDFEM for problem (5.1)-(5.2) and analyze it on *ar-*

*bitrary* grids $\mathcal{T}_N$. With a special choice of the stabilization bubble function, we will prove that the new SDFEM approximation $\tilde{u}_N$ is nearly optimal (or so-called quasi-optimal) in the maximum norm, namely

$$\|u - \tilde{u}_N\|_\infty \leq C \inf_{v_N \in V^N} \|u - v_N\|_\infty, \tag{5.4}$$

where $V^N$ is the linear finite element space based on $\mathcal{T}_N$ with appropriate boundary conditions. We would like to explicitly point out again that here $C$ is a constant that is independent on both $\varepsilon$ and $N$.

The estimate (5.4) is the most desirable estimate we may expect to obtain for problem (5.1)-(5.2). Such types of estimates have been known for "diffusion dominated" problem in both one and two dimensions [166, 167]. But we have not seen such an estimate for singularly perturbed problems. The added difficulty is of course the uniformity with respect to $\varepsilon$. Effort to obtain such type of result can also be found in [164, 165].

## Convergence of the new SDFEM and optimal monitor function

Thanks to (5.4) the convergence of the new SDFEM becomes an approximation problem which is well studied in the literature (see, e.g. [62, 46]). If, for example, the function $|u''|^{1/2}$ is monotone, there exists a grid such that

$$\|u - u_I\|_\infty \leq C\|u''\|_{1/2} N^{-2}, \tag{5.5}$$

and thus by (5.4)

$$\|u - \tilde{u}_N\|_\infty \leq C\|u - u_I\|_\infty \leq C\|u''\|_{1/2} N^{-2}, \tag{5.6}$$

where $\|u''\|_{1/2} := (\int_0^1 |u''|^{1/2} dx)^2$. Note that $\|u''\|_{1/2}$ is $\varepsilon$-uniformly bounded in many cases, the convergence (5.6) is indeed $\varepsilon$-uniform.

A commonly used approach to constructing such a grid is the use of monitor function $M(x)$ and the equidistribution principle. The grid $\{0 = x_0 < ... < x_{N+1} =$

1} is chosen such that

$$\int_{x_i}^{x_{i+1}} M(x)dx = \text{constant}, \quad i = 0, 1, 2, ..., N.$$

In the literature, the monitor function resulting a first order uniform convergence is the arc-length function $M = \sqrt{1 + |u'|^2}$ or its discrete analogue [154, 117, 51, 52]. The optimal choice of the monitor function for a second order uniform convergent scheme remains open. Based on our convergence result (5.6) $M = |u''|^{1/2}$ is evidently a monitor function that leads to the optimal rate of convergence.

## 5.1 Error analysis of the standard finite element method

In this section we will study the stability of the standard FEM applied to equation (5.1)-(5.2) on arbitrary grids. The approach used here mainly follows the work of Kopteva for central difference discretization in [116].

Let us first introduce some notation. For a positive integer $N$, let $\mathcal{T}_N := \{x_i \,|\, 0 = x_0 < x_1 < ... < x_{N+1} = 1\}$ be an arbitrary grid and let $\varphi_i$ be the nodal basis function at point $x_i$. The linear finite element space $V^N := \{v_N | v_N = \sum_{i=0}^{N+1} v_i\varphi_i\}$. For a function $u \in C[0,1]$, let $u_i := u(x_i)$ be the nodal values and let $u_I := \sum_{i=0}^{N+1} u_i\varphi_i$ be the nodal interpolant of $u$. The discrete maximum norm of $u$ is denoted by $\|u\|_{\infty,\mathcal{T}_N} := \max_{0\le i\le N+1} |u_i|$. For an index set $I \subset \{0, 1, ..., N + 1\}$, $\|u\|_{\infty,I} := \max_{i\in I} |u_i|$. On the other hand, given a discrete function $\{v_i, i = 0, 1, ..., N + 1\}$, the same letter without the subindex will denote the piecewise linear and global continuous function $V^h$, i.e. $v := \sum_{i=0}^{N+1} v_i\varphi_i$. Thus the discrete maximum norm for the discrete function $v_i$ will be written as $\|v\|_{\infty,\mathcal{T}_N}$.

### 5.1.1 Basic error equation

Let $H^1 = \{v | v \in L^2(0,1), v' \in L^2(0,1)\}$ and $H_0^1 = \{v | v \in H^1, v(0) = v(1) = 0\}$. The weak solution to (5.1)-(5.2) is a function $u \in H^1$ satisfying $u(0) = g_0, u(1) = g_1$ and

$$a(u, v) = (f, v), \forall v \in H_0^1, \tag{5.7}$$

where $(\cdot,\cdot)$ is the $L^2$ inner product and $a(u,v)=\varepsilon(u',v')+(bu,v')$. The existence and uniqueness of the weak solution are easy to establish.

The finite element discretization of (5.7) is to find a $u_N \in V^N$ such that $u_N(0)=g_0, u_N(1)=g_1$ and

$$a(u_N,v_N)=(f,v_N), \forall\, v_N \in V^N \cap H_0^1. \tag{5.8}$$

Let $e=u_I-u_N=\sum i=0^{N+1} e_i\varphi_i$. Since $a(u-u_N,v_N)=0$, we obtain the error equation

$$a(e,\varphi_i)=a(u_I-u,\varphi_i),\ i=1,2,...,N, \tag{5.9}$$

$$e_0=e_{N+1}=0. \tag{5.10}$$

Let $a_{i,j}=a(\varphi_j,\varphi_i)$ and $h_i=x_i-x_{i-1}$. It is easy to get

$$a(e,\varphi_i)=a_{i,i-1}e_{i-1}+a_{i,i}e_i+a_{i,i+1}e_{i+1},\ \text{for } i=1,2,...,N,$$

where

$$a_{i,i-1}=-\frac{\varepsilon}{h_i}+\frac{b}{2},\ a_{i,i}=\frac{\varepsilon}{h_i}+\frac{\varepsilon}{h_{i+1}},\ a_{i,i+1}=-\frac{\varepsilon}{h_{i+1}}-\frac{b}{2}.$$

It is well known that if $h=\max\limits_i h_i \le 2\varepsilon/b$, the matrix $A=(a_{i,j})$ will be an $M$-matrix and thus the scheme satisfies a discrete maximum principle. We are more interested in the case $\varepsilon \ll h$ where in general the discrete maximum principle is not valid. We will solve the error equation directly. This procedure is essentially an $LU$ factorization of a tridiagonal system.

**Lemma 5.1.** *The error equation (5.9)-(5.10) can be written as*

$$(D^N e)_i-(D^N e)_{i+1}=r_i-r_{i+1},\ i=1,2,...,N, \tag{5.11}$$

$$e_0=e_{N+1}=0, \tag{5.12}$$

*where*

$$(D^N e)_i \;=\; (\frac{\varepsilon}{bh_i}+\frac{1}{2})e_i-(\frac{\varepsilon}{bh_i}-\frac{1}{2})e_{i-1},$$

$$r_i \;=\; \frac{1}{h_i}\int_{x_{i-1}}^{x_i}(u_I-u)(x)dx.$$

*Proof.* Since $a(\varphi_i,1)=0$, we get $a(\varphi_i,\varphi_{i-1}+\varphi_{i+1})=-a(\varphi_i,\varphi_i)$, namely $a_{i,i}=-a_{i-1,i}-a_{i+1,i}$. Therefore

$$\begin{aligned} a(e,\varphi_i) &= a_{i,i-1}e_{i-1}+a_{i,i}e_i+a_{i,i+1}e_{i+1} \\ &= a_{i,i-1}e_{i-1}-a_{i-1,i}e_i-a_{i+1,i}e_i+a_{i,i+1}e_{i+1} \\ &= b\left[(D^Ne)_i-(D^Ne)_{i+1}\right]. \end{aligned}$$

On the other hand

$$\int_{x_{i-1}}^{x_i}(u_I-u)'\varphi_i'(x)dx=(u_I-u)|_{x_{i-1}}^{x_i}-\int_{x_{i-1}}^{x_i}(u_I-u)\varphi_i''(x)dx=0,$$

since $(u_I-u)(x_k)=0$ for $k=i-1,i$ and $\varphi_i''=0$ in $[x_{i-1},x_i]$. The right hand side of (5.9) becomes

$$a(u_I-u,\varphi_i)=\int_{x_{i-1}}^{x_{i+1}}b(u_I-u)\varphi_i'=b(r_i-r_{i+1}).$$

The desired result then follows. □

It is easy to see that $(D^Ne)_i=r_i-C$ with an appropriate constant $C$ such that $e_0=e_{N+1}=0$. However it is difficult to determine $C$ explicitly. Instead we use the following splitting of $e_i$.

**Lemma 5.2.**

$$e_i=W_i-\frac{W_{N+1}}{V_{N+1}}V_i,\quad i=1,2,...,N,$$

*where $V_i$ solves the difference equation*

$$(D^NV)_i=1, i=1,2,...,N+1,\ V_0=0,$$

*and $W_i$ solves the difference equation*

$$(D^NW)_i=r_i, i=1,2,...,N+1,\ W_0=0.$$

*Proof.* It is clear that $e_i = W_i - CV_i$. Since $e_{N+1} = 0$, we get $C = W_{N+1}/V_{N+1}$. □

**Lemma 5.3.** *Let*

$$\lambda_i = (\frac{\varepsilon}{bh_i} - \frac{1}{2})(\frac{\varepsilon}{bh_i} + \frac{1}{2})^{-1},\ S_j^i = \prod_{k=j}^{i} \lambda_k,\ i, j = 1, 2, ..., N+1,$$

*(with the convention that if $j > i, S_j^i = 1$) then for $i = 0, 1, ..., N+1$,*

$$V_i = 1 - S_1^i,$$

*and*

$$W_i = \sum_{j=1}^{i} \left[r_j(1-\lambda_j)S_{j+1}^i\right] = r_i - r_1 S_1^i + \sum_{j=1}^{i-1} \left[(r_j - r_{j+1})S_{j+1}^i\right].$$

*Proof.* By the definition of $W_i$, we have

$$(\frac{\varepsilon}{bh_i} + \frac{1}{2})W_i - (\frac{\varepsilon}{bh_i} - \frac{1}{2})W_{i-1} = r_i, \quad i = 1, 2, ..., N+1,$$

and thus

$$W_i = \lambda_i W_{i-1} + (1-\lambda_i)r_i, \quad i = 1, 2, ..., N+1.$$

Here we use the relation $1 - \lambda_i = (\frac{\varepsilon}{bh_i} + \frac{1}{2})^{-1}$. Since $W_0 = 0$, we get

$$W_i = \sum_{j=1}^{i} \left[r_j(1-\lambda_j)S_{j+1}^i\right] = \sum_{j=1}^{i} r_j(S_{j+1}^i - S_j^i).$$

By the discrete version of integration by parts (summation by parts), we get

$$W_i = r_i - r_1 S_1^i + \sum_{j=1}^{i-1} \left[(r_j - r_{j+1})S_{j+1}^i\right].$$

The formula of $V_i$ can be obtained by replacing $r_i = 1$ in the above identity. □

The following two lemmas concern the stability of the scheme.

**Lemma 5.4.** *If $\mathcal{T}_N$ satisfy the condition*

$$|V_{N+1}|^{-1} = |1 - \prod_{i=1}^{N+1} \lambda_i|^{-1} \leq \rho_{stab}, \tag{5.13}$$

*then*

$$\|e\|_{\infty,\mathcal{T}_N} \leq 2(\rho_{stab} + 1)\|W\|_{\infty,\mathcal{T}_N}.$$

*Proof.* It is easy to see $|\lambda_i| \leq 1$ and thus $|V_i| \leq 2$. If (5.13) holds, then from Lemma 5.2, we can easily get $|e_i| \leq 2\rho_{stab}|W_{N+1}| + |W_i|$, which leads to the lemma. □

**Lemma 5.5.** *Let $I$ be an index set and $l(I) := \sum_{i\in I} h_i$. If $\lambda_i \geq 0$, for $i \in I$, then $\mathcal{T}_N$ satisfies condition (5.13) with*

$$\rho_{stab} = (1 - e^{-bl(I)/(2\varepsilon)})^{-1}.$$

*Proof.* We note that, if $\lambda_i \geq 0$, then $bh_i \leq 2\varepsilon$. Using the simple inequality that $\ln(1-x) \leq -x$ for $x \in (0,1)$, we have

$$\begin{aligned} \sum_{i\in I} \ln \lambda_i &= \sum_{i\in I} \ln(1 - \frac{2bh_i}{2\varepsilon + bh_i}) \leq -\sum_{i\in I} \frac{2bh_i}{2\varepsilon + bh_i} \\ &\leq -\frac{b\sum_{i\in I} h_i}{2\varepsilon} = -\frac{bl(I)}{2\varepsilon}. \end{aligned}$$

Therefore

$$|1 - \prod_{i=1}^{N+1} \lambda_i| \geq 1 - |\prod_{i\in I} \lambda_i| \geq 1 - e^{-bl(I)/(2\varepsilon)},$$

as desired. □

When all $\lambda_i \geq 0$, the resulting matrix is an $M$-matrix. By Lemma 5.5, it is stable with $\rho_{stab} \sim 1$. To stabilize the scheme, according to Lemma 5.5, we only need $l(I) = O(\varepsilon)$. Note that a local grid refinement usually produces such grids. It justifies that grid adaptation will enhance the stability of the scheme which is often observed in the numerical computation.

### 5.1.2 Smooth solutions

In this subsection we will consider the case when the solution to (5.1)-(5.2) is smooth and the grid is uniform.

**Lemma 5.6.** *If $\|u^{(k)}\|_\infty$, $k = 1, 2, 3$ are uniformly bounded, for a uniform grid $\mathcal{T}_N$, we have:*

$$\|r\|_{\infty,\mathcal{T}_N} \le CN^{-2}, \quad \textit{and} \quad \max_{1\le i\le N} |r_i - r_{i+1}| \le CN^{-3},$$

*and thus*

$$\|W\|_{\infty,\mathcal{T}_N} \le CN^{-2}.$$

*Proof.* It is easy to see that

$$r_i = \frac{1}{2} h_i^2 u''(\xi_i), \text{ with some } \xi_i \in [x_{i-1}, x_i].$$

Therefore $|r_i| \le CN^{-2}$, and

$$|r_i - r_{i+1}| = |h_i^2 u''(\xi_i) - h_{i+1}^2 u''(\xi_{i+1})| \le N^{-3} \|u^{(3)}\|_\infty.$$

For the estimate of $\|W\|_{\infty,\mathcal{T}_N}$, we use the fact that $|S_j^i| \le 1$ to get

$$|W_i| \le 2\|r\|_{\infty,\mathcal{T}_N} + \sum_{i=1}^{N} |r_i - r_{i+1}| \le CN^{-2}.$$

□

**Lemma 5.7.** *For a uniform grid $\mathcal{T}_N$ with $N$ interior points, we have*

1. *if $N$ is even, then*
$$|1 - \prod_{i=1}^{N+1} \lambda_i|^{-1} \le C;$$

2. *if $N$ is odd, for a fixed $N$, then* $\lim_{\varepsilon\to 0} (1 - \prod_{i=1}^{N+1} \lambda_i)^{-1} = \infty.$

*Proof.* (1) If $\lambda_i \ge 0$, the stability result follows from Lemma 5.5.

If $\lambda_i < 0$, since $N$ is even,

$$1 - \prod_{i=1}^{N+1} \lambda_i = 1 - (-1)^{N+1} \prod_{i=1}^{N+1} |\lambda_i| = 1 + \prod_{i=1}^{N+1} |\lambda_i| > 1.$$

(2) When $N$ is odd,

$$1 - \prod_{i=1}^{N+1} \lambda_i = 1 - (-1)^{N+1} \prod_{i=1}^{N+1} |\lambda_i| = 1 - \prod_{i=1}^{N+1} |\lambda_i|.$$

Note that $\lim_{\varepsilon \to 0} |\lambda_i| = 1$, we conclude that $\lim_{\varepsilon \to 0} (1 - \prod_{i=1}^{N+1} \lambda_i)^{-1} = \infty$. □

With Lemma 5.6 and 5.7, we immediately get the following result.

**Theorem 5.8.** *Let $u$ be the solution of (5.7) and let $u_N$ be the standard finite element approximation of $u$ based on a uniform grid $\mathcal{T}_N$. Suppose $\|u^{(k)}\|_\infty$, $k = 1, 2, 3$ are uniformly bounded. Then $\|u - u_N\|_\infty \leq CN^{-2}$ if the number of unknowns is even, while the method is not $\varepsilon$-uniformly stable if the number of unknowns is odd.*

Lemma 5.7 says that for smooth solution and uniform grids, when $\varepsilon$ is small, the stability of the scheme depends on the parity of the grid. An intuitive way to understand this interesting phenomenon is to consider the limiting matrix as $\varepsilon$ goes to zero which is for the reduced problem $-bu' = f$. Both the PDE operator and the corresponding matrix are skew symmetric. Thus if the dimension of the matrix is odd, it has a zero eigenvalue. The corresponding eigenvector spans the kernel of the discrete problem which makes the scheme unstable. For this simple example, the eigenvector is $(1, 0, 1, ..., 0, 1)$. Indeed, Lenferink [129] eliminates every other unknown to stabilize the scheme.

It is interesting to note that if we modify the grid such that one element has $O(\varepsilon)$ mesh size, the scheme will be stabilized.

**Lemma 5.9.** *Let $Pe_i := bh_i/(2\varepsilon)$ be the grid Peclét number. If there exists an element $[x_{k-1}, x_k]$ in the grid $\mathcal{T}_N$ such that*

$$\rho_0 \leq Pe_k \leq \rho_0^{-1}, \text{ for some } \rho_0 \in (0, 1], \tag{5.14}$$

*then $\mathcal{T}_N$ satisfies condition (5.13) with $\rho_{stab} = (1+\rho_0^{-1})/2$.*

*Proof.* Note that the function $|x-1|/(x+1)$ is increasing for $x > 1$ and decreaseing for $x \le 1$. With the assumption (5.14), if $Pe_k > 1$ then

$$|\lambda_k| = \frac{Pe_k - 1}{Pe_k + 1} \le \frac{\rho_0^{-1} - 1}{\rho_0^{-1} + 1} = \frac{1-\rho_0}{1+\rho_0}.$$

Otherwise

$$|\lambda_k| = \frac{1 - Pe_k}{Pe_k + 1} \le \frac{1-\rho_0}{1+\rho_0}.$$

Therefore

$$\Big|1 - \prod_{i=1}^{N+1} \lambda_i\Big| > 1 - |\lambda_k| \ge \frac{2\rho_0}{1+\rho_0},$$

as desired. □

Since the local mesh refinement will generate some grids with the same scaling of $\varepsilon$, in view of Lemma 5.9 it will stabilize the standard FEM.

**Lemma 5.10.** *Let the grid $\mathcal{T}_N$ satisfy (1) $h_k$ satisfying (5.14) for some $1 \le k \le N+1$ and (2) $h_i = CN^{-1}$ for $i = 1, ..., N+1, i \ne k$. Then*

$$\|W\|_{\infty,\mathcal{T}_N} \le CN^{-2}.$$

*Proof.* By the proof of Lemma 5.6, it is easy to see

$$\|r\|_{\infty,\mathcal{T}_N} \le CN^{-2}, \text{ and } |r_i - r_{i+1}| \le CN^{-3} \text{ for } i \ne k-1, k.$$

Therefore

$$\begin{aligned} |W_i| &\le 2\|r\|_{\infty,\mathcal{T}_N} + |r_{k-1} - r_k| + |r_k - r_{k+1}| + \sum_{i=0, i\ne k-1,k}^{N} |r_i - r_{i+1}| \\ &\le 6\|r\|_{\infty,\mathcal{T}_N} + \sum_{i=0, i\ne k-1,k}^{N} |r_i - r_{i+1}| \le CN^{-2}. \end{aligned}$$

□

Combining Lemmas 5.9 and 5.10, we get the following result.

**Theorem 5.11.** *Let grid $\mathcal{T}_N$ satisfy conditions in Lemma 5.10 (namely it is uniform except for one element that has size of $\mathcal{O}(\varepsilon)$). If $\|u^{(k)}\|_\infty$, $k = 1, 2, 3$ are uniformly bounded, then the standard FEM approximation $u_N$ based on $\mathcal{T}_N$ satisfy*

$$\|u - u_N\|_\infty \leq CN^{-2}.$$

In the proof of Lemma 5.10, we use the uniformity of the grid to bound the summation of $|r_i - r_{i+1}|$ with few exceptions. We will show the uniformity of the grid is crucial for the second order convergence by the following example.

**Example 5.12.** *There exist a sequence of quasiuniform grids $\{\mathcal{T}_N\}$ such that the standard finite element approximation $u_N$ to the following equation:*

$$-\varepsilon u''(x) - u'(x) = -2\varepsilon - 2x,\ x \in (0, 1), \tag{5.15}$$

$$u(0) = 0, u(1) = 1, \tag{5.16}$$

*is only first order provided $\varepsilon$ small enough.*

The real solution of (5.15)-(5.16) is $u = x^2$. Let $N$ be an odd integer and $\mathcal{T}_N$ be the uniform grid with equal size $h$. We modify

$$x_{2i+1} = (i + 0.25)h, \quad i = 0, 1, ..., \frac{N+1}{2} - 1.$$

In this case

$$r_i - r_{i+1} = (-1)^i h^2, \quad i = 1, ..., N.$$

Let us choose small $\varepsilon$ such that $\lambda_i < 0$. Note that $S_i^{N+1} = (-1)^i |S_i^{N+1}|$ also oscillates, we have

$$(r_i - r_{i+1})S_i^{N+1} = h^2 |S_i^{N+1}| > h^2 q_\varepsilon^{N-i+1},$$

where $q_\varepsilon = (5 - 8\varepsilon N)(5 + 8\varepsilon N)^{-1}$. Therefore

$$|W_{N+1}| \geq h^2 (\sum_{i=1}^{N+1} |S_i^{N+1}| - 1) \geq h^2 (\frac{1 - q_\varepsilon^{N+2}}{1 - q_\varepsilon} - 1).$$

Since $\lim_{\varepsilon \to 0} (1 - q^N)(1 - q)^{-1} = N$, we may choose $\varepsilon$ small enough such that $|W_{N+1}| \geq$

$(N+1)h^2$. Note that $W_1 = (1-\lambda_1)h_1^2 \geq 5/2h^2$, we have

$$\|u_I - u_N\|_\infty \geq |e_1| \geq |W_{N+1}| - |W_1| \geq (N-4)h^2 \geq CN^{-1}.$$

For smooth functions, we know that, on a quasi-uniform grid, the interpolation error is still of second order, namely $\|u - u_I\| \leq CN^{-2}$. The optimal convergence rate we would like to expect for the numerical solution is also of second order. Example 1 tells us for singularly perturbed problems, when $\varepsilon \ll 1$, we may lose one order of accuracy for the standard FEM if we perturb the uniform grid to be a quasi-uniform one. In other words, the standard FEM is not stable with respect to the perturbation of the grid.

### 5.1.3 Solutions with boundary layers

In this subsection, we will consider the solutions with boundary layers and prove the convergence of standard FEM on two special layer adapted grids: Bakhvalov grid and Shishkin grid. The convergence result is known; see for example [116]. However we will show the accuracy of the standard FEM depends crucially on the uniformality of the grid in the smooth part.

More speficially we will consider equation (5.1)-(5.2) with homogeneous boundary condition and smooth source term $f$. Namely

$$-\varepsilon u'' - bu' = f \ \text{ in } (0,1), \tag{5.17}$$

$$u(0) = u(1) = 0. \tag{5.18}$$

The solution to (5.17)-(5.18) typically has a boundary layer near $x = 0$. The following *a priori* bound of the derivatives of the solution is well known in the literature, see for example [114, 163, 148] or [146].

**Lemma 5.13.** *Let $u$ be the solution to equation (5.17)-(5.18), we have the following a priori bound:*

$$|u^{(k)}(x)| \leq C(1 + \varepsilon^{-k} e^{-bx/\varepsilon}), \ \forall x \in [0,1], \ k = 0,1,2,3.$$

Layer-adapted grids are needed to capture the boundary layer. The first such

grid is the Bakhvalov grid [15] . Let $N$ be an even integer and

$$\theta = \min\{\frac{1}{2}, \frac{2\varepsilon \ln \varepsilon^{-1}}{b}\}.$$

In $[0, \theta]$ we put $N/2$ elements such that

$$\int_{x_{i-1}}^{x_i} \varepsilon^{-1} e^{-bx/(2\varepsilon)} dx = \frac{2}{N} \int_0^{\theta} \varepsilon^{-1} e^{-bx/(2\varepsilon)} dx,$$

namely

$$x_i = -\frac{2\varepsilon}{b} \ln(1 - 2(1-\varepsilon)\frac{i}{N}), i = 0, 1, ..., N/2.$$

In $[\theta, 1]$, we put $N/2$ uniform grids and denote the mesh size $H = 2(1-\theta)/N \leq CN^{-1}$.

The interpolation error estimate on Bakhvalov grid is known [134] which can be also derived from Theorem 5.24 in Section 3.2.

**Lemma 5.14.** *Let $u$ be the solution to (5.17)-(5.18). For Bakhvalov grid,*

$$\|u - u_I\|_{L^\infty} \leq CN^{-2}.$$

To prove the convergence, we need the following technical lemma.

**Lemma 5.15.** *For $j < i$, let $I = \{j, j+1, ..., i\}$ and $h_I = \max_{k \in I} h_k$. We have*

1. $$|W_i| \leq |W_j| + |r_i| + |r_j| + \sum_{k \in I \setminus \{i\}} |r_k - r_{k+1}|, \tag{5.19}$$

2. *If $\lambda_k \geq 0$, $k \in I$, then*

$$|W_i| \leq |W_j| + 2 \max_{k \in I \setminus \{j\}} |r_k|. \tag{5.20}$$

*Proof.* By Lemma 5.3, we get

$$W_i = S_{j+1}^i W_j + \sum_{k=j+1}^{i} r_k (S_{k+1}^i - S_k^i)$$

$$= \quad S^i_{j+1}W_j + r_i - r_j S^i_j + \sum_{k=j}^{i-1} \big[(r_k - r_{k+1})S^i_{k+1}\big].$$

Hence (5.19) follows from the fact $|S^i_j| \le 1$.

If $\lambda_k \ge 0$, $k \in I$, then $S^i_k$ is monotone decreasing with respect to $k$, and thus

$$|W_i| \le |W_j| + \max_{j+1\le k\le i} |r_k| \Big|\sum_{k=j}^{i-1}(S^i_k - S^i_{k+1})\Big| \le |W_j| + 2\max_{k\in I\setminus\{j\}} |r_k|.$$

□

**Theorem 5.16.** *For Bakhvalov grid, the standard FEM approximation is uniformly optimal. Namely*

$$\|u - u_N\|_\infty \le CN^{-2}.$$

*Proof.* We divide $\mathcal{T}_N$ into boundary layer $I_1 = \{0, 1, ..., N/2\}$ and smooth part $I_2 = \{N/2 + 1, ..., N + 1\}$ by the transition point $\theta$. In $I_1$, it is easy to see $h_i/\varepsilon = CN^{-1}$ by the mean value theorem. Thus $\lambda_i > 0, i \in I_1$ for sufficient large $N$ (independent of $\varepsilon$). Since $l(I_1) = \theta$, the stability of the scheme follows from Lemma 5.5.

With $W_0 = 0$ and (5.20) we have

$$\|W\|_{\infty,I_1} \le \|r\|_{\infty,I_1} \le C\|u - u_I\|_\infty \le CN^{-2}. \tag{5.21}$$

In the smooth part $[\theta, 1]$

$$\begin{aligned} \|W\|_{\infty,I_2} &\le |W_{N/2+1}| + |r_{N/2+1}| + |r_{N+1}| + \sum_{i\in I_2\setminus\{N+1\}} |r_i - r_{i+1}| \\ &\le |W_{N/2} + C\|r\|_{\infty,\mathcal{T}_N} + \sum_{i\in I_2} C\|u^{(3)}\|_{\infty,[x_{i-1},x_{i+1}]}H^3. \end{aligned}$$

By (5.21) the first two terms are bounded by $CN^{-2}$. We now estimate the third term

$$\sum_{i\in I_2} \|u^{(3)}\|_{\infty,[x_{i-1},x_{i+1}]}H^3 \quad \le \quad H^2\Big(\sum_{i\in I_2}(1 + \varepsilon^{-3}e^{-bx/\varepsilon})H\Big)$$

$$\begin{aligned} &\leq N^{-2}\int_{\theta}^{1}(1+\varepsilon^{-3}e^{-bx/\varepsilon})dx \\ &\leq CN^{-2}. \end{aligned}$$

Combining those estimates together, we get

$$\|u_I - u_N\|_\infty \leq C\|W\|_{\infty,\mathcal{T}_N} \leq CN^{-2}.$$

The result then follows from the triangle inequality

$$\|u - u_N\|_\infty \leq \|u - u_I\|_\infty + \|u_I - u_N\|_\infty,$$

and the interpolation error estimate for $\|u - u_I\|_\infty$. □

Another simple layer adapted grid is Shishkin grid [171]. Let $N$ be an even integer and the transition point

$$\theta = \min\{\frac{1}{2}, \frac{2\varepsilon \ln N}{b}\}.$$

In practice, $\varepsilon$ is so small that $\theta = 2b^{-1}\varepsilon \ln N$. Then $[0,\theta]$ and $[\theta,1]$ are divided into $N/2$ equidistant subintervals. The following interpolation error estimate for Shishkin grid is well known [133, 163, 146, 134].

**Lemma 5.17.** *Let $u$ be the solution to (5.17)-(5.18). For Shishkin grid,*

$$\|u - u_I\|_{L^\infty(x_{i-1},x_i)} \leq \begin{cases} CN^{-2}\ln^2 N, & i = 1,2,...,N/2, \\ \\ CN^{-2}, & i = N/2+1,...,N+1. \end{cases}$$

Similar to the proof of Theorem 5.16, we can get the convergence of the standard FEM approximation on Shishkin grid.

**Theorem 5.18.** *For Shishkin grid, the standard FEM approximation $u_N$ is an almost seconder order approximation*

$$\|u - u_N\|_\infty \leq CN^{-2}\ln^2 N,$$

From the proof of Theorem 5.16, we see inside the boundary layer, we use $|r_i| \leq \|u - u_I\|_{\infty,[x_i,x_{i+1}]}$ and only need to bound the interpolation error. Therefore the grid inside the boundary layer can be relaxed to be quasiuniform. While in the smooth part, we need the uniformity of the grid to ensure $|r_i - r_{i+1}| \leq CN^{-3}$. Actually, the convergence rate highly depends on the uniformality of the grid in the smooth part.

We can construct an example similar to Example 5.12 to show that when the smooth part of layer adapted grids is only quasiuniform, the convergence rate will degrade to the first order for small $\varepsilon$.

**Example 5.19.** *There exist a sequence of Bakhvalov type grids $\{\mathcal{T}_N\}$ such that the standard FEM approximation $u_N$ to the following equation:*

$$-\varepsilon u''(x) - u'(x) = -2\varepsilon - 2x,\ x \in (0,1), \tag{5.22}$$

$$u(0) = 1, u(1) = 1, \tag{5.23}$$

*is only of first order provided $\varepsilon$ is small enough. Namely*

$$\|u - u_N\|_\infty \geq CN^{-1},\ \textit{but}\ \|u - u_I\|_\infty \leq CN^{-2}.$$

The real solution to (5.22)-(5.23) is

$$u = \frac{e^{-x/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}} + x^2,$$

which contains a boundary layer near $x = 0$. We modify the Bakhvalov grid in the smooth part by moving all odd grid points (except $x_{N+1}$) with a right offset $h/4$. In this case, $r_i - r_{i+1} = (-1)^i N^{-2}$ and for small $\varepsilon$, the error in the smooth part will accumulate to $N^{-1}$. The proof is the same as that in Example 5.12.

Note that for such a modified Bakhvalov grid, the interpolation errror is still of second order, namely $\|u - u_I\|_\infty \leq CN^{-2}$. Example 2 implies that the accuracy of the adapted approximation of standard FEM is very sensitive to the perturbation of grid points in the region where the solution is smooth.

## 5.2 Error analysis of a new streamline diffusion finite element method

In this section, we will propose a new SDFEM based on a special choice of the stabilization bubble functions and prove that the new method produces a nearly optimal approximation.

### 5.2.1 Uniform stability

To introduce the SDFEM, we first modify our bilinear form to be

$$\tilde{a}(u,v) := a(u,v) - \sum_{i=1}^{N} \int_{x_{i-1}}^{x_i} \delta_i(-\varepsilon u'' - bu')bv',$$

where $\delta_i$ is a stabilization function in $[x_{i-1}, x_i]$. We will discuss the choice of $\delta_i$ in a moment. For the exact solution $u$ of (5.1)-(5.2), it satisfies

$$\tilde{a}(u,v) = \tilde{f}(v), \text{ for all } v \in H_0^1,$$

where $\tilde{f}(v) = (f,v) - \sum_{k=1}^{N} \int_{x_{k-1}}^{x_k} \delta_k f b v'$. The SDFEM is to find $\tilde{u}_N \in V^N$ such that $\tilde{u}_N(0) = g_0, \tilde{u}_N(1) = g_1$ and

$$\tilde{a}(u_N, v_N) = \tilde{f}(v_N), \quad \forall\, v_N \in V^N \cap H_0^1. \tag{5.24}$$

In the traditional SDFEM, $\delta_i$ is chosen to be a proper constant, such as $h_i$, on each interval $[x_{i-1}, x_i]$. The key in the new SDFEM is that $\delta_i$ is chosen to be a bubble function on each $[x_{i-1}, x_i]$ defined as follows

$$\delta_i = \min\{\frac{h_i}{2\varepsilon}, \frac{1}{b}\} h_i(\varphi_i \varphi_{i-1})(x). \tag{5.25}$$

Recall that $\varphi_i$ is the nodal basis function at point $x_i$, so $\delta_i$ is a quadratic bubble function with scale $h_i$. It is interesting to note that the bubble function is used in some general stabilized method such as residual-free bubble finite element method [34, 90] and multiscale variational methods [108, 110].

With such a special choice of $\delta_i$, we have the most desirable stability estimate stated in the following theorem.

**Theorem 5.20.** *Let $\tilde{u}_N$ be the SDFEM approximation to equation (5.1)-(5.2) with stabilization function $\delta_i$ determined by (5.25) and let $V_D^N := \{v_N \in V^N, v_N(0) = g_0, v_N(1) = g_1\}$, we have*

$$\|u - \tilde{u}_N\|_\infty \leq C \inf_{v_N \in V_D^N} \|u - v_N\|_\infty.$$

The rest of this section is devoted to the proof of Theorem 5.20. Similar to the standard FEM, we write the error equation for $e = u_I - \tilde{u}_N$ as:

$$\tilde{a}(e, \varphi_i) = \tilde{a}(u_I - u, \varphi_i), \ i = 1, 2, ..., N \tag{5.26}$$

$$e_0 = e_{N+1} = 0. \tag{5.27}$$

**Lemma 5.21.** *The error equation (5.26) can be written as*

$$(\tilde{D}^N e)_i - (\tilde{D}^N e)_{i+1} = \tilde{r}_i - \tilde{r}_{i+1}, \ i = 1, 2, ..., N$$
$$e_0 = e_{N+1} = 0,$$

*where*

$$\begin{aligned} (\tilde{D}^N e)_i &= (\frac{\varepsilon + b^2\bar{\delta}_i}{bh_i} + \frac{1}{2})e_i - (\frac{\varepsilon + b^2\bar{\delta}_i}{bh_i} - \frac{1}{2})e_{i-1}, \\ \bar{\delta}_i &= \frac{1}{h_i}\int_{x_{i-1}}^{x_i} \delta_i(x)dx, \end{aligned}$$

*and*

$$\begin{aligned} \tilde{r}_i &= \frac{1}{h_i}\Big[\int_{x_{i-1}}^{x_i} (u_I - u)(x)dx && (5.28) \\ &\quad + \int_{x_{i-1}}^{x_i} \delta_i \varepsilon u''dx && (5.29) \\ &\quad - \int_{x_{i-1}}^{x_i} b\delta_i(u_I - u)'dx\Big]. && (5.30) \end{aligned}$$

Let $\tilde{\varepsilon}_i = \varepsilon + b^2\bar{\delta}_i$,

$$\tilde{\lambda}_i = (\frac{\tilde{\varepsilon}_i}{bh_i} - \frac{1}{2})(\frac{\tilde{\varepsilon}_i}{bh_i} + \frac{1}{2})^{-1},$$

and $\tilde{S}^i_j, \tilde{W}_i, \tilde{V}_i$ be defined similarly. We can follow the same lines in Section 2 to solve the error equation and get similar results in the last section.

The following lemma is crucial to obtain our main theorem.

**Lemma 5.22.** *For $\delta_i$ given by (5.25), we have*

$$|\tilde{r}_i| \le \frac{C}{h_i}\int_{x_{i-1}}^{x_i} |u - u_I|dx, \quad \textit{and thus } \|\tilde{r}\|_{\infty,\mathcal{T}_N} \le \|u - u_I\|_\infty. \tag{5.31}$$

*Proof.* We will prove (5.31) by estimating the three terms (5.28)-(5.30) respectively. The proof for (5.28) is trivial. For (5.29), we have

$$\begin{aligned}
|\frac{1}{h_i}\int_{x_{i-1}}^{x_i} \varepsilon\delta_i u''dx| &\le |\int_{x_{i-1}}^{x_i} \min\{\frac{1}{2}, \frac{\varepsilon}{bh_i}\}h_i(\varphi_i\varphi_{i-1})u''dx| \\
&\le |\frac{1}{2h_i}\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_i)u''(x)dx| \\
&\le |\frac{1}{2h_i}\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_i)(u - u_I)''(x)dx| \\
&= |\frac{1}{h_i}\int_{x_{i-1}}^{x_i} (u_I - u)(x)dx|.
\end{aligned}$$

The last step follows from integration by parts twice.

For (5.30), we have

$$\begin{aligned}
|\frac{1}{h_i}\int_{x_{i-1}}^{x_i} b\delta_i(u_I - u)'dx| &= |\frac{1}{h_i}\int_{x_{i-1}}^{x_i} b\delta_i'(u_I - u)dx| \\
&\le |\int_{x_{i-1}}^{x_i} (u_I - u)(\varphi_{i-1}\varphi_i)'dx| \\
&\le \frac{1}{h_i}\int_{x_{i-1}}^{x_i} |u_I - u|dx.
\end{aligned}$$

□

**Theorem 5.23.** *For the SDFEM with $\delta_i$ determined by (5.25), we have*

$$\|u_I - \tilde{u}_N\|_\infty \le C\|u - u_I\|_\infty,$$

*and thus*

$$\|u - \tilde{u}_N\|_\infty \leq C\|u - u_I\|_\infty.$$

*Proof.* If $\dfrac{h_i}{2\varepsilon} < \dfrac{1}{b}$, then

$$\bar{\delta}_i = \frac{h_i^2}{4\varepsilon}, \text{ and } \frac{\varepsilon + b^2\bar{\delta}_i}{h_i} = \frac{\varepsilon}{h_i} + \frac{b^2 h_i}{4\varepsilon} \geq \frac{b}{2}.$$

Otherwise

$$\bar{\delta}_i = \frac{h_i}{2b}, \text{ and } \frac{\varepsilon + b^2\bar{\delta}_i}{h_i} > \frac{b^2\bar{\delta}_i}{h_i} \geq \frac{b}{2}.$$

Thus $\tilde{\lambda}_i \geq 0$, for all $i = 1, ..., N+1$. By Lemma 5.4, we have $\|u_I - \tilde{u}_N\|_\infty \leq C\|\tilde{W}\|_{\infty,\mathcal{T}_N}$ and by Lemma 5.15 we have $\|\tilde{W}\|_{\infty,\mathcal{T}_N} \leq C\|\tilde{r}\|_{\infty,\mathcal{T}_N}$.

Now using Lemma 5.22, we have:

$$\|u_I - \tilde{u}_N\|_\infty \leq C\|\tilde{r}_i\|_{\infty,\mathcal{T}_N} \leq C\|u - u_I\|_\infty.$$

The second inequality in the theorem is obtained by the triangle inequality. □

We are now in a position to prove the main theorem in this section.

**Proof of Theorem 5.20** Let us first consider the case $g_0 = g_1 = 0$. We denote the corresponding finite element space by $V_0^N$. We define the projection operator

$$P_N : H_0^1 \to V_0^N \text{ by } P_N u = \tilde{u}_N.$$

By Theorem 5.23,

$$\|u - \tilde{u}_N\|_\infty \leq C\|u - u_I\|_\infty \leq C(\|u\|_\infty + \|u_I\|_\infty) \leq C\|u\|_\infty.$$

Thus

$$\|P_N u\|_\infty = \|\tilde{u}_N\|_\infty \leq \|u\|_\infty + \|u - \tilde{u}_N\|_\infty \leq C\|u\|_\infty.$$

With the property $P_N^2 = P_N$, for any $v_N \in V_0^N$, we have

$$\|u - \tilde{u}_N\|_\infty = \|(I - P_N)(u - v_N)\|_\infty \leq C\|u - v_N\|_\infty.$$

Since it is true for any $v_N \in V_0^N$, the optimality result for homogeneous boundary

condition is then obtained.

For general boundary conditions, we define $u_N^* = (g_1 - g_0)x + g_0$ which belongs to $V_D^N$. Note that $u - u_N^* \in H_0^1$ solving the following equation

$$\begin{aligned} -\varepsilon v'' - bv' &= f + bu_N^* \ \text{ in } (0,1), \\ v(0) = 0, v(1) &= 0. \end{aligned}$$

Thus $P_N(u - u_N^*)$ is well defined. On the other hand $\tilde{u}_N - u_N^*$ is also a SDFEM approximation of the above equation. By the uniqueness, we have $\tilde{u}_N - u_N^* = P_N(u - u_N^*)$. Therefore

$$\begin{aligned} \|u - \tilde{u}_N\|_\infty &= \|(u - u_N^*) - P_N(u - u_N^*)\|_\infty \\ &\leq C \inf_{v_N \in V_0^N} \|u - u_N^* - v_N\|_\infty \\ &= C \inf_{v_N \in V_D^N} \|u - v_N\|_\infty. \end{aligned}$$

In [49] we extend Theorem 5.20 to a slightly more general variable coefficients $b(x)$ by means of discrete Green functions.

### 5.2.2 Uniform convergence

We first discuss how to adapt the grid to get optimal interpolation error estimates. Given a function $u \in C^2[0,1]$, a positive function $H(x)$ is called a majorant of the second order derivative of $u$, if $|u''(x)| \leq H(x)$, $x \in (0,1)$. For an element $\tau_i = [x_{i-1}, x_i]$, its length in the metric $H$ are denoted by $|\tau_i|_H$, namely

$$|\tau_i|_H = \int_{x_{i-1}}^{x_i} H^{1/2}(x)dx.$$

We need two basic assumptions to get a nearly optimal interpolation error estimate.

**(A1)** $H$ is monotone in each element $\tau_i$, $\quad i = 1, 2, ..., N+1$.

**(A2)** $|\tau_i|_H$ is nearly equidistributed in the sense that

$$\max_{1 \leq i \leq N+1} |\tau_i|_H \leq \frac{C}{N} \sum_{i=1}^{N+1} |\tau_i|_H.$$

**Theorem 5.24.** *[59, 62] Let $u \in C^2[0,1]$ and the mesh $\mathcal{T}_N$ satisfy assumptions (A1) and (A2), the following error estimate holds:*

$$\|u - u_I\|_\infty \leq C\|H\|_{1/2}N^{-2}, \tag{5.32}$$

*where*

$$\|H\|_{L^{1/2}} := \Big(\int_0^1 H^{1/2}dx\Big)^2.$$

**Remark 5.25.** *This error estimate is optimal in the sense that for a strictly convex (or concave) function, the above inequality holds asymptotically in a reversed direction with $H = |u''|$.*

The assumption (A2) can be used to direct our construction of the nearly optimal mesh. In the context of the so-called moving mesh method [104, 107, 37], it can be done by the equidistribution of a monitor function. A monitor function $M = M(u, u', u'', ...)$ is a function involving $u$ and its derivatives. We say that the grid $\mathcal{T}_N$ nearly equidistributes the monitor function $M$ if

$$\int_{x_i}^{x_{i+1}} Mdx \leq \frac{C}{N}\int_0^1 Mdx, \quad i = 0, 1, 2, ..., N.$$

Based on the interpolation error estimates, an optimal monitor function for linear interpolant is $M = H^{1/2}$.

**Theorem 5.26.** *Let $\tilde{u}_N$ be the SDFEM approximation to the solution to (5.1)-(5.2) on a grid obtained by nearly equidistributing a monotone majorant $H$ of the second derivative of $u$ and $\delta_i$ is determined by (5.25), then*

$$\|u - \tilde{u}_N\|_\infty \leq C\|H\|_{1/2}N^{-2}.$$

The convergence of the new SDFEM on different interesting cases can be obtained as corollaries of Theorem 5.26 and the interpolation error estimates. The convergence rate is not sensitive to the perturbation of the grid since it is controlled by the interpolation error. For equation (5.17)-(5.18), the convergence of the new SDFEM on different layer adapted grids is straightforward.

**Corollary 5.27.** *The SDFEM approximation $\tilde{u}_N$ with $\delta_i$ determined by (5.25) to (5.17)-(5.18) on the grid obtained by the nearly equidistribution of monitor function*

$M = \sqrt{1 + \varepsilon^{-2} e^{-bx/\varepsilon}}$ *satisfies*

$$\|u - \tilde{u}_N\|_\infty \leq CN^{-2}.$$

**Corollary 5.28.** *The SDFEM approximation $\tilde{u}_N$ with $\delta_i$ determined by (5.25) to (5.17)-(5.18) on Bakhvalov grid satisfies*

$$\|u - \tilde{u}_N\|_\infty \leq CN^{-2}.$$

**Corollary 5.29.** *The SDFEM approximation $\tilde{u}_N$ with $\delta_i$ determined by (5.25) to (5.17)-(5.18) on Shishkin grid satisfies*

$$\|u - \tilde{u}_N\|_\infty \leq CN^{-2} \ln^2 N.$$

We would like to emphasis again that in the proof of the uniform optimality of the new SDFEM, we do not make use of the *a priori* information about $|u''|$ and the structure of the grid. The $\varepsilon$-uniform stability result Theorem 5.26 can be applied for non-smooth data $f$ also.

To show the convergence, all we need to do is to get *a priori* information of the second derivative and adapt the grid to get a good interpolant. For example, let us consider the following equation as studied in [164].

$$-\varepsilon u'' - bu' = f + \delta(\cdot - d) \ \text{ on } \Omega^- \cup \Omega^+, \tag{5.33}$$

$$u(0) = u(1) = 0, \tag{5.34}$$

where $\Omega = (0,1), d \in \Omega, \Omega^- = (0,d), \Omega^+ = (d,1)$ and $\delta(\cdot - d)$ denotes the Dirac-delta function at point $d$. Function $f$ is sufficiently smooth on $\bar{\Omega}$. The equation (5.33)-(5.34) should be understood in the distribution sense and it is well known that it has a unique solution $u \in H_0^1(\Omega)$ which has an exponential interior layer at $x = d$ and boundary layer at $x = 0$. Furthermore, the following *a priori* estimate of the second derivative can be found at [164]:

$$\begin{aligned} |u^{(k)}(x)| &\leq C(1 + \varepsilon^{-k} e^{-bx/\varepsilon}),\ x \in \Omega^-, k = 0,1,2,3, \text{ and} \\ |u^{(k)}(x)| &\leq C(1 + \varepsilon^{-k} e^{-b(x-d)/\varepsilon}),\ x \in \Omega^+, k = 0,1,2,3. \end{aligned}$$

With this information we can construct the corresponding layer-adapted grid to get an optimal interpolant $u_I$ and thus obtain the optimal convergence of the new SDFEM. This example illustrates the usefulness of the near optimality of the SDFEM (Theorem 5.23) considering the fact that results for singularly perturbed problem with discontinuous right-hand side are relatively rare [164, 84].

In this chapter, we have shown the stabilization effect of the adaptive grid for the standard finite element method and developed an optimal streamline diffusion finite element method. The main results are listed below.

1. We found that the uniformity of the grid in the smooth part of the solution plays a crucial role for the optimality of the approximation.

2. In contrast, the new streamline diffusion finite element method that we developed inherits a quasi-optimal approximation property which is uniform with respect to $\varepsilon$.

3. With the optimal interpolation error estimate, we have answered an open question about the optimal choice of the monitor function for the singularly perturbed problem.

The above results raise many interesting questions for singularly perturbed problems in multiple dimensions. It is natural to expect that similar results should be still valid, but a rigorous theoretical analysis are still lacking and further research is still required. But at least our one dimensional results should provide some guidance to adaptive finite elements in high dimensions. For example, it is easy to construct a grid which is quasiuniform in the smooth part such that the convergence rate is deteriorated. Since for general domains in high dimensions, it is not easy to get uniform grids, the stabilization of the standard FEM is needed.

Appendix A

# Related Problems on Optimal Delaunay triangulations

In this appendix, we will discuss several related problems on optimal Delaunay triangulations. We first present an optimal Voronoi tessellation which is known as Centroidial Voronoi Tessellations (CVTs). Then we apply two special ODTs which minimize $Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty)$ and $Q(\mathcal{T}, \|\mathbf{x}\|^2, 1)$to the sphere covering problem and the optimal polytope approximation of convex bodies, respectively.

## A.1 Centroidal Voronoi tessellations

In this section, we shall understand the Voronoi tessellations as circumscribe polytopes approximation of the paraboloid. We measure the approximation error by the volume difference. The optimal one is called a centroidal Voronoi tessellation (CVT) and it is, more or less, the dual of an ODT.

We begin with the classic definition of Voronoi tessellations (or Voronoi diagrams).

**Definition A.1.** *Let $\Omega$ be an open set in $\mathbb{R}^n$ and $S = \{\mathbf{x}_i\}_{i=1}^N \subset \Omega$. For any $\mathbf{x}_i \in S$, we define the* Voronoi region *of $\mathbf{x}_i$ as*

$$V_i = \{\mathbf{x} \in \Omega, s.t. \|\mathbf{x} - \mathbf{x_i}\| < \|\mathbf{x} - \mathbf{x_j}\|\}.$$

*Then $\bar{\Omega} = \sum \bar{V}_i$. We call this partition $\mathcal{V}$ a* Voronoi tessellation *or* Voronoi diagram

*of* $\Omega$ *and points* $\{\mathbf{x}_i\}$ generators.

If we lift generators to the paraboloid $(\mathbf{x}, \|\mathbf{x}\|^2)$, we can characterize the Voronoi tessellation as the vertical projection of an upper convex envelope of tangential hyperplanes at those points [89]. Note that the envelope will form a circumscribed polytope $P^c$ of $C$. Thus we can understand the Voronoi tessellation as a circumscribe polytope approximation. The duality of Voronoi tessellation and Delaunay triangulation can be understand as the polar duality [12] of the inscribed and circumscribe polytopes.

**Theorem A.2.** *The volume difference between* $P^c$ *and* $C$ *is*

$$D(\mathcal{V}, \|\mathbf{x}\|^2, 1) := \sum_{i=1}^{N} \int_{V_i} \|\mathbf{x} - \mathbf{x}_i\|^2 d\mathbf{x}. \tag{A.1}$$

*Proof.* Let $\{\mathbf{x}_i\}_{i=1}^{n+1}$ be vertices of a simplex $\tau$ and $TM_{\mathbf{x}_i'}$ the tangential hyperplane of paraboloid at $\mathbf{x}_i'$ which is

$$x_{n+1} = \|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{x}_i\|^2. \tag{A.2}$$

It is clear that the point $(\mathbf{c}_o, \|\mathbf{c}_o\|^2 - R^2)$ satisfies (A.2) for $i = 1, 2, ...n+1$, where $\mathbf{c}_o$ and $R$ are the center and radius of the circumscribe sphere of $\tau$. The vertical projection of the upper convex envelope $\mathcal{V}'$ of $TM_{\mathbf{x}_i'}$ is the Voronoi tessellation.

By the construction of VT, we see that the part of boundary of $P^c$ which is projected to Voronoi region $V_i$ is supported by the tangent hyperplane $TM_{\mathbf{x}_i'}$. Thus by (A.2) the difference of the volume is:

$$\sum_{i=1}^{N} \int_{V_i} \left( \|\mathbf{x}\|^2 - x_{n+1} \right) d\mathbf{x} = \sum_{i=1}^{N} \int_{V_i} \|\mathbf{x} - \mathbf{x}_i\|^2 d\mathbf{x}.$$

□

We can generalize this quality with respect to any density function $\rho(\mathbf{x})$.

**Definition A.3.** *Let* $\rho(\mathbf{x})$ *be a density function on* $\Omega$*. For a Voronoi tessellation*

*$\mathcal{V}$ of $\Omega$ corresponding to generators $\{\mathbf{x}_i\}_{i=1}^N$, we define*

$$D(\mathcal{V},\rho,1)=\sum_{i=1}^{N}\int_{V_i}\|\mathbf{x}-\mathbf{x}_i\|^2\rho(\mathbf{x})d\mathbf{x}. \tag{A.3}$$

A dual concept of the optimal Delaunay triangulations or the optimal inscribed polytope approximations is the optimal Voronoi tessellations or the optimal circumscribe polytope approximations by minimizing $D(\mathcal{V},\rho,1)$.

**Definition A.4.** *$\mathcal{V}^*$ is a centroidal Voronoi tessellation if and only if*

$$D(\mathcal{V}^*,\rho,1)=\inf_{\mathcal{V}\in\mathcal{P}_N}D(\mathcal{V},\rho,1).$$

*Here $\mathcal{P}_N$ stands for the set of all Voronoi tessellation with at most $N$ generators.*

Why is it called centroidal Voronoi tessellation? Because for a CVT, the generator $\mathbf{x}_i$ is also the centroid of its Voronoi region $V_i$, i.e.

$$\mathbf{x}_i=\frac{\int_{V_i}\mathbf{x}\rho(\mathbf{x})d\mathbf{x}}{\int_{V_i}\rho(\mathbf{x})d\mathbf{x}}.$$

The proof is very simple. Let $\mathbf{x}_i$ be the centroid of $V_i$. For any point $\mathbf{z}_i\in V_i$, we have

$$\begin{aligned}
&\int_{V_i}||\mathbf{x}-\mathbf{x_i}||^2\rho(\mathbf{x})d\mathbf{x}\\
=\;&\int_{V_i}(\mathbf{x}-\mathbf{x_i})\cdot(\mathbf{x}-\mathbf{z_i})\rho(\mathbf{x})d\mathbf{x}\\
\le\;&\Big(\int_{V_i}||\mathbf{x}-\mathbf{x_i}||^2\rho(\mathbf{x})\Big)^{1/2}d\mathbf{x}\Big(\int_{V_i}||\mathbf{x}-\mathbf{z_i}||^2\rho(\mathbf{x})\Big)^{1/2}d\mathbf{x}.
\end{aligned}$$

Thus

$$\int_{V_i}||\mathbf{x}-\mathbf{x_i}||^2\rho(\mathbf{x})d\mathbf{x}\le\int_{V_i}||\mathbf{x}-\mathbf{z_i}||^2\rho(\mathbf{x})d\mathbf{x}.$$

For various important and interesting applications of CVTs, we refer to a nice review of Du et. al. [70]. Nowadays the theories and algorithms of CVTs are successfully applied to mesh generation and adaptation [71], general surface grid generation [73], anisotropic mesh generation [76] and mesh optimization in three

dimensions [77]. We believe that ODT shall also play an important role in the mesh generation and adaptation.

## A.2 Sphere covering problem

Roughly speaking, sphere covering problem is to seek the most economical way to cover a domain $\Omega$ in $\mathbb{R}^n$ with overlapping balls of equal size. Let us denote $B_n(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| \leq r\}$ and $S_n(\mathbf{x}, r)$ its boundary. If center is $\mathbf{o}$ or radius is 1, it will be omitted. For a convex domain $\Omega \subset \mathbb{R}^n$, we define the thickness $\theta_n$ as

$$\theta_n = \liminf_{r \to 0} \frac{N_r |B_n(r)|}{|\Omega|},$$

where $N_r$ is the minimum number of balls with radius $r$ needed to cover the domain and $|\cdot|$ is the standard Lebesgue measure. In the literature, the thickness is always defined as the limit when the domain goes to $\mathbb{R}^n$ while using the unit ball [192]. However it is equivalent to let the radius go to zero by the scaling argument. The choice of the convex domain $\Omega$ in the definition above is somewhat arbitrary since we have a theorem by Hlawka (see Zong [192] p.4) which says any convex domain leads to an equivalent definition. In other words, it is saying that in the asymptotic sense we can neglect the affection of the boundary of $\Omega$.

Now we consider the problem in the other way around. Let $V = \{\mathbf{x}_i\}_{i=1}^N$ be a finite point set such that the convex hull of $V$ is $\Omega$. We use these points as the centers of balls and denote the minimum radius needed to cover $\Omega$ by $R_V^c$. If we let $R_N^c = \inf_{\#V=N} R_V^c$, by the standard $\epsilon - N$ argument, it is easy to show that

$$\theta_n = \liminf_{N \to \infty} \frac{N |B_n(R_N^c)|}{|\Omega|}.$$

The sphere covering problem is then translated into finding the optimal distribution of $N$ points which will coincide with the vertices of an ODT. More precisely, we shall prove that

$$(R_N^c)^2 = \inf_{\mathcal{T} \in \mathcal{P}_N} Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty).$$

We then derive a lower bound for the interpolation error $Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty)$ which

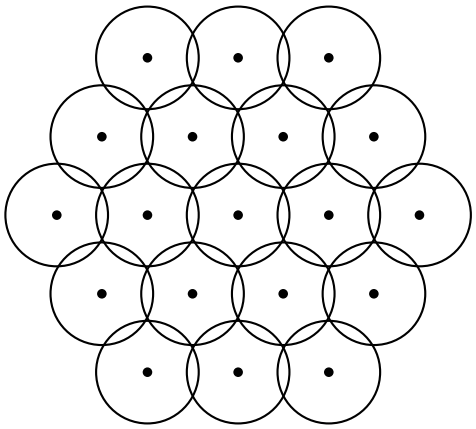


**Figure A.1.** Optimal sphere covering in two dimensions.

results a new approach to obtain Coxeter-Few-Rogers lower bound of $\theta_n$ [56]. Let

$$\tau_n = \phi_n \frac{n!}{\sqrt{n+1}} (\frac{n}{n+1})^{n/2-1},$$

where $\phi_n$ is the solid angle of a vertex of the $n$-regular simplex.

**Theorem A.5.**

$$\theta_n \geq \tau_n,\ 1 \leq n < \infty.$$

*Furthermore, we can only achieve this lower bound by regular triangulation for* $n = 1, 2$*, and thus*

$$\theta_1 = 1, \theta_2 = \frac{2\pi}{3\sqrt{3}}.$$

The regular triangulation in the theorem above means the triangulation with all simplices are regular, i.e. all the edge lengths in the triangulation are equal. It is well known that only for for $n = 1, 2$, we can have a regular tessellation of $\mathbb{R}^n$(see, for example, [55]). This is the reason why we have so many open problems in dimensions higher than two.

Let $V$ be a finite point set such that the convex hull of $V$ is $\Omega$. With the same point set $V$, we have an Delaunay triangulation. The following lemma reveals the connection between sphere covering problems and Delaunay triangulations.

**Lemma A.6.**

$$(R_V^c)^2 = Q(DT, \|\mathbf{x}\|^2, \infty) = \min_{\mathcal{T}\in\mathcal{P}_V} Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty).$$

*Proof.* Let us look at a simplex $\tau$ with vertices $\{\mathbf{x}_i\}_{i=1}^{n+1}$. For $u(\mathbf{x}) = \|\mathbf{x}\|^2$, by the multiple points Taylor expansion ([53] p.128), we know

$$u_I(\mathbf{x}) - u(\mathbf{x}) = \sum_{j=1}^{n+1} \lambda_j(\mathbf{x})\|\mathbf{x} - \mathbf{x}_j\|^2 \leq E_{\max}, \tag{A.4}$$

where $\lambda_i(\mathbf{x})$ is the barycenter coordinate of $\mathbf{x}$ in $\tau$ and $E_{\max}$ denotes $Q(\tau, \|\mathbf{x}\|^2, \infty)$. Since $\sum \lambda_i(\mathbf{x}) = 1$, (A.4) implies that, first, there exists a vertex $\mathbf{x}_i$ such that $\|\mathbf{x}-\mathbf{x}_i\|^2 \leq E_{\max}$ which means $\tau \subset \cup_{i=1}^{n+1} B(\mathbf{x}_i, E_{\max}^{1/2})$, and, secondly for $\mathbf{x}^*$ at which the error attains the maximum value, $\sum_{j=1}^{n+1} \|\mathbf{x}^* - \mathbf{x}_j\|^2 = E_{\max}$, which means to cover $\tau$ with balls of equal size centered at its vertices, the minimum radius is $E_{\max}^{1/2}$.

We thus proved that for any triangulation $\mathcal{T} \in \mathcal{P}_V$, if we use vertices as centers of balls, the square of the minimum radius needed to cover the domain is $Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty)$. By the optimality of Delaunay triangulations we finish the proof. □

As a direct consequence, the optimal distribution of the centers of the covering balls coincides with the vertices of an optimal Delaunay triangulation.

**Corollary A.7.**

$$(R_N^c)^2 = \inf_{\mathcal{T}\in\mathcal{P}_N} Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty).$$

We then derive a lower bound for the interpolation error $Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty)$.

**Lemma A.8.**

$$Q(\tau, \|\mathbf{x}\|^2, \infty) \geq \frac{n}{n+1}\frac{n!^{2/n}}{(n+1)^{1/n}}|\tau|^{2/n}, \tag{A.5}$$

*where the equality holds if and only if $\tau$ is regular.*

*Proof.* By (A.4), $E(\mathbf{x}) := u_I(\mathbf{x}) - u(\mathbf{x})$ only depends on the quadratic part of the function. Hence we may consider $g(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}_o\|^2$, where $\mathbf{c}_o$ is the circum center

of $\tau$. By looking at this way, we get

$$E(\mathbf{x}) = u_I(\mathbf{x}) - u(\mathbf{x}) = g_I(\mathbf{x}) - g(\mathbf{x}) = R_\tau^2 - \|\mathbf{x} - \mathbf{c}_o\|^2, \tag{A.6}$$

where $R_\tau$ is the radius of the circum sphere of $\tau$. If $\mathbf{c}_o \in \tau$, then $E_{\max} = R_\tau^2$. (A.5) is a well known geometric inequality for a simplex, for example see [147] (p.515) and the equality holds if and only if $\tau$ is regular.

Otherwise $E(\mathbf{x})$ attains its maximum at $\mathbf{x}^*$, the projection of $\mathbf{c}_o$ to $\tau$, i.e. $E_{\max} = R_\tau^2 - \|\mathbf{c}_o - \mathbf{x}^*\|^2$. In this case $\mathbf{x}^*$ is on some facet $\sigma$ of $\tau$, which is an $(n-1)$-simplex. By the definition of the projection, for $\mathbf{x} \in \sigma$

$$\|\mathbf{x} - \mathbf{x}^*\|^2 + \|\mathbf{x}^* - \mathbf{c}_o\|^2 = \|\mathbf{x} - \mathbf{c}_o\|^2. \tag{A.7}$$

Without lose of generality, we may assume $\sigma$ is opposite to vertex $\mathbf{x}_{n+1}$, namely it is made up by $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$. By (A.7), all the distances between $\mathbf{x}_i\,(1 \le i \le n)$ and $\mathbf{x}^*$ are equal. Thus $\mathbf{x}^*$ is the circum center of $\sigma$ and $E_{\max}$ is the square of the radius of the circum sphere of $\sigma$. By the characterization of the projection $(\mathbf{c}_o - \mathbf{x}^*) \cdot (\mathbf{x}_{n+1} - \mathbf{x}^*) \le 0$, we get

$$\begin{aligned} \|\mathbf{x}_{n+1} - \mathbf{x}^*\|^2 &= \|\mathbf{x}_{n+1} - \mathbf{c}_o\|^2 + \|\mathbf{c}_o - \mathbf{x}^*\|^2 + 2(\mathbf{x}_{n+1} - \mathbf{c}_o) \cdot (\mathbf{c}_o - \mathbf{x}^*) \\ &= R_\tau^2 - \|\mathbf{c}_o - \mathbf{x}^*\|^2 + 2(\mathbf{x}_{n+1} - \mathbf{x}^*) \cdot (\mathbf{c}_o - \mathbf{x}^*) \\ &\le R_\tau^2 - \|\mathbf{c}_o - \mathbf{x}^*\|^2. \end{aligned}$$

Thus $\tau \subset B_n(\mathbf{x}^*, E_{\max}^{1/2})$.

We then construct a simplex $\tau'$ with $|\tau'| \ge |\tau|$ which is inscribed to $B(\mathbf{x}^*, E_{\max}^{1/2})$. Let us choose a coordinate such that $\mathbf{x}^*$ is the origin and $\sigma$ is on $x^{n+1} = 0$. Suppose the coordinate of the vertex which opposites to $\sigma$ is $\mathbf{v} = (v^1, v^2, ..., v^{n+1})$. We change it to $\mathbf{v}' = (v^1, v^2, ..., (E_{\max} - \sum_{i=1}^n (v^i)^2)^{1/2})$. Then $\mathbf{v}'$ and $\sigma$ gives us an inscribed simplex $\tau'$ and obviously $|\tau'| \ge |\tau|$. Applying the first case to $\tau'$, we finish the proof. □

**Theorem A.9.** *Let $N_T$ be the number of simplices in the triangulation, we have*

$$Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty) \ge C_{n,\infty} N_T^{-\frac{2}{n}} |\Omega|^{\frac{2}{n}},$$

*where*

$$C_{n,\infty} = \frac{n}{n+1}\frac{n!^{2/n}}{(n+1)^{1/n}}.$$

*The equality holds if and only if $\mathcal{T}$ is a regular triangulation, namely all edges of $\mathcal{T}$ are equal.*

*Proof.* By Lemma A.5 and the Cauchy inequality,

$$\begin{aligned} Q(\mathcal{T}, \|\mathbf{x}\|^2, \infty) &= \max_{\tau\in\mathcal{T}} Q(\tau, \|\mathbf{x}\|^2, \infty) \geq \sum_{\tau\in\mathcal{T}} Q(\tau, \|\mathbf{x}\|^2, \infty)/N_T \\ &\geq C_{n,\infty}\sum_{\tau\in\mathcal{T}} |\tau|^{2/n}/N_T \geq C_{n,\infty} {N_T}^{-\frac{2}{n}}\Omega^{\frac{2}{n}}. \end{aligned}$$

The equality holds if and only if $Q(\tau, \|\mathbf{x}\|^2, \infty) = C_{n,\infty}|\tau|^{2/n} = \text{constant}$, $\forall\tau \in \mathcal{T}$, i.e. $\mathcal{T}$ is a regular triangulation. □

Now we are going to connect the number of simplices $N_T$ and number of vertices $N$. For $n = 1$, it is trivial to show $\lim_{N\to\infty} N/N_T = 1$. Let us consider triangulations in two dimension. We sum angles $\phi_{\tau,k}$ of triangles in two different ways, namely elementwise and pointwise. We can easily show

$$2\pi N > \sum_{i=1}^{N_T}\left(\sum_{k=1}^{3}\phi_{\tau_i,k}\right) = \pi N_T,$$

and

$$\lim_{N\to\infty}\frac{N}{N_T} = \frac{1}{2}.$$

Thus with Theorem A.9 we proved that $\theta_1 = 1, \theta_2 = 2\pi/3\sqrt{3}$.

To deal with higher dimensions, we shall introduce the concept of the solid angle. The following definition and lemma are adopted from Zong's book [192].

**Definition A.10.** *Let $P$ be a polytope in $\mathbb{R}^n$ with vertices $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$, and write*

$$V_i = \{\mathbf{v}_i + \lambda(\mathbf{x} - \mathbf{v}_i) : \mathbf{x} \in P, \lambda \geq 0\}.$$

*Then we call*

$$\phi(\mathbf{v}_i) = |S_n(\mathbf{v}_i, 1) \cap V_i|_s$$

*the* solid angle *of $P$ at $\mathbf{v}_i$, where $|\cdot|_s$ means the surface area.*

For a regular simplex, all the solid angles are equal. We denote it by $\phi_n$. Let $\kappa_n := |S_n|_s/\phi_n$ be the number of equilateral simplices surrounding a vertex. The integer $\kappa_n$ is corresponding to a regular triangulation which is only possible for $n = 1, 2$; See for example [55].

The following important lemma was introduced by Coxeter, Few and Rogers [56]. The proof can be found in [192].

**Lemma A.11.** *Let $\tau$ be an $n$-dimensional simplex, with vertices $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{n+1}$. Then*

$$\sum_{i=1}^{n+1} \phi(\mathbf{v}_i) \geq (n+1)\phi_n,$$

*where the equality holds when $\tau$ is a regular simplex.*

With this lemma, we apply the same argument as that in the two dimensions to get an inequality between $N$ and $N_T$.

**Corollary A.12.** *For a triangulation $\mathcal{T}$,*

$$\frac{N}{N_T} \geq \frac{n+1}{\kappa_n}.$$

Now we are in the position to prove Theorem A.5

*Proof.* **Proof of Theorem A.5** We will use abbreviation $Q_p(\mathcal{T}) = Q(\mathcal{T}, \|\mathbf{x}\|^2, p)$, $1 \leq p \leq \infty$. Since

$$\theta_n = \liminf_{N\to\infty} \frac{N|B_n(Q_\infty^{1/2}(T))|}{|\Omega|} = \liminf_{N\to\infty} |B_n| \frac{N}{N_T} \frac{N_T Q_\infty^{n/2}(T)}{|\Omega|},$$

the result follows from Theorem A.9 and Corollary A.12. □

However it is not easy to get a computable formula for $\tau_n$. Here we list an asymptotic formula obtained by Rogers [161].

$$\tau_n \sim \frac{n}{e\sqrt{e}} \text{ when } n \to \infty. \tag{A.8}$$

The proof can be found at [192].

## A.3 Optimal polytope approximation of convex bodies

Let us now discuss briefly the second problems on the optimal polytope approximation to a convex body. A convex body is a compact convex subset of $\mathbb{R}^n$ with non-empty interior. We denote $\mathcal{C}$ the space of all convex bodies in $\mathbb{R}^n$ and $\delta^V(\cdot,\cdot)$ the volume difference metric on the space $\mathcal{C}$, i.e. $\delta^V(C,D) = |C \cup D| - |C \cap D|$. For a given convex body $C \in \mathcal{C}$, let $\mathcal{P}_N^i$ be the set of all polytopes inscribed to $C$ with at most $N$ vertices and $\delta^V(C,\mathcal{P}_N^i) = \inf_{P\in\mathcal{P}_N} \delta^V(C,P)$. In [100], Gruber showed that for a convex body $C$ whose boundary is of class $\mathcal{C}^2$ with Gauss curvature $\kappa_C > 0$ in $\mathbb{R}^{n+1}$, there exists a constant $\mathrm{del}_n$ depending only on $n$ such that

$$\lim_{N\to\infty} N^{\frac{2}{n}} \delta^V(C,\mathcal{P}_N^i) = \frac{1}{2}\mathrm{del_n} \left( \int_{\partial C} \kappa_C(\mathbf{x})^{\frac{1}{n+2}} d\sigma(\mathbf{x}) \right)^{\frac{n+2}{n}}, \tag{A.9}$$

where $\sigma$ is the ordinary surface area measure in $\mathbb{R}^n$. Further $\mathrm{del}_1 = 1/6$, and $\mathrm{del}_2 = 1/(2\sqrt{3})$. Again for $n \geq 3$, it is difficult, if it is not impossible, to get the exact value of $\mathrm{del_n}$. There are some estimates about $\mathrm{del_n}$ [96, 142, 143]. We shall present a sharper estimate for the constant $\mathrm{del}_n$ in this paper.

For $n = 1$, (3) was indicated by L. Fejes Tóth [179] and proved by McClure and Vitale [144]. Proof for $n = 2$ is due to Gruber [98] and the general case was obtained by Gruber [100]. For other forms of optimal approximating polytopes with respect to other metrics, we refer to [95, 99, 113].

**Remark A.13.** *In view of the characterization theory of the nonlinear approximation [61], to retain the asymptotic formula for $\delta^V(C,\mathcal{P}_N^i)$, $\partial C$ must have certain regularity in terms of Besov norms.*

Since Gauss curvature only appears in the last term of (A.9), to estimate $\mathrm{del}_n$ we can choose any convenient convex body we want. By considering the paraboloid $(\mathbf{x}, \|\mathbf{x}\|^2)$, it is easy to show that (c.f.[100])

$$\mathrm{del_n} = \lim_{\mathrm{N}\to\infty} \mathrm{N}^{2/\mathrm{n}} \inf_{\mathcal{T}\in\mathcal{P}_\mathrm{N}} \mathrm{Q}(\mathcal{T}, \|\mathbf{x}\|^2, 1)/|\Omega|^{2/\mathrm{n}+1},$$

for any convex domain $\Omega \subset \mathbb{R}^n$. With the lower and upper bound of $Q(\mathcal{T}, \|\mathbf{x}\|^2, 1)$,

we get an estimate of the constant $\mathrm{del_n}$.

**Theorem A.14.**

$$\frac{n+1}{n+2}(\frac{\tau_n}{|B_n|})^{2/n} \leq \mathrm{del_n} \leq \frac{n+1}{n+2}(\frac{\theta_n}{|B_n|})^{2/n},$$

*where $|B_n|$ is the volume of the unit ball in $\mathbb{R}^n$.*

Our estimate is asymptotic exact when dimension $n$ goes to infinity.

**Corollary A.15.**

$$\lim_{n\to\infty} \frac{\mathrm{del}_n}{n} = \frac{1}{2\pi e}.$$

The asymptotic exact estimate is also obtained in [143]. However our approach here is simpler and more straightforward. For $n = 1, 2$ since $\theta_n = \tau_n$, the estimate is exact, i.e. we obtain $\mathrm{del}_1 = 1/6$, and $\mathrm{del}_2 = 1/(2\sqrt{3})$ by our estimate. Although the thickness in the upper bound are not known for $n \geq 3$, any reasonable upper bound of $\theta_n$ can be used to bound $\mathrm{del_n}$ above. For example, by choosing a special lattice sphere covering scheme (see [54], p.36), which is the thinnest covering known in all dimensions $n \leq 23$, we get a computable formula for the upper bound that is

$$\mathrm{del}_n \leq (n+1)^{1/n}\frac{n}{12}. \tag{A.10}$$

When $n$ large, we may use the upper bound obtained by Rogers [160],

$$\theta_n < n\ln n + n\ln\ln n + 5n, \text{ for } n \geq 3.$$

Comparing with the result of Mankiewicz and Schütt [143],

$$\frac{n}{n+2}\frac{1}{|B_n|^{2/n}} \leq \mathrm{del_n} \leq \frac{n}{n+2}\frac{1}{|B_n|^{2/n}}\frac{\Gamma(n+2+2/n)}{(n+1)!}, n \geq 2$$

our lower bound is sharper and the upper bound (A.10) is sharper in lower dimensions ($n \leq 13$). The reason for the upper bound (4) becomes worse when $n \geq 14$ is that the sphere covering scheme we choose are away from the optimal one especially when $n$ is large. Actually $\lim_{n\to\infty}(n+1)^{1/n}/12 = e/12 > 1/(2\pi e)$.

We will follow the same line in section 2 to estimate the constant $\mathrm{del_n}$. We first

present an explicit formula for the interpolation error $Q(\mathcal{T}, \|\mathbf{x}\|^2, 1)$. The following lemma is a special case of Lemma 2.10 in Chapter 1.

**Lemma A.16.**

$$Q(\mathcal{T}, \|\mathbf{x}\|^2, 1) = \frac{1}{(n+2)(n+1)} \sum_{\tau \in \mathcal{T}} |\tau| \sum_{k=1}^{n(n+1)/2} d_{\tau,k}^2,$$

*where $d_{\tau,k}$ is the $k$-th edge length of $\tau$.*

The following geometric inequality can be found at [147] (p.517). For two dimensions, it is a direct consequence of the well-known Heron's formula of a triangle.

**Lemma A.17.** *For an $n$-simplex $\tau$, we have:*

$$\sum_{k=1}^{n(n+1)/2} d_k^2 \geq \frac{n(n+1)n!^{2/n}}{(n+1)^{1/n}} |\tau|^{2/n},$$

*and the equality holds if and only if $\tau$ is regular.*

**Theorem A.18.** *For a triangulation $\mathcal{T}$ of a bounded domain $\Omega$ with $N_T$ simplices, we have*

$$Q(\mathcal{T}, \|\mathbf{x}\|^2, 1) \geq C_{n,1} {N_T}^{-\frac{2}{n}} |\Omega|^{\frac{n+2}{n}}$$

*where*

$$C_{n,1} = \frac{n}{n+2} \frac{n!^{2/n}}{(n+1)^{1/n}}.$$

*The equality holds if and only if $\mathcal{T}$ is an regular triangulation, namely all edges of $\mathcal{T}$ are equal.*

*Proof.* By Lemma A.16 and Lemma A.17, we have

$$Q(\mathcal{T}, \|\mathbf{x}\|^2, 1) \geq C_{n,1} \sum_{i=1}^{N_T} |\tau_i|^{2/n+1} \geq C_{n,1} N_T^{-2/n} |\Omega|^{\frac{n+2}{n}}.$$

First equality holds if and only if $\tau_i$'s are regular and the second one holds if and only if $|\tau_i|$'s are equal. Thus equality holds if and only if all edges are equal. □

In general, we have

$$Q(\mathcal{T}, \|\mathbf{x}\|^2, p) \geq C_{n,p} {N_T}^{-\frac{2}{n}} |\Omega|^{\frac{n+2}{n}}, 1 \leq p \leq \infty.$$

The expression of $C_{n,p}$ can be found at Chapter 1.

*Proof.* **Proof of Theorem A.14**. Combing Theorem A.18 with the lower bound of $N/N_T$ (see Corollary A.12), we will prove the lower bound for $\mathrm{del}_n$, i.e.

$$\mathrm{del}_n \geq \frac{n+1}{n+2} (\frac{\tau_n}{|B_n|})^{2/n}.$$

Without loss of generality, we may choose $\Omega$ such that $|\Omega| = 1$. For any triangulation $\mathcal{T}$ of $\Omega$, we have

$$N^{2/n} Q_1(\mathcal{T}) = (\frac{N}{N_T})^{2/n} N_T^{2/n} Q_1(\mathcal{T}) \geq (\frac{n+1}{\kappa_n})^{2/n} C_{n,1} = \frac{n+1}{n+2} (\frac{\tau_n}{|B_n|})^{2/n}.$$

The desired result is obtained by sending $N$ to $\infty$.

To prove the upper bound

$$\mathrm{del}_n \leq \frac{n+1}{n+2} (\frac{\theta_n}{|B_n|})^{\frac{2}{n}},$$

we use a geometric inequality for a simplex $\tau$ (see [147],p.515)

$$\sum_{i=1}^{n+1} d_{\tau,i}^2 \leq (n+1)^2 R_\tau^2.$$

$R_\tau^2$ in the right side can be modified to $Q(\tau, \|\mathbf{x}\|^2, \infty)$ by the same argument as that in Lemma A.5. Combining with Lemma A.16, we know

$$Q_1(\mathcal{T}) = \frac{1}{(n+2)(n+1)} \sum_{\tau \in \mathcal{T}} (|\tau| \sum_{i=1}^{n(n+1)/2} d_{\tau,i}^2) \leq \frac{n+1}{n+2} Q_\infty(T) |\Omega|.$$

For simplicity, we choose $|\Omega| = 1$. For any $\mathcal{T}$ with $N$ vertices we have

$$\mathrm{del}_n \leq N^{2/n} Q_1(\mathcal{T}) \leq \frac{n+1}{n+2} (N Q_\infty^{\frac{n}{2}}(\mathcal{T}))^{2/n}.$$

The desired result then follows. □

*Proof.* Proof of Corollary A.15 By the asymptotic formula of $\tau_n$ (A.8), we know $\lim_{n\to\infty} \tau_n^{2/n} = 1$. On the other hand, Rogers [160] gives an upper bound for $\theta_n$,

$$\theta_n < n \ln n + n \ln \ln n + 5n, \text{ for } n \geq 3.$$

Thus $\lim_{n\to\infty} \theta_n^{2/n} = 1$.

It is well known that

$$|B_n| = \frac{\pi^{n/2}}{\Gamma(n/2+1)}.$$

With Stirling's formula

$$\Gamma(n/2+1) \sim \sqrt{2\pi} e^{-n/2} (\frac{n}{2})^{(n+1)/2},$$

we get

$$\lim_{n\to\infty} \frac{\mathrm{del}_n}{n} = \lim_{n\to\infty} \frac{1}{n|B_n|^{2/n}} = \frac{1}{2\pi e}.$$

□

# Bibliography


[1] S. Adjerid and J. E. Flaherty. A moving-mesh finite element method with local refinement for parabolic partial differential equations. *Comp. Meht. Appl. Mech. Engry.*, 55:3–26, 1986.

[2] A. Agouzal, K. Lipnikov, and Y. Vassilevski. Adaptive generation of quasi-optimal tetrahedral meshes. *East–West J. Numer. Math.*, 7:223–244, 1999.

[3] A. Agouzal and Y. Vassilevski. On a discrete Hessian recovery for $P_1$ finite elements. *J. Numer. Math.*, 10:1–12, 2002.

[4] M. Aiffa and J. E. Flaherty. A geometrical approach to mesh smoothing. *Comput. Methods Appl. Mech. Engrg.*, 192:4497–4514, 2003.

[5] P. Alliez, D. Cohen-Steiner, M. Yvinec, and M. Desbrun. Variational tetrahedral meshing. *ACM Transactions on Graphics*, 24(3):617–625, 2005.

[6] N. Amenta, M. Bern, and D. Eppstein. Optimal point placement for mesh smoothing. *Journal of Algorithms*, pages 302–322, February 1999.

[7] T. Apel. *Anisotropic finite elements: Local estimates and applications*. Book Series: Advances in Numerical Mathematics. Stuttgart: Teubner, 1999.

[8] T. Apel, M. Berzins, P. K. Jimack, G. Kunert, A. Plake, I. Tsukerman, and M. Walkley. Mesh shape and anisotropic elements: Theory and practice. *In J. R. Whiteman (ed. ): The Mathematics of Finite Elements and Applications X, Elsevier, Amsterdam,*, pages 367–376, 2000.

[9] T. Apel, S. Grosman, P. K. Jimack, and A. Meyer. A new methodology for anisotropic mesh refinement based upon error gradients. *Prepint*, 2001.

[10] T. Apel and J. Schoberl. Multigrid methods for anisotropic edge refinement. *SIAM Journal on Numerical Analysis*, 40(5):1993–2006, 2000.

[11] D. C. Arney and J. E. Flaherty. An adaptive mesh-moving and local refinement method for time-dependentt partial differential equatioins. *ACM Trans. Math. Software*, 16:48–71, 1990.

[12] F. Aurenhammer and R. Klein. *Handbook of Computational Geometry*. Amsterdam, Netherlands: North-Holland, 2000.

[13] I. Babuška and A. K. Aziz. On the angle condition in the finite element method. *SIAM Journal on Numerical Analysis*, 13(2):214–226, 1976.

[14] I. Babuška and T. Strouboulis. *The Finite Element Method and Its Reliability*. Numerical Mathematics and Scientific Computation, Oxford Science Publications, 2001.

[15] N. S. Bakhalov. Towards optimization of methods for solving boundary value problems in the presence of boundary layers. *Zh. Vychisl. Mater. Mater. Fiz.*, 9:841–859, 1969. in Russian.

[16] R. E. Bank and A. H. Sherman. An adaptive multilevel method for elliptic boundary value problems. *Computing*, 26:91–105, 1981.

[17] R. E. Bank, A. H. Sherman, and A. Weiser. Refinement algorithms and data structures for regular local mesh refinement. In R. S. et al., editor, *Scientific Computing*, pages 3–17. IMACS/North-Holland Publishing Company, Amsterdam, 1983.

[18] R. E. Bank and R. K. Smith. Mesh smoothing using a posteriori error estimates. *SIAM Journal on Numerical Analysis*, 34:979–997, 1997.

[19] R. E. Bank and J. Xu. Asymptotically exact a posteriori error estimators, Part I: Grids with superconvergence. *SIAM Journal on Numerical Analysis*, 41(6):2294–2312, 2003.

[20] R. E. Bank and J. Xu. Asymptotically exact a posteriori error estimators, Part II: General unstructured grids. *SIAM Journal on Numerical Analysis*, 41(6):2313–2332, 2003.

[21] S. Bartels and C. Carstensen. Each averaging technique yields reliable a posteriori error control in fem on unstructured grids. II. higher order fem. *Mathematics of Computation*, 71(239):971–994, 2002.

[22] G. Beckett, J. A. Mackenzie, A. Ramage, and D. M. Sloan. Computational solution of two-dimensional unsteady PDEs using moving mesh methods. *J. Comput. Phys.*, 182(2):478–495, November 2002.

[23] M. Bern and D. Eppstein. Mesh generation and optimal triangulation. In *Computing in Euclidean Geometry, Edited by Ding-Zhu Du and Frank Hwang*. World Scientific, Lecture Notes Series on Computing – Vol. 1, 1992.

[24] M. Berzins. A solution-based triangular and tetrahedral mesh quality indicator. *SIAM J. Scientific Computing*, 19:2051–2060, 1998.

[25] M. Berzins, P. K. Jimack, M. Walkley, and L. J. K. Durbeck. Mesh quality and moving meshes for 2d and 3d unstructured mesh flow solvers. *VKI Lecture Series 2000-05, 31st Computational Fluid Dynamics*, 2000.

[26] P. Binev, W. Dahmen, R. DeVore, and P. Petrushev. Approximation classes for adaptive methods. *Serdica Math. J*, 28:391–416, 2002.

[27] H. Borouchaki, M. J. Castro-Diaz, P. L. George, F. Hecht, and B. Mohammadi. Anisotropic adaptive mesh generation in two dimensions for CFD. In *5th International Conference On Numerical Grid Generation in Computational Field Simulations*, volume 3, pages 197–206. Mississppi State University, 1996.

[28] H. Borouchaki, P. L. George, F. Hecht, P. Laug, and E. Saltel. Delaunay mesh generation governed by metric specifications. I. algorithms. *Finite Elem. Anal. Des.*, 25(1-2):61–83, 1997.

[29] H. Borouchaki, P. L. George, and B. Mohammadi. Delaunay mesh generation governed by metric specifications. II. applications. *Finite Elem. Anal. Des.*, 25(1-2):85–109, 1997.

[30] J. U. Brackbill and J. S. Saltzman. Adaptive zoning for singular problems in two dimensions. *J. Comput. Phys.*, 46:342–368, 1982.

[31] F. Breezi, L. P. Franca, T. J. R. Hughes, and A. Russo. $b = \int g$. *Comput. Methods Appl. Mech. Engrg.*, 145:329–339, 1997.

[32] F. Brezzi, T. J. R. Hughes, L. D. Marini, A. Russo, and E. Süli. A priori error analysis of residual-free bubbles for advection-diffusion problems. *SIAM Journal on Numerical Analysis*, 36(4):1933–1948, 1999.

[33] F. Brezzi, D. Marini, and E. Süli. Residual-free bubbles for advection-diffusion problems: the general error analysis. *Numer. Math.*, 85:31–47, 2000.

[34] F. Brezzi and A. Russo. Choosing bubbles for advection-diffusion problems. *Mathematical Models and Methods in Applied Science*, 4:571–587, 1994.

[35] K. Q. Brown. Voronoi diagrams from convex hulls. *Inform. Process. Lett.*, 9:223–228, 1979.

[36] G. C. Buscaglia and E. A. Dari. Anisotropic mesh optimization and its application in adaptivity. *nternational Journal for Numerical Methods in Engineering*, 40(22):4119–4136, 1997.

[37] W. Cao, W. Huang, and R. D. Russell. A study of monitor functions for two dimensional adaptive mesh generation. *SIAM J. Sci. Comput.*, 20:1978–1994, 1999.

[38] W. Cao, J. Lang, W. Huang, and R. D. Russell. A two-dimensional moving finite element method with local refinement based on a posteriori error estimates. *Appl. Numer. Math.*, 46:75–94, 2003.

[39] G. F. Carey and H. T. Dinh. Grading functions and mesh redistribution. *SIAM Journal on Numerical Analysis*, 22(5):1028–1040, 1985.

[40] C. Carstensen and S. Bartels. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. I. low order conforming, nonconforming, and mixed FEM. *Mathematics of Computation*, 71(239):945–969, 2002.

[41] C. M. Chen and Y. Huang. *High accuracy theory of finite element methods*. Hunan, Science Press, Hunan, China (in Chinese), 1995.

[42] L. Chen. Mesh smoothing schemes based on optimal Delaunay triangulations. In *13th International Meshing Roundtable*, pages 109–120, Williamsburg, VA, 2004. Sandia National Laboratories.

[43] L. Chen. New analysis of the sphere covering problems and optimal polytope approximation of convex bodies. *Journal of Approximation Theory*, 133(1):134–145, March 2005.

[44] L. Chen. Superconvergence of tetrahedral linear finite elements. *International Journal of Numerical Analysis and Modeling*, 3(3):273–282, 2006.

[45] L. Chen, P. Sun, and J. Xu. Multilevel homotopic adaptive finite element methods for convection dominated problems. In *The Proceedings for 15th Conferences for Domain Decomposition Methods*, Lecture Notes in Computational Science and Engineering 40, pages 459–468. Springer, 2004.

[46] L. Chen, P. Sun, and J. Xu. Optimal anisotropic simplicial meshes for minimizing interpolation errors in $L^p$-norm. *Accepted by Mathematics of Computation*, 2004.

[47] L. Chen, J. Z. Wang, and J. Xu. Asymptotically optimal and liner-time algorithm for polygonal curve simplification. *Submitted to IEEE*, 2005.

[48] L. Chen and J. Xu. Optimal Delaunay triangulations. *Journal of Computational Mathematics*, 22(2):299–308, 2004.

[49] L. Chen and J. Xu. An optimal streamline diffusion finite element method for a singularly perturbed problem. In *AMS Contemporary Mathematics Series: Recent Advances in Adaptive Computation*, volume 383, pages 236–246, Hangzhou, 2005.

[50] L. Chen and J. Xu. Stability and accuracy of adapted finite element methods for singularly perturbed problems. *Submitted to Numer. Math.*, 2005.

[51] Y. Chen. Uniform pointwise convergence for a singularly perturbed problem using arc-length equidistribution. proceedings of the 6th japan-china joint seminar on numerical mathematics (tsukuba, 2002). *J. Comput. Appl. Math. 159*, 159(1):25–34, 2003.

[52] Y. Chen. Uniform convergence analysis of finite difference approximations for singular perturbation problems on an adaptive grid. *Accepted by Advances in Computational Mathematics*, 2005.

[53] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 4 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978.

[54] J. H. Conway and N. J. A. Sloane. *Sphere Packing, Lattices and Groups*. 2nd ed. New York, NY:Spring-Verlag, 1999.

[55] H. S. M. Coxeter. *Regular polytopes*. New York : Macmillan, 1963.

[56] H. S. M. Coxeter, L. Few, and C. A. Rogers. Covering space with equal spheres. *Mathematika*, 6:147–157, 1959.

[57] E. F. D'Azevedo. Optimal triangular mesh generation by coordinate transformation. *SIAM J. Sci. Statist. Comput.*, 12:755–786, 1991.

[58] E. F. D'Azevedo and R. B. Simpson. On optimal interpolation triangle incidences. *SIAM J. Sci. Statist. Comput.*, 6:1063–1075, 1989.

[59] C. de Boor. Good approximation by splines with variable knots. *Int. Seines Numer. Math, Birkhauser Verlag, Basel*, 21:57–72, 1973.

[60] C. de Boor. Good approximation by splines with variables knots II. In G. A. Watson, editor, *Proceedings of the Eleventh International Conference on Numerical Methods in Fluid Dynamics*, volume 363, pages 12–20. Springer–Verlag, Dundee, Scotland, 1974.

[61] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.

[62] R. A. Devore and G. G. Lorentz. *Constructive Approximation*. New York, NY:Spring-Verlag, 1993.

[63] Y. Di, R. Li, T. Tang, and P. Zhang. Moving mesh finite element methods for the incompressible navier-stokes equations. *to appear in SIAM J. Sci. Comput.*, 2003.

[64] V. Dolejsi. Anisotropic mesh adaptation for finite volume and finite element methods on triangular meshes. *Computing and Visualization in Science*, 1:165–178, 1998.

[65] V. Dolejsi. Anisotropic mesh adaptation technique for viscous flow simulation. *East-West Journal of Numerical Mathematics, 1-24,2001*, 9(1):1–24, 2001.

[66] V. Dolejsi and J. Felcman. Anisotropic mesh adaptation for numerical solution of boundary value problems. *Submitted to Numerical Methods for Partial Differential Equations*, 2001.

[67] J. Dompierre, M.-G. Vallet, P. Labbe, and F. Guibault. On simplex shape measures with extension for anisotropic meshes. *Presented at Workshop on Mesh Quality and Dynamic Meshing*, more, CA, janvier 2003. Sandia National Laboratories.:46–71, 2003.

[68] Q. Du and M. Emelianenko. Uniform convergence of an energy-based multilevel quantization scheme. *preprint*, 2004.

[69] Q. Du, M. Emelianenko, and L. Ju. Convergence properties of the Lloyd algorithm for computing the centrodial Voronoi tessellations. *SIAM Journal on Numerical Analysis*, To appear, 2004.

[70] Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):637–676, 1999.

[71] Q. Du and M. Gunzburger. Grid generation and optimization based on centroidal Voronoi tessellations. *Appl. Mathematics of Computation*, 133:591–607, 2002.

[72] Q. Du, M. Gunzburger, and L. Ju. Meshfree probabilistc determinate of point sets and support regions for meshless computing. *Comput. Methods Appl. Mech. Engrg.*, 191:1349–1366, 2002.

[73] Q. Du, M. Gunzburger, and L. Ju. Constrained CVTs in general surfaces. *SIAM J. Scientific Computing*, 24:1488–1506, 2003.

[74] Q. Du, M. Gunzburger, and L. Ju. Voronoi-based finite volume methods, optimal voronoi meshes, and pdes on the sphere. *Comput. Methods Appl. Mech. Engrg.*, 192:3993–3957, 2003.

[75] Q. Du, M. Gunzburger, L. Ju, and V. Faber. Finite volume methods on a sphere based on the constrained centoridal voronoi tessellations. *Comput. Methods Appl. Mech. Engrg.*, to appear.

[76] Q. Du and D. Wang. Anisotropic centroidal voronoi tessellations and their applications. *SIAM J. Scientific Computing*, Accepted, 2003.

[77] Q. Du and D. Wang. Tetrahedral mesh generation and optimization based on centroidal voronoi tessellations. *International Journal for Numerical Methods in Engineering*, 56:1355–1373, 2003.

[78] Q. Du and D. Wang. Theoretically guaranteed constrained boundary recovery for three dimensional delaunay triangulations. *report*, 2003.

[79] T. Dupont and Y. Liu. Symmetric error estimates for moving mesh galerkin methods for advection-diffusion equations. *SIAM Journal on Numerical Analysis*, 40(3):914–927, 2002.

[80] A. S. Dvinsky. Adaptive grid generation from harmonic maps on Riemannian manifolds. *J. Comput. Phys.*, 95(2):450–476, 1991.

[81] H. Edelsbrunner. Triangulations and meshes in computational geometry. *Acta Numerica*, pages 1–81, 2000.

[82] H. Edelsbrunner and R. Seidel. Voronoi diagrams and arrangements. *Disc. and Comp. Geom.*, 8(1):25–44, 1986.

[83] M. Emelianenko. *Multilevel and adaptive methods for some nonlinear optimization problems*. PhD thesis, Department of Mathematics, The Pennsylvanlia State University, 2005.

[84] P. A. Farrell, A. F. Hegarty, J. J. H. Miller, E. O'Riordan, and G. I. Shishkin. Singularly perturbed convection-diffusion problems with boundary and weak interior layers. *Journal of Computational and Applied Mathematics*, 166:131–151, 2004.

[85] D. A. Field. Laplacian smoothing and Delaunay triangulation. *Communications in Applied Numerical Methods*, 4:709–712, 1988.

[86] D. A. Field. Qualitative measures for initial meshes. *International Journal for Numerical Methods in Engineering*, 47:887–906, 2000.

[87] L. Formaggia and S. Perotto. New anisotropic a priori error estimates. *Numer. Math.*, 89(4):641–667, 2001.

[88] L. Formaggia and S. Perotto. Anisotropic error estimates for elliptic problems. *Numer. Math.*, pages 67–92, 2003.

[89] S. Fortune. Voronoi diagrams and Delaunay triangulations. In *Computing in Euclidean Geometry, Edited by Ding-Zhu Du and Frank Hwang*. World Scientific, Lecture Notes Series on Computing - Vol. 1, 1992.

[90] L. P. Franca and A. Russo. Deriving upwinding, mass lumping and selective reduced integration by residual-free bubbles. *Applied Mathematics Letters*, 9:83–88, 1996.

[91] L. Freitag. On combining laplacian and optimization-based mesh smoothing techniques. *AMD Trends in Unstructured Mesh Generation, ASME*, 220:37–43, July 1997.

[92] L. A. Freitag, M. T. Jones, and P. E. Plassmann. An efficient parallel algorithm for mesh smoothing. In *4th International Meshing Roundtable*, pages 47–58. Sandia National Laboratories, 1995.

[93] L. A. Freitag, M. T. Jones, and P. E. Plassmann. A parallel algorithm for mesh smoothing. *SIAM J. Sci. Comput.*, 20(6):2023–2040, 1999.

[94] I. Fried. Condition of finite element matrices generated from nonuniform meshes. *AIAA J.*, 10:219–221, 1972.

[95] S. Glasauer and P. M. Gruber. Asymptotic estimates for best and stepwise approximation of convex bodies III. *Forum Math.*, 9:383–404, 1997.

[96] Y. Grodon, S. Reinsner, and C. Schütt. Umbrellas and polytopal approximation of the Euclidean ball. *J. Approx. Theory*, 90:9–22, 1997.

[97] W. D. Gropp. Local uniform mesh refinement with moving grids. *SIAM J. Sci. Statist. Comput.*, 8:292–304, 1987.

[98] P. M. Gruber. Volume approximation of convex bodies by inscribed polytopes. *Math. Ann.*, 281:229–245, 1988.

[99] P. M. Gruber. Aspects of approximation of convex bodies. In P. M. Gruber and J. M. Wills, editors, *Handbook of Convex Geometry.*, volume A, pages 319–345. Amsterdam: North-Holland, 1993.

[100] P. M. Gruber. Asymptotic estimates for best and stepwise approximation of convex bodies II. *Forum Math*, 5:521–538, 1993.

[101] W. G. Habashi, M. Fortin, J. Dompierre, M. G. Vallet, D. Ait-Ali-Yahia, Y. Bourgault, M. P. Robichaud, A. Tam, and S. Boivin. *Anisotropic mesh optimization for structured and unstructured meshes*. In 28th Computational Fluid Dynamics Lecture Series. von Karman Institute, March 1997.

[102] R. Hamilton. *Harmonic Maps of Manifolds with Boundary*, volume 471 of *Lecture Notes in Math*. American Mathematical Society, Springer-Verlag, New York, 1975.

[103] W. Hoffmann, A. H. Schatz, L. B. Wahlbin, and G. Wittum. Asymptotically exact a posteriori estimators for the pointwise gradient error on each element in irregular meshes I: A smooth problem and globally quasi-uniform meshes. *Mathematics of Computation*, 70:897–909, 2001.

[104] W. Huang. Practical aspects of formulation and solution of moving mesh partial differential equations. *J. Comput. Phys.*, 171:753–775, 2001.

[105] W. Huang. Variational mesh adaptation: isotropy and equidistribution. *J. Comput. Phys.*, 174:903–924, 2001.

[106] W. Huang and R. D. Russell. Moving mesh strategy based on a gradient flow equation for two-dimensional problems. *SIAM J. Sci. Comput.*, 20:998–1015, 1999.

[107] W. Huang and W. Sun. Variational mesh adaptation II: Error estimates and monitor functions. *J. Comput. Phys.*, 184:619–648, 2003.

[108] T. J. R. Hughes. Multiscale phenomena: Green's functions, the dirichlet-to-neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Comput. Methods Appl. Mech. Engrg.*, pages 127, no. 1–4, 387–401., 1995.

[109] T. J. R. Hughes and A. Brooks. A multidimensional upwind scheme with no crosswind diffusion. In T. J. R. Hughes, editor, *Finite Element Methods for Convection Dominated Flows*, AMD,vol. 34,ASME, pages 19–35. New York, 1979.

[110] T. J. R. Hughes, G. Feijoo, L. Mazzei, and J. B. Quincy. The variational multiscale method - a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, 166:3–24, 1998.

[111] C. Johnson and U. Nvert. An analysis of some finite element methods for advection-diffusion problems. In A. v. d. S. O. Axelsson, L. S. Frank, editor, *Analytical and Numerical Approaches to Asymptotic Problems in Analysis*, pages 99–116. NorthHolland, Amsterdam, 1981.

[112] C. Johnson, A. H. Schatz, and L. B. Wahlbin. Crosswind smear and pointwise errors in streamline diffusion finite element methods. *Mathematics of Computation*, 49:25–38, 1987.

[113] J. K. Böröczky and M. Ludwig. Approximation of convex bodies and a momentum lemma for power diagrams. *Mh. Math.*, 127:101–110, 1999.

[114] R. B. Kellogg and A. Tsan. Analysis of some difference approximations for a singular perturbation problem without turning points. *Math. Comput*, 32:1025–1039, 1978.

[115] P. Knupp. Mesh generation using vector-fields. *J. Comput. Phys.*, 119:142–148, 1995.

[116] N. V. Kopteva. Uniform convergence with respect to a small parameter of a scheme with central difference on refining grids. *Comput. Math. Math. Phys.*, 39 (10):1594–1610, 1999.

[117] N. V. Kopteva and M. Stynes. A robust adaptive method for quasi-linear one-dimensional convection-diffusion problem. *SIAM Journal on Numerical Analysis*, 39:1446–1467, 2001.

[118] R. Kornhuber and R. Roitzsch. On adaptive grid refinement in the presence of internal or boundary layers. *IMPACT Comput. Sci. Engrg.*, 2:40–72, 1990.

[119] G. Kunert. A posteriori error estimation for anisotropic tetrahedral and triangular finite element meshes. *Ph. D. Thesis*, 1999.

[120] G. Kunert. An a posteriori residual error estimator for the finite element method on anisotropic tetrahedral meshes. *Numer. Math.*, 86(3):471–490, 2000.

[121] G. Kunert. A local problem error estimator for anisotropic tetrahedral finite element meshes. *SIAM Journal on Numerical Analysis*, 39(2):668–689, 2001.

[122] G. Kunert. A posteriori $l^2$ error estimation on anisotropic tetrahedral finite element meshes. *IMA Journal of Numerical Analysis*, 21(2):503–523, 2001.

[123] G. Kunert. Toward anisotropic mesh construction and error estimation in the finite element method. *Numerical Methods for Partial Differential Equations*, 189(5):625 – 648, 2001.

[124] G. Kunert. A posteriori error estimation for convection dominated problems on anisotropic meshes. *Math. Meth. Appl. Sci.*, 267:589–617, 2003.

[125] G. Kunert and S. Nicaise. Zienkiewicz-zhu error estimators on anisotropic tetrahedral and triangular finite element meshes. *M2AN Math. Model. Numer. Anal.*, 37(6):1013–1043, 2003.

[126] G. Kunert and R. Verfürth. Edge residuals dominate a posteriori error estimates for linear finite element methods on anisotropic triangular and tetrahedral meshes. *Numer. Math.*, 86(2):283–303, 2000.

[127] T. Lamber. The Delaunay triangulation maximize the mean inradius. In *Proc. 6th Canad. Conf. Comput. Geom*, pages 201–206, 1994.

[128] C. L. Lawson. Software for $C^1$ surface interpolation. *Mathematical Software III*, pages 161–194 J. R. Rice, ed. , Academic Press, 1977.

[129] W. Lenferink. Pointwise convergence of approximations to a convection-diffusion equation on a Shishkin mesh. *Appl. Numer. Math.*, 32 (1):69–86, 2000.

[130] B. Li and Z. M. Zhang. Analysis of a class of superconvergence patch recovery techniques for linear and bilinear finite elements. *Numer. Meth. PDEs*, 15:151–167, 1999.

[131] R. Li, T. Tang, and P. Zhang. Moving mesh methods in multiple dimensions based on harmonic maps. *J. Comput. Phys.*, 170(2):562–588, July 2001.

[132] R. Li, T. Tang, and P. Zhang. A moving mesh finite element algorithm for singular problems in two and three space dimensions. *J. Comput. Phys.*, 177(2):365–393, April 2002.

[133] T. Linß. Sufficient conditions for uniform convergence on layer-adapted grids. *Applied Numerical Mathematics*, 37:241–255, 2001.

[134] T. Linß. Layer-adapted meshes for convectioni-diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 192:1061–1105, 2003.

[135] T. Linß and M. Stynes. The SDFEM on Shishkin meshes for linear convection-diffusion problems. *Numer. Math.*, 87:457–484, 2001.

[136] K. Lipnikov and Y. Vassilevski. Optimal triangulations: Existence, approximation and double differentiation of $P_1$ finite element functions. *Comput. Math. Math. Phys.*, 43(6):827–835, 2003.

[137] K. Lipnikov and Y. Vassilevski. Parallel adaptive solution of 3d boundary value problems by hessian recovery. *Comput. Methods Appl. Mech. Engrg.*, 192:1495–1513, 2003.

[138] K. Lipnikov and Y. Vassilevski. Error estimates for Hessian-based mesh adaptation algorithms with control of adaptivity. In *13th International Meshing Roundtable*, pages 345–351. Sandia National Laboratories, 2004.

[139] V. D. Liseikin. *Grid Generation Methods*. Springer Verlag, Berlin, 1999.

[140] Y. Liu, R. E. Bank, T. F. Dupont, S. Garcia, and R. F. Santos. Symmetric error estimates for moving mesh mixed methods for advection-diffusion equations. *SIAM Journal on Numerical Analysis*, 40(6):2270–2291, 2003.

[141] J. A. Mackenzie and M. L. Robertson. A moving mesh method for the solution of the one-dimensional phase-field equations. *J. Comput. Phys.*, 181(2):526–544, Sept. 2002.

[142] P. Mankiewicz and C. Schütt. A simple proof of an estimate for the approximation of the Euclidean ball and the Delone triangulation numbers. *J. Approx. Theory*, 107:268–280, 2000.

[143] P. Mankiewicz and C. Schütt. Note on the Delone triangulation numbers. *J. Approx. Theory*, 111:139–142, 2001.

[144] D. E. McClure and R. E. Vitale. Polygonal approximation of plane convex bodies. *J. Math. Anal. Appl.*, 51:326–358, 1975.

[145] S. Micheletti, S. Perotto, and M. Picasso. Stabilized finite elements on anisotropic meshes :a priori error estimates for the advantion-diffusion and the stokes problem. *SIAM Journal on Numerical Analysis*, 41(3):1131–1162, 2003.

[146] J. J. H. Miller, E. O'Riordan, and G. I. Shishkin. *Fitted Numerical Methods For Singular Perturbation Problems*. World Scientific, 1996.

[147] D. S. Mitrinovic, J. E. Pecaric, and V. Volenec. *Recent Advances in Geometric Inequalities*. Mathematics and its applications: East European Series 28, 1989.

[148] K. W. Morton. *Numerical Solution of Convection-Diffusion Problems*, volume 12 of *Applied Mathematics and Mathematical Computation*. Chapman & Hall, 1996.

[149] E. Nadler. Piecewise linear best $L_2$ approximation on triangulations. In C. K. Chui, L. L. Schumaker, and J. D. Ward, editors, *Approximation Theory*, volume V, pages 499–502. Academic Press, 1986.

[150] K. Niijima. Pointwise error estimates for a streamline diffusion finite element scheme. *Numer. Math*, 56:707–719, 1990.

[151] M. Picasso. An anisotropic error indicator based on zienkiewicz-zhu error estimator:application to elliptic and parabolic problems. *SIAM J. SCI. COMPUT.*, 24(4):1328–1355, 2003.

[152] M. Picasso. Numerical study of the effectivity index for an anisotropic error indicator based on zienkiewicz-zhu error estimator. *COMMUNICATION IN NUMERICAL METHODS IN ENGINEERING*, 19:13–23, 2003.

[153] H. Pottmann, R. Krasauskas, B. Hamann, K. Joy, and W. Seibold. On piecewise linear approximation of quadratic functions. *J. Geometry and Graphics*, 4(1):31–53, 2000.

[154] Y. Qiu, D. M. Sloan, and T. Tang. Convergence analysis of an adaptive finite difference method for a singular perturbation problem. *J. Comput. Appl. Math.*, 116:121–143, 2000.

[155] V. T. Rajan. Optimality of the Delaunay triangulation in $r^d$. *Proc. of the Seventh Annual Symp. on Comp. Geom*, pages 357–363, 1991.

[156] W. Ren and X. Wang. An iterative grid redistribution method for singular problems in multiple dimensions. *J. Comput. Phys.*, 159:246–273, 2000.

[157] S. Rippa. Minimal roughness property of the Delaunay triangulation. *Comput. Aided Geom. Design*, 7:489–497, 1990.

[158] S. Rippa. Long and thin triangles can be good for linear interpolation. *SIAM Journal on Numerical Analysis*, 29:257–270, 1992.

[159] M. C. Rivara. Mesh refinement processes based on the generalized bisection of simplices. *SIAM Journal on Numerical Analysis*, 21:604–613, 1984.

[160] C. A. Rogers. A note on coverings. *Mathematika*, 4:1–6, 1957.

[161] C. A. Rogers. An asymptotic expansion for certain Schlafli functions. *J. London Math. Soc.*, 36:78–80, 1961.

[162] H. G. Roos. Layer-adapted grids for singular perturbation problems. *ZAMM,Z. Angew. Math. Mech.*, 78 (5):291–309, 1998.

[163] H. G. Roos, M. Stynes, and L. Tobiska. *Numerical Methods for Singularly Perturbed Differential Equations*, volume 24 of *Springer series in Computational Mathematics*. Springer Verlag, 1996.

[164] H. G. Roos and H. Zarin. The streamline-diffusion method for a convection-diffusion problem with a point source. *J. Comput. Appl. Math.*, 150:109–128, 2003.

[165] G. Sangalli. Quasi optimality of the supg method for the one-dimensional adavection-diffusion problem. *SIAM Journal on Numerical Analysis*, 41(4):1528–1542, 2003.

[166] A. H. Schatz and L. B. Wahlbin. On the quasi-optimality in $L_\infty$ of the $\overset{\circ}{H}{}^1$-projection into finite element spaces. *Mathematics of Computation*, 38(157):1–22, 1982.

[167] A. H. Schatz and L. B. Wahlbin. On the finite element method for singularly perturbed reaction-diffusion problems in two and one dimensions. *Mathematics of Computation*, 40(161):47–89, 1983.

[168] A. H. Schatz and L. B. Wahlbin. Asymptotically exact a posteriori estimators for the pointwise gradient error on each element in irregular meshes. part II: The piecewise linear case. *Mathematics of Computation*, 73:517–523, 2003.

[169] R. Schoen and S. Y. Yau. On univalent harmonic maps between surfaces. *Invent. Math.*, 44:265–278, 1978.

[170] M. Shephard and M. Georges. Automatic three-dimensional mesh generation by the finite octree technique. *Internat. J. Numeri. Methods Engrg.*, 32:709–749, 1991.

[171] G. I. Shishkin. *Grid approximation of singularly perturbed elliptic and parabolic equations.* PhD thesis, Second doctorial thesis,Keldysh Institute,Moscow, 1990. in Russian.

[172] R. Sibson. Locally equiangular triangulations. *Computer Journal*, 21:243–245, 1978.

[173] R. B. Simpson. Anisotropic mesh transformations and optimal error control. *Applied Numerical Mathematics*, 14:183–198, 1994.

[174] J. H. Smith and A. M. Stuart. Analysis of continuous moving mesh equations. *SIAM J. Sci. Statist. Comput*, 13:1265–1286, 1992.

[175] M. Stynes and L. Tobiska. A finite difference analysis of a streamline diffusion method on a Shishkin mesh. *Numerical Algorithms*, 18:337–360, 1998.

[176] M. Stynes and L. Tobiska. The SDFEM for a convection-diffusion problem with a boundary layer: Optimal error analysis and enhancement of accuracy. *SIAM Journal on Numerical Analysis*, 41, No. 5:1620–1642, 2003.

[177] X. Tai and J. Xu. Global convergence of subspace correction methods for convex optimization problems. *Mathematics of Computation*, 71(237):105–124, 2002.

[178] S. H. Teng and C. W. Wong. Unstructured mesh generation: Theory, practice, and perspectives. *International Journal of Computational Geometry and Applications*, 10(3):227–266, 2000.

[179] L. F. Tóth. *Lagerungen in der Ebene, auf der Kugel und im Raum*. Springer, Berlin, reprinted 1972.

[180] L. B. Wahlbin. *Superconvergence in Galkerkin finite element methods*. Springer Verlag, Berlin, 1995.

[181] A. B. White. On selection of equidistributing meshes for two-point boundary-value problems. *SIAM Journal on Numerical Analysis*, 16:472–502, 1979.

[182] G. Xia. *Anisotropic adaptation on unstructured grids*. PhD thesis, University of Tennessee, 2003.

[183] J. Xu and Z. M. Zhang. Analysis of recovery type a posteriori error estimators for mildly structured grids. *Mathematics of Computation*, pages 781–801, 2003.

[184] P. A. Zegeling. $r$-refinement for evolutionary PDEs with finite elements or finite differences. *Appl. Numer. Math.*, 26B:97–104, January 1998.

[185] P. A. Zegeling. Moving grid techniques. In B. K. S. F. Thompson and N. P. Weatherill, editors, *Handbook of Grid Generation*, pages 37–1 – 37–18. CRC Press LLC, 1999.

[186] Z. M. Zhang. Ultraconvergence of the patch recovery technique II. *Mathematics of Computation*, 69(229):141–158, 1999.

[187] Z. M. Zhang and A. Naga. A new finite element gradient recovery method: Superconvergence property. *SIAM J. Sci. Comput.*, Accepted.

[188] G. Zhou and R. Rannacher. Pointwise superconvergence of the streamline diffusion finite element method. *Numer. Meth. PDEs*, 12, CMP 96:05:123–145, 1996.

[189] T. Zhou and K. Shimada. An angle-based approach to two-dimensional mesh smoothing. In *9th International Meshing Roundtable*, pages 373–384. Sandia National Laboratories, October 2000.

[190] O. C. Zienkiewicz and J. Z. Zhu. The superconvergence patch recovery and a posteriori error estimates. Part 1: The recovery techniques. *International Journal for Numerical Methods in Engineering*, 33:1331–1364, 1992.

[191] O. C. Zienkiewicz and J. Z. Zhu. The superconvergence patch recovery and a posteriori error estimates. Part 2: Error estimates and adaptivity. *International Journal for Numerical Methods in Engineering*, 33:1365–1382, 1992.

[192] C. M. Zong. *Sphere packing*. New York, NY:Spring-Verlag, 1999.

# Vita

## Long Chen

Long Chen was born on October 9, 1974 in Leshan, Sichuan, P. R. CHINA. He received a B.S. degree in mathematics from Nanjing University in 1997 and M.S. degree in mathematics from Beijing University in 2000. In 2000 he enrolled in the Ph. D. program in mathematics at the Pennsylvania State University.