

SPECTRAL METHODS FOR BOUNDARY VALUE
PROBLEMS IN COMPLEX DOMAINS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yiqi Gu

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Jie Shen, Chair

Department of Mathematics

Dr. Suchuan Dong

Department of Mathematics

Dr. Guang Lin

Department of Mathematics

Dr. Jianlin Xia

Department of Mathematics

Dr. Xiangxiong Zhang

Department of Mathematics

Approved by:

David Goldberg

Head of Graduate Program

I dedicate this dissertation,
to my beloved family for their unconditional support.

ACKNOWLEDGMENTS

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my advisor, Dr. Jie Shen, whose expertise was invaluable in the formulating of the research topic and methodology in particular.

I would like to thank my committee members, Dr. Suchuan Dong, Dr. Guang Ling, Dr. Jianlin Xia and Dr. Xiangxiong Zhang for taking their time and energy to go through my thesis work and coordinate with me.

I would also like to acknowledge my colleagues from the research group at Purdue for their wonderful collaboration. You supported me greatly and were always willing to help me. I would particularly like to point out Dr. Jie Xu and Dr. Zhiguo Yang, I want to thank you for your excellent research presentations and ideas, and for all of the comments I was assisted in my research and further my dissertation.

In addition, I would like to thank my family and friends, who were of great support in deliberating over our problems and findings, as well as providing happy distraction to rest my mind outside of my research.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	viii
ABSTRACT	xi
1 OVERVIEW OF THESIS	1
2 SPECTRAL METHOD WITH RECTANGULAR EMBEDDING	4
2.1 Introduction	4
2.2 Embedding formulation	6
2.3 The first method	6
2.3.1 Well-posedness	7
2.3.2 A non-conforming Petrov-Galerkin spectral method	8
2.4 The second method	15
2.4.1 Weak formulation	16
2.4.2 Spectral approximation	16
2.4.3 General elliptic equations with non-constant coefficients	17
2.5 Numerical implementation	17
2.5.1 Derivation of the linear system	17
2.5.2 Fast and robust algorithm for the linear system	20
2.5.3 Fast matrix-vector multiplication	24
2.6 Numerical examples	27
2.7 Conclusion	40
3 SPECTRAL METHOD WITH CIRCULAR EMBEDDING	42
3.1 Introduction	42
3.2 Problems in simply connected domains	43
3.2.1 Dimension reduction	43
3.2.2 Approximation	45

	Page
3.2.3	Algorithm 48
3.3	Bessel-type equation solver 49
3.3.1	Homogeneous case 49
3.3.2	Inhomogeneous case 50
3.4	Error estimates 54
3.4.1	Notations 54
3.4.2	The 2-D Petrov-Galerkin formulation 55
3.4.3	Error estimate when $\rho(\theta) = \rho_0$ 58
3.5	Numerical examples 65
3.6	Conclusion 72
4	SPECTRAL METHOD WITH ANNULAR EMBEDDING 73
4.1	Introduction 73
4.2	Problem in non-simply connected domains 73
4.2.1	Dimension reduction 73
4.2.2	Approximation 75
4.2.3	Algorithm 76
4.3	Bessel-type equation solver 76
4.3.1	Homogeneous case 77
4.3.2	Inhomogeneous case 77
4.4	Numerical examples 79
4.5	Conclusion 83
5	APPLICATION I: FRACTIONAL LAPLACIAN EQUATION 86
5.1	Introduction 86
5.2	Fractional Laplacian Problem 86
5.3	Caffarelli-Silvestre extension 88
5.3.1	Extended problem 88
5.3.2	Galerkin formulation 89
5.3.3	Numerical examples 93

	Page
5.4 Dunford-Taylor representation	95
5.4.1 Representation by integrals	95
5.4.2 Numerical examples	96
5.5 Conclusion	96
6 APPLICATION II: HELMHOLTZ EQUATION	100
6.1 Introduction	100
6.2 Dimension reduction	100
6.3 Approximation	102
6.4 Numerical examples	104
6.5 Conclusion	107
REFERENCES	110
VITA	114

LIST OF FIGURES

Figure	Page	
2.1	<i>The structure of \mathbf{A} for $N = 15$ (around 5×10^4 total entries)</i>	21
2.2	l^2 error for computing $\mathbf{B}\mathbf{e}$ by interpolation method for different N_0 and N	26
2.3	<i>Number of arithmetic operations for different N</i>	26
2.4	$\kappa_1(\mathbf{R}_k)$ for the first example ($N=51$)	29
2.5	<i>the problem domain, extended domain and the boundary nodes for the first example ($N=51$)</i>	29
2.6	<i>the numerical solution (left) and the error (right) for the first example ($N=51$)</i>	30
2.7	$\ u - u_N\ _{L^2(\Omega)}$ versus N for the first example	30
2.8	$\ u - u_N\ _{L^2(\Omega)}$ versus N for the first example of different α	31
2.9	$\kappa_1(\mathbf{R}_k)$ for the second example ($N=35$)	32
2.10	<i>the problem domain, extended domain and the boundary nodes for the second example ($N=35$)</i>	32
2.11	<i>the numerical solution (left) and the error (right) for the second example ($N=35$)</i>	33
2.12	$\ u - u_N\ _{L^2(\Omega)}$ versus N for the second example	33
2.13	$\ u - u_N\ _{L^2(\Omega)}$ versus N for the second example of different α	34
2.14	<i>the problem domain, extended domain and the boundary nodes for the third example ($N=55$)</i>	35
2.15	<i>the error for the third example ($N=55$)</i>	35
2.16	$\ u - u_N\ _{L^2(\Omega)}$ versus N for the third example	36
2.17	<i>the problem domain, extended domain and the boundary nodes for the fourth example ($N=55$)</i>	37
2.18	<i>the error for the fourth example ($N=55$)</i>	37
2.19	$\ u - u_N\ _{L^2(\Omega)}$ versus N for the fourth example	38
2.20	<i>the problem domain, extended domain and the boundary nodes for the fifth example ($N=35$)</i>	39

Figure	Page
2.21 L^2 error $\ u - u_N\ _{L^2(\Omega)}$ versus N for the fifth example	39
3.1 The original domain Ω and the enclosing circle $\tilde{\Omega}$	44
3.2 the original domain, enclosing domain(left) and the error profile(right) for $M = 24$ in the first example	66
3.3 L^2 and max error versus M in the first example	66
3.4 the original domain, enclosing domain(left) and the error profile(right) for $M = 30$ in the second example	67
3.5 L^2 and max error versus M in the second example	67
3.6 the original domain, enclosing domain(left) and the error mesh(right) for $M = 64$ in the third example	68
3.7 decomposed domain by SEM package in the third example	69
3.8 L^2 error versus M by Algorithm 3.2.1 (left) and the square root of degree of freedom by general SEM package (right) in the third example	69
3.9 the original domain, enclosing domain(left) and the error profile(right) for $M = 96$ in the fourth example	70
3.10 decomposed domain by SEM package in the fourth example	70
3.11 L^2 error versus M by Algorithm 3.2.1 (left) and the square root of degree of freedom by general SEM package (right) in the fourth example	71
4.1 the original domain, enclosing domain and sampling nodes for $M = 8$ in the first example	80
4.2 L^2 error versus M in the first example	81
4.3 the original domain, enclosing domain and sampling nodes for $M = 8$ in the second example	82
4.4 L^2 error versus M in the second example	82
4.5 the error mesh for $M = 12$ in the third example	83
4.6 the original domain, enclosing domain and sampling nodes for $M = 12$ in the third example	84
4.7 L^2 error versus M in the third example	84
5.1 $\ u - u_{MN}\ _{L^2(\Omega)}$ versus N for the example solved by Caffarelli-Silvestre extension	94
5.2 $\ u - u_{MN}\ _{L^2(\Omega)}$ versus N for the example solved by Dunford-Taylor representation using sinc quadrature	97

Figure	Page
5.3 $\ u - u_{MN}\ _{L^2(\Omega)}$ versus N for the example solved by Dunford-Taylor representation using global adaptive quadrature	98
6.1 the original domain, embedding annulus and collocation nodes for $M = 15$ in the first example	105
6.2 L^2 errors versus M for $k = 10, 50, 100$ in the first example	105
6.3 L^∞ errors versus M for $k = 10, 50, 100$ in the first example	106
6.4 the original domain, embedding annulus and collocation nodes for $M = 15$ in the second example	107
6.5 L^2 errors versus M for $k = 10, 50, 100$ in the second example	108
6.6 L^∞ errors versus M for $k = 10, 50, 100$ in the second example	108

ABSTRACT

Gu, Yiqi Ph.D., Purdue University, August 2019. Spectral Methods For Boundary Value Problems in Complex Domains. Major Professor: Jie Shen.

Spectral methods for partial differential equations with boundary conditions in complex domains are developed with the help of a fictitious domain approach. For rectangular embedding, spectral-Galerkin formulations with special trial and test functions are presented and discussed, as well as the well-posedness and the error analysis. For circular and annular embedding, dimension reduction is applied and a sequence of 1-D problems with artificial boundary values are solved. Applications of our methods include the fractional Laplace problem and the Helmholtz equations. In numerical examples, our methods show good performance on the boundary value problems in both smooth and polygonal complex domains, and the L^2 errors decay exponentially for smooth solutions. For singular problems, high-order convergence rates can also be obtained.

1. OVERVIEW OF THESIS

Spectral method is one of the approximation techniques for solving partial differential equations (PDEs), and has been widely studied and applied in recent decades ([1–7]). The critical idea of spectral method is to expand the solution by globally supported smooth functions and to solve for the corresponding coefficients. For Dirichlet, Neumann and Robin boundary conditions, the basis functions are usually chosen by orthogonal polynomials, and for periodic boundary conditions, Fourier basis or spherical harmonic functions are always adopted. The accuracy of spectral method depends on the number of expansion terms and the regularity of the solutions.

Compared to other techniques for PDEs such as finite difference method and finite element method, spectral approximation enjoys two advantages. One is the regularity-dependent convergence rate. Especially for C^∞ solutions, the decay of numerical errors is usually able to attain to the exponential level. The other is the possibility of employing the fast Fourier or Chebyshev transformation when using Fourier or Chebyshev basis, which reduces the complexity from $O(N^2)$ to $O(N \ln N)$ for 1-D cases and from $O(N^3)$ to $O(N^2 \ln N)$ for 2-D cases.

One of the main restrictions of spectral method is the strict requirement on problem domains. In multi-dimensional cases, we can find spectral basis functions having closed forms only for separable domains, where the basis is usually constructed by tensor-product. In the thesis, some variants of spectral method have been developed carefully and they are specifically effective for boundary value problems in the domains of complex geometry.

In Chapter 2, spectral method for elliptic PDEs in complex domains is developed with the help of a fictitious domain approach. The original 2-D domain is embedded in a square, in which a new extended problem is solved. Two types of Petrov-Galerkin formulations with special trial and test functions are presented and discussed, as well

as the well-posedness and the error analysis. The corresponding linear system is solved by a fast algorithm. In numerical examples, this method shows good performance on the boundary value problems in both smooth and polygonal complex domains, and the L^2 errors decay exponentially. For singular problems, high-order convergence rates are also obtained.

In Chapter 3, we put forward the spectral method with circular embedding for elliptic PDEs, which transforms the multi-dimensional extended problem to a sequence of one-dimensional differential equations. Thanks to the dimension reduction, the extended problem can be weakly solved through a specific spectral-Galerkin formulation. The error estimate of this method is also studied and displayed. Finally, numerical examples demonstrate the convergence rate can reach to the exponential level for smooth solutions. For highly singular problems, the order of convergence is observed to be no less than that of classic spectral element solvers.

In Chapter 4, we deal with the elliptic PDEs in non-simply connected domains. The corresponding spectral method with annular embedding is proposed, and it works as well as the circular embedding scheme. In numerical examples, exponential error decay is obtained for analytic solutions and algebraic error decay of high orders for singular problems.

In Chapter 5, one application of the spectral methods with fictitious domain embedding is introduced. We will consider the fractional Laplacian equation, which can be solved by two approaches, Caffarelli-Silvestre extension or Dunford-Taylor representation. Both approaches find the solutions by solving a sequence of general Poisson-type Dirichlet problems. Our method shows special effectiveness for fractional Laplacian problems in complex domains.

In Chapter 6, another application will be presented, that we extend the spectral method with annular embedding scheme from Poisson-type problem to Helmholtz problem for acoustic scattering. We put forward an algorithm for the problems with respect to obstacles of complex geometry, and its complexity is no greater than the classic spectral method for circular obstacles. Moreover, the numerical examples

show the behaviors of error decay are consistent with that from the case of circular obstacles.

2. SPECTRAL METHOD WITH RECTANGULAR EMBEDDING

2.1 Introduction

Solving elliptic partial differential equations (PDEs) in multi-dimensional domains of complex geometry is a challenging problem. This is especially complicated by using spectral methods, since different from the solvers which discretize differential operators by local information, such as finite difference method or finite element method, spectral solver has to read the data globally. The irregularity of the domain contributes to difficulties of constructing global basis. A popular group of approaches are based on the fictitious domain method [8], which embeds the actual domain into a larger and regular one hence classical spectral solvers can be applied. One type of these approaches is to eliminate the original boundary condition by modifying the PDE in the extended domain, and to add an artificial condition on the extended boundary. This type includes the penalty method, where a penalty term is introduced into the PDE to force the original boundary condition to be satisfied approximately [9–11], and the diffuse domain method, where a phase-field function is used to characterize the original irregular domain [12–14]. However, the low order of regularity of the coefficient terms in these methods impedes high convergence rates of the spectral solvers. Although some techniques have been developed to improve the regularity [15–17], the corresponding computational cost is also increased significantly. Another type of approaches is to preserve the PDE but manipulate the original boundary condition, such as Lagrange multiplier approach [8, 18] and boundary integral method [19–21].

For fictitious domain methods, one has to extend the coefficient and data functions of the PDEs from the original domain to the enlarged one, and a smoother extension

usually leads to a higher convergence rate. The extension (or continuation) of a given function by using truncated Fourier series in 1D is widely studied [22–24], and in higher dimensional cases, the Fourier extension is usually implemented by performing 1D extension on a fixed direction [25–28]. We assume in the following a continuous extension can always be done on a given function (see (2.2)).

Spectral-collocation methods are applied to solve elliptic PDEs in complex domains in [29], where the boundary condition are satisfied at finite equispaced nodes along the boundary. The discrete spectral differential operator and the boundary constraints are combined to form the linear system. In spite of the ill-conditioning, it can be solved by Schur complement and truncated singular value decomposition. In [30], a spectral-Galerkin formulation with Lagrange multipliers is presented, and the boundary conditions are manipulated by using internal forcing functions compactly supported inside the fictitious domain. This method is improved in [31] by modifying the Dirac delta function basis for the Lagrange multipliers to be Fourier basis, thus both the PDE solution and the Lagrange multiplier are solved in the frequency space and the spectral accuracy is recovered.

A spectral-Petrov-Galerkin approach based on fictitious domain methods will be presented in this work. This approach follows [32] that a specific trial space is developed by which the original boundary condition is approximated by finite boundary constraints, and a general test space is chosen to preserve the well-posedness. Two methods will be discussed for the Poisson type problem $-\Delta u + \alpha u = f$ ($\alpha \geq 0$) with Dirichlet condition, and a particular algorithm for the (ill-conditioned) linear system will also be described. Furthermore, both the error analysis and numerical examples show the exponential convergence rate of this approach.

The organization of this chapter is as follows. In section 2.2 the embedding process and weak formulations of the extended problem are illustrated. In section 2.3 the Galerkin formulation, well-posedness analysis and error estimates of one spectral method are discussed. In section 2.4, another spectral-Galerkin method for more general cases is formulated. In section 2.5, the numerical implementation and the fast

algorithm are presented in detail. Numerical examples are shown in section 2.6 and the conclusion is drawn in section 2.7.

2.2 Embedding formulation

Given a simply connected domain $\Omega \in \mathbb{R}^d$, we consider the following PDE with Dirichlet boundary condition:

$$\begin{aligned} Lu &= f \text{ in } \Omega, \\ u &= h \text{ on } \partial\Omega, \end{aligned} \tag{2.1}$$

where L is a linear elliptic operator and $f \in C(\overline{\Omega})$. We consider first the case with homogeneous boundary condition ($h = 0$), and the inhomogeneous case will be discussed in the end of Sec. 2.5.1.

Our strategy for solving (2.1) is to embed Ω into a larger, regular domain. More precisely, we choose a suitable rectangular domain $\tilde{\Omega} \in \mathbb{R}^d$ s.t. $\Omega \subset \tilde{\Omega}$, and solve the following extended problem:

$$\begin{aligned} L\tilde{u} &= \tilde{f} \text{ in } \tilde{\Omega}, \\ \tilde{u} &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{2.2}$$

where $\tilde{f} \in L^2(\tilde{\Omega})$, with $\tilde{f} = f$ in Ω , is a smooth extension of f onto $\tilde{\Omega}$. It is clear that we have $\tilde{u} = u$ in Ω . Note that (2.2) is not a classical boundary value problem since the PDE solution is governed by prescribing its value on a $(d - 1)$ -dimensional manifold inside $\tilde{\Omega}$. In the following, we will simply denote \tilde{u} and \tilde{f} by u and f without ambiguity.

2.3 The first method

We restrict our attention to the Poisson equation, i.e. $L = -\Delta$. Let the trial space X and test space Y be defined as

$$X := \{u \in H^2(\tilde{\Omega}) : u = 0 \text{ on } \partial\Omega\}, \quad Y := L^2(\tilde{\Omega}). \tag{2.3}$$

X and Y are both Banach spaces with

$$\|u\|_X := \left(\int_{\tilde{\Omega}} |\Delta u|^2 \right)^{\frac{1}{2}}, \forall u \in X, \quad (2.4)$$

$$\|v\|_Y := \left(\int_{\tilde{\Omega}} |v|^2 \right)^{\frac{1}{2}}, \forall v \in Y. \quad (2.5)$$

It is clear that the norm defined in (2.4) is indeed a norm, since $\|u\|_X = 0$ implies u is harmonic, so by the maximum principle, we have $u = 0 \in \Omega$, and by unique continuation of harmonic function, we have $u = 0 \in \tilde{\Omega}$.

Our weak formulation is to find $u \in X$ s.t.

$$a_1(u, v) := - \int_{\tilde{\Omega}} \Delta u v = \int_{\tilde{\Omega}} f v, \quad \forall v \in Y. \quad (2.6)$$

2.3.1 Well-posedness

First, it is obvious that $a_1(\cdot, \cdot)$ is a continuous bilinear form on $X \times Y$. For the well-posedness, we need the following lemma.

Lemma 2.3.1 *Suppose Ω satisfies an interior cone condition [33, p.27]. Then under the definition in (2.4), (2.5) and (2.6), we have*

$$\inf_{u \in X} \sup_{v \in Y} \frac{a_1(u, v)}{\|u\|_X \|v\|_Y} \geq 1; \quad (2.7)$$

and

$$\sup_{0 \neq u \in X} a_1(u, v) > 0, \quad \forall 0 \neq v \in Y. \quad (2.8)$$

Specifically, (2.7) and (2.8) holds if Ω is a C^1 domain or a polygon.

Proof Given $u \in X$, we have $\Delta u \in Y$, so

$$\sup_{v \in Y} \frac{a_1(u, v)}{\|u\|_X \|v\|_Y} \geq \frac{a_1(u, \Delta u)}{\|u\|_X \|\Delta u\|_Y} = 1. \quad (2.9)$$

Next, for any $0 \neq v \in Y$, the Dirichlet problem

$$\begin{aligned} \Delta u &= v \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \quad (2.10)$$

admits a solution $u \in H^2(\Omega)$, denoted by u_1 . On the other hand, since $\tilde{\Omega} \setminus \bar{\Omega}$ satisfies an exterior cone condition, the Dirichlet problem

$$\begin{aligned} \Delta u &= v \quad \text{in } \tilde{\Omega} \setminus \bar{\Omega}, \\ u &= 0 \quad \text{on } \partial\Omega \cup \partial\tilde{\Omega}, \end{aligned} \tag{2.11}$$

admits a solution in $H^2(\tilde{\Omega} \setminus \bar{\Omega})$, denoted by u_2 [33, Theorem 2.14]. Let

$$u = \begin{cases} u_1 & \text{in } \Omega \\ u_2 & \text{in } \tilde{\Omega} \setminus \bar{\Omega} \\ 0 & \text{on } \partial\Omega \cup \partial\tilde{\Omega} \end{cases}, \tag{2.12}$$

then $u \in X$, and

$$\begin{aligned} a_1(u, v) &= \int_{\tilde{\Omega}} (\Delta u)v = \int_{\Omega} (\Delta u_1)v + \int_{\tilde{\Omega} \setminus \bar{\Omega}} (\Delta u_2)v \\ &= \int_{\Omega} v^2 + \int_{\tilde{\Omega} \setminus \bar{\Omega}} v^2 = \|v\|_Y^2 > 0. \end{aligned} \tag{2.13}$$

■

We then derive from the Banach-Necăs-Babuška theorem [34, p.112] that

Theorem 2.3.1 *Under the hypothesis of Lemma 2.3.1, the problem (2.6) admits a unique solution u satisfying*

$$\|u\|_X \leq \|f\|_Y, \quad \forall f \in Y. \tag{2.14}$$

2.3.2 A non-conforming Petrov-Galerkin spectral method

Let N be an odd integer, and P_N the polynomial space of degree no greater than N . Let $\xi_i : C(\partial\Omega) \rightarrow \mathbb{R}$ with $i = 1, \dots, 2N + 2$ represents $2N + 2$ independent constraints placed on u to approximate the original boundary condition $u = 0$ on $\partial\Omega$ in (2.3). This is similar to the boundary element used in boundary integral method ([20]). For example, one simple choice for ξ_i is

$$\xi_i(u_N) := u_N(\mathbf{z}_i), \quad i = 1, \dots, 2N + 2, \tag{2.15}$$

where $\{\mathbf{z}_i\}$ are a set of prescribed points on $\partial\Omega$. Another choice is

$$\xi_i(u_N) := \int_{\partial\Omega} u_N \chi_i ds, \quad i = 1, \dots, 2N + 2, \quad (2.16)$$

where $\{\chi_i\}$ are a set of linearly independent functions defined on $\partial\Omega$, and they play a similar role to the Lagrange multipliers (see [30]).

To simplify the presentation, we shall consider only the 2-D case although extension to 3D is straightforward. We also assume that the problem domain Ω in (2.6) is scaled so that it can be enclosed in $\tilde{\Omega} = (-1, 1) \times (-1, 1)$. We define

$$X_N := \{u_N \in P_N \times P_N, \xi_i(u_N) = 0, i = 1, \dots, 2N + 2\}, \quad (2.17)$$

and

$$Y_N := \text{span}\{\Delta(x^i y^j)\}_{i,j=0}^N. \quad (2.18)$$

Note that X_N is not a subspace of X . It is clear that

$$\dim(X_N) = (N + 1)^2 - (2N + 2) = N^2 - 1. \quad (2.19)$$

On the other hand, we have

Lemma 2.3.2 $\dim(Y_N) = N^2 - 1$ if N is odd and $\dim(Y_N) = N^2$ if N is even.

Proof We use the following table T to describe $\{\Delta(x^i y^j)\}_{i,j=0}^N$:

	0	1	2	3	...	N
0	0	0	1	x	...	x^{N-2}
1	0	0	y	xy	...	$x^{N-2}y$
2	1	x	(x^2, y^2)	(x^3, xy^2)	...	$(x^N, x^{N-2}y^2)$
3	y	xy	(x^2y, y^3)	(x^3y, xy^3)	...	$(x^N y, x^{N-2}y^3)$
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
N	y^{N-2}	xy^{N-2}	(x^2y^{N-2}, y^N)	(x^3y^{N-2}, xy^N)	...	$(x^N y^{N-2}, x^{N-2}y^N)$

In the above table, $T(j, i)$ is filled by $\Delta(x^i y^j)$ without its coefficient, and the parenthesis (\cdot, \cdot) means the linear combination of two terms with nonzero coefficients. From the table it is straightforward to see that, if N is odd,

$$T(0, i) \in \text{span}\{T(2, i - 2), T(4, i - 4), \dots, T(i - 1, 1)\} \quad (2.20)$$

and

$$T(1, i) \in \text{span}\{T(3, i - 2), T(5, i - 4), \dots, T(i, 1)\} \quad (2.21)$$

for $i = 2, \dots, N$. Hence by removing the first two rows of T , the reduced table $\{T(i, j)\}_{i=0, j=2}^N$ is still a spanning set of Y_N . To show $\{T(i, j)\}_{i=0, j=2}^N$ is linearly independent, note for $i = -(N - 2), -(N - 1), \dots, 2N - 2$, each anti-diagonal $\{T(N, i), T(N - 1, i + 1), \dots, T(3, i + N - 3)\}$ consists of all the entries of order $N - 2 + i$ in the reduced table (ignore the entries with indices which are negative or greater than $N + 1$), so distinct anti-diagonals are linearly independent. Also, every anti-diagonal itself is linearly independent since each entry in it has a special term that cannot be obtained by linearly combining other entries. Therefore, $\dim(Y_N)$ is equal to the number of entries in $\{T(i, j)\}_{i=0, j=2}^N$, which is $(N + 1)(N - 1) = N^2 - 1$.

The case of N even is essentially the same as the odd case except for one entry in (2.21), that is

$$T(1, N) \notin \text{span}\{T(3, N - 2), T(5, N - 4), \dots, T(N - 1, 2)\}. \quad (2.22)$$

Hence $T(1, N) \cup \{T(i, j)\}_{i=3, j=1}^N$ form a basis for Y_N and $\dim(Y_N) = N^2$. ■

Note $\dim(X_N) = \dim(Y_N)$ for odd N . Since X_N is not a subspace of X , we define

$$\|u_N\|_{X_N} := \left(\int_{\tilde{\Omega}} |\Delta u_N|^2 \right)^{\frac{1}{2}}, \quad (2.23)$$

which is consistent with (2.4). The norm of X_N in (2.23) is indeed a norm, as long as $\{\xi_i\}_{i=1}^{2N+2}$ in (2.17) are specifically chosen s.t. $\Delta : X_N \rightarrow Y_N$ has a trivial nullspace (this can always be satisfied in numerical implementation, and we assume this hypothesis holds in the remaining context).

Let $\mathbf{I}_N : L^2(\tilde{\Omega}) \rightarrow P_N \times P_N$ be the 2D tensorial polynomial interpolation operator at the Legendre-Gauss-Lobatto points. Our Petrov-Galerkin spectral method for (2.6) is: find $u_N \in X_N$ s.t.

$$a_1(u_N, v_N) = \int_{\tilde{\Omega}} \mathbf{I}_N f v_N, \quad \forall v_N \in Y_N. \quad (2.24)$$

To consider the well-posedness of (2.24), we need

Lemma 2.3.3

$$\inf_{u_N \in X_N} \sup_{v_N \in Y_N} \frac{a_1(u_N, v_N)}{\|u_N\|_{X_N} \|v_N\|_{Y_N}} \geq 1, \quad (2.25)$$

and

$$\sup_{u_N \in X_N} |a_1(u_N, v_N)| > 0, \quad \forall 0 \neq v_N \in Y_N. \quad (2.26)$$

Proof (2.25) can be proven by the exactly same argument as in the proof of Lemma 2.3.1. And (2.26) follows the fact $\dim(X_N) = \dim(Y_N)$ and [35, Proposition 2.21]. ■

Finally, by Lemma 2.3.3 we obtain

Theorem 2.3.2 *The approximate problem (2.24) admits a unique solution u_N , which satisfies the a priori estimate*

$$\|u_N\|_{X_N} \leq \|\mathbf{I}_N f\|_{L^2(\tilde{\Omega})}. \quad (2.27)$$

Error estimates for the solution

The first part is about the approximation property of X_N to X . We need

Lemma 2.3.4 *For any odd integer N ,*

$$P_{\frac{N-3}{2}} \times P_{\frac{N-3}{2}} \subset Y_N. \quad (2.28)$$

Proof By virtue of the proof of Theorem 3.1, the following reduced table consists of a basis for Y_N if N is odd.

	0	1	2	3	...	N
2	1	x	(x^2, y^2)	(x^3, xy^2)	...	$(x^N, x^{N-2}y^2)$
3	y	xy	(x^2y, y^3)	(x^3y, xy^3)	...	$(x^Ny, x^{N-2}y^3)$
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
N	y^{N-2}	xy^{N-2}	(x^2y^{N-2}, y^N)	(x^3y^{N-2}, xy^N)	...	$(x^Ny^{N-2}, x^{N-2}y^N)$

Denote $T_k := \{T(k+2, 0), T(k+1, 1), T(k, 2), \dots, T(2, k)\}$, for $k = 2, \dots, N-3$, which consists exactly of $k+1$ independent entries of order k . Hence T_k spans the space of 2D monomial of degree k . Therefore for $i \leq \frac{N-3}{2}, j \leq \frac{N-3}{2}$, $x^i y^j \in \text{span} T_{i+j}$, which implies $P_{\frac{N-3}{2}} \times P_{\frac{N-3}{2}} \subset Y_N$. \blacksquare

Next we introduce the error estimate for 2D tensorial polynomial interpolation, which is given by

Lemma 2.3.5 *Suppose the interpolation nodes for $\mathbf{I}_N : L^2(\tilde{\Omega}) \rightarrow P_N \times P_N$ are the roots of the Legendre polynomial of degree N for each variable, and let $u \in H^r(\tilde{\Omega})$ with $2 \leq r \leq N+1$, then*

$$\|\mathbf{I}_N u - u\|_{L^2(\tilde{\Omega})} \leq c \sqrt{\frac{(N-r+1)!}{N!}} (N+r)^{-\frac{r+1}{2}} |u|_{H^r(\tilde{\Omega})} \quad (2.29)$$

with a constant c . If r is fixed, then for large N ,

$$\|\mathbf{I}_N u - u\|_{L^2(\tilde{\Omega})} \leq c N^{-r} |u|_{H^r(\tilde{\Omega})}. \quad (2.30)$$

More detail about Lemma 2.3.5 can be found in [1]. Furthermore, we have the approximation result as follows.

Theorem 2.3.3 *Under the hypothesis for \mathbf{I}_N in Lemma 2.3.5 and the assumption that $u \in X \cap H^r(\tilde{\Omega})$ with fixed $r \geq 4$, it has*

$$\inf_{u_N \in X_N} \|\Delta(u - u_N)\|_{L^2(\tilde{\Omega})} \leq \left(\frac{N-3}{2}\right)^{-(r-2)} |u|_{H^r(\tilde{\Omega})}. \quad (2.31)$$

Proof Let $q := \mathbf{I}_{\frac{N-3}{2}}(\Delta u) \in P_{\frac{N-3}{2}} \times P_{\frac{N-3}{2}} \subset Y_N$ by Lemma 2.3.4. Note the linear problem

$$\text{find } w_N \in X_N \text{ s.t. } \Delta w_N = q, \quad (2.32)$$

admits a unique solution since $\dim(X_N) = \dim(Y_N)$ and Δ has a trivial nullspace.

Therefore

$$\begin{aligned} \inf_{u_N \in X_N} \|\Delta(u - u_N)\|_{L^2(\tilde{\Omega})} &\leq \|\Delta u - \Delta w_N\|_{L^2(\tilde{\Omega})} = \|\Delta u - \mathbf{I}_{\frac{N-3}{2}}(\Delta u)\|_{L^2(\tilde{\Omega})} \\ &\leq \left(\frac{N-3}{2}\right)^{-(r-2)} |\Delta u|_{H^{-(r-2)}(\tilde{\Omega})} \leq \left(\frac{N-3}{2}\right)^{-(r-2)} |u|_{H^r(\tilde{\Omega})}. \end{aligned} \quad (2.33)$$

■

Finally, the error estimate for (2.24) is given by

Theorem 2.3.4 *Suppose the solution u of problem (2.6) with $L = -\Delta$ and $f \in H^s(\tilde{\Omega})$ for some $s \geq 2$ satisfies the regularity hypothesis $u \in X \cap H^r(\tilde{\Omega})$ for some $r \geq 4$, then the solution u_N of the approximate problem (2.24) satisfies*

$$\|u - u_N\|_X \leq c \left(\left(\frac{N-3}{2}\right)^{-(r-2)} |u|_{H^r(\tilde{\Omega})} + N^{-s} |f|_{H^s(\tilde{\Omega})} \right) \quad (2.34)$$

for some constant $c > 0$.

Proof Due to the discrete inf-sup condition (2.25) and the continuity of $a(\cdot, \cdot)$ on $(X + X_N) \times Y$, (2.24) satisfies the hypothesis of the Second Strang Lemma ([36]), which gives

$$\begin{aligned} \|\Delta(u - u_N)\|_{L^2(\tilde{\Omega})} &\leq (1 + \|a\|) \inf_{u_N \in X_N} \|\Delta u - \Delta u_N\|_{L^2(\tilde{\Omega})} \\ &\quad + \sup_{v_N \in Y_N} \frac{|\int_{\tilde{\Omega}} \mathbf{I}_N f v_N - a(u, v_N)|}{\|v_N\|_{Y_N}}. \end{aligned} \quad (2.35)$$

Note if $f \in H^s(\tilde{\Omega})$, then by (2.30)

$$\begin{aligned} \left| \int_{\tilde{\Omega}} \mathbf{I}_N f v_N - a(u, v_N) \right| &= \left| \int_{\tilde{\Omega}} \mathbf{I}_N f v_N - \int_{\tilde{\Omega}} f v_N \right| \\ &\leq \|\mathbf{I}_N f - f\|_{L^2(\tilde{\Omega})} \|v_N\|_{L^2(\tilde{\Omega})} \leq c N^{-s} |f|_{H^s(\tilde{\Omega})} \|v_N\|_{L^2(\tilde{\Omega})} \end{aligned} \quad (2.36)$$

for some constant $c > 0$. Hence together with (2.31), the inequality (2.34) follows from (2.35). ■

Error estimates by using H^2 norm

We hope to finally obtain H^2 norm error for $e_N := u - u_N$ which is more convincing, but $\|e_N\|_{H^2(\tilde{\Omega})}$ is not bounded by $\|\Delta(e_N)\|_{L^2(\tilde{\Omega})}$ since no restriction is placed on the data near the boundary $\partial\tilde{\Omega}$. However, note the original PDE is considered only inside Ω , in which the error estimate is really worth studying. Moreover, the following well-known inequality holds

$$\|w\|_{H^2(\Omega)} \leq c\|Lw\|_{L^2(\Omega)}, \quad \forall w \in H^2(\Omega) \cap H_0^1(\Omega) \quad (2.37)$$

where $Lw := a^{ij}(x)D_{ij}w + b^i(x)D_iw + c(x)w$ is a strictly elliptic operator in Ω with coefficients $a^{ij} \in C(\bar{\Omega})$, $b^i, c \in L^\infty(\Omega)$ and $c \leq 0$. Note $e_N \in \bar{X}_N := \{w \in H^2(\Omega), \xi_i(w) = 0, i = 1, \dots, 2N + 2\}$, and \bar{X}_N approximates $H^2(\Omega) \cap H_0^1(\Omega)$ as $N \rightarrow \infty$ in some sense, it should have a similar form to (2.37) such as

$$\|e_N\|_{H^2(\Omega)} \leq \epsilon(e_N) + c\|Le_N\|_{L^2(\Omega)}, \quad (2.38)$$

for all $e_N \in \bar{X}_N$ with some specific requirement on e_N , where $\epsilon(e_N)$ is an infinitesimal term as $N \rightarrow \infty$ which depends on e_N . For example, we assume e_N has the following estimate on $\partial\Omega$,

$$\sup_{\partial\Omega} |e_N| < \delta(N), \quad (2.39)$$

and also assume Ω is regular enough so that the boundary value problem

$$\begin{aligned} \Delta e &= 0 \quad \text{in } \Omega, \\ e &= e_N \quad \text{on } \partial\Omega, \end{aligned} \quad (2.40)$$

admits a solution, denoted by \bar{e}_N . Note then $e_N - \bar{e}_N \in H^2(\Omega) \cap H_0^1(\Omega)$, hence by (2.37), it has

$$\|e_N - \bar{e}_N\|_{H^2(\Omega)} \leq c\|\Delta(e_N - \bar{e}_N)\|_{L^2(\Omega)} = c\|\Delta e_N\|_{L^2(\Omega)}. \quad (2.41)$$

Now let $\Omega' \subset \Omega$ be a domain such that $d' := \text{dist}(\Omega', \partial\Omega) > 0$, then by the regularity estimate for harmonic functions,

$$\sup_{\Omega'} |D^\alpha u| \leq \left(\frac{d|\alpha|}{d'}\right)^{|\alpha|} \sup_{\Omega} |u|, \quad \forall u \text{ harmonic in } \Omega, \quad (2.42)$$

we have

$$\begin{aligned}\sup_{\Omega'} |D\bar{e}_N| &\leq \frac{2}{d'} \sup_{\Omega} |\bar{e}_N| < \frac{2\delta(N)}{d'}, \\ \sup_{\Omega'} |D^2\bar{e}_N| &\leq \frac{16}{d'^2} \sup_{\Omega} |\bar{e}_N| < \frac{16\delta(N)}{d'^2}.\end{aligned}\tag{2.43}$$

Hence

$$\begin{aligned}\|\bar{e}_N\|_{H^2(\Omega')} &= \left(\int_{\Omega'} |\bar{e}_N|^2 + |D\bar{e}_N|^2 + |D^2\bar{e}_N|^2 \right)^{\frac{1}{2}} \\ &< c|\Omega'|^{\frac{1}{2}} \left(1 + \frac{1}{d'^2} + \frac{1}{d'^4} \right)^{\frac{1}{2}} \delta(N).\end{aligned}\tag{2.44}$$

Therefore by combining (2.41) and (2.44), we obtain the following H^2 estimate for e_N in Ω' ,

$$\begin{aligned}\|e_N\|_{H^2(\Omega')} &\leq \|\bar{e}_N\|_{H^2(\Omega')} + \|e_N - \bar{e}_N\|_{H^2(\Omega')} \\ &\leq \|\bar{e}_N\|_{H^2(\Omega')} + \|e_N - \bar{e}_N\|_{H^2(\Omega)} \\ &\leq c \left(|\Omega'|^{\frac{1}{2}} \left(1 + \frac{1}{d'^2} + \frac{1}{d'^4} \right)^{\frac{1}{2}} \delta(N) + \|\Delta e_N\|_{L^2(\Omega)} \right).\end{aligned}\tag{2.45}$$

In (2.45), the bound for the H^2 error grows up as $\Omega' \rightarrow \Omega$, that implies the error is likely to become significantly as it approaches to the boundary of Ω . If only L^2 estimate is required, then by modifying (2.44) for $\|\bar{e}_N\|_{L^2(\Omega)}$, it directly leads to the L^2 estimate for e_N in Ω ,

$$\|e_N\|_{L^2(\Omega)} \leq c \left(|\Omega|^{\frac{1}{2}} \delta(N) + \|\Delta e_N\|_{L^2(\Omega)} \right).\tag{2.46}$$

2.4 The second method

Although the method presented in the last section can be applied to more general elliptic equations, it is only mathematically justified for the Poisson equation. In fact, numerical evidence indicates that the convergence rate deteriorates as $\alpha \rightarrow 0$ if the method is applied to the Poisson problem

$$\begin{aligned}(-\Delta + \alpha I)u &= f, \quad \text{in } \Omega, \\ u &= 0, \quad \text{on } \partial\Omega.\end{aligned}\tag{2.47}$$

Therefore, we shall present another Petrov-Galerkin method which does not have this problem.

2.4.1 Weak formulation

We restrict our attention to the special case that $L = -\Delta + \alpha I$ ($\alpha > 0$), and set the trial and test spaces to be

$$X := \{u \in H^1(\tilde{\Omega}), \text{tr}(u) = 0 \text{ on } \partial\Omega\}, \quad \|u\|_X := \left(\int_{\tilde{\Omega}} u^2 + |\nabla u|^2 \right)^{\frac{1}{2}}, \quad (2.48)$$

$$Y := H_0^1(\tilde{\Omega}), \quad \|v\|_Y := \left(\int_{\tilde{\Omega}} v^2 + |\nabla v|^2 \right)^{\frac{1}{2}}. \quad (2.49)$$

Then the weak problem is to find $u \in X$ s.t.

$$a_2(u, v) := \int_{\tilde{\Omega}} \nabla u \cdot \nabla v + \alpha uv = \int_{\tilde{\Omega}} f v, \quad \forall v \in Y. \quad (2.50)$$

Here X differs from Y by forcing the functions contained in it vanish on the interior boundary $\partial\Omega$ rather than the outer boundary $\partial\tilde{\Omega}$.

2.4.2 Spectral approximation

We set the approximate trial and test spaces as

$$X_N := \{u_N \in P_N \times P_N, \xi_i(u_N) = 0, i = 1, \dots, 4N\}, \quad (2.51)$$

and

$$Y_N := P_N^0 \times P_N^0, \quad (2.52)$$

where $P_N^0 := \{p \in P_N, p(\pm 1) = 0\}$. The sampling points $\{\xi_i\}$ are still distributed on $\partial\Omega$ as in the last section but the number here is increased to $4N$ to force $\dim(X_N) = \dim(Y_N) = (N - 1)^2$. Our Petrov-Galerkin method is to find $u_N \in X_N$ s.t.

$$a_2(u_N, v_N) = \int_{\tilde{\Omega}} \mathbf{I}_N f v_N, \quad \forall v_N \in Y_N, \quad (2.53)$$

where $a_2(\cdot, \cdot)$ is defined in (2.50).

The numerical experiments show the second spectral method (2.53) works better than the preceding one described in Sec. 2.3 for problem (2.47) with a nonzero α (see Sec. 2.5).

2.4.3 General elliptic equations with non-constant coefficients

Furthermore, the second method (2.53) can be employed to solve elliptic equations with non-constant coefficients. Considering the following Dirichlet problem,

$$\begin{aligned} -\nabla \cdot (\beta(x, y)\nabla u) + \alpha(x, y)u &= f \quad \text{in } \Omega, \\ u &= h \quad \text{on } \partial\Omega, \end{aligned} \tag{2.54}$$

where $\alpha(x, y) \geq 0$ and $\beta(x, y) \geq \beta_0 > 0$, we can write its weak formulation by

$$\begin{cases} \text{find } u \in X \text{ s.t.} \\ a_{2,g}(u, v) := \int_{\tilde{\Omega}} \beta \nabla u \cdot \nabla v + \alpha uv = \int_{\tilde{\Omega}} f v, \quad \forall v \in Y. \end{cases} \tag{2.55}$$

Correspondingly, the Galerkin formulation is given by

$$\begin{cases} \text{find } u_N \in X_N \text{ s.t.} \\ a_{2,g}(u_N, v_N) = \int_{\tilde{\Omega}} I_N f v_N, \quad \forall v_N \in Y_N. \end{cases} \tag{2.56}$$

2.5 Numerical implementation

2.5.1 Derivation of the linear system

We shall use Legendre polynomials to construct basis functions for X_N and Y_N . Recall the Legendre polynomials $\{L_k\}_{k=0}^N$ form an orthogonal basis for P_N satisfying

$$\int_{-1}^1 L_n(x)L_m(x)dx = \frac{2}{2n+1}\delta_{mn}. \tag{2.57}$$

Hence, we define $\bar{L}_n(x)$ be the polynomial that has a second derivative equal to $L_{n-2}(x)$ for $n \geq 2$, namely

$$\bar{L}_0(x) = 1, \quad \bar{L}_1(x) = x, \quad \bar{L}_2(x) = x^2/2, \quad \bar{L}_3(x) = x^3/6, \tag{2.58}$$

and

$$\begin{aligned}\bar{L}_n(x) &:= \int_{-1}^x \int_{-1}^t L_n(s) ds dt \\ &= \frac{1}{(2n-3)(2n-5)} L_{n-4}(x) - \frac{2}{(2n-1)(2n-5)} L_{n-2}(x) + \frac{1}{(2n-1)(2n-3)} L_n(x),\end{aligned}\tag{2.59}$$

for $n \geq 4$. It can be verified that $\{\bar{L}_n(x)\}_{n=0}^N$ form a basis for P_N and

$$\frac{d^2}{dx^2} \bar{L}_n(x) = L_{n-2}(x) \text{ for } n \geq 2.\tag{2.60}$$

We start with basis functions for Y_N . For the first spectral method described in Sec. 2.3,

$$Y_N = \text{span}\{\tilde{L}_{mn}\}_{m=0, n=2}^N, \text{ with } \tilde{L}_{mn} = L_{m-2}(x)L_n(y) + L_m(x)L_{n-2}(y),\tag{2.61}$$

where $L_{-2} = L_{-1} := 0$. And for the second method described in Sec. 2.4,

$$Y_N = \text{span}\{\tilde{L}_{mn}\}_{m, n=0}^{N-2}, \text{ with } \tilde{L}_{mn} = \tilde{L}_m(x)\tilde{L}_n(y),\tag{2.62}$$

where $\tilde{L}_m(t) := L_{m+2}(t) - L_m(t) \in P_N^0$. Generally, we denote $Y_N = \text{span}\{\psi_j\}_{j=1}^{M'}$, where $M' = N^2 - 1$ for the first one and $M' = (N-1)^2$ for the second one is the dimension of Y_N and X_N .

Next we consider how to construct basis functions $\{\phi_i\}_{i=1}^{M'}$ for X_N . Due to complexity of domain boundary $\partial\Omega$ and the prescribed constraints $\{\xi_k(u_N) = 0\}$ in the definition of X_N , it is not possible to write these basis functions in a closed form, so we write

$$\phi_i = \sum_{s, t=0}^N d_{st}^i \bar{L}_s(x) \bar{L}_t(y) \text{ such that } \xi_k(\phi_i) = 0 \quad \forall k = 1, \dots, M,\tag{2.63}$$

where $M = 2N + 2$ for the first one and $M = 4N$ for the second one is the number of sampling points on the boundary of Ω .

For each ϕ_i , the M constraints $\{\xi_k(\phi_i) = 0\}_{k=1}^M$ defined in (2.17) can be written in a matrix-vector form:

$$\mathbf{B} \mathbf{d}^i = 0,\tag{2.64}$$

where $\mathbf{B} \in \mathbb{R}^{M \times (M+M')}$, independent of i , with the k -th row corresponding to the k -th constraint $\{\xi_k(\phi_j) = 0\}_{j=1}^{M'}$, and

$$\mathbf{d}^i := [d_{00}^i \ d_{01}^i \ \cdots \ d_{NN}^i]^T, \quad (2.65)$$

which is a long vector consisting all the coefficients of ϕ_i in (2.63) lexicographically. It is to be observed that \mathbf{B} is determined by Ω , $\tilde{\Omega}$ and the choice for ξ_k , and is independent of the PDE operator L and the data f .

It is now evident that $\{\phi_i\}_{i=1}^{M'}$ can be constructed by finding a basis for $\text{null}(\mathbf{B})$, since the basis contains exactly M' vectors, each of which corresponds to one element of $\{\phi_i\}_{i=1}^{M'}$. More precisely, let

$$\mathbf{D} := [\mathbf{d}^1 \ \mathbf{d}^2 \ \cdots \ \mathbf{d}^{M'}] \in \mathbb{R}^{(M+M') \times M'} \quad (2.66)$$

with linearly independent columns such that $\mathbf{B}\mathbf{D} = 0$, and denote

$$\bar{\mathbf{L}}(x, y) := [\bar{L}_0(x)\bar{L}_0(y) \ \bar{L}_0(x)\bar{L}_1(y) \ \cdots \ \bar{L}_N(x)\bar{L}_N(y)], \quad (2.67)$$

then formally we have

$$[\phi_1 \ \phi_2 \ \cdots \ \phi_{M'}] = \bar{\mathbf{L}}(x, y)\mathbf{D}. \quad (2.68)$$

Let $u_N = \sum_{i=1}^{M'} \tilde{u}_i \phi_i$, then (2.24) (or (2.53)) leads to the following linear system,

$$\sum_{i=1}^{M'} a(\phi_i, \psi_j) \tilde{u}_i = \int_{\tilde{\Omega}} \mathbf{I}_N f \psi_j := f_j, \text{ for } j = 1, \dots, M'. \quad (2.69)$$

Denoting

$$\mathbf{A} := [a(\bar{L}_s(x)\bar{L}_t(y), \psi_j)] \in \mathbb{R}^{M' \times (M+M')} \quad (2.70)$$

with row indices $j = 1, \dots, M'$ and column indices $s, t = 0, \dots, N$, and with the notation in (2.68), we can rewrite (2.69) as

$$\mathbf{A}\mathbf{D}\mathbf{u} = \mathbf{f}, \quad (2.71)$$

where

$$\mathbf{u} := [\tilde{u}_1 \ \tilde{u}_2 \ \cdots \ \tilde{u}_{M'}]^T, \quad \mathbf{f} := [f_1 \ f_2 \ \cdots \ f_{M'}]^T.$$

Note that for any given point (x_p, y_p) at which the solution is evaluated,

$$u_N(x_p, y_p) = [\phi_1 \ \phi_2 \ \cdots \ \phi_{M'}]_{|(x_p, y_p)} \mathbf{u} = \tilde{\mathbf{L}}(x_p, y_p) \mathbf{D} \mathbf{u} := \tilde{\mathbf{L}}(x_p, y_p) \mathbf{y}, \quad (2.72)$$

which means the evaluation of u_N only depends on $\tilde{\mathbf{L}}(x_p, y_p)$ and \mathbf{y} . Hence, instead of solving (2.71) for \mathbf{u} explicitly, we can solve

$$\mathbf{A} \mathbf{y} = \mathbf{f}, \quad (2.73)$$

directly.

Note (2.73) has M' equations for $M + M'$ unknowns. The rest of equations are from the boundary constraints

$$\mathbf{B} \mathbf{y} = 0. \quad (2.74)$$

Hence, the final linear system to be solved is

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}. \quad (2.75)$$

For problems with non-homogeneous boundary condition $u|_{\partial\Omega} = h$, it suffices to let

$$\mathbf{h} = \left[\int_{\partial\Omega} h \chi_1 ds \quad \int_{\partial\Omega} h \chi_2 ds \quad \cdots \quad \int_{\partial\Omega} h \chi_M ds \right]^T, \quad (2.76)$$

and replace the right vector in (2.75) by

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{f} \\ \mathbf{h} \end{bmatrix}. \quad (2.77)$$

2.5.2 Fast and robust algorithm for the linear system

Unfortunately, it is numerically observed that (2.75) is very ill-conditioned so a direct solver is not feasible. Note that the upper part $\mathbf{A} \mathbf{y} = \mathbf{f}$ is the approximation to the PDE $Lu = f$, while the lower part $\mathbf{B} \mathbf{y} = 0$ describes the boundary constraints. The idea is to solve the upper part accurately and relax the accuracy requirement for the lower part. More precisely, we aim to reduce the residue of $\mathbf{B} \mathbf{y} = \mathbf{h}$ as much as

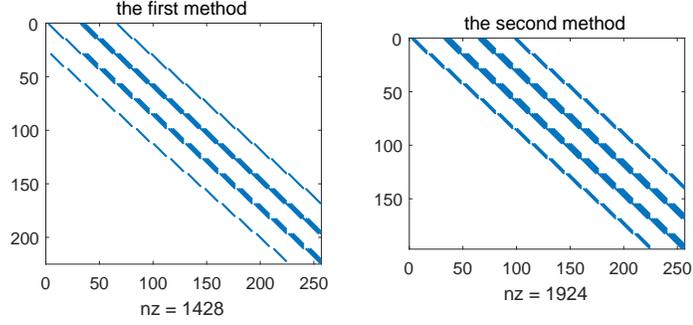


Fig. 2.1.: The structure of \mathbf{A} for $N = 15$ (around 5×10^4 total entries)

possible subject to $\mathbf{A}\mathbf{y} = \mathbf{f}$. A straightforward approach is to solve the least square problem

$$\min_{\mathbf{y} \in \mathbf{y}_s + \mathbf{Y}_K} \|\mathbf{h} - \mathbf{B}\mathbf{y}\|_2, \quad (2.78)$$

where \mathbf{y}_s is a particular solution of $\mathbf{A}\mathbf{y} = \mathbf{f}$ and \mathbf{Y}_K is a K -dimensional subspace of $\text{null}(\mathbf{A})$ with $K \leq M$. Note that if $K = M$, (2.78) is equivalent to (2.75). Hence, to avoid the ill-conditioning, K should not be too close to M in practical computation.

For a fixed $K < M$, we first find a particular solution \mathbf{y}_s of $\mathbf{A}\mathbf{y} = \mathbf{f}$ by letting \mathbf{y}_s being in the row space of \mathbf{A} , i.e.

$$\mathbf{y}_s = \mathbf{A}^T \mathbf{x}. \quad (2.79)$$

Hence it follows

$$(\mathbf{A}\mathbf{A}^T)\mathbf{x} = \mathbf{f}, \quad (2.80)$$

where $\mathbf{A}\mathbf{A}^T$ is symmetric positive-definite, so the above can be easily solved.

Thanks to the orthogonality of the Legendre polynomials, \mathbf{A} is a sparse block band matrix with 4 block bands (the structure of \mathbf{A} for $N = 15$ is shown in Fig. 2.1). So we can find easily an orthonormal set $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K$ in $\text{null}(\mathbf{A})$. Denote

$$\mathbf{Y}_K = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_K] \in \mathbb{R}^{(M+M') \times K}, \quad (2.81)$$

then (2.78) can be rewritten as

$$\min_{\mathbf{z}_K \in \mathbb{R}^K} \|\mathbf{h} - \mathbf{B}(\mathbf{Y}_K \mathbf{z}_K + \mathbf{y}_s)\|_2. \quad (2.82)$$

Therefore it suffices to compute the least square solution \mathbf{z}_K to (2.82) so that the solution to (2.78) is given by

$$\mathbf{y} = \mathbf{Y}_K \mathbf{z}_K + \mathbf{y}_s. \quad (2.83)$$

The choice of K is of critical importance, since large K may cause a large condition number, and small K may lead to large errors for the boundary constraints $\mathbf{B}\mathbf{y} = 0$. Therefore, we employ an adaptive procedure to choose K which better balances the ill-conditioning and the errors for the boundary constraints $\mathbf{B}\mathbf{y} = 0$.

We now describe how to solve the problem (2.82). We first rewrite it as the following over-determined linear system

$$\mathbf{B}\mathbf{Y}_K \mathbf{z}_K = \mathbf{g} := \mathbf{h} - \mathbf{B}\mathbf{y}_s. \quad (2.84)$$

We start by using the QR factorization with Householder transformation to (2.84). In the $(k-1)$ -th iteration, we have the following form

$$\tilde{\mathbf{Q}}_{k-1} \tilde{\mathbf{Q}}_{k-2} \cdots \tilde{\mathbf{Q}}_1 \mathbf{B}\mathbf{Y}_{k-1} = \mathbf{R}_{k-1}, \quad (2.85)$$

where $\tilde{\mathbf{Q}}_{k-1}, \tilde{\mathbf{Q}}_{k-2}, \dots, \tilde{\mathbf{Q}}_1 \in \mathbb{R}^{M \times M}$ is an orthogonal matrix and $\mathbf{R}_{k-1} \in \mathbb{R}^{M \times (k-1)}$ is upper-triangular. Note in the k -th iteration,

$$\mathbf{B}\mathbf{Y}_k = \mathbf{B} [\mathbf{Y}_{k-1} \ \mathbf{y}_k] = [\mathbf{B}\mathbf{Y}_{k-1} \ \mathbf{B}\mathbf{y}_k], \quad (2.86)$$

which is obtained by adding a new column $\mathbf{B}\mathbf{y}_k$ to $\mathbf{B}\mathbf{Y}_{k-1}$. Hence

$$\tilde{\mathbf{Q}}_{k-1} \tilde{\mathbf{Q}}_{k-2} \cdots \tilde{\mathbf{Q}}_1 \mathbf{B}\mathbf{Y}_k = [\mathbf{R}_{k-1} \ \mathbf{r}_k], \quad (2.87)$$

where $\mathbf{r}_k = \tilde{\mathbf{Q}}_{k-1} \tilde{\mathbf{Q}}_{k-2} \cdots \tilde{\mathbf{Q}}_1 \mathbf{B}\mathbf{y}_k$. Write $\mathbf{r}_k = \begin{bmatrix} \mathbf{r}_k^t \\ \mathbf{r}_k^b \end{bmatrix}$ with $\mathbf{r}_k^t \in \mathbb{R}^{k-1}$ and $\mathbf{r}_k^b \in$

\mathbb{R}^{M-k+1} , and let \mathbf{H}_k be the Householder reflector associated with \mathbf{r}_k^b , then $\tilde{\mathbf{Q}}_k := \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix}$ will make

$$\tilde{\mathbf{Q}}_k \tilde{\mathbf{Q}}_{k-1} \cdots \tilde{\mathbf{Q}}_1 \mathbf{B}\mathbf{Y}_k = \mathbf{R}_k, \quad (2.88)$$

which is upper-triangular. So far we can estimate the condition number of the k -step least square system (2.84) by estimating the condition number $\kappa(\mathbf{R}_k)$ (it suffices to

consider $\kappa(\tilde{\mathbf{R}}_k)$, where $\tilde{\mathbf{R}}_k := \mathbf{R}_k(1 : k, :)$ is the top square part of \mathbf{R}_k) and decide whether to continue the iteration or not. Given a threshold $\epsilon > 0$, the k -th iteration stops if $\kappa(\tilde{\mathbf{R}}_k) > \epsilon^{-1}$. Actually, $\kappa(\tilde{\mathbf{R}}_k)$ can also be computed iteratively, that is, we can update $\kappa(\tilde{\mathbf{R}}_k)$ by the information of $\tilde{\mathbf{R}}_{k-1}$. For example, one simple approach is to use 1 or ∞ -condition number $\kappa_*(\tilde{\mathbf{R}}_k)$ for $*$ = 1 or ∞ . Suppose we have evaluated $\tilde{\mathbf{R}}_{k-1}^{-1}$ by the $(k-1)$ -th iteration, and obtained $\tilde{\mathbf{R}}_k$ in the k -th iteration as following form

$$\tilde{\mathbf{R}}_k = \begin{bmatrix} \tilde{\mathbf{R}}_{k-1} & \mathbf{r}_k \\ 0 & \sigma_k \end{bmatrix}, \quad (2.89)$$

then $\tilde{\mathbf{R}}_k^{-1}$ can be evaluated by

$$\tilde{\mathbf{R}}_k^{-1} = \begin{bmatrix} \tilde{\mathbf{R}}_{k-1}^{-1} & -\sigma_k^{-1} \tilde{\mathbf{R}}_{k-1}^{-1} \mathbf{r}_k \\ 0 & \sigma_k^{-1} \end{bmatrix}, \quad (2.90)$$

which only costs $O(k^2)$ flops. Next $\kappa_*(\tilde{\mathbf{R}}_k) = \|\tilde{\mathbf{R}}_k\|_* \|\tilde{\mathbf{R}}_k^{-1}\|_*$ can be updated from the information of $\tilde{\mathbf{R}}_{k-1}$ and $\tilde{\mathbf{R}}_{k-1}^{-1}$ by $O(k)$ flops. Hence the total flops for computing $\kappa_*(\mathbf{R}_k)$ in all iterations will be no greater than $O(K^3)$ flops, where K is the total number of iterations.

After the QR factorization, (2.84) can be rewritten as

$$\mathbf{Q}_K \mathbf{R}_K \mathbf{z}_K \approx \mathbf{g}, \quad (2.91)$$

and then the least square solution \mathbf{z}_K is computed by applying back-substitution to

$$\tilde{\mathbf{R}}_K \mathbf{z}_K = (\mathbf{Q}_K^T \mathbf{g})(1 : K). \quad (2.92)$$

All in all, the whole algorithm for (2.75) can be depicted as follows.

1. find a particular solution \mathbf{y}_s by (2.79) and (2.80), and let $\mathbf{g} := \mathbf{h} - \mathbf{B}\mathbf{y}_s$;
2. define $\mathbf{R} = []$ which is an empty matrix in the beginning;
3. for $k = 1 : M$
4. find $\mathbf{y}_k \in \text{null}(\mathbf{A})$ which is orthonormal to $\mathbf{y}_1, \dots, \mathbf{y}_{k-1}$;

5. $\mathbf{r}_k = \mathbf{B}\mathbf{y}_k$;
6. $\mathbf{r}_k = \tilde{\mathbf{Q}}_{k-1} \cdots \tilde{\mathbf{Q}}_1 \mathbf{r}_k$; (if $k = 1$, skip this line)
7. define $\mathbf{r}_k^t = \mathbf{r}_k(1 : k - 1)$, $\mathbf{r}_k^b = \mathbf{r}_k(k : M)$;
8. $\mathbf{s}_k = -\text{sign}((\mathbf{r}_k^b)_1) \|\mathbf{r}_k^b\| \mathbf{e}_1$,
9. $\mathbf{v}_k = (\mathbf{s}_k - \mathbf{r}_k^b) / \|\mathbf{s}_k - \mathbf{r}_k^b\|$;
10. $\mathbf{R} = \begin{bmatrix} \mathbf{R} & \begin{bmatrix} \mathbf{r}_k^t \\ \mathbf{s}_k \end{bmatrix} \end{bmatrix}$;
11. if $\kappa(\mathbf{R}(1 : k, :)) > \epsilon^{-1}$, break;
12. end for
13. $\mathbf{g} = \tilde{\mathbf{Q}}_k \cdots \tilde{\mathbf{Q}}_1 \mathbf{g}$;
14. solve $\mathbf{R}(1 : k, :)\mathbf{z} = \mathbf{g}(1 : k)$ for \mathbf{z} by back-substitution;
15. $\mathbf{y} = [\mathbf{y}_1 \cdots \mathbf{y}_k] \mathbf{z} + \mathbf{y}_s$;

In Line 6, $\mathbf{r}_k = \tilde{\mathbf{Q}}_{k-1} \cdots \tilde{\mathbf{Q}}_1 \mathbf{r}_k$ can be computed implicitly by

1. for $i = 1 : k - 1$
2. $\mathbf{r}_k(i : M) = \mathbf{r}_k(i : M) - 2\mathbf{v}_i (\mathbf{v}_i^T \mathbf{r}_k(i : M))$;
3. end for

and in Line 13, \mathbf{g} can be computed by the same way.

2.5.3 Fast matrix-vector multiplication

The majority of computation in the algorithm is occupied by Line 5, namely, computing $\mathbf{B}\mathbf{y}_k$. Since \mathbf{B} is of size $O(N) \times O(N^2)$, a direct matrix-vector multiplication $\mathbf{B}\mathbf{y}$ costs $O(N^3)$ arithmetic operations. Fortunately, the specific data array

of \mathbf{B} allows a fast multiplication. Note the adjacent rows of \mathbf{B} are highly linearly dependent, and each row varies smoothly from previous ones. We first consider the boundary constraints (2.15), where \mathbf{B} has the following form

$$\mathbf{B} = \begin{bmatrix} \bar{L}_0(x_1)\bar{L}_0(y_1) & \bar{L}_0(x_1)\bar{L}_1(y_1) & \cdots & \bar{L}_N(x_1)\bar{L}_N(y_1) \\ \bar{L}_0(x_2)\bar{L}_0(y_2) & \bar{L}_0(x_2)\bar{L}_1(y_2) & \cdots & \bar{L}_N(x_2)\bar{L}_N(y_2) \\ \cdots & \cdots & \cdots & \cdots \\ \bar{L}_0(x_M)\bar{L}_0(y_M) & \bar{L}_0(x_M)\bar{L}_1(y_M) & \cdots & \bar{L}_N(x_M)\bar{L}_N(y_M) \end{bmatrix}, \quad (2.93)$$

where $(x_i, y_i) = \mathbf{z}_i$, $i = 1, \dots, M$ are the points spaced on $\partial\Omega$. Given

$$\mathbf{y} = [y_{00} \ y_{01} \ \cdots \ y_{N+1, N+1}]^T \in \mathbb{R}^{(N+1)^2}, \quad (2.94)$$

then

$$\mathbf{B}(i, :)\mathbf{y} = \sum_{j,k} \bar{L}_j(x_i)\bar{L}_k(y_i)y_{jk} \quad (2.95)$$

evaluates the expansion with base functions $\bar{L}_j\bar{L}_k$ and coefficients y_{jk} at point \mathbf{z}_i . Hence, the plot of $\mathbf{B}\mathbf{y}$ shows the profile of $\sum \bar{L}_j(x)\bar{L}_k(y)y_{jk}$ defined on $\partial\Omega$, which is usually (piecewise) smooth as long as $\partial\Omega$ is (piecewise) smooth.

Due to its smoothness, instead of evaluating the whole product $\mathbf{B}\mathbf{y}$, it suffices to choose several sampling nodes on $\partial\Omega$ (namely, several rows of \mathbf{B}) and do multiplication on them. After that, the value at non-sampling points on $\partial\Omega$ can be interpolated based on the data at sampling nodes. Fortunately, the complexity of evaluation at a point by usual interpolation techniques is much less than doing a direct vector multiplication. Therefore, when computing the product $\mathbf{B}\mathbf{y}$, we can only multiply a fixed number N_0 rows of \mathbf{B} by \mathbf{y} , and estimate other part of $\mathbf{B}\mathbf{y}$ by interpolation, for instance, the cubic spline interpolation which costs $O(N_0)$ for a solo entry and $O(N_0N)$ for all entries. By this method, the total complexity for computing $\mathbf{B}\mathbf{y}$ is $O(N_0N^2)$. In practical implementation, N_0 is determined by the accuracy requirement and is independent of N . We demonstrate it by the following example, in which $\partial\Omega$ is set by $r = 0.65 + 0.25 \sin(3\theta)$ and \mathbf{B} is multiplied by an all-one vector \mathbf{e} . In Fig. 5.2, the l^2 errors of computing $\mathbf{B}\mathbf{e}$ by our interpolation method versus N are

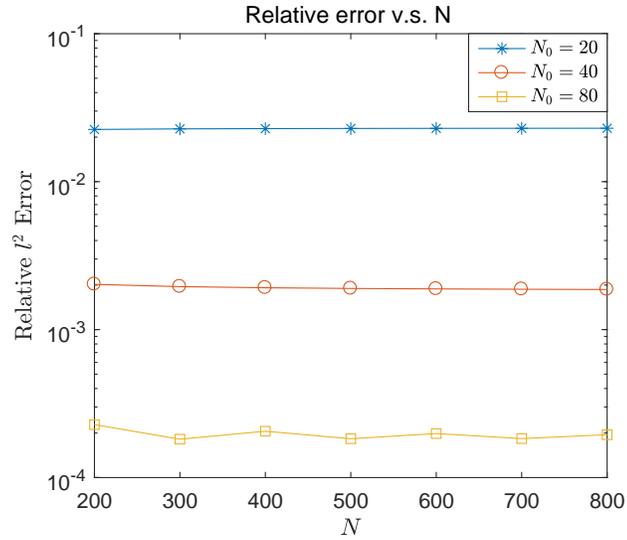


Fig. 2.2.: l^2 error for computing \mathbf{Be} by interpolation method for different N_0 and N

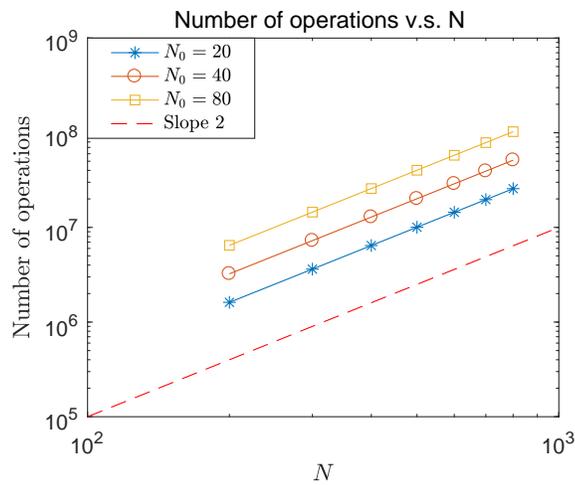


Fig. 2.3.: Number of arithmetic operations for different N

shown, and it is observed the errors only depend on the number of sampling nodes, rather than the size of \mathbf{B} . Furthermore, the numbers of operations for different N and N_0 are estimated and presented in Fig. 5.3, from which we see the complexity for the matrix-vector multiplication on \mathbf{B} is indeed about $O(N^2)$.

For boundary constraints (2.16), we suppose the number of test functions $\{\chi_i\}$ and the quadrature nodes are set by $O(N)$, then in this case \mathbf{B} is formed as the product of a $O(N) \times O(N)$ matrix related to $\{\chi_i\}$ and another $O(N) \times O(N^2)$ matrix of the form in (2.93). Hence $\mathbf{B}\mathbf{y}$ is computed by first applying the preceding fast multiplication technique, and then doing a usual $O(N)$ by $O(N)$ matrix-vector multiplication. Thus the total number of operations is also $O(N^2)$.

Now we can determine the complexity of the preceding algorithm. First note Line 4 can be precomputed since it does not depend on the domain and the data. Due to the orthogonality, $\mathbf{A} \in \mathbb{R}^{O(N^2) \times O(N^2)}$ is sparsely structured with $O(1)$ nonzero entries in each row. So sparse solvers can be applied to compute an orthonormal set of $\text{null}(\mathbf{A})$ in advance. Then for other lines relate to computation, note Line 1 costs $O(sN^2)$ if an iterative solver is used for (2.80) and executes s times of iteration. Inside the for-loop, Line 5 costs $O(N^2)$ by fast computation and Line 6,8,9,11 and 13 each costs no more than $O(N^2)$, hence each loop costs $O(N^2)$, and the whole for-loop costs at most $O(N^3)$ due to $M = O(N)$. Finally the cost of Line 14 and 15 is within $O(N^3)$. Therefore, the total process costs $O(N^3) + O(sN^2)$ operations.

2.6 Numerical examples

The accuracy of the preceding two methods will be demonstrated by five numerical examples. In the first and second ones, we solve smooth Poisson-type problems in a smooth domain and a convex polygon. In the third and fourth ones, singular problems in a square and an L-shaped domain are studied. And in the last example, the elliptic problem with non-constant α is considered. For all examples, (2.15) is chosen as the approximate boundary condition.

In the first example, the first method is applied to the following Poisson-type equation

$$\begin{aligned} -\Delta u + \alpha u &= f \quad \text{in } \Omega, \\ u &= h \quad \text{on } \partial\Omega, \end{aligned} \tag{2.96}$$

where $\partial\Omega$ is characterized by the polar expression

$$r = r_0 + \delta \sin(n\theta). \quad (2.97)$$

And the exact solution is set by

$$u = r^3(r_0 + \delta \sin(n\theta) - r), \quad (2.98)$$

with homogeneous Dirichlet boundary condition, where $r_0 = 0.65$, $\delta = 0.25$, $n = 3$. Recall N is the degree of tensorial polynomial space specified in (2.17)-(2.18), namely, the degree of freedom. First, we let $\alpha = 0$, and for $N = 51$ ($M = 104$) we run out the whole for-loop in the algorithm presented in Section 2.5 without break and compute $\kappa_1(\mathbf{R}_k)$ in each loop. The growth tendency of $\kappa_1(\mathbf{R}_k)$ is shown in Fig. 2.4, from which we can see $\kappa_1(\mathbf{R}_k)$ increases rapidly from the beginning, and reaches an acceptable level of 10^6 when k is around $\frac{1}{3}M$, therefore in this example we choose $K = \lfloor \frac{1}{3}M \rfloor = \lfloor \frac{2N+2}{3} \rfloor$ as a prescribed number of iterations for the for-loop, that means the solution to the least square problem (2.84) is searched in a K -dimensional subspace of $\text{null}(\mathbf{A})$. The original domain Ω , extended domain $\tilde{\Omega}$ and the sampling nodes $\{\mathbf{z}_i\}$ defined in (2.15) for $N = 51$ are shown in Fig. 2.5. Moreover, the numerical solution u_N and the error mesh $u - u_N$ are presented in Fig. 2.6, from which we see the error dominates near the boundary. This is consistent with the error estimate (2.45). Finally, (2.96) is solved for $N = 35, 43, \dots, 91$, and the decay of L^2 error $\|u - u_N\|_{L^2(\Omega)}$ is shown in Fig. 2.7, where the exponential convergence rate is observed.

For testing the robustness of the first method on the cases $\alpha \neq 0$, we repeat computing the same problem with different positive α under the same parameter setting. The L^2 error is shown in Fig. ???. From the result, we can see the error can only reach the level of 10^{-1} to 10^{-2} , as long as $\alpha \neq 0$.

The second example is to solve the problem (2.96) by the second method, where Ω is a pentagon with vertices $(0, 0.9)$, $(-0.9, 0.2)$, $(-0.7, -0.8)$, $(0.7, -0.8)$, and $(0.9, 0.2)$. The exact solution is chosen as

$$u = \exp\left(-\frac{x^2 + y^2}{2}\right). \quad (2.99)$$

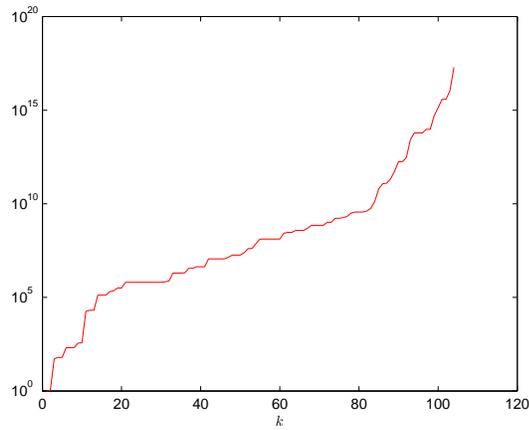


Fig. 2.4.: $\kappa_1(\mathbf{R}_k)$ for the first example ($N=51$)

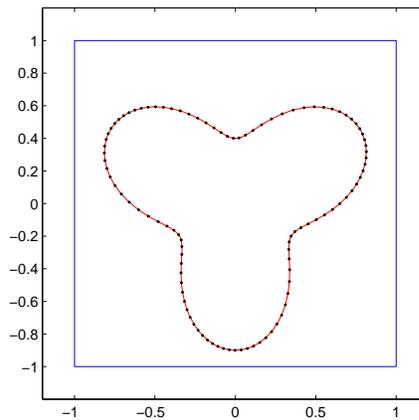


Fig. 2.5.: the problem domain, extended domain and the boundary nodes for the first example ($N=51$)

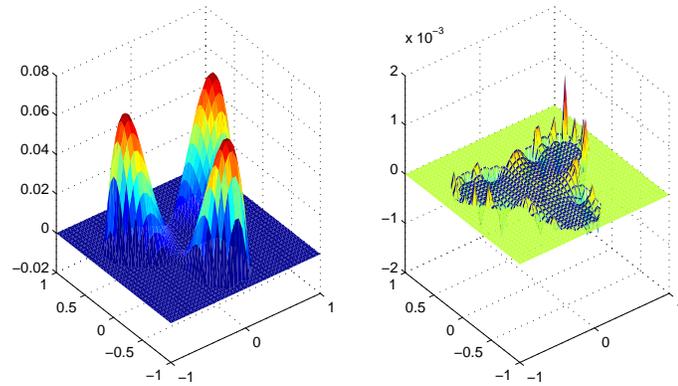


Fig. 2.6.: *the numerical solution (left) and the error (right) for the first example ($N=51$)*

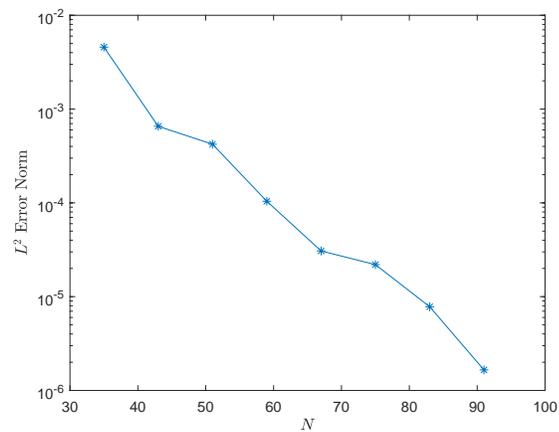


Fig. 2.7.: $\|u - u_N\|_{L^2(\Omega)}$ versus N for the first example

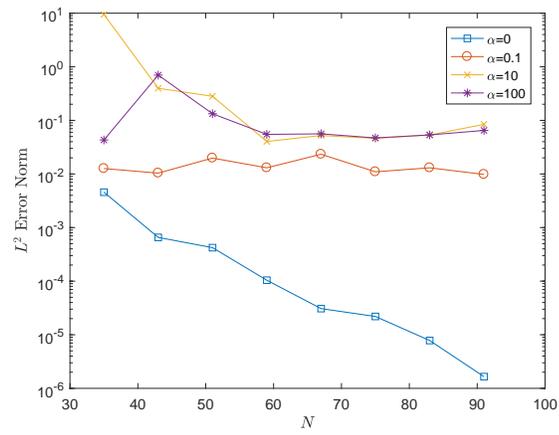


Fig. 2.8.: $\|u - u_N\|_{L^2(\Omega)}$ versus N for the first example of different α

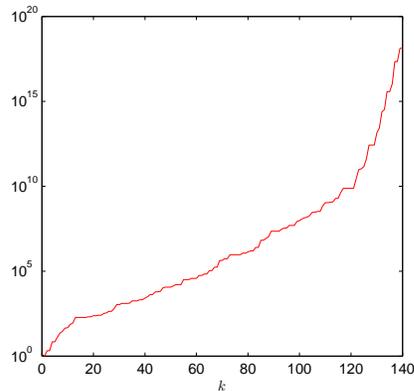


Fig. 2.9.: $\kappa_1(\mathbf{R}_k)$ for the second example ($N=35$)

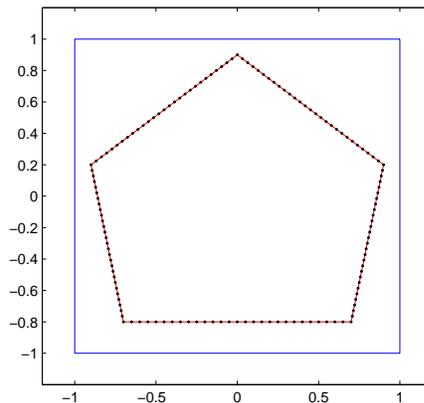


Fig. 2.10.: the problem domain, extended domain and the boundary nodes for the second example ($N=35$)

First we consider $\alpha = 10$. For $N = 35$, $\kappa_1(\mathbf{R}_k)$ for different k is shown in Fig. 2.9, together with the original domain Ω , extended domain $\tilde{\Omega}$ and the sampling nodes $\{\mathbf{z}_i\}$ shown in Fig. 2.10. In this example, $\epsilon^{-1} = 10^{1+N/10}$ is chosen as the stopping criterion for Line 11 in the algorithm. The numerical solution and the error mesh for $N = 35$ are depicted in Fig. 2.11. Finally the L^2 error for $N = 15, 25, \dots, 95$ is presented in Fig. 2.12. The exponential error decay is also observed for this method.

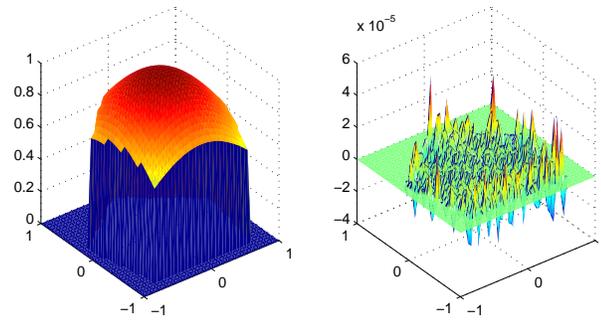


Fig. 2.11.: *the numerical solution (left) and the error (right) for the second example ($N=35$)*

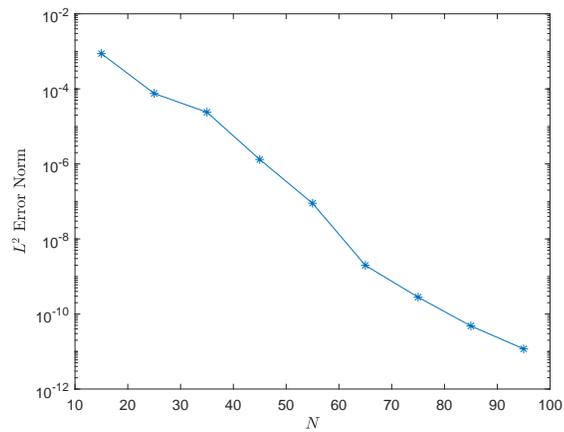


Fig. 2.12.: $\|u - u_N\|_{L^2(\Omega)}$ versus N for the second example

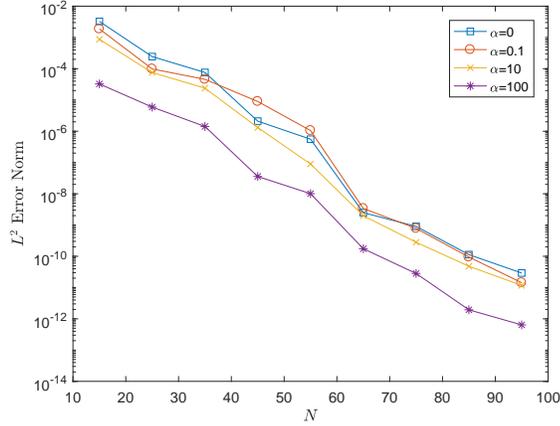


Fig. 2.13.: $\|u - u_N\|_{L^2(\Omega)}$ versus N for the second example of different α

The effectiveness of the second method on the cases of smaller α is also tested. We repeat computing the same problem with different nonnegative α under the same parameter setting. The L^2 error is shown in Fig. 2.13, in which the exponential error decay is observed for all α . Hence the second method indeed works for smooth Poisson-type problem independent of α .

In the following, the problems with singularities will be tested by our methods. In the third example, the first method is applied to the Poisson problem

$$\begin{aligned} -\Delta u &= 1 \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{2.100}$$

where Ω is a square with vertices (T, T) , $(T, -T)$, $(-T, -T)$, $(-T, T)$ (here T is chosen by 0.8). The exact solution is given by

$$u(x, y) = -\frac{64T^2}{\pi^4} \sum_{\substack{n,m=1 \\ n,m \text{ odd}}}^{\infty} (-1)^{\frac{n+m}{2}} \frac{\cos(\frac{n\pi x}{2T}) \cos(\frac{m\pi y}{2T})}{nm(n^2 + m^2)}, \tag{2.101}$$

which has singularities at the four corners. For such problems the exponential error decay is hardly obtained by usual numerical approaches (see [37, 38]). Now the problem is solved by the first method. For $N = 55$, the original domain Ω , extended domain $\tilde{\Omega}$ and the boundary nodes are shown in Fig. 2.14, together with the error

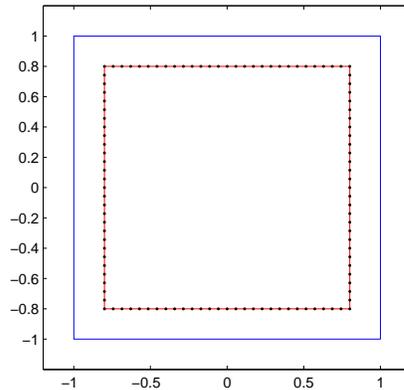


Fig. 2.14.: *the problem domain, extended domain and the boundary nodes for the third example ($N=55$)*

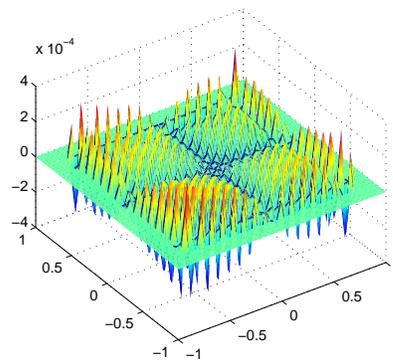


Fig. 2.15.: *the error for the third example ($N=55$)*

meshed in Fig. 2.15, from which we can see the error is mainly distributed near the corners. Then for $N = 35, 45, \dots, 105$, we compute the L^2 errors and present them in Fig. 2.16, in which the convergence rate is observed between 4th order to 5th order.

For the fourth example, we consider (2.100) in an L-shaped domain, namely, Ω is an L-shaped polygon with vertices $(0, 0)$, $(T, 0)$, $(T, -T)$, $(-T, -T)$, $(-T, T)$ and $(0, T)$ (here T is chosen by 0.8). The solution of the PDE has singularity at the origin. In [39], a finite difference scheme with domain decomposition is proposed, but the convergent rate is at most 1st order. We apply the second method with stopping

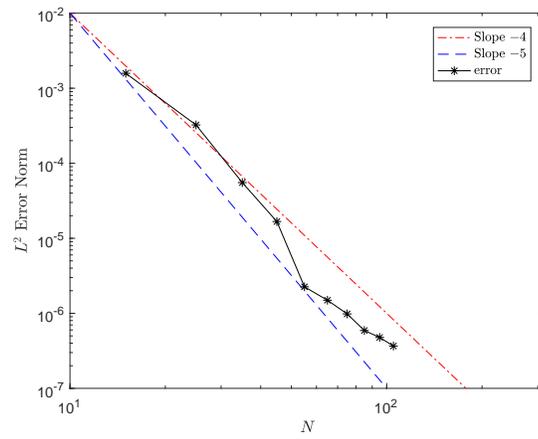


Fig. 2.16.: $\|u - u_N\|_{L^2(\Omega)}$ versus N for the third example

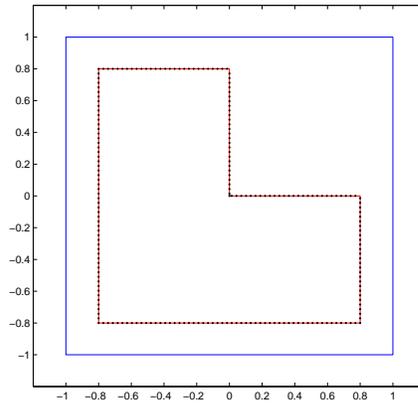


Fig. 2.17.: *the problem domain, extended domain and the boundary nodes for the fourth example ($N=55$)*

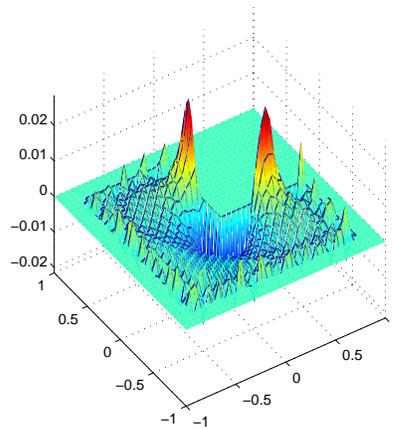


Fig. 2.18.: *the error for the fourth example ($N=55$)*

criteria $\epsilon^{-1} = 10^{N/20}$ to this problem. The original domain Ω , extended domain $\tilde{\Omega}$ and the boundary nodes $\{\mathbf{z}_i\}$ for $N = 55$ is shown in Fig. 2.17. A high-resolution solution of $N = 255$ is computed as the exact solution, and the error mesh for $N = 55$ is shown in Fig. 2.18, from which we see the error accumulates in the vicinity of the singular corner $(0,0)$. Finally the L^2 error for $N = 25, 35, \dots, 135$ is presented in Fig. 2.19, where the convergence rate is observed to be between 2nd and 3rd order.

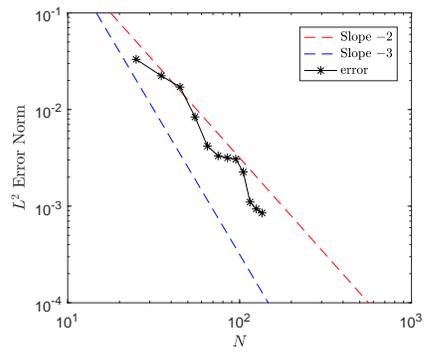


Fig. 2.19.: $\|u - u_N\|_{L^2(\Omega)}$ versus N for the fourth example

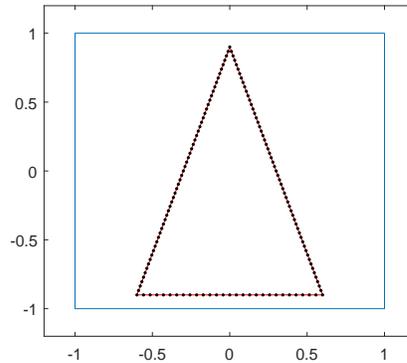


Fig. 2.20.: *the problem domain, extended domain and the boundary nodes for the fifth example ($N=35$)*

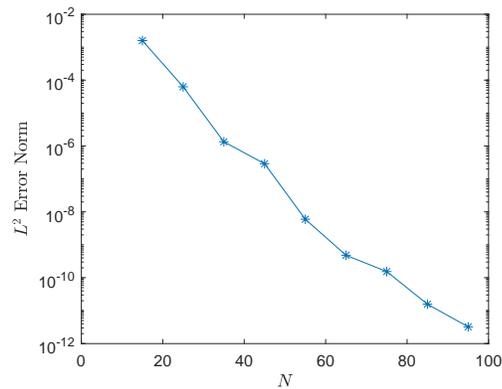


Fig. 2.21.: L^2 error $\|u - u_N\|_{L^2(\Omega)}$ versus N for the fifth example

In the fifth example, we apply the second method on the general elliptic problem (2.54), where $\alpha(x, y) = (\sin x + 1)(\cos x + 1)$ and $\beta(x, y) = \exp(x + y)$ are non-constant coefficients. The exact solution is also chosen by (2.99), and Ω is set as a triangle with vertices $(0, 0.9), (0.6, -0.9), (-0.6, -0.9)$. The problem domains for $N = 35$ is shown in Fig. 6.9, and the L^2 error for $N = 15, 25, \dots, 95$ can be seen in Fig. 6.10. Due to the smoothness of the coefficient and solution, the exponential decay of error is still observed.

2.7 Conclusion

In this work, we have developed spectral methods for elliptic PDEs defined in complex domains. These methods are performed by embedding the original domain into a larger and regular one, and continuously extending the original equation. Specific trial and test functions are chosen to build up weak formulations. The corresponding spectral Petrov-Galerkin formulations are then presented by approximating the trial and test spaces by finite dimensional tensorial polynomial spaces, in which the Legendre polynomials are employed to construct the basis functions.

Actually, we put forward two types of Petrov-Galerkin formulations for the extended problem. One is specifically designed for the Poisson-type equation (2.96) with $\alpha = 0$, by choosing H^2 -type trial space and L^2 -type test space, which are separately approximated by $P_N \times P_N$ and $\Delta(P_N \times P_N)$. The well-posedness and error estimate for the first method are discussed completely in Section 2.3. However, the application of the first method to the cases $\alpha \neq 0$ is not as good as the case $\alpha = 0$, where the error only decays to the level between 10^{-1} to 10^{-2} in the numerical results. The other one is using H^1 -type trial space and H_0^1 -type test space, which are separately approximated by $P_N \times P_N$ and $P_N^0 \times P_N^0$. We observe that the second method can deal with the equation (2.96) with all $\alpha \geq 0$ elegantly, but the well-posedness is subtle and depends on the geometry of the original and enclosing domains, as well as the value of α . And we have no theoretical error estimate for the second method.

Furthermore, we solve the derived ill-conditioned linear system by a particular algorithm. The process divides the system to two parts (one is from the differential equation and the other is from the boundary condition) and deals with them under different mechanisms, instead of solving the linear system directly. The total complexity for both methods are $O(N^3)$.

In numerical examples, our methods work for problems in smooth domains as well as polygons, and the exponential convergence rate has been obtained for smooth solutions. Moreover, for the singular Poisson problem in a square domain, the conver-

gence rate is observed to be between 4th and 5th, and for L-shaped domain problem with singularity at the concave corner, the error decays within between 2nd and 3rd order. Finally, an example of elliptic equation with non-constant smooth coefficient is presented. In Section 2.3.2, we have estimated that the error is mostly distributed near the boundary, which is also observed in these examples. Some future work may include applying the method to PDEs of other types, developing trial and test functions of better consistence and approximate property or devising boundary constraints in trial spaces more carefully to reduce the boundary error.

3. SPECTRAL METHOD WITH CIRCULAR EMBEDDING

3.1 Introduction

Numerical methods for solving partial differential equations (PDEs) given in irregular domains have been studied in the past decade. One group of the methods is based on fictitious domain concept, by which the original problem domain is embedded in a larger and regular one, and the equation is also extended in the new fictitious domain. Such methods include adding penalty to the extended equations [9–11], introducing Lagrange multipliers on the boundary [8, 18, 40], diffuse domain method [12–14], boundary integral method [19–21] etc. Furthermore, some techniques combine fictitious domain concept and the general finite (or spectral) element methods, which are referred to as the finite (or spectral) cell method. The main drawbacks of these methods are the low order of convergence rate for singular solutions and the high computational cost for highly irregular domains [40–43].

Our preceding work is to apply general spectral solvers after domain embedding to solving two-dimensional second-order elliptic PDEs. The original complex domain is enclosed by a larger rectangular domain, and the problem is redefined in the new domain without removing the original boundary condition. We have built a specific Petrov-Galerkin formulation for the extended problem, and take advantage of orthogonal polynomial-type basis functions to approximate the weak solution. The complexity of the whole process is within $O(N^3)$, including solving a $O(N^2) \times O(N^2)$ sparse linear system, where N is the degree of freedom for each variable.

However, the flexibility of the domain embedding process allows us to insert the complex domain into a larger and circular one. Afterwards, the extended problem can be set up in the outer circle. Thanks to the polar transformation in circular

domains $x = r \cos(\theta)$, $y = r \sin(\theta)$, the 2-D problem can be separated into a sequence of ordinary differential equations (ODEs), which are actually Bessel-type equations. The boundary values of such ODEs are implicitly determined by the original condition on the boundary of the 2-D complex domain, that is, one needs to look for appropriate boundary values for these ODEs such that the 2-D boundary condition is satisfied in some sense. During this approach, one only needs to solve a sequence of $O(N) \times O(N)$ sparse linear systems and a $O(N) \times O(N)$ full one. Compared to the preceding rectangular embedding, the complexity is not increased and the implementation is simplified to a great extent.

We study the circular embedding approach in this work. In Section 3.2, we describe the fictitious domain setting, and then build the extended problem and the corresponding Galerkin formulation for the 1-D Bessel-type equations. In Section 3.3, we discuss the spectral solver for Bessel ODEs. In Section 3.4, we estimate the solution error for the whole algorithm for a specific problem domain. In Section 3.5, we present several numerical examples to show the convergence rate for problems of different type. The whole work is concluded in Section 3.6.

3.2 Problems in simply connected domains

3.2.1 Dimension reduction

Let us consider the following Poisson-type model problem

$$\begin{aligned} \alpha U - \Delta U &= F, & \text{in } \Omega, \\ U &= H, & \text{on } \partial\Omega. \end{aligned} \tag{3.1}$$

where $\alpha \geq 0$ and Ω is a two-dimensional simply connected domain of complex geometry, and $F \in C(\overline{\Omega})$, $H \in C(\partial\Omega)$. Now let $\tilde{\Omega}$ be a circular domain which encloses Ω ,

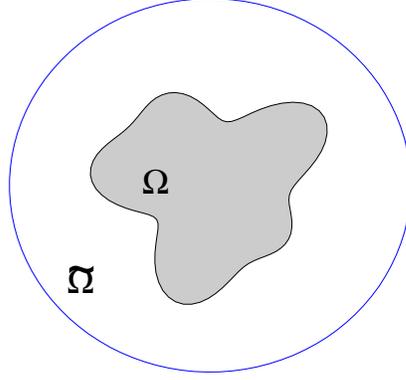


Fig. 3.1.: The original domain Ω and the enclosing circle $\tilde{\Omega}$

namely, $\Omega \subset\subset \tilde{\Omega}$ (see Fig. 3.1). We are then to solve the following extended problem in $\tilde{\Omega}$,

$$\begin{aligned} \alpha U - \Delta U &= F, & \text{in } \tilde{\Omega}, \\ U &= H, & \text{on } \partial\Omega. \end{aligned} \quad (3.2)$$

Without loss of generality, in the following text we assume $\tilde{\Omega} = \{(x, y) : x^2 + y^2 < 1\}$ which is the interior of the unit circle. Let $\Pi := (0, 1) \times [0, 2\pi)$, by applying the polar transformation

$$\mathbb{T} : \bar{\Pi} \rightarrow \tilde{\Omega}, \quad \mathbb{T}(r, \theta) = (r \cos \theta, r \sin \theta), \quad (3.3)$$

to (3.2) and denoting

$$u(r, \theta) := U(r \cos \theta, r \sin \theta), \quad (3.4)$$

$$f(r, \theta) := F(r \cos \theta, r \sin \theta), \quad (3.5)$$

$$h(r, \theta) := H(r \cos \theta, r \sin \theta), \quad (3.6)$$

we have

$$\begin{aligned}
\alpha u - \frac{1}{r}(ru_r)_r - \frac{1}{r^2}u_{\theta\theta} &= f, \quad (r, \theta) \in \Pi, \\
u_\theta(0, \theta) &= 0, \\
u &= h, \quad \text{on } \Gamma, \\
u &\text{ is periodic in } \theta,
\end{aligned} \tag{3.7}$$

where $\Gamma = \mathbb{T}^{-1}(\partial\Omega)$ is the preimage of $\partial\Omega$. Now the two-dimensional problem (3.7) can be reduced to a sequence of one-dimensional problems by Fourier transformation in θ direction. Actually, by expanding u and f as

$$u(r, \theta) = \sum_{|m|=0}^{\infty} u^m(r) e^{im\theta}, \tag{3.8}$$

$$f(r, \theta) = \sum_{|m|=0}^{\infty} f^m(r) e^{im\theta}, \tag{3.9}$$

and substituting (3.8) and (3.9) into (3.7) leads to a sequence of Bessel-type ordinary differential equations for $u^m(r)$, namely,

$$-\frac{1}{r}\partial_r(r\partial_r u^m) + \left(\frac{m^2}{r^2} + \alpha\right)u^m = f^m(r), \quad 0 < r < 1, \tag{3.10}$$

with one-sided pole conditions

$$u^m(0) = 0 \quad \text{if } m \neq 0. \tag{3.11}$$

Note the system (3.10) and (3.11) are underdetermined, unless we add an artificial boundary condition for each ODE such as

$$u^m(1) = t^m. \tag{3.12}$$

3.2.2 Approximation

The numerical method is performed by prescribing a cut-off number $M > 0$ and let the approximate solution to (3.7) be

$$u_M(r, \theta) = \sum_{|m|=0}^M u^m(r) e^{im\theta}. \tag{3.13}$$

In order to determine u^m , we need to determine the boundary values $\{t^m\}_{|m|=0}^M$ that concerns to the boundary condition $u = h$ on Γ in the original problem (3.7). Specifically, we look for desired values for $\{t^m\}_{|m|=0}^M$ such that $u_M \approx h$ on Γ . For this purpose, we first explore the relation between u^m and t^m . Note for the general two-point boundary value problems, the following property holds.

Theorem 3.2.1 *Suppose the ordinary differential equation*

$$Lu(x) = f(x), \quad x \in (a, b), \quad (3.14)$$

with boundary condition

$$u(a) = 0, \quad u(b) = t, \quad (3.15)$$

is well-defined and admits a unique solution $u(x; t)$. Let $\phi(x)$ be the solution to

$$\begin{cases} L\phi(x) = 0, & x \in (a, b), \\ \phi(a) = 0, & \phi(b) = 1, \end{cases} \quad (3.16)$$

and $\psi(x)$ be the solution to

$$\begin{cases} L\psi(x) = f(x), & x \in (a, b), \\ \psi(a) = 0, & \psi(b) = 0, \end{cases} \quad (3.17)$$

then $u(x; t) = t\phi(x) + \psi(x)$.

Remark 3.2.1 *If the two-point boundary condition (3.15) is replaced by one-sided condition $u(b) = t$ in the hypothesis, the result still holds for the ϕ and ψ satisfying the corresponding one-sided boundary problems.*

Now by virtue of Theorem 3.2.1, let $\{\phi^m\}$ be the solution to

$$\begin{aligned} -\frac{1}{r}\partial_r(r\partial_r\phi^m) + \left(\frac{m^2}{r^2} + \alpha\right)\phi^m &= 0, \quad 0 < r < 1, \\ \phi^m(0) &= 0 \quad \text{if } m \neq 0, \quad \phi^m(1) = 1, \end{aligned} \quad (3.18)$$

and $\{\psi^m\}$ be the solution to

$$\begin{aligned} -\frac{1}{r}\partial_r(r\partial_r\psi^m) + \left(\frac{m^2}{r^2} + \alpha\right)\psi^m &= f^m, \quad 0 < r < 1, \\ \psi^m(0) &= 0 \quad \text{if } m \neq 0, \quad \psi^m(1) = 0, \end{aligned} \quad (3.19)$$

then we have

$$u^m(r; t^m) = t^m \phi^m(r) + \psi^m(r). \quad (3.20)$$

To determine $\{t^m\}_{|m|=0}^M$, a straightforward idea is to prescribe K equispaced sampling nodes $\{(\hat{r}_k, \hat{\theta}_k)\}_{k=1}^K$ on Γ , and minimize the discrete L^2 error on these nodes in terms of variables $\{t^m\}_{|m|=0}^M$, namely, to solve

$$\min_{t^m} \sum_{k=1}^K |u_M(\hat{r}_k, \hat{\theta}_k) - h(\hat{r}_k, \hat{\theta}_k)|^2, \quad (3.21)$$

or

$$\min_{t^m} \sum_{k=1}^K \left| \sum_{|m|=0}^M (t^m \phi^m(\hat{r}_k) + \psi^m(\hat{r}_k)) e^{im\hat{\theta}_k} - h(\hat{r}_k, \hat{\theta}_k) \right|^2, \quad (3.22)$$

Note (3.22) is a least square problem that can be solved by general linear solvers. However, adjacent rows of the coefficient matrix are highly parallel due to the closeness of the adjacent sampling nodes, which causes the ill-conditioning when K becomes larger. Another approach is to force u_M to be the projector of h onto a finite subspace of $L^2(\Gamma)$. Specifically, suppose Γ is parametrized by

$$\Gamma = \{(\rho(\theta), \theta) : 0 \leq \theta < 2\pi\}, \quad (3.23)$$

and let $\{\xi_k\}_{|k|=0}^M$ be a set of periodic functions defined in $[0, 2\pi)$, then it requires

$$\int_0^{2\pi} u_M(\rho(\theta), \theta) \xi_k(\theta) d\theta = \int_0^{2\pi} h(\rho(\theta), \theta) \xi_k(\theta) d\theta, \quad (3.24)$$

or

$$\sum_{|m|=0}^M \int_0^{2\pi} (t^m \phi^m(\rho(\theta)) + \psi^m(\rho(\theta))) e^{im\theta} \xi_k(\theta) d\theta = \int_0^{2\pi} h(\rho(\theta), \theta) \xi_k(\theta) d\theta, \quad (3.25)$$

for $|k| = 0, \dots, M$. An appropriate choice of $\{\xi_k\}_{|k|=0}^M$ can remarkably reduce the condition number of the derived linear system. For example, a decent choice is the scaled Fourier basis

$$\xi_k(\theta) := \rho_{\text{avg}}^{-|k|} e^{-ik\theta} \quad (3.26)$$

where $\rho_{\text{avg}} \in (0, 1)$ represents the average radius of the domain boundary. A straightforward formula for ρ_{avg} is

$$\rho_{\text{avg}} = \frac{1}{2\pi} \int_0^{2\pi} \rho(\theta) d\theta. \quad (3.27)$$

The exponentially growing factor $\rho_{\text{avg}}^{-|k|}$ added here plays the role of a preconditioner, which will be discussed in more detail in Section 3.3.1. This choice for $\{\xi_k\}$ can bring desirable simpleness both for numerical implementation and error analysis.

3.2.3 Algorithm

To sum up, we can derive the following algorithm from the preceding discussion

Algorithm 3.2.1 *Solve the Poisson-type problem (3.1)*

1. *perform the domain embedding $\Omega \subset \subset \tilde{\Omega}$ and extend f from Ω to $\tilde{\Omega}$ smoothly;*
2. *compute the truncated Fourier expansion of $f(r, \theta) := F(r \cos \theta, \sin \theta)$ with respect to θ , obtaining (3.9);*
3. *compute $\{\phi^m\}$ and $\{\psi^m\}$ which satisfy the Bessel-type equations (3.18) and (3.19);*
4. *solve the optimization (3.22) or the linear system (3.25), to obtain $\{t^m\}$;*
5. *compute u_M by (3.13) and (3.20);*
6. *the final solution $U_M(x, y) = u_M(\mathbb{T}^{-1}(x, y))$.*

let us estimate the computational cost for this algorithm. Suppose the Bessel-type problem (3.18) is solved by the representation (3.28), and (3.19) is solved by the spectral-Galerkin formulation with degree of freedom N (see Section 3.3.1), then for each m , ϕ^m and ψ^m are computed with $O(N^2)$ flops. Hence Step 2 costs totally $O(MN^2)$ flops. For Step 1, Fast Fourier Transformation(FFT) is performed on $f(r, \theta)$ at $O(N)$ collocation points in r direction, therefore total $O(NM \log(M))$ flops are needed. For Step 3, forming the matrix in (3.22) costs $O(KMN)$ flops and solving

the least square problem costs $O(KM^2)$ flops. Similarly, if we suppose the integral in (3.25) are evaluated by numerical quadrature with $O(K)$ nodes and weights, then the total flops for dealing with (3.25) is $O(KMN + KM^2 + M^3)$. All in all, by the fact $K \geq M$, the complexity for Algorithm 3.2.1 is $O(MN^2 + KMN + KM^2)$. If we set $K = O(M)$ and $N = O(M)$, then the total complexity is $O(M^3)$.

Note that the crucial step in Algorithm 3.2.1 is to compute $\{\phi^m\}$ and $\{\psi^m\}$, which will be discussed in the following section.

3.3 Bessel-type equation solver

In this section, we will present the solver for the 1-D Bessel-type equation (3.18) and (3.19). Actually, for the homogeneous one (3.18) an explicit representation of the solution can be provided, while for the inhomogeneous one (3.19), classical ODE solvers for boundary value problems can be applied. In this work, a spectral-Galerkin formulation is put forward and implemented on it, and the corresponding error analysis is also presented.

3.3.1 Homogeneous case

It can be verified the solutions to (3.18) have the following expression,

$$\phi^m(r) = \begin{cases} r^{|m|}, & \alpha = 0, \\ \frac{I_{|m|}(\sqrt{\alpha}r)}{I_{|m|}(\sqrt{\alpha})}, & \alpha > 0 \end{cases} \quad (3.28)$$

for all m , where $I_{|m|}(z)$ are the modified Bessel functions of the first kind. In practical computation, $\{\phi^m\}$ can be evaluated by using series representation, or by solving the Bessel-type equations numerically.

Now let us consider the linear system (3.25) with test functions defined by (3.26), which is a finite approximation to the original boundary condition $U = H$ on $\partial\Omega$ in (3.1). Rewriting (3.25) by using (3.20) gives

$$\sum_{|m|=0}^M \left(\rho_{\text{avg}}^{-|k|} \int_0^{2\pi} \phi^m(\rho(\theta)) e^{i(m-k)\theta} d\theta \right) t^m = \rho_{\text{avg}}^{-|k|} \left(\int_0^{2\pi} h(\rho(\theta), \theta) e^{-ik\theta} d\theta - \sum_{|m|=0}^M \int_0^{2\pi} \psi^m(\rho(\theta)) e^{i(m-k)\theta} d\theta \right). \quad (3.29)$$

For the trivial case $\alpha = 0$ and $\rho(\theta) = \rho_0 \in (0, 1)$, the coefficient matrix in (3.29) is exactly the identity thanks to the exponentially growing coefficient $\rho_{\text{avg}}^{-|k|}$. For general domains, our numerical examples show that the condition number of the linear systems may grow exponentially as M increases however with a base number very close to 1.

3.3.2 Inhomogeneous case

Notations

Denote $I = (-1, 1)$. Now define the Jacobi weight function $\omega^{\alpha, \beta}(x)$ by

$$\omega^{\alpha, \beta}(x) = (1-x)^\alpha (1+x)^\beta, \quad (3.30)$$

and the weighted L^2 inner product is given by

$$(u, v)_{\omega^{\alpha, \beta}} = \int_I u(x) \bar{v}(x) \omega^{\alpha, \beta}(x) dx, \quad (3.31)$$

with which the weighted L^2 space is defined by

$$L_{\omega^{\alpha, \beta}}^2(I) = \{u(x) : (u, u)_{\omega^{\alpha, \beta}} < \infty\}, \quad (3.32)$$

associated with norm

$$\|u\|_{\omega^{\alpha, \beta}; I} = (u, u)_{\omega^{\alpha, \beta}}^{\frac{1}{2}}. \quad (3.33)$$

Furthermore, the weighted H^1 space is defined by

$$H_{\omega^{\alpha, \beta}}^1(I) = \{u(x) : u, \partial_x u \in L_{\omega^{\alpha, \beta}}^2(I)\} \quad (3.34)$$

equipped with norm

$$\|u\|_{1,\omega^{\alpha,\beta};I} = \left(\|u\|_{\omega^{\alpha,\beta};I}^2 + \|\partial_x u\|_{\omega^{\alpha,\beta};I}^2 \right)^{\frac{1}{2}}, \quad (3.35)$$

Spectral-Galerkin formulation

Now let us consider the spectral methods for (3.19). Note $\{\psi^m\}$ satisfy the model problems

$$\begin{aligned} -\frac{1}{r}(ru_r)_r + \left(\frac{m^2}{r^2} + \alpha\right)u &= f, \quad 0 < r < 1, \\ u(0) = 0 \quad \text{if } m \neq 0, \quad u(1) &= 0, \end{aligned} \quad (3.36)$$

By introducing the transformation

$$r = \frac{1+t}{2}, \quad (3.37)$$

and denoting

$$v(t) := u\left(\frac{1+t}{2}\right), \quad (3.38)$$

(3.36) will be changed to

$$\begin{aligned} -\frac{1}{1+t}((1+t)v_t)_t + \left(\frac{m^2}{(1+t)^2} + \frac{\alpha}{4}\right)v &= \frac{1}{4}f\left(\frac{1+t}{2}\right), \quad t \in I, \\ v(-1) = 0 \quad \text{if } m \neq 0, \quad v(1) &= 0, \end{aligned} \quad (3.39)$$

Define the solution space by

$$X^m = \begin{cases} \left\{ v : v(1) = 0, \int_I (1+t)|v|^2 + (1+t)|v_t|^2 dt < \infty \right\}, & m = 0, \\ \left\{ v : v(\pm 1) = 0, \int_I \frac{1}{1+t}|v|^2 + (1+t)|v_t|^2 dt < \infty \right\}, & m \neq 0, \end{cases} \quad (3.40)$$

equipped with norms

$$\|v\|_{X^m} = \begin{cases} \left(\int_I (1+t)|v|^2 + (1+t)|v_t|^2 dt \right)^{\frac{1}{2}}, & m = 0, \\ \left(\int_I \frac{1}{1+t}|v|^2 + (1+t)|v_t|^2 dt \right)^{\frac{1}{2}}, & m \neq 0. \end{cases} \quad (3.41)$$

Also, define

$$a_m(v, w) = (v_t, w_t)_{\omega^{0,1}} + m^2 (v, w)_{\omega^{0,-1}} + \frac{\alpha}{4} (v, w)_{\omega^{0,1}}, \quad (3.42)$$

then the weak formulation for (3.39) is

$$\begin{cases} \text{To find } v \in X^m \text{ such that} \\ a_m(v, w) = \frac{1}{4}(f(\frac{1+t}{2}), w)_{\omega^{0,1}}, \quad \forall w \in X^m. \end{cases} \quad (3.43)$$

To show the well-posedness of the weak problem (3.43), first note the following continuity holds

$$|a_m(v, w)| \leq \begin{cases} \max(1, \alpha/4) \|v\|_{X^m} \|w\|_{X^m}, & m = 0, \\ (m^2 + \alpha/2) \|v\|_{X^m} \|w\|_{X^m}, & m > 0, \end{cases} \quad (3.44)$$

for all $v, w \in X^m$.

Furthermore, we introduce the following Hardy-type inequality.

Lemma 3.3.1

$$\int_{-1}^1 (1+t)|v(t)|^2 dt \leq \int_{-1}^1 (1+t)|v'(t)|^2 dt, \quad (3.45)$$

for all $v \in H_{\omega^{0,1}}^1(I)$ with $v(1) = 0$.

Proof Since $v(1) = 0$,

$$v(t) = - \int_t^1 v'(s) ds. \quad (3.46)$$

Hence

$$\begin{aligned} \int_{-1}^1 (1+t)|v(t)|^2 dt &= \int_{-1}^1 (1+t) \left| \int_t^1 v'(s) ds \right|^2 dt \\ &= \int_{-1}^1 (1+t) \left| \int_t^1 (1+s)^{\frac{1}{2}} (1+s)^{-\frac{1}{2}} v'(s) ds \right|^2 dt \\ &\leq \int_{-1}^1 (1+t) \left(\int_t^1 (1+s)^{-1} ds \right) \left(\int_t^1 (1+s) |v'(s)|^2 ds \right) dt \\ &\leq \int_{-1}^1 (1+t) \left(\int_t^1 (1+s)^{-1} ds \right) \left(\int_{-1}^1 (1+s) |v'(s)|^2 ds \right) dt \\ &= \int_{-1}^1 (1+t) |v'(s)|^2 ds \cdot \int_{-1}^1 (1+t) \int_t^1 (1+s)^{-1} ds dt \\ &= \int_{-1}^1 (1+t) |v'(t)|^2 dt. \end{aligned} \quad (3.47)$$

■

Remark 3.3.1 *By using similar argument, we can show (3.45) also holds for all $v \in H_{\omega^{0,1}}^1(I)$ with $v(-1) = 0$.*

By Lemma 3.3.1, we directly have

$$a_0(v, v) \geq \int_I (1+t)|v_t|^2 dt \geq \frac{1}{2} \|v\|_{X^0}^2. \quad (3.48)$$

Also,

$$a_m(v, v) \geq \|v\|_{X^m}^2, \quad \text{for } m > 0, \quad (3.49)$$

therefore $a_m(\cdot, \cdot)$ is a coercive bilinear form. By Lax-Milgram Lemma, (3.43) admits a unique solution.

For building the Galerkin formulation, we denote all the polynomials with complex coefficients on $[-1, 1]$ of degree no greater than N by P_N , and define the finite-dimensional subspace as

$$X_N^m = X^m \cap P_N, \quad (3.50)$$

then Galerkin formulation is

$$\begin{cases} \text{To find } v_N \in X_N^m \text{ such that} \\ a_m(v_N, w_N) = \frac{1}{4} (f(\frac{1+t}{2}), w_N)_{\omega^{0,1}}, \quad \forall w_N \in X_N^m, \end{cases} \quad (3.51)$$

Remark 3.3.2 *By Céa Lemma, the problem (3.51) admits a unique solution $v_N \in X_N^m$ such that*

$$\|v_N\|_{X^m} \leq \frac{1}{2} \|f(\frac{1+t}{2})\|_{\omega^{0,1}; I}. \quad (3.52)$$

Moreover, if v is the solution to (3.43), then

$$\|v - v_N\|_{X^m} \leq c_\alpha(m) \inf_{w_N \in X_N^m} \|v - w_N\|_{X^m}, \quad (3.53)$$

where

$$c_\alpha(m) = \begin{cases} \max(2, \alpha/2), & m = 0, \\ m^2 + \alpha/2, & m > 0. \end{cases} \quad (3.54)$$

Remark 3.3.3 *The basis functions for X_N^m can be chosen by $\{L_j - L_{j+1}\}_{j=0}^{N-1} \cup \{(L_j - L_{j+1})i\}_{j=0}^{N-1}$ for $m = 0$, and $\{L_j - L_{j+2}\}_{j=0}^{N-2} \cup \{(L_j - L_{j+2})i\}_{j=0}^{N-2}$ for $|m| \geq 1$. Here $\{L_j\}_{j=0}^N$ is the set of the Legendre polynomials which form an orthogonal basis for P_N with respect to L^2 inner product. By this means, the corresponding matrix is symmetric positive definite and has 7 non-zero diagonals, thus the linear system can be solved within $O(N^2)$ complexity by most solvers.*

Remark 3.3.4 *In practical computation, the integral $(f(\frac{1+t}{2}), w_N)_{\omega^{0,1}}$ on the right hand side is usually computed approximately by numerical quadrature. However, for simplicity, we assume it can be evaluated accurately in the following error analysis. Actually, most adaptive quadrature schemes evaluate the integral up to the machine precision, whose error is overwhelmed by that from the Galerkin method, as long as f is continuous in $[0, 1]$. Hence we ignore the error coming from evaluating $(f(\frac{1+t}{2}), w_N)_{\omega^{0,1}}$ in the error analysis in Section 3.4. On the other hand, Gauss quadrature with fixed number of nodes can also be applied to computing this term, and the numerical error depends on the regularity of f (see [1]). Suppose the Legendre polynomial-type basis functions and the Gauss quadrature with $O(N)$ nodes and weights are used for solving (3.51) numerically, then the total complexity is $O(N^2)$.*

3.4 Error estimates

3.4.1 Notations

We define the scaled Jacobi weight function $\hat{\omega}^{\alpha,\beta}(r)$ by

$$\hat{\omega}^{\alpha,\beta}(r) = (1-r)^\alpha r^\beta, \quad (3.55)$$

and the weighted L^2 inner product is given by

$$(u, v)_{\hat{\omega}^{\alpha,\beta}} = \int_0^1 u(r)\bar{v}(r)\hat{\omega}^{\alpha,\beta}(r)dr. \quad (3.56)$$

The associated scaled weighted L^2 and H^1 spaces are defined by

$$L_{\hat{\omega}^{\alpha,\beta}}^2 = \left\{ u(r) : \int_0^1 |u(r)|^2 \hat{\omega}^{\alpha,\beta}(r) dr < \infty \right\}, \quad (3.57)$$

$$H_{\hat{\omega}^{\alpha,\beta}}^1 = \left\{ u(r) : u, \partial_r u \in L_{\hat{\omega}^{\alpha,\beta}}^2 \right\} \quad (3.58)$$

equipped with norms

$$\|u\|_{\hat{\omega}^{\alpha,\beta}} = \left(\int_0^1 |u(r)|^2 \hat{\omega}^{\alpha,\beta}(r) dr \right)^{\frac{1}{2}}, \quad (3.59)$$

$$\|u\|_{1,\hat{\omega}^{\alpha,\beta}} = \left(\|u\|_{\hat{\omega}^{\alpha,\beta}}^2 + \|\partial_r u\|_{\hat{\omega}^{\alpha,\beta}}^2 \right)^{\frac{1}{2}}. \quad (3.60)$$

Furthermore, we introduce the space $H_{\mathbb{p}}^{1,s'}(\Pi)$ with $s' \geq 0$, which contains functions defined in Π that are 2π -periodic with respect to θ . We say

$$u(r, \theta) = \sum_{|m|=0}^{\infty} u^m(r) e^{im\theta} \in H_{\mathbb{p}}^{1,s'}(\Pi) \quad (3.61)$$

if and only if

$$u^m(r) \in H_{\hat{\omega}^{0,1}}^1 \cap L_{\hat{\omega}^{0,-1}}^2 \quad (3.62)$$

and

$$\sum_{|m|=0}^{\infty} m^{2s'} (m^2 \|u^m\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r u^m\|_{\hat{\omega}^{0,1}}^2) < \infty. \quad (3.63)$$

And the norm is defined by

$$\|u\|_{H_{\mathbb{p}}^{1,s'}(\Pi)} = \left(\sum_{|m|=0}^{\infty} m^{2s'} (m^2 \|u^m\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r u^m\|_{\hat{\omega}^{0,1}}^2) \right)^{\frac{1}{2}}. \quad (3.64)$$

Finally, we define the non-uniformly weighted Sobolev space by

$$\hat{B}_{\alpha,\beta}^s = \left\{ u(r) : \partial_r^k u \in L_{\hat{\omega}^{\alpha+k,\beta+k}}^2, \quad 0 \leq k \leq s \right\}. \quad (3.65)$$

3.4.2 The 2-D Petrov-Galerkin formulation

In this section, we have to recover the Petrov-Galerkin formulation corresponding to the method described in Section 3.2 and 3.3 for the extended 2-D problem (3.7),

and then apply the inf-sup conditions and space approximation property to obtain the error estimates. Suppose Γ is parametrized by (3.23). For simplicity, we only consider $\alpha = 0$, in which case $\phi^m = r^{|m|}$. For $\alpha > 0$, taking $\phi^m = \frac{J_m(\sqrt{\alpha}r)}{J_m(\sqrt{\alpha})}$ will leads to similar results.

First let us define the 1-D function spaces by

$$\begin{aligned} \hat{W}^m &= \begin{cases} \left\{ u(r) : \int_0^1 (|u|^2 + |\partial_r u|^2) r dr < \infty \right\}, & m = 0, \\ \left\{ u(r) : u(0) = 0, \int_0^1 \frac{1}{r} |u|^2 + r |\partial_r u|^2 dr < \infty \right\}, & m \neq 0, \end{cases} \\ \hat{Y}^m &= \{u \in \hat{W}^m : u(1) = 0\}, \end{aligned} \quad (3.66)$$

and the finite-dimensional subspaces $\hat{W}_N^m := \hat{W}^m \cap \hat{P}_N$, $\hat{Y}_N^m := \hat{Y}^m \cap \hat{P}_N$, where \hat{P}_N is the set of all the polynomials with complex coefficients on $[0, 1]$ of degree no greater than N . \hat{W}^m can be equipped with the following norm

$$\|u\|_{\hat{W}^m} := (m^2 \|u\|_{\tilde{\omega}^{0,-1}}^2 + \|\partial_r u\|_{\tilde{\omega}^{0,1}}^2)^{\frac{1}{2}}. \quad (3.67)$$

Now define the solution space by

$$\begin{aligned} X := \left\{ u(r, \theta) = \sum_{|m|=0}^{\infty} u^m(r) e^{im\theta} : u^m \in \hat{W}^m, \right. \\ \left. \sum_{|m|=0}^{\infty} \|u^m\|_{\hat{W}^m}^2 < \infty, u|_{\partial\Omega} = 0 \text{ in the trace sense} \right\}. \end{aligned} \quad (3.68)$$

with norm

$$\|u\|_X := \left(\sum_{|m|=0}^{\infty} \|u^m\|_{\hat{W}^m}^2 \right)^{\frac{1}{2}}. \quad (3.69)$$

And the test space is introduced by

$$Y := \left\{ v(r, \theta) = \sum_{|m|=0}^{\infty} v^m(r) e^{im\theta} : v^m \in \hat{Y}^m, \sum_{|m|=0}^{\infty} \|v^m\|_{\hat{Y}^m}^2 < \infty \right\}. \quad (3.70)$$

with norm

$$\|v\|_Y := \left(\sum_{|m|=0}^{\infty} \|v^m\|_{\hat{Y}^m}^2 \right)^{\frac{1}{2}}. \quad (3.71)$$

For building Galerkin framework, the finite dimensional approximate space to X is given by

$$X_{MN} := \left\{ u_{MN}(r, \theta) = \sum_{|m|=0}^M u_N^m(r) e^{im\theta} : u_N^m \in \hat{W}_N^m, \right. \\ \left. (u_{MN}, \xi_k)_{\partial\Omega} = 0, \text{ for } |k| = 0, \dots, M \right\}. \quad (3.72)$$

with norm

$$\|u_{MN}\|_{X_{MN}} := \left(\sum_{|m|=0}^M \|u_N^m\|_{\hat{W}_N^m}^2 \right)^{\frac{1}{2}}, \quad (3.73)$$

where $(u_{MN}, \xi_k)_{\partial\Omega} := \int_0^{2\pi} u_{MN}(\rho(\theta), \theta) \xi_k(\theta) d\theta$, and $\{\xi_k\}_{k=0}^M$ are defined by (3.26).

Similarly, the finite dimensional subspace of Y is given by

$$Y_{MN} := \left\{ v_{MN}(r, \theta) = \sum_{|m|=0}^M v_N^m(r) e^{im\theta} : v_N^m \in \hat{Y}_N^m, \right\}, \quad (3.74)$$

with norm

$$\|v_{MN}\|_{Y_{MN}} := \left(\sum_{|m|=0}^M \|v_N^m\|_{\hat{Y}_N^m}^2 \right)^{\frac{1}{2}}. \quad (3.75)$$

Now let $X_s := X + X_{MN}$ be the sum space, and suppose $u = \sum_{|m|=0}^{\infty} u^m(r) e^{im\theta} \in X_s$ and $v = \sum_{|m|=0}^{\infty} v^m(r) e^{im\theta} \in Y$, define the bilinear form $A(\cdot, \cdot) : X_s \times Y \rightarrow \mathbb{C}$ by

$$A(u, v) := \sum_{|m|=0}^{\infty} m^2 (u, v)_{\hat{\omega}^{0,-1}} + (\partial_r u, \partial_r v)_{\hat{\omega}^{0,1}}. \quad (3.76)$$

$A(\cdot, \cdot)$ is bounded since

$$\begin{aligned} |A(u, v)| &\leq \sum_{|m|=0}^{\infty} m^2 |(u, v)_{\hat{\omega}^{0,-1}}| + |(\partial_r u, \partial_r v)_{\hat{\omega}^{0,1}}| \\ &\leq \sum_{|m|=0}^{\infty} m^2 \|u\|_{\hat{\omega}^{0,-1}} \|v\|_{\hat{\omega}^{0,-1}} + \|\partial_r u\|_{\hat{\omega}^{0,1}} \|\partial_r v\|_{\hat{\omega}^{0,1}} \leq \|u\|_{X_s} \|v\|_Y, \quad \forall u \in X_s, v \in Y. \end{aligned} \quad (3.77)$$

Under the preceding settings, the weak formulation for the problem (3.7) is given by

$$\begin{cases} \text{To find } u \in X \text{ such that} \\ A(u, v) = \langle f, v \rangle, \quad \forall v \in Y, \end{cases} \quad (3.78)$$

where

$$\langle f, v \rangle = \left\langle \sum_{|m|=0}^{\infty} f^m(r) e^{im\theta}, \sum_{|m|=0}^{\infty} v^m(r) e^{im\theta} \right\rangle := \sum_{|m|=0}^{\infty} (f^m, v^m)_{\hat{\omega}^{0,1}}. \quad (3.79)$$

And by applying finite-dimensional space approximation, we can derive the following Petrov-Galerkin formulation,

$$\begin{cases} \text{To find } u_{MN} \in X_{MN} \text{ such that} \\ A(u_{MN}, v_{MN}) = \langle f, v_{MN} \rangle, \quad \forall v_{MN} \in Y_{MN}. \end{cases} \quad (3.80)$$

It can be verified that applying the method described in Section 3.2 and 3.3 for solving (3.7) is equivalent to solving (3.80) directly. Therefore it suffices to make analysis on (3.78) and (3.80) to estimate the solution error obtained by Algorithm 3.2.1 and the Bessel-type solver in Sec.3.

3.4.3 Error estimate when $\rho(\theta) = \rho_0$

We only consider the special case $\rho(\theta) = \rho_0$, in which the solution spaces X and X_{MN} can be characterized more explicitly. First, define the following ρ_0 -dependent subspace of \hat{W}^m

$$\hat{X}^m = \{u \in \hat{W}^m : u(\rho_0) = 0\}, \quad (3.81)$$

Denote $u = \sum_{|m| \geq 0} u^m(r) e^{im\theta}$. In the case $\rho(\theta) = \rho_0$, $u|_{\partial\Omega} = 0$ is equivalent to $u^m(\rho_0) = 0$, thus X can be rewritten as

$$X := \left\{ u = \sum_{|m| \geq 0} u^m(r) e^{im\theta} : u^m \in \hat{X}^m, \sum_{|m| \geq 0} \|u^m\|_{\hat{W}^m}^2 < \infty \right\}. \quad (3.82)$$

Similarly,

$$X_{MN} := \left\{ u_{MN} = \sum_{|m|=0}^M u_N^m(r) e^{im\theta} : u_N^m \in \hat{X}_N^m, |m| = 0, \dots, M \right\}. \quad (3.83)$$

Note in the special case X_{MN} is a subspace of X .

We introduce the 1-D bilinear operator $\hat{b}_m(\cdot, \cdot) : \hat{X}^m \times \hat{Y}^m \rightarrow \mathbb{C}$ by

$$\hat{b}_m(u, v) = m^2(u, v)_{\hat{\omega}^{0,-1}} + (\partial_r u, \partial_r v)_{\hat{\omega}^{0,1}}. \quad (3.84)$$

$\hat{b}_m(\cdot, \cdot)$ is bounded since

$$|\hat{b}_m(u, v)| \leq \|u\|_{\hat{W}^m} \|v\|_{\hat{W}^m}. \quad (3.85)$$

Now suppose $u = \sum_{|m| \geq 0} u^m(r) e^{im\theta}$ and $u_{MN} = \sum_{|m|=0}^M u_N^m(r) e^{im\theta}$ are the solutions to (3.78) and (3.80), then

$$\begin{cases} u^m \in \hat{X}^m \text{ satisfies} \\ \hat{b}_m(u^m, v^m) = (f^m, v^m)_{\hat{\omega}^{0,1}}, \quad \forall v^m \in \hat{Y}^m, \end{cases} \quad (3.86)$$

and

$$\begin{cases} u_N^m \in \hat{X}_N^m \text{ satisfies} \\ \hat{b}_m(u_N^m, v_N^m) = (f^m, v_N^m)_{\hat{\omega}^{0,1}}, \quad \forall v_N^m \in \hat{Y}_N^m, \end{cases} \quad (3.87)$$

In the following, we deduce the inf-sup condition for $\hat{b}_m(\cdot, \cdot)$, and then estimate $\|u^m - u_N^m\|$, by accumulating which we obtain the total error $\|u - u_{MN}\|$.

Inf-sup condition for $\hat{b}_m(\cdot, \cdot)$

First, we need the following result which estimates the boundary value for the functions in \hat{X}^m .

Lemma 3.4.1 *For $|m| \neq 0$, suppose $u \in \hat{X}^m$, then $u(1)$ is finite. Specifically, if $\|u\|_{\hat{W}^m} = 1$, then*

$$|u(1)| < \hat{c}(\rho_0; m) |m|^{-\frac{1}{2}}, \quad (3.88)$$

where

$$\hat{c}(\rho_0; m) = (1 - \rho_0^{2|m|})^{\frac{1}{2}} (1 + \rho_0^{2|m|})^{-\frac{1}{2}} < 1. \quad (3.89)$$

Proof We need to estimate

$$\sup |u(1)|, \quad \text{s.t.} \quad \|u\|_{\hat{W}^m} = 1, \quad (3.90)$$

which is clearly equal to

$$\sup |u(1)|, \quad \text{s.t.} \quad \|u\|_{\hat{W}^m} = 1, \quad u \equiv 0 \text{ in } (0, \rho_0), \quad (3.91)$$

or

$$\sup |u(1)|, \quad \text{s.t.} \quad \int_{\rho_0}^1 \frac{m^2}{r} |u|^2 + r |\partial_r u|^2 dr = 1, \quad u(\rho_0) = 0. \quad (3.92)$$

It suffices to consider the following variational problem

$$\inf \int_{\rho_0}^1 \frac{m^2}{r} |u|^2 + r |\partial_r u|^2 dr, \quad \text{s.t.} \quad u(\rho_0) = 0, \quad u(1) = 1. \quad (3.93)$$

Note the Euler-Lagrange equation of (3.93) is given by

$$\frac{m^2}{r} u - \partial_r (r \partial_r u) = 0, \quad u(\rho_0) = 0, \quad u(1) = 1, \quad (3.94)$$

whose solution is

$$u^*(r) = (1 - \rho_0^{2m})^{-1} r^m + (1 - \rho_0^{-2m})^{-1} r^{-m}. \quad (3.95)$$

If we define

$$u^{**} = \begin{cases} 0, & \text{in } (0, \rho_0), \\ u^* & \text{in } (\rho_0, 1), \end{cases} \quad (3.96)$$

then

$$\begin{aligned} \inf_{u \in \hat{X}^m, u(1)=1} \|u\|_{\hat{W}^m} &= \|u^{**}\|_{\hat{W}^m} \\ &= \left(\int_{\rho_0}^1 \frac{m^2}{r} |u^*|^2 + r |\partial_r u^*|^2 dr \right)^{\frac{1}{2}} \\ &= |m|^{\frac{1}{2}} \left((1 - \rho_0^{2|m|})^{-1} - (1 - \rho_0^{-2|m|})^{-1} \right)^{\frac{1}{2}}. \end{aligned} \quad (3.97)$$

Hence

$$\begin{aligned} \sup_{\|u\|_{\hat{W}^m}=1} |u(1)| &= \left(\inf_{u \in \hat{X}^m, u(1)=1} \|u\|_{\hat{W}^m} \right)^{-1} \\ &= |m|^{-\frac{1}{2}} (1 - \rho_0^{2|m|})^{\frac{1}{2}} (1 + \rho_0^{2|m|})^{-\frac{1}{2}}. \end{aligned} \quad (3.98)$$

■

Now we have

Theorem 3.4.1 *the following inequality holds*

$$\inf_{u^m \in \hat{X}^m} \sup_{v^m \in \hat{Y}^m} \frac{\hat{b}_m(u^m, v^m)}{\|u^m\|_{\hat{W}^m} \|v^m\|_{\hat{W}^m}} > \frac{1}{2} \rho_0^{2|m|}. \quad (3.99)$$

Proof Given $u^m \in \hat{X}^m$, and suppose $\|u^m\|_{\hat{W}^m} = 1$ without loss of generality. Let $w^m = u(1)r^{|m|}$ and $v^m = u^m - w^m \in \hat{Y}^m$. Then by Lemma 3.4.1 it follows

$$\|w^m\|_{\hat{W}^m} = |u(1)||m|^{\frac{1}{2}} < \hat{c}(\rho_0; m), \quad (3.100)$$

therefore

$$\|v^m\|_{\hat{W}^m} \leq \|u^m\|_{\hat{W}^m} + \|w^m\|_{\hat{W}^m} < 1 + \hat{c}(\rho_0; m). \quad (3.101)$$

Also, since

$$\frac{m^2}{r} w^m - \partial_r(r \partial_r w^m) = 0, \quad (3.102)$$

by integrating by parts and (3.100) we have

$$\begin{aligned} \hat{b}_m(u^m, v^m) &= m^2(u^m, u^m - w^m)_{\omega^{0,-1}} + (\partial_r u^m, \partial_r u^m - \partial_r w^m)_{\omega^{0,1}} \\ &= \|u^m\|_{\hat{W}^m}^2 - \int_0^1 \left(\frac{m^2}{r} \overline{w^m} - \partial_r(r \partial_r \overline{w^m}) \right) u^m dr - u^m(r \partial_r \overline{w^m}) \Big|_0^1 \\ &= 1 - |u^m(1)|^2 |m| > 1 - \hat{c}^2(\rho_0; m) > \frac{1}{2} \rho_0^{2|m|}, \end{aligned} \quad (3.103)$$

Combining (3.103) and (3.101) leads to (3.99). ■

Approximation property

Now let us consider the degree to which \hat{X}_N^m approximates \hat{X}^m in the norm sense. It is necessary to introduce the orthogonal projection $\hat{\pi}_N^m : \hat{Y}^m \rightarrow \hat{Y}_N^m$ by

$$\langle \hat{\pi}_N^m v - v, w_N \rangle_{\hat{Y}^m} = 0, \quad \forall w_N \in \hat{Y}_N^m, \quad (3.104)$$

where the inner product $\langle \cdot, \cdot \rangle_{\hat{Y}^m} : \hat{Y}^m \times \hat{Y}^m \rightarrow \mathbb{C}$ is defined by

$$\langle v, w \rangle_{\hat{Y}^m} = m^2(v, w)_{\omega^{0,-1}} + (v', w')_{\omega^{0,1}}. \quad (3.105)$$

The approximation property from \hat{Y}_N^m to \hat{Y}^m is discussed in [1], which is

Theorem 3.4.2 For any $v \in \hat{Y}^m \cap \hat{B}_{-1,-1}^s$ with $1 \leq s \leq N+1$,

$$\begin{aligned} m^2 \|\hat{\pi}_N^m v - v\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r(\hat{\pi}_N^m v - v)\|_{\hat{\omega}^{0,1}}^2 \\ \leq c(1 + m^2 N^{-2}) \frac{(N-s+1)!}{N!} (N+s)^{1-s} \|\partial_r^s v\|_{\hat{\omega}^{s-1,s-1}}^2, \end{aligned} \quad (3.106)$$

where c is a positive constant independent of m , N and v .

Actually, a more general result also exists.

Proposition 3.4.1 Suppose $u \in \hat{W}^m \cap \hat{B}_{-1,-1}^s$ with $1 \leq s \leq N+1$, then there exists a linear operator $\tilde{\pi}_N^m : \hat{W}^m \rightarrow \hat{W}_N^m$ such that

$$\begin{aligned} m^2 \|\tilde{\pi}_N^m u - u\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r(\tilde{\pi}_N^m u - u)\|_{\hat{\omega}^{0,1}}^2 \\ \leq c(1 + m^2 N^{-2}) \frac{(N-s+1)!}{N!} (N+s)^{1-s} \|\partial_r^s u\|_{\hat{\omega}^{s-1,s-1}}^2, \end{aligned} \quad (3.107)$$

where c is a positive constant independent of m , N and u .

Proof Denote $u^* = (1-r)u(0) + ru(1)$, then $u - u^* \in \hat{Y}^m$. Define the operator by

$$\tilde{\pi}_N^m u := \hat{\pi}_N^m(u - u^*) + u^*. \quad (3.108)$$

Since $u(r) - u(1)r \in \hat{Y}^m \cap \hat{B}_{-1,-1}^s$, by using Theorem 4.1,

$$\begin{aligned} m^2 \|\tilde{\pi}_N^m u - u\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r(\tilde{\pi}_N^m u - u)\|_{\hat{\omega}^{0,1}}^2 \\ = m^2 \|\hat{\pi}_N^m(u - u^*) - (u - u^*)\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r(\hat{\pi}_N^m(u - u^*) - (u - u^*))\|_{\hat{\omega}^{0,1}}^2 \\ \leq c(1 + m^2 N^{-2}) \frac{(N-s+1)!}{N!} (N+s)^{1-s} \|\partial_r^s(u - u^*)\|_{\hat{\omega}^{s-1,s-1}}^2, \end{aligned} \quad (3.109)$$

For $s \geq 2$, $\|\partial_r^s(u - u^*)\|_{\hat{\omega}^{s-1,s-1}} = \|\partial_r^s u\|_{\hat{\omega}^{s-1,s-1}}$. And for $s = 1$, note

$$|\partial_r u^*| = |u(1) - u(0)| = \left| \int_0^1 \partial_r u dr \right| \leq \|\partial_r u\|_{\hat{\omega}^{0,0}}, \quad (3.110)$$

so

$$\|\partial_r(u - u^*)\|_{\hat{\omega}^{0,0}} \leq \|\partial_r u\|_{\hat{\omega}^{0,0}} + \|\partial_r u^*\|_{\hat{\omega}^{0,0}} \leq 2\|\partial_r u\|_{\hat{\omega}^{0,0}}. \quad (3.111)$$

Therefore (3.107) follows immediately from (3.109). ■

Now given $u^m \in \hat{X}^m$, let

$$\tilde{u}_N^m(r) = \tilde{\pi}_N^m u^m(r) + s^m r, \quad (3.112)$$

where $s^m = -\frac{1}{\rho_0} \tilde{\pi}_N^m u^m(\rho_0)$, Hence by the facts $u^m(1) - \tilde{\pi}_N^m u^m(1) = 0$, $u^m(\rho_0) = 0$ and Cauchy-Schwarz inequality, we have

$$\begin{aligned} |s^m| &= \frac{1}{\rho_0} |(u^m(1) - \tilde{\pi}_N^m u^m(1)) - (u^m(\rho_0) - \tilde{\pi}_N^m u^m(\rho_0))| \\ &= \frac{1}{\rho_0} \left| \int_{\rho_0}^1 \partial_r u^m - \partial_r \tilde{\pi}_N^m u^m dr \right| \\ &\leq \frac{1}{\rho_0} \left(\int_{\rho_0}^1 |\partial_r u^m - \partial_r \tilde{\pi}_N^m u^m|^2 r dr \right)^{\frac{1}{2}} \left(\int_{\rho_0}^1 \frac{1}{r} dr \right)^{\frac{1}{2}} \\ &\leq \frac{\sqrt{-\ln \rho_0}}{\rho_0} \|\partial_r u^m - \partial_r \tilde{\pi}_N^m u^m\|_{\hat{\omega}^{0,1}}. \end{aligned} \quad (3.113)$$

Finally by using (3.107) and (3.113) we obtain

$$\begin{aligned} \|u^m - \tilde{u}_N^m\|_{\hat{W}^m} &= m^2 \|u^m - \tilde{\pi}_N^m u^m - s^m r\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r(u^m - \tilde{\pi}_N^m u^m - s^m r)\|_{\hat{\omega}^{0,1}}^2 \\ &\leq c \left(m^2 \|u^m - \tilde{\pi}_N^m u^m\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r(u^m - \tilde{\pi}_N^m u^m)\|_{\hat{\omega}^{0,1}}^2 + \right. \\ &\quad \left. m^2 \|s^m r\|_{\hat{\omega}^{0,-1}}^2 + \|\partial_r(s^m r)\|_{\hat{\omega}^{0,1}}^2 \right) \\ &\leq c(1 + m^2 N^{-2}) \frac{(N-s+1)!}{N!} (N+s)^{1-s} \left(1 - m^2 \frac{\ln \rho_0}{\rho_0}\right) \|\partial_r^s u^m\|_{\hat{\omega}^{s-1,s-1}}^2, \end{aligned} \quad (3.114)$$

where c is a positive constant independent of m , N , ρ_0 and u .

Total error estimation

First let us introduce the ρ_0 -dependent space $B_{p,\rho_0}^{s,s'}(\Pi)$ with $s \geq 1$, $s' \geq 0$, which contains functions defined in Π that are 2π -periodic with respect to θ . We say

$$u(r, \theta) = \sum_{|m|=0}^{\infty} u^m(r) e^{im\theta} \in B_{p,\rho_0}^{s,s'}(\Pi) \quad (3.115)$$

if and only if

$$u^m \in \hat{X}^m \cap \hat{B}_{-1,-1}^s \quad (3.116)$$

and

$$\|u\|_{B_{p,\rho_0}^{s,s'}(\Pi)} := \left(\sum_{|m|=0}^{\infty} |m|^{2s'} \rho_0^{-2|m|} \|\partial_r^s u^m\|_{\hat{\omega}^{s-1,s-1}}^2 \right)^{\frac{1}{2}} < \infty. \quad (3.117)$$

Now suppose $u^m \in \hat{X}^m \cap \hat{B}_{-1,-1}^s$ is the solution to (3.86) and $u_N^m \in \hat{X}_N^m$ is the solution to (3.87). By applying Second Strang Lemma with (3.99) and (3.114), it immediately follows

$$\begin{aligned} \|u^m - \tilde{u}_N^m\|_{\hat{W}^m} &\leq \left(1 + \frac{2}{\rho_0^{2|m|}}\right) \inf_{u_N^m \in \hat{X}_N^m} \|u^m - u_N^m\|_{\hat{W}^m} \\ &\leq c(1 + m^2 N^{-2}) \frac{(N-s+1)!}{N!} (N+s)^{1-s} \left(1 - m^2 \frac{\ln \rho_0}{\rho_0}\right) \rho_0^{-2|m|} \|\partial_r^s u^m\|_{\hat{\omega}^{s-1,s-1}}^2, \end{aligned} \quad (3.118)$$

and if $u = \sum_{|m|\geq 0} u^m(r) e^{im\theta}$ and $u_{MN} = \sum_{|m|=0}^M u_N^m(r) e^{im\theta}$, we have

$$\begin{aligned} \|u - u_{MN}\|_X^2 &= \sum_{|m|=0}^M \|u^m - u_N^m\|_{\hat{W}^m}^2 + \sum_{|m|>M} \|u^m\|_{\hat{W}^m}^2 \\ &\leq c(1 + M^2 N^{-2}) \frac{(N-s+1)!}{N!} (N+s)^{1-s} \left(\sum_{|m|\geq 0} \rho_0^{-2|m|} \|\partial_r^s u^m\|_{\hat{\omega}^{s-1,s-1}}^2 - \right. \\ &\quad \left. \frac{\ln \rho_0}{\rho_0} \sum_{|m|\geq 0} |m|^2 \rho_0^{-2|m|} \|\partial_r^s u^m\|_{\hat{\omega}^{s-1,s-1}}^2 \right) + m^{-2s'} \sum_{|m|>M} m^{2s'} \|u^m\|_{\hat{W}^m}^2 \\ &\leq c(1 + M^2 N^{-2}) \frac{(N-s+1)!}{N!} (N+s)^{1-s} \left(\|u\|_{B_{p,\rho_0}^{s,0}(\Pi)}^2 - \right. \\ &\quad \left. \frac{\ln \rho_0}{\rho_0} \|u\|_{B_{p,\rho_0}^{s,1}(\Pi)}^2 \right) + M^{-2s'} \|u\|_{H_p^{1,s'}(\Pi)^2}, \end{aligned} \quad (3.119)$$

which leads to the following error estimate

Theorem 3.4.3 *Suppose $u \in X$ and $u_{MN} \in X_{MN}$ are the solutions to (3.78) and (3.80). Furthermore, suppose $u \in B_{p,\rho_0}^{s,1}(\Pi) \cap H_p^{1,s'}(\Pi)$ with $1 \leq s \leq N+1$ and $s' \geq 1$. Then*

$$\begin{aligned} \|u - u_{MN}\|_X &\leq c(1 + MN^{-1}) \sqrt{\frac{(N-s+1)!}{N!}} (N+s)^{\frac{1-s}{2}} \left(\|u\|_{B_{p,\rho_0}^{s,0}(\Pi)} + \right. \\ &\quad \left. \sqrt{-\frac{\ln \rho_0}{\rho_0}} \|u\|_{B_{p,\rho_0}^{s,1}(\Pi)} \right) + M^{-s'} \|u\|_{H_p^{1,s'}(\Pi)}. \end{aligned} \quad (3.120)$$

Specifically, if s is fixed, then

$$\begin{aligned} & \|u - u_{MN}\|_X \\ & \leq c(1 + MN^{-1})N^{1-s} \left(\|u\|_{B_{p,\rho_0}^{s,0}(\Pi)} + \sqrt{-\frac{\ln \rho_0}{\rho_0}} \|u\|_{B_{p,\rho_0}^{s,1}(\Pi)} \right) + M^{-s'} \|u\|_{H_p^{1,s'}(\Pi)}. \end{aligned} \quad (3.121)$$

From (3.121) we see if $\rho_0 \rightarrow 1$, the solution error estimate is simplified as

$$\|u - u_{MN}\|_X \leq c(1 + MN^{-1})N^{1-s} \|u\|_{B_{p,1}^{s,0}(\Pi)} + M^{-s'} \|u\|_{H_p^{1,s'}(\Pi)}, \quad (3.122)$$

which is exactly the one from applying standard spectral-Galerkin method to 2D-Laplace equation in a circular domain [1]. On the other hand, if $\rho_0 \rightarrow 0^+$, the coefficient term $\sqrt{-\ln \rho_0/\rho_0}$ approaches to $+\infty$, which implies the solution error may grow rapidly. Indeed, the error of the solution solved by embedding method will be smaller as the embedding domain becomes closer to the original one in the common sense.

3.5 Numerical examples

Algorithm 3.2.1 are tested on the model problem (3.1) by four examples. In the first two ones, analytic solutions are looked for in smooth and polygonal domains. Theorem 3.4.3 implies the numerical solutions should converge in an exponential rate. In the last two ones, the solutions are created with singularities, and the convergence rate by Algorithm 3.2.1 are compared with that by the general spectral element method (SEM) package.

In the first example, α is set by 10, and an analytic solution

$$U = \exp(y/(x + 2)) \quad (3.123)$$

inside a smooth domain

$$\Omega = \{(r, \theta) : r < 0.7 + 0.2 \sin(3\theta)\}, \quad (3.124)$$

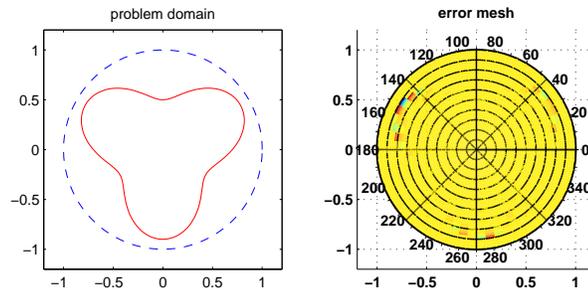


Fig. 3.2.: the original domain, enclosing domain(left) and the error profile(right) for $M = 24$ in the first example

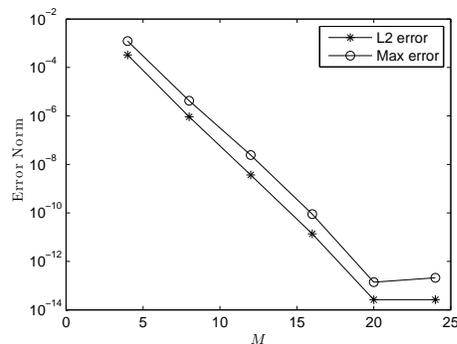


Fig. 3.3.: L^2 and max error versus M in the first example

is chosen. Here $\{t^m\}$ are computed by (3.25), and we set $N = 2M$. The problem domains and error profile when $M = 24$ are depicted in Fig. 3.2, and the relative L^2 and max errors versus M for $M = 4, 8, \dots, 24$ are plotted in Fig. 3.3. The errors decaying exponentially to the machine precision are observed.

In the second example, we solve for the same solution (3.123) in a convex polygonal domain whose vertices are $(0, 0.7)$, $(-0.7, 0.2)$, $(-0.5, -0.6)$, $(0.5, -0.6)$, $(0.7, 0.2)$. The setting is same as the first one except that $\{t^m\}$ are computed by (3.22) and K is set by $4M$. The numerical results are shown in Fig. 3.4 and Fig. 3.5, in which we can see the error is mainly accumulated in the center and near the boundary. The exponential convergence rate is also observed.

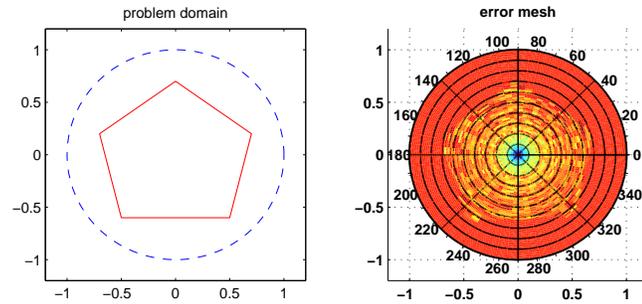


Fig. 3.4.: the original domain, enclosing domain(left) and the error profile(right) for $M = 30$ in the second example

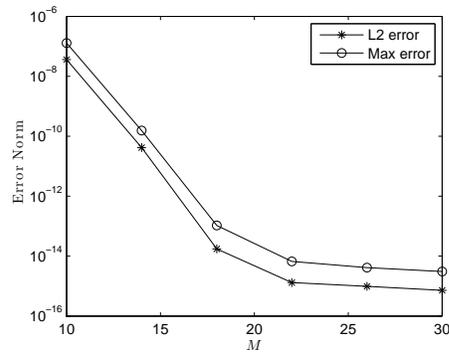


Fig. 3.5.: L^2 and max error versus M in the second example

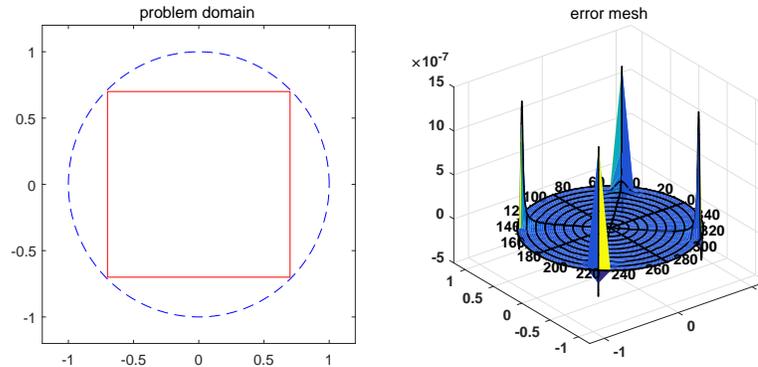


Fig. 3.6.: *the original domain, enclosing domain(left) and the error mesh(right) for $M = 64$ in the third example*

Now let us consider the problem (3.1) with $F \equiv 1$ and Ω being a polygon, in which case the solution has singularities at the corners of the domain. In the third example, $\alpha = 0$ and Ω is set by a square with vertices (T, T) , $(T, -T)$, $(-T, -T)$, $(-T, T)$ and T is chosen by 0.7. The exact solution in this case is given by

$$u(x, y) = -\frac{64T^2}{\pi^4} \sum_{\substack{n,m=1 \\ n,m \text{ odd}}}^{\infty} (-1)^{\frac{n+m}{2}} \frac{\cos(\frac{n\pi x}{2T})\cos(\frac{m\pi y}{2T})}{nm(n^2 + m^2)}, \quad (3.125)$$

First we perform Algorithm 3.2.1 on it with $\{t^m\}$ being computed by (3.25). The domains and error mesh for $M = 64$ are shown in Fig. 3.6, from which we see the error is mainly distributed near the four corners. Also, we take advantage of the general SEM package on it, and the domain decomposition is shown in Fig. 3.7. The error convergence rate of these two approaches are presented in Fig. 3.8. We show the relative L^2 error by Algorithm 3.2.1 versus M ($N = 2M$) in the left figure. For comparison, the same error by SEM package versus the square root of degree of freedom is plotted in the right figure. It can be observed that both errors converge in orders between 4th and 5th.

The fourth example takes the same setting as the third one, except that the domain is a concave L-shaped polygon. A numerical solution solved with high-accuracy is used as the exact solution for error computation. Algorithm 3.2.1 with $\{t^m\}$ being

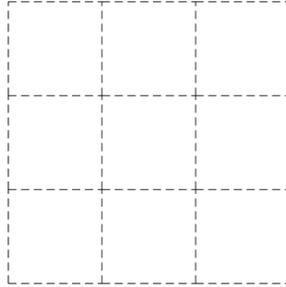


Fig. 3.7.: *decomposed domain by SEM package in the third example*

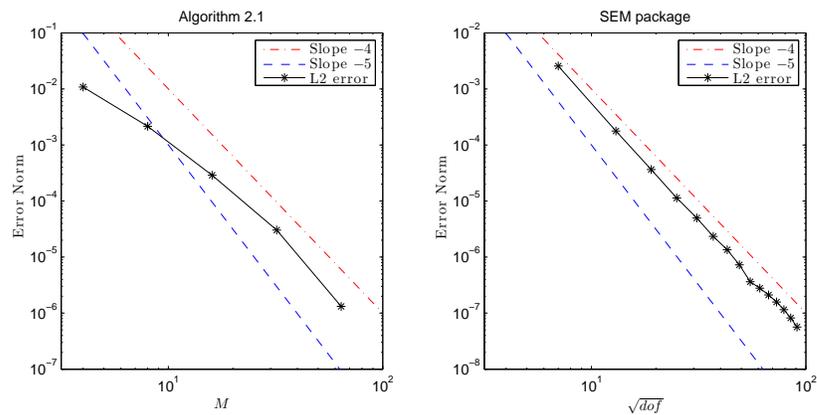


Fig. 3.8.: *L^2 error versus M by Algorithm 3.2.1 (left) and the square root of degree of freedom by general SEM package (right) in the third example*

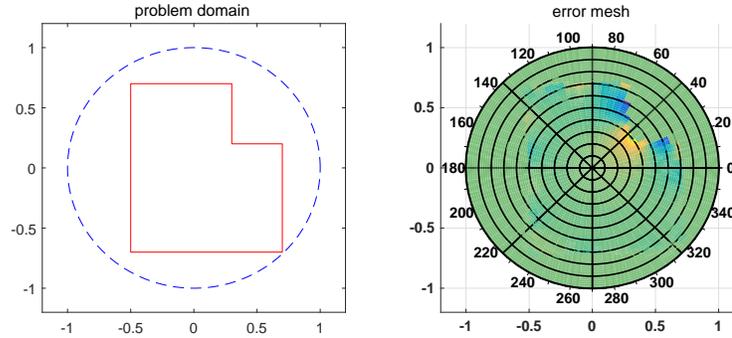


Fig. 3.9.: *the original domain, enclosing domain(left) and the error profile(right) for $M = 96$ in the fourth example*

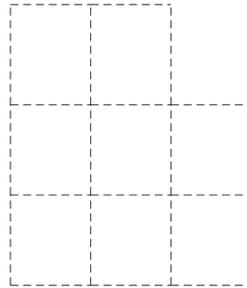


Fig. 3.10.: *decomposed domain by SEM package in the fourth example*

computed by (3.22) is performed here. To enhance the accuracy near the concave corner where the singularity most dominates, we put more nodes $\{\hat{r}_k, \hat{\theta}_k\}$ near the corner. The domains and error profile for $M = 96$ is demonstrated in Fig. 3.9, where the majority of the error is located near concave corner and the upper-right corners. The relative L^2 error versus M ($N = 2M$) is shown on the left in Fig.5.10, and the convergence rate is observed to be between 1st and 2nd order. On the other hand, we apply the SEM package on the same problem. The domain is decomposed by the means shown in Fig. 3.10, and the decaying error is plotted on the right in Fig. 3.11, from which we can see its convergence rate is less than 1st order.

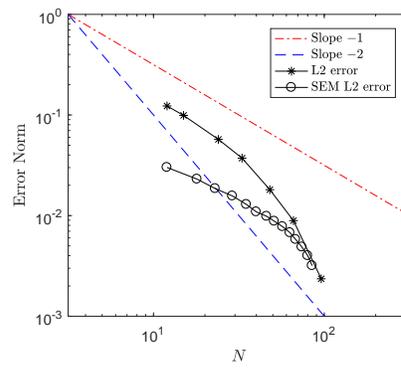


Fig. 3.11.: L^2 error versus M by Algorithm 3.2.1 (left) and the square root of degree of freedom by general SEM package (right) in the fourth example

3.6 Conclusion

In this chapter, we develop a domain-embedding approach for solving second-order elliptic problems with Dirichlet boundary condition in 2-D complex domains. By inserting the original domain into a circle, we obtain an extend problem and solve for a weak solution. Thanks to the circular geometry, the polar transformation can be applied, leading the 2-D problem to a sequence of 1-D Bessel-type equations, which can be solved by general ODE solvers. Specifically, we put forward a spectral-Galerkin formulation for such ODEs, and perform error analysis for the whole algorithm.

Although the error estimate is studied and presented by Theorem 3.4.3 only for the special case where the domain is characterized by $\rho(\theta) = \rho_0$, it can be foreseen that the result also applies to the general cases $\rho(\theta) = \rho_0 + \epsilon(\theta)$ where $\epsilon(\theta)$ is a small perturbation. Indeed, the first and second numerical examples in Section 3.5 show the convergence rate can reach the exponential level as long as the solution is smooth enough. For the problems whose solution have singularities, the algorithm is also able to provide decent numerical results. The third example in Section 3.5 is working out a slightly singular problem in a square domain, and both our algorithm and the standard spectral element package obtain the convergence rate of 4th to 5th order. For the highly singular problem in the L-shaped domain described in the last example, our algorithm obtain higher order (greater than the 1st) than the standard package (less than the 1st), and the implementation is also much more straightforward (without domain decomposition).

Compared to the approach presented in the last Chapter, which encloses the domain by a rectangle, this circular embedding scheme is less time-consuming and simpler to implement because it suffices to solve a sequence of ODEs instead of a multi-dimensional PDE. Furthermore, the circular embedding framework can be applied to other problems in complex domains, for example, the Helmholtz equation for acoustic scattering, that will be our future work.

4. SPECTRAL METHOD WITH ANNULAR EMBEDDING

4.1 Introduction

In the last two chapters, we discuss the spectral methods and fictitious domain framework for the problems in 2D simply connected complex domains. However, for a non-simply connected domain, embedding by a rectangle or a circle is not always applicable, since smooth extensions of the problem may not be found easily in the new regular domain. For example, a data function $f(x) := 1/(x^2 + y^2)$ defined in $\Omega := [-1, 1] \times [-1, 1] \setminus [-0.5, 0.5] \times [-0.5, 0.5]$ can not be extended analytically to a circle enclosing Ω because of the singularity at origin. Hence, similarity to the procedure of the circular embedding discussed in the last chapter, we put forward the annular embedding scheme for the case of non-simply connected domains.

In Section 4.2, we describe the problem briefly and put forward the corresponding spectral method. In Section 4.3, the Galerkin method for the Bessel-type equation derived from dimension reduction is discussed. In Section 4.4 several examples are presented to show the effectiveness of the algorithm. And we conclude for this chapter in Section 4.5.

4.2 Problem in non-simply connected domains

4.2.1 Dimension reduction

Let us consider the following Poisson-type model problem

$$\begin{aligned} \alpha U - \Delta U &= F, & \text{in } \Omega, \\ U &= H, & \text{on } \partial\Omega. \end{aligned} \tag{4.1}$$

where $\alpha \geq 0$ and Ω is a two-dimensional complex domain formed by removing a simply connected interior part from a simply connected domain, and $F \in C(\bar{\Omega})$, $H \in C(\partial\Omega)$. For convenience, we suppose $(0, 0) \notin \bar{\Omega}$. Now let $\tilde{\Omega} := \{(x, y) : a^2 < x^2 + y^2 < b^2\}$ with $0 < a < b$ be an annular domain which encloses Ω , namely, $\Omega \subset\subset \tilde{\Omega}$ (see Fig.), and we solve the following extended problem in $\tilde{\Omega}$,

$$\begin{aligned} \alpha U - \Delta U &= F, \quad \text{in } \tilde{\Omega}, \\ U &= H, \quad \text{on } \partial\Omega. \end{aligned} \tag{4.2}$$

Denote $\Pi := (a, b) \times [0, 2\pi)$. By applying the following polar transformation

$$u(r, \theta) := U(r \cos \theta, r \sin \theta), \tag{4.3}$$

$$f(r, \theta) := F(r \cos \theta, r \sin \theta), \tag{4.4}$$

$$h(r, \theta) := H(r \cos \theta, r \sin \theta), \tag{4.5}$$

we have the equation for u , i.e.

$$\begin{aligned} \alpha u - \frac{1}{r}(ru_r)_r - \frac{1}{r^2}u_{\theta\theta} &= f, \quad (r, \theta) \in \Pi, \\ u &= h, \quad \text{on } \Gamma, \\ u &\text{ is periodic in } \theta, \end{aligned} \tag{4.6}$$

where Γ is the polar image of $\partial\Omega$. By expanding u and f as

$$u(r, \theta) = \sum_{|m|=0}^{\infty} u^m(r) e^{im\theta}, \tag{4.7}$$

$$f(r, \theta) = \sum_{|m|=0}^{\infty} f^m(r) e^{im\theta}, \tag{4.8}$$

and substituting (4.7) and (4.8) to (4.6), we can obtain a sequence of Bessel-type ordinary differential equations for $u^m(r)$, namely,

$$-\frac{1}{r}\partial_r(r\partial_r u^m) + \left(\frac{m^2}{r^2} + \alpha\right)u^m = f^m(r), \quad a < r < b, \tag{4.9}$$

Same as the idea of circular embedding, we add artificial boundary conditions for each ODE, which are

$$u^m(a) = t_a^m, \quad u^m(b) = t_b^m. \tag{4.10}$$

4.2.2 Approximation

Given a cut-off number $M > 0$ and suppose the approximate solution to (4.6) is built by

$$u_M(r, \theta) = \sum_{|m|=0}^M u^m(r) e^{im\theta}. \quad (4.11)$$

For computing u^m , We employ the following result by virtue of Theorem 3.2.1. Let $\{\phi_a^m\}$ be the solution to

$$\begin{aligned} -\frac{1}{r} \partial_r (r \partial_r \phi_a^m) + \left(\frac{m^2}{r^2} + \alpha\right) \phi_a^m &= 0, \quad a < r < b, \\ \phi_a^m(a) &= 1, \quad \phi_a^m(b) = 0, \end{aligned} \quad (4.12)$$

$\{\phi_b^m\}$ be the solution to

$$\begin{aligned} -\frac{1}{r} \partial_r (r \partial_r \phi_b^m) + \left(\frac{m^2}{r^2} + \alpha\right) \phi_b^m &= 0, \quad a < r < b, \\ \phi_b^m(a) &= 0, \quad \phi_b^m(b) = 1, \end{aligned} \quad (4.13)$$

and $\{\psi^m\}$ be the solution to

$$\begin{aligned} -\frac{1}{r} \partial_r (r \partial_r \psi^m) + \left(\frac{m^2}{r^2} + \alpha\right) \psi^m &= f^m, \quad a < r < b, \\ \psi^m(a) &= 0, \quad \psi^m(b) = 0, \end{aligned} \quad (4.14)$$

then we have

$$u^m(r; t_a^m, t_b^m) = t_a^m \phi_a^m(r) + t_b^m \phi_b^m(r) + \psi^m(r). \quad (4.15)$$

In determining $\{t_a^m, t_b^m\}$, the optimization scheme described in the circular embedding can be utilized. That is, we prescribe K equispaced sampling nodes $\{(\hat{r}_k, \hat{\theta}_k)\}_{k=1}^K$ on Γ , and minimize the discrete L^2 error on these nodes in terms of variables $\{t_a^m, t_b^m\}_{|m|=0}^M$, namely, to solve

$$\min_{t_a^m, t_b^m} \sum_{k=1}^K \left| u_M(\hat{r}_k, \hat{\theta}_k) - h(\hat{r}_k, \hat{\theta}_k) \right|^2, \quad (4.16)$$

or

$$\min_{t_a^m, t_b^m} \sum_{k=1}^K \left| \sum_{|m|=0}^M (t_a^m \phi_a^m(\hat{r}_k) + t_b^m \phi_b^m(\hat{r}_k) + \psi^m(\hat{r}_k)) e^{im\hat{\theta}_k} - h(\hat{r}_k, \hat{\theta}_k) \right|^2, \quad (4.17)$$

4.2.3 Algorithm

To sum up, the spectral method with annular embedding can be performed by the following algorithm.

Algorithm 4.2.1 *Solve the extended problem (4.2)*

1. *compute the truncated Fourier expansion of $f(r, \theta) := F(r \cos \theta, \sin \theta)$ with respect to θ , obtaining (4.8);*
2. *compute $\{\phi_a^m, \phi_b^m, \psi^m\}$ which satisfy the Bessel-type equations (4.12), (4.13) and (4.14);*
3. *solve the optimization (4.17) to obtain $\{t_a^m, t_b^m\}$;*
4. *compute u_M by (3.22) and (4.15);*
5. *the final solution $U_M(x, y) = u_M(\sqrt{x^2 + y^2}, \text{angle}(x, y))$.*

Same as the case of circular embedding, the whole complexity of the algorithm is $O(N^3)$, which is competitive to the spectral methods for standard separable domains.

4.3 Bessel-type equation solver

In this section, we will present the solver for the 1-D Bessel-type equation (4.12), (4.13) and (4.14). Same as the circular case discussed in last chapter, the homogeneous equations (4.12) and (4.13) admit a solution of explicit representation. And for the inhomogeneous equation (4.14), we will first shift it to the interval $(-1, 1)$ and then present a spectral-Galerkin formulation for it.

4.3.1 Homogeneous case

It can be verified the solutions to (4.12) have the following expression,

$$\phi_a^m(r) = \begin{cases} \left\{ \begin{array}{l} \frac{\ln a - \ln r}{\ln a - \ln b}, \quad m = 0 \\ \frac{r^{|m|} b^{-|m|} - b^{|m|} r^{-|m|}}{a^{|m|} b^{-|m|} - b^{|m|} a^{-|m|}}, \quad m \neq 0 \end{array} \right. , \quad \alpha = 0, \\ \frac{K_{-|m|}(\sqrt{\alpha b}) I_{-|m|}(\sqrt{\alpha r}) - I_{-|m|}(\sqrt{\alpha b}) K_{-|m|}(\sqrt{\alpha r})}{K_{-|m|}(\sqrt{\alpha b}) I_{-|m|}(\sqrt{\alpha a}) - I_{-|m|}(\sqrt{\alpha b}) K_{-|m|}(\sqrt{\alpha a})}, \quad \alpha > 0 \end{cases} \quad (4.18)$$

for all m , where $I_{-|m|}(z)$ and $K_{-|m|}(z)$ are the modified Bessel functions of the first and second kind. In practical computation, $\{\phi_a^m\}$ can be evaluated by using series representation, or by solving the Bessel-type equations numerically.

4.3.2 Inhomogeneous case

Spectral-Galerkin formulation

Now let us consider the spectral methods for (4.14). Note $\{\psi^m\}$ satisfy the model problems

$$\begin{aligned} -\frac{1}{r}(ru_r)_r + \left(\frac{m^2}{r^2} + \alpha\right)u &= f, \quad a < r < b, \\ u(a) &= 0 \quad u(b) = 0, \end{aligned} \quad (4.19)$$

Denote $p := b - a$ and $q := a + b$. By introducing the transformation

$$r = \frac{pt + q}{2}, \quad (4.20)$$

and denoting

$$v(t) := u\left(\frac{pt + q}{2}\right), \quad (4.21)$$

(4.19) will be changed to

$$\begin{aligned} -\frac{1}{p^2(pt + q)} [(pt + q)v_t]_t + \left(\frac{m^2}{(pt + q)^2} + \frac{\alpha}{4}\right)v \\ = \frac{1}{4}f\left(\frac{pt + q}{2}\right), \quad t \in I, \quad v(-1) = 0 \quad v(1) = 0, \end{aligned} \quad (4.22)$$

Define the solution space by

$$X^m = \begin{cases} \{v : v(\pm 1) = 0, \int_I (pt + q) (|v|^2 + |v_t|^2) dt < \infty\}, & m = 0, \\ \{v : v(\pm 1) = 0, \int_I \frac{1}{pt+q} |v|^2 + (pt + q) |v_t|^2 dt < \infty\}, & m \neq 0, \end{cases} \quad (4.23)$$

equipped with norms

$$\|v\|_{X^m} = \begin{cases} \left(\int_I (pt + q) (|v|^2 + |v_t|^2) dt \right)^{\frac{1}{2}}, & m = 0, \\ \left(\int_I \frac{1}{pt+q} |v|^2 + (pt + q) |v_t|^2 dt \right)^{\frac{1}{2}}, & m \neq 0. \end{cases} \quad (4.24)$$

Also, define

$$a_m(v, w) = \frac{1}{p^2} ((pt + q)v_t, w_t) + m^2 \left(\frac{1}{pt + q} v, w \right) + \frac{\alpha}{4} ((pt + q)v, w), \quad (4.25)$$

then the weak formulation for (4.22) is

$$\begin{cases} \text{find } v \in X^m \text{ such that} \\ a_m(v, w) = \frac{pt+q}{4} (f(\frac{pt+q}{2}), w), \quad \forall w \in X^m. \end{cases} \quad (4.26)$$

The well-posedness of (4.26) can be shown by the similar argument in last chapter for the weak problem (3.43). For building the Galerkin formulation, we define the finite-dimensional subspace as

$$X_N^m = X^m \cap P_N, \quad (4.27)$$

then Galerkin formulation is

$$\begin{cases} \text{find } v_N \in X_N^m \text{ such that} \\ a_m(v_N, w_N) = \frac{pt+q}{4} (f(\frac{1+t}{2}), w_N), \quad \forall w_N \in X_N^m, \end{cases} \quad (4.28)$$

Remark 4.3.1 *From Céa Lemma, the problem (4.28) admits a unique solution $v_N \in X_N^m$ such that*

$$\|v_N\|_{X^m} \leq \left\| \frac{\sqrt{pt+q}}{2} f\left(\frac{pt+q}{2}\right) \right\|. \quad (4.29)$$

Moreover, if v is the solution to (4.26), then

$$\|v - v_N\|_{X^m} \leq c_\alpha(m) \inf_{w_N \in X_N^m} \|v - w_N\|_{X^m}, \quad (4.30)$$

where $c_\alpha(m)$ is a constant only depending on m .

Remark 4.3.2 *Similar to the circular embedding case, the basis functions for X_N^m can be chosen by $\{L_j - L_{j+2}\}_{j=0}^{N-2} \cup \{(L_j - L_{j+2})i\}_{j=0}^{N-2}$, and the corresponding matrix is symmetric positive definite and has 7 non-zero diagonals, thus the linear system can be solved within $O(N^2)$ complexity by most solvers.*

4.4 Numerical examples

Three examples are presented to show the annular embedding algorithm, which are employed to solve (4.1). In all examples, we set the relation $N = 2M$, and the number of sampling nodes is chosen by $K = 16M$.

In the first one, $\alpha = 10$ and the solution is given by

$$U = \exp(y/(x + 2)). \quad (4.31)$$

The problem domain Ω is non-simply connected and bounded by two smooth curves,

$$\rho_{\text{outer}}(\theta) = 0.9 + 0.1 \sin(4\theta), \quad \rho_{\text{inner}}(\theta) = 0.3 + 0.1 \sin(3\theta). \quad (4.32)$$

We let Ω be embedded into the annulus $\tilde{\Omega} = \{(r, \theta), 0.2 < r < 1\}$, and then apply Algorithm 4.2.1 to the extended problem. The problem domains and sampling nodes on the boundary for $M = 8$ are depicted in Fig. 4.1, and the relative L^2 versus M for $M = 8, 8, \dots, 32$ are plotted in Fig. 3.3. The error decaying exponentially to the machine precision is observed.

In the second example, we let $\alpha = 0$, and the same solution in (4.31) is solved for. Here Ω is bounded by two polygons with the following vertices,

$$\text{outer: } (0, 0.9), (0.8, 0.1), (0.4, -0.8), (-0.5, -0.7), (-0.8, 0.1); \quad (4.33)$$

$$\text{inner: } (0, 0.4), (0.4, 0), (0, -0.4), (-0.4, 4). \quad (4.34)$$

The same annulus as the first example is used for embedding. The resulting figures are shown in Fig. 4.3 and 4.4. It can be seen from the results the error is still decaying exponentially for polygonal domains.

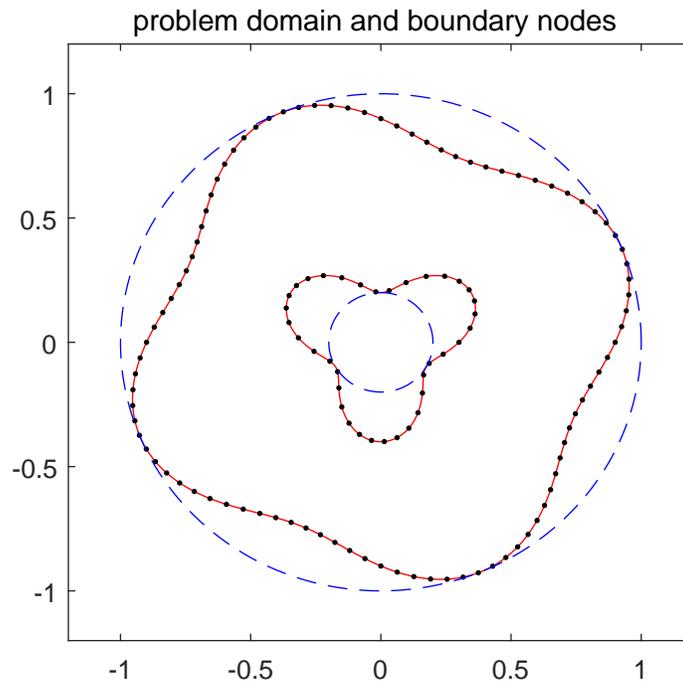


Fig. 4.1.: *the original domain, enclosing domain and sampling nodes for $M = 8$ in the first example*

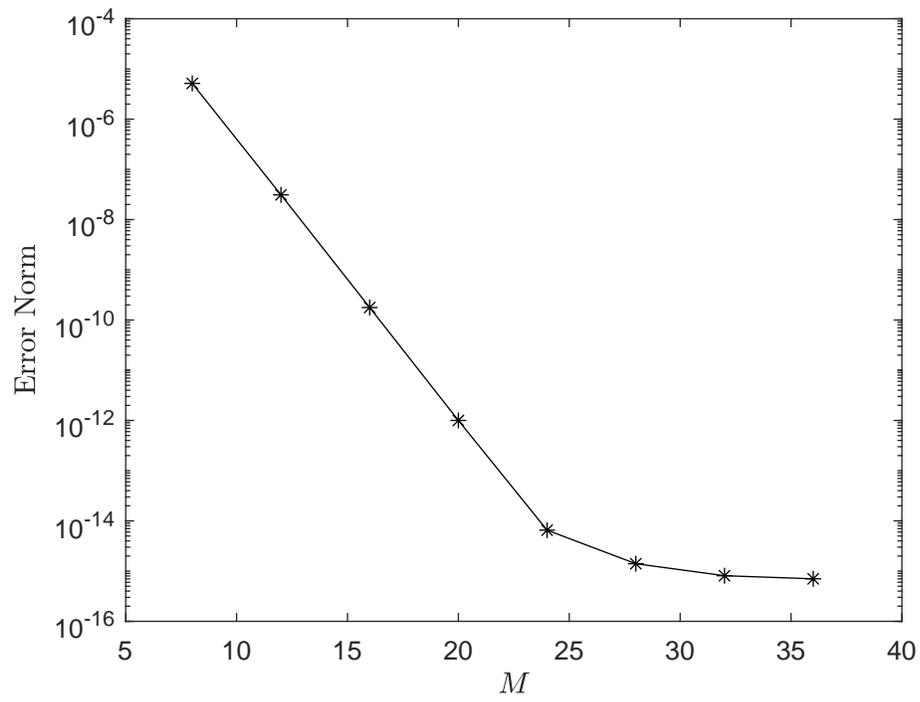


Fig. 4.2.: L^2 error versus M in the first example

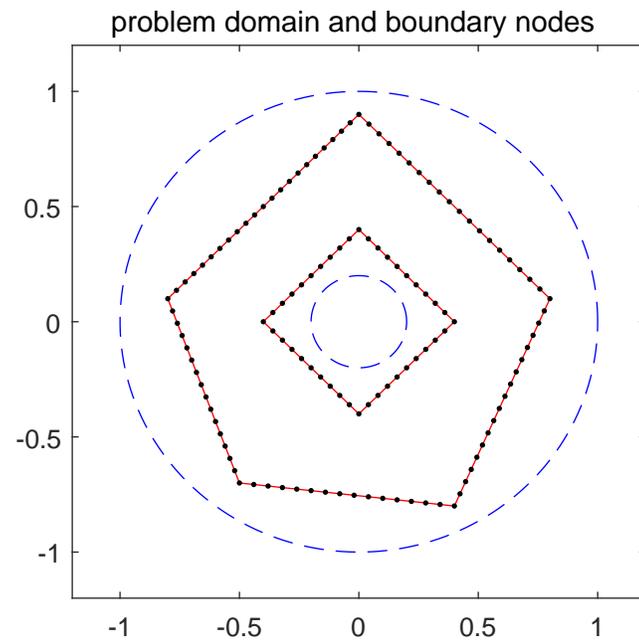


Fig. 4.3.: the original domain, enclosing domain and sampling nodes for $M = 8$ in the second example

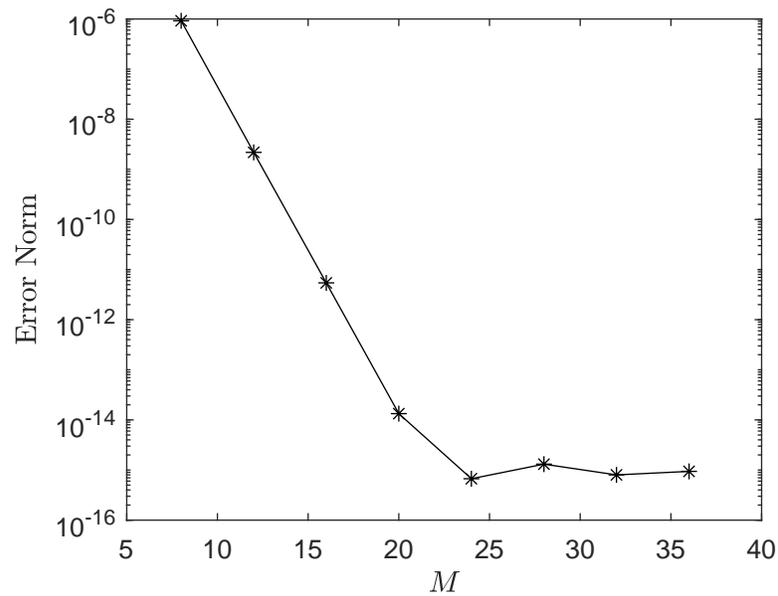


Fig. 4.4.: L^2 error versus M in the second example

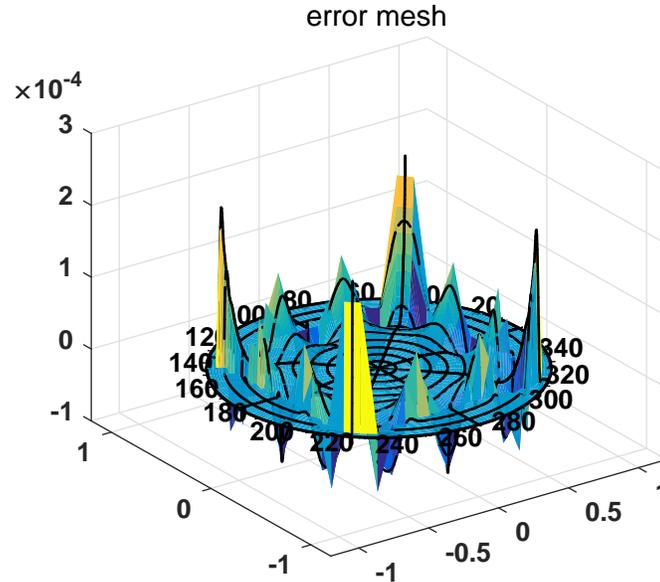


Fig. 4.5.: the error mesh for $M = 12$ in the third example

In the third example, we solve the problem with $\alpha = 0$ and $F = 1$ in the domain bounded by the square having vertices

$$(0.7, 0.7), (0.7, -0.7), (-0.7, -0.7), (-0.7, 0.7), \quad (4.35)$$

and the curve

$$\rho(\theta) = 0.3 + 0.05 \sin(5\theta). \quad (4.36)$$

The solution is expressed by (3.125), which has four singularities at the four outer corners. We use the same annulus to enclose the domain as the first example. The resulting figures are shown in Fig. 4.5, 4.6 and 4.7. We can see from the figures the error is mainly distributed near the singularity for the singular problem. Also, it is observed the error decays by the order between 4th and 5th.

4.5 Conclusion

In this chapter, we extend the spectral method of circular embedding discussed in last chapter for the Poisson-type problem in non-simply connected domains. The

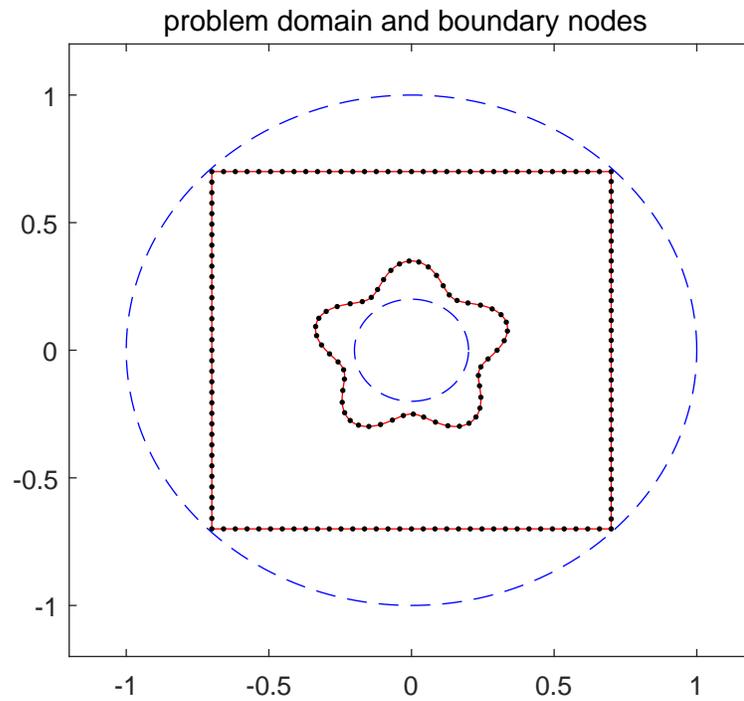


Fig. 4.6.: the original domain, enclosing domain and sampling nodes for $M = 12$ in the third example

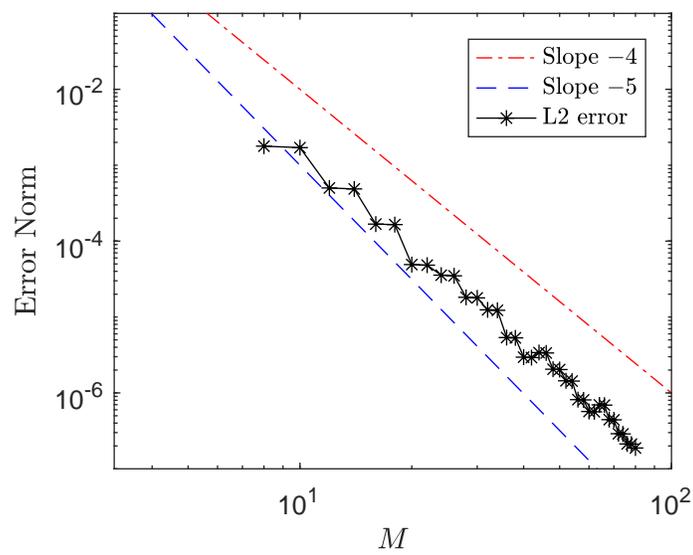


Fig. 4.7.: L^2 error versus M in the third example

corresponding spectral method with annular embedding works as well as the circular one. In numerical examples, we obtain exponential error decay for analytic solutions and algebraic error decay of high orders for singular problems.

5. APPLICATION I: FRACTIONAL LAPLACIAN EQUATION

5.1 Introduction

In this chapter, one application of the spectral methods with fictitious domain embedding will be introduced. We will consider the fractional Laplacian equation, which can be solved by two approaches, Caffarelli-Silvestre extension or Dunford-Taylor representation. Both approaches find the solutions by solving a sequence of general Poisson-type equations $-\Delta u + \alpha u = f$ with Dirichlet boundary conditions. Our method is especially effective for fractional Laplacian problems in complex domains.

The basic problem will be described in Section 5.2. The approaches by Caffarelli-Silvestre extension and Dunford-Taylor representation will be presented separately in Section 5.3 and 5.4, which both include related numerical examples. The conclusion is given in Section 5.5.

5.2 Fractional Laplacian Problem

In recent decades, topics in non-local partial differential equations have received a great attention. The fractional Laplacian equation is one of the most basic and important type of the non-local diffusion and have been widely studied so far [44–46]. Specifically, given a function $v : \mathbb{R}^d \rightarrow \mathbb{R}$, the operator with power $0 < s < 1$ can be defined by using Fourier transform, namely,

$$\mathcal{F}[(-\Delta)^s](\omega) = |\omega|^{2s} \mathcal{F}[v](\omega), \quad (5.1)$$

and this can be also formulated equivalently by the following integral form

$$(-\Delta)^s v(x) = c_{d,s} \text{ p.v. } \int_{\mathbb{R}^d} \frac{v(x) - v(\tau)}{|x - \tau|^{d+2s}} d\tau, \quad (5.2)$$

where p.v. stands for the Cauchy principal value, and $c_{d,s}$ is a constant defined by

$$c_{d,s} = \frac{2^{2s} s \Gamma(\frac{2s+d}{2})}{\pi^{d/2} (1-s)}. \quad (5.3)$$

Note in (5.2) the evaluation of $(-\Delta)^s v(x)$ is non-local, that it depends the information in the whole \mathbb{R}^d instead of the neighbourhood of x .

Now let us consider the fractional Laplacian operator with respect to the homogeneous Dirichlet condition in a smooth domain $\Omega \subset \mathbb{R}^d$. First, note the following Dirichlet problem is well-defined,

$$\begin{aligned} -\Delta v &= f, & \text{in } \Omega, \\ v &= 0, & \text{on } \partial\Omega. \end{aligned} \quad (5.4)$$

Furthermore, the inverse operator $-\Delta : \mathcal{D}(-\Delta) = H_0^1(\Omega) \cap H^2(\Omega) \subset L^2(\Omega) \rightarrow L^2(\Omega)$ is an unbounded, positive and closed operator with compact inverse, it has countably many eigenpairs $\{\mu_k, \varphi_k\}_{k \in \mathbb{N}^+} \subset \mathbb{R}^+ \times H_0^1(\Omega)$ such that $\{\varphi_k\}_{k \in \mathbb{N}^+}$ forms an orthonormal basis for $H_0^1(\Omega)$. Under the preceding setting, the normed space $\mathbb{H}^s(\Omega)$ is defined by

$$\mathbb{H}^s(\Omega) = \left\{ v = \sum_{k=1}^{\infty} c_k \varphi_k : \|v\|_{\mathbb{H}^s(\Omega)}^2 = \sum_{k=1}^{\infty} \lambda_k^s c_k^2 < \infty \right\}, \quad (5.5)$$

for $-1 \leq s \leq 1$. Now the Laplacian operator with Dirichlet condition of power s can be defined by

$$(-\Delta)^s v := \sum_{k=1}^{\infty} \mu_k^s \tilde{v}_k \varphi_k, \quad \tilde{v}_k := \int_{\Omega} v \varphi_k dx, \quad (5.6)$$

for all $v \in C_0^\infty(\Omega)$. The domain $\mathcal{D}((-\Delta)^s)$ can be extended to $\mathbb{H}^s(\Omega)$ by density theorem.

Based on the definition (5.6), let us consider the following problem

$$\begin{aligned} (-\Delta)^s u &= f, & \text{in } \Omega, \\ u &= 0, & \text{on } \partial\Omega. \end{aligned} \quad (5.7)$$

where Ω is a smooth and simply-connected d -dimensional domain. Due to the non-local property of the fractional Laplacian operator, the classical discretization strategy can not work efficiently. Hence, it is desirable to transform the original fractional problems to non-fractional ones by specific manipulations.

5.3 Caffarelli-Silvestre extension

5.3.1 Extended problem

L. Caffarelli and L. Silvestre showed the fractional power of the Laplacian operator in \mathbb{R}^d can be determined as an operator that maps a Dirichlet boundary condition to a Neumann-type condition via an extension problem on the upper half-space $\mathbb{R}^d \times [0, \infty)$ (see [44]). For bounded domains, the extension is further adapted in [47–49]. In detail, let $U(x, y) : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ be the solution of the following $(d+1)$ -dimensional problem,

$$\begin{aligned} \nabla \cdot (y^\alpha \nabla U(x, y)) &= 0, \quad \text{in } \mathcal{D} := \Omega \times (0, \infty), \\ U &= 0, \quad \text{on } \partial_L \mathcal{D} := \partial\Omega \times [0, \infty), \\ \lim_{y \rightarrow 0} y^\alpha U_y(x, y) &= -d_s f(x), \\ \lim_{y \rightarrow \infty} U(x, y) &= 0, \end{aligned} \tag{5.8}$$

with $\alpha := 1 - 2s$ and $d_s := 2^{1-2s}\Gamma(1-s)/\Gamma(s)$, then the solution to (5.7) is given by

$$u(x) = U(x, 0). \tag{5.9}$$

Before deducing the weak formulation for (5.8), we introduce the following space

$$H_{y^\alpha}^1(\mathcal{D}) := \{\nabla u \in L_{y^\alpha}^2(\mathcal{D}) : \lim_{y \rightarrow \infty} y^\alpha u(x, y) = 0, u = 0 \text{ on } \partial_L \mathcal{D}\}, \tag{5.10}$$

with norm

$$\|u\|_{H_{y^\alpha}^1(\mathcal{D})} = \|\nabla u\|_{L_{y^\alpha}^2(\mathcal{D})}, \tag{5.11}$$

where $L_{y^\alpha}^2(\mathcal{D})$ is the L^2 space in \mathcal{D} with weight function y^α .

Now the weak formulation of (5.8) is given by

$$\begin{cases} \text{find } U \in H_{y^\alpha}^1(\mathcal{D}) \text{ such that} \\ (y^\alpha \nabla U, \nabla V)_{\mathcal{D}} = d_s(f, V(x, 0))_{\Omega}, \quad \forall V \in H_{y^\alpha}^1(\mathcal{D}). \end{cases} \quad (5.12)$$

The weak problem (5.12) can be proven to be well-posed with the help of following result

Lemma 5.3.1 ([50], **Proposition 2.5**) *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain, then $\{u(x, 0) : u \in H_{y^\alpha}^1(\mathcal{D})\} = \mathbb{H}^s(\Omega)$ and*

$$\|u(x, 0)\|_{\mathbb{H}^s(\Omega)} \leq c \|u\|_{H_{y^\alpha}^1(\mathcal{D})}, \quad \forall u \in H_{y^\alpha}^1(\mathcal{D}). \quad (5.13)$$

The Hilbert space \mathbb{H}^s is defined in

5.3.2 Galerkin formulation

Let X_M be a finite-dimensional space approximate to $H_0^1(\Omega)$, and

$$X_M = \text{span}\{\phi_m^x\}_{m=1}^M, \quad (5.14)$$

where ϕ_m^x is a set of basis functions. For example, if $\Omega = (-1, 1)^d$, then it can be chosen that

$$X_M = (P_M^0)^d, \quad (5.15)$$

where P_M^0 consists of all polynomials in $(-1, 1)$ that vanish at the ending points ± 1 . For spectral methods, the basis ϕ_m^x are often chosen by the variants of orthogonal polynomials.

Besides, let Y_N be a finite-dimensional space approximate to $H_{y^\alpha}^1(0, \infty) := \{v(y) : v' \in L_{y^\alpha}^2(0, \infty), \lim_{y \rightarrow \infty} y^\alpha v = 0\}$, and $\{\phi_n^y\}_{n=1}^N$ be a set of basis of Y_N . Then the tensor-product space $X_M \times Y_N$ is a space approximate to $H_{y^\alpha}^1(\mathcal{D})$. Furthermore, we can employ the generalized Laguerre functions ([1, 51]) to build the basis for Y_N . Specifically, define

$$\widehat{\mathcal{L}}_n^\alpha(y) := e^{-\frac{y}{2}} \mathcal{L}_n^\alpha(y), \quad (5.16)$$

where $\mathcal{L}_n^\alpha(y)$ is the generalized Laguerre polynomial. Then $\{\mathcal{L}_n^\alpha(y)\}_{n=1}^N$ forms a basis for Y_N . Generally, we denote $Y_N = \text{span}\{\phi_n^y\}_{n=1}^N$.

Under the preceding settings, the Galerkin formulation is given by

$$\begin{cases} \text{find } U_{MN} \in X_M \times Y_N \text{ such that} \\ (y^\alpha \nabla U_{MN}, \nabla V_{MN})_{\mathcal{D}} = d_s(f, V_{MN}(x, 0))_{\Omega}, \quad \forall V_{MN} \in X_M \times Y_N. \end{cases} \quad (5.17)$$

In (5.17), by expanding the solution by

$$u_{MN} = \sum_{m=1}^M \sum_{n=1}^N \tilde{u}_{mn} \phi_m^x(x) \phi_n^y(y) \quad (5.18)$$

and taking $V_{MN} = \phi_i^x(x) \phi_j^y(y)$ for $i = 1, \dots, M$ and $j = 1, \dots, N$, we immediately obtain the following linear system of matrices,

$$\mathbf{S}^x \tilde{\mathbf{U}} \mathbf{M}^y + \mathbf{M}^x \mathbf{U} \mathbf{S}^y = \mathbf{F}, \quad (5.19)$$

where

$$\begin{aligned} \mathbf{S}^x &= [(\nabla_x \phi_i^x, \nabla_x \phi_m^x)_{\Omega}]_{M \times M}, & \mathbf{M}^x &= [(\phi_i^x, \phi_m^x)_{\Omega}]_{M \times M}, \\ \mathbf{S}^y &= [(\nabla_y \phi_j^y, \nabla_y \phi_n^y)_{\Omega}]_{N \times N}, & \mathbf{M}^y &= [(\phi_j^y, \phi_n^y)_{\Omega}]_{N \times N}, \\ \tilde{\mathbf{U}} &= [\tilde{u}_{mn}]_{M \times N}, & \mathbf{F} &= d_s[\phi_n^y(0)(f, \phi_m^x)_{\Omega}]_{M \times N} \end{aligned} \quad (5.20)$$

Thanks to the orthogonality of the generalized Laguerre functions, the matrices with respect to y are sparse and symmetric positive definite. In detail, it satisfies

$$\mathbf{S}_{n,j}^y = \frac{1}{4} \begin{cases} \gamma_{n-2}^{(\alpha)}, & j = n-1, \\ \gamma_{n-2}^{(\alpha)} + \gamma_{n-1}^{(\alpha)}, & j = n, \\ \gamma_{n-1}^{(\alpha)}, & j = n+1, \\ 0 & \text{others,} \end{cases} \quad \mathbf{M}_{n,j}^y = \frac{1}{4} \begin{cases} -\gamma_{n-2}^{(\alpha)}, & j = n-1, \\ \gamma_{n-2}^{(\alpha)} + \gamma_{n-1}^{(\alpha)}, & j = n, \\ -\gamma_{n-1}^{(\alpha)}, & j = n+1, \\ 0 & \text{others,} \end{cases} \quad (5.21)$$

where $\gamma_{-1}^{(\alpha)} = 0$ and $\gamma_k^{(\alpha)} = \Gamma(k + \alpha + 1)/\Gamma(k + 1)$.

Note (5.19) can be solve by performing eigen-decomposition on the sparse matrices \mathbf{M}^y and \mathbf{S}^y . Suppose we have

$$\mathbf{M}^y \mathbf{E} = \mathbf{S}^y \mathbf{E} \mathbf{\Lambda}, \quad (5.22)$$

where $\mathbf{E} := [\mathbf{e}_1, \dots, \mathbf{e}_N]$ with B_i being the generalized eigenvector with respect to \mathbf{M}^y and \mathbf{S}^y , and $\mathbf{\Lambda} := \text{diag}(\lambda_1, \dots, \lambda_N)$ with λ_i being a corresponding eigenvalue for $i = 1, \dots, N$. Hence, by letting $\tilde{\mathbf{U}} = \tilde{\mathbf{V}}\mathbf{E}^T$, (5.19) can be rewritten by

$$\mathbf{S}^x \tilde{\mathbf{V}} \mathbf{\Lambda} \mathbf{E}^T \mathbf{S}^y + \mathbf{M}^x \tilde{\mathbf{V}} \mathbf{E}^T \mathbf{S}^y = \mathbf{F}, \quad (5.23)$$

or

$$\mathbf{S}^x \tilde{\mathbf{V}} \mathbf{\Lambda} + \mathbf{M}^x \tilde{\mathbf{V}} = \mathbf{G} := \mathbf{F}(\mathbf{S}^y)^{-1} \mathbf{E}. \quad (5.24)$$

Denote $\tilde{\mathbf{V}} := [\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_N]$ and $\mathbf{G} := [\mathbf{g}_1, \dots, \mathbf{g}_N]$, then (5.24) can be separated by a sequence of linear systems,

$$(\lambda_i \mathbf{S}^x + \mathbf{M}^x) \tilde{\mathbf{v}}_i = \mathbf{g}_i \quad (5.25)$$

for $i = 1, \dots, N$.

Actually, it is not necessary to solve for $\tilde{\mathbf{v}}_i$ through (5.25). Especially, if Ω is a domain of complex geometry, then we can hardly find the closed form for ϕ^x , hence the direct computation of \mathbf{S}^x and \mathbf{M}^x is infeasible. Instead, we consider the following Poisson-type problem

$$\begin{aligned} -\Delta v_i + \frac{1}{\lambda_i} v_i &= \frac{d_s}{\lambda_i} \phi^y(0)^T \mathbf{t}_k f(x), \quad \text{in } \Omega, \\ v_i &= 0, \quad \text{on } \partial\Omega, \end{aligned} \quad (5.26)$$

where $\lambda_i > 0$, $\phi^y(y) := [\phi_1^y(y), \dots, \phi_M^y(y)]^T$ and \mathbf{t}_i is the i -th column of $\mathbf{T} := (\mathbf{S}^y)^{-1} \mathbf{E}$.

Note the variational form of (5.26) can be described by

$$\begin{cases} \text{find } v_i \in H_0^1(\Omega) \text{ such that} \\ \lambda_i (\nabla v_i, \nabla w)_\Omega + (v_i, w)_\Omega = d_s \phi^y(0)^T \mathbf{t}_k (f, w)_\Omega \quad \forall w \in H_0^1(\Omega). \end{cases} \quad (5.27)$$

Also, the Galerkin formulation of (5.27) corresponding to X_M is given by

$$\begin{cases} \text{find } v_{i,M} \in X_M = \text{span}\{\phi_m^x\}_{m=1}^M \text{ such that} \\ \lambda_i (\nabla v_{i,M}, \nabla w_M)_\Omega + (v_{i,M}, w_M)_\Omega = d_s \phi^y(0)^T \mathbf{t}_k (f, w_M)_\Omega \quad \forall w \in X_M(\Omega). \end{cases} \quad (5.28)$$

Therefore, by letting $v_{i,M} = \sum_{m=1}^M \tilde{v}_{i,m} \phi_m^x$ and taking $w_M = \phi_m^x$ in (5.28), as well as denoting $\tilde{\mathbf{v}}_i = [\tilde{v}_{i,1}, \dots, \tilde{v}_{i,M}]^T$, the derived linear system is exactly (5.25).

Furthermore, we should be attentive that our final objective is to evaluate $U_{MN}(x, 0)$ at any given sampling point $\{x_p\}$, namely, computing

$$U_{MN}(x_p, 0) = \phi_x(x_p)^T \tilde{\mathbf{U}} \phi^y(0) = \phi_x(x_p)^T \tilde{\mathbf{V}} \mathbf{E}^T \phi^y(0) = \mathbf{v}_M(x_p)^T \mathbf{E}^T \phi^y(0), \quad (5.29)$$

where $\mathbf{v}_M(x) := [v_{1,M}(x), \dots, v_{N,M}(x)]^T$ are formed by the solutions of (5.28). Note $v_{i,M}(x)$ is an approximate solution of the continuous problem (5.26), hence it is implied from (5.29) that the computing $U_{MN}(x, 0)$ approximately at sampling points only requires us to solve (5.26) (by any feasible finite-dimensional approximations) rather than employing the Galerkin formulation (5.28) exclusively. For example, the spectral methods with fictitious domain techniques presented in previous chapters are especially effective for the Poisson-type problems in complex Ω . Suppose we have solved (5.26) for an approximate solution $\hat{v}_{i,M}$ for each i , then

$$U_{MN}(x_p, 0) \approx \hat{\mathbf{v}}_M(x_p)^T \mathbf{E}^T \phi^y(0), \quad (5.30)$$

where $\hat{\mathbf{v}}_M(x) := [\hat{v}_{1,M}(x), \dots, \hat{v}_{N,M}(x)]^T$.

Finally, the whole algorithm can be summarized by

Algorithm 5.3.1 *Given a set of sampling points $\{x_p\}$, compute $u(x_p)$ where u is the solution of the fractional Laplacian problem (5.7),*

1. *compute the generalized eigenvalue problem (5.22) by referring to the matrix information given in (5.21), obtaining $\mathbf{\Lambda}$ and \mathbf{E} ;*
2. *compute the right hand side of the Poisson-type problem (5.26) for $i = 1, \dots, N$;*
3. *solve (5.26) at $\{x_p\}$ by any effective methods for $i = 1, \dots, N$;*
4. *evaluate $u(x_p) \approx U_{MN}(x_p, 0)$ which is approximated by (5.30).*

5.3.3 Numerical examples

One model problem of fractional Laplacian equation (5.7) is solved by the preceding algorithm to demonstrate the effectiveness, in which we choose Ω to be the square $(-T, T) \times (-T, T)$ with $T = 0.7$, and take

$$f(x, t) = \left(\frac{2\pi^2}{T^2}\right)^s \sin\left(\frac{\pi x}{T}\right) \sin\left(\frac{\pi y}{T}\right). \quad (5.31)$$

The exact solution is given by

$$u(x, y) = \sin\left(\frac{\pi x}{T}\right) \sin\left(\frac{\pi y}{T}\right). \quad (5.32)$$

For solving the Poisson-type problem (5.26), although the general PDE solvers for 2-D square domains can be directly utilized, we employ the spectral methods with circular embedding discussed in previous chapters to assess the effects of Algorithm 5.3.1 on the fractional Laplacian problems in complex domains. The unit circle is introduced to enclose Ω , and we discretize the problem with degree of freedom M on both r -direction and θ -direction. The number of sampling nodes on $\partial\Omega$ for boundary constraints is chosen by $K = 3M$. Also, we set $N = M$ be the degree of freedom in the eigen-decomposition. The numerical example is performed for $N = 10, 20, \dots, 100$ and for $s = 0.3, 0.4, \dots, 0.8$. The L^2 error versus N for each s are shown in Fig. 5.1. From the figure, it is observed the convergence rate for $s = 0.5$ is much higher than other values of s , which attains a decay of exponential rate. The error decays much slowly as s goes far away from 0.5. This is because of the singularity in y direction introduced by Caffarelli-Silvestre extension.

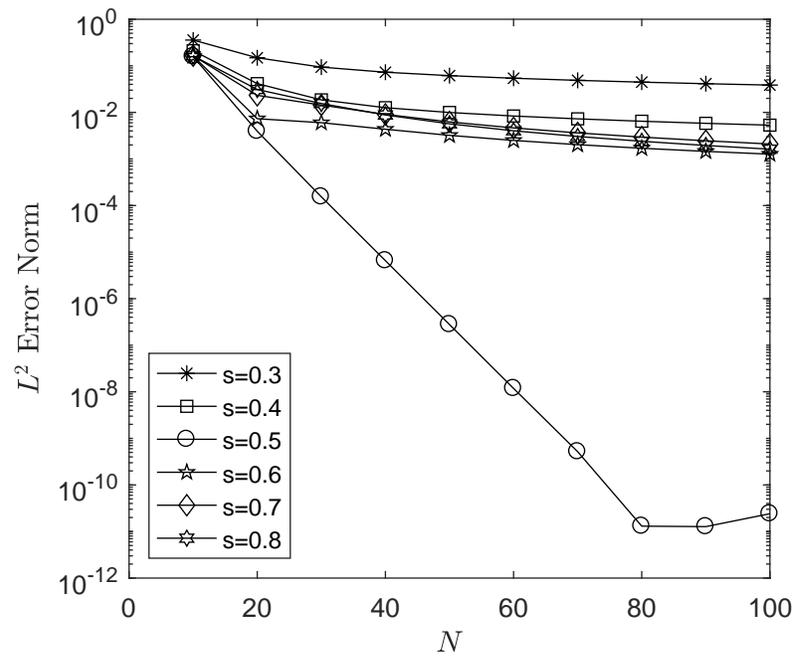


Fig. 5.1.: $\|u - u_{MN}\|_{L^2(\Omega)}$ versus N for the example solved by Caffarelli-Silvestre extension

5.4 Dunford-Taylor representation

5.4.1 Representation by integrals

Another way to transform the fractional Laplacian operator with homogeneous Dirichlet condition to variants of the general Laplacian is by using Dunford-Taylor integrals [52, 53]. For $s \in (0, 1)$ and $f \in \mathbb{H}^{-s}$ we have

$$(-\Delta)^{-s} f = \frac{1}{2\pi i} \int_{\mathcal{C}} z^{-s} (z + \Delta)^{-1} f dz, \quad (5.33)$$

where \mathcal{C} is a Jordan curve oriented to have the spectrum of $-\Delta$ to its right. The formula (5.33) can be simplified to an integral in \mathbb{R}^+ , namely the so-called Balakrishnan formula,

$$(-\Delta)^{-s} f = \frac{\sin(s\pi)}{\pi} \int_0^\infty \mu^{-s} (\mu - \Delta)^{-1} f d\mu. \quad (5.34)$$

Therefore, if the integral (5.34) is approximated by numerical quadrature, then the evaluation of $u = (-\Delta)^{-s} f$ only relies on computing a sequence of Poisson-type problems. In [54], the sinc quadrature is employed and proven to have exponential error decay. Specifically, sinc quadrature technique is performed by introducing change of variable $\mu = e^y$ in (5.34) so that

$$(-\Delta)^{-s} f = \frac{\sin(s\pi)}{\pi} \int_{-\infty}^\infty e^{(1-s)y} (e^y - \Delta)^{-1} f dy. \quad (5.35)$$

Given the step length $h > 0$, set

$$N_+ := \lceil \frac{\pi^2}{4sh^2} \rceil, \quad N_- := \lceil \frac{\pi^2}{4(1-s)h^2} \rceil, \quad (5.36)$$

and $y_i := ih$. Then the integral is approximated by the sum

$$(-\Delta)^{-s} f \approx \frac{\sin(s\pi)}{\pi} h \sum_{i=-N_-}^{N_+} e^{(1-s)y_i} (e^{y_i} - \Delta)^{-1} f. \quad (5.37)$$

The term $(e^{y_i} - \Delta)^{-1} f$ in (5.37) can be evaluated by solving the Poisson-type problem

$$\begin{aligned} -\Delta v_i + e^{y_i} v_i &= f, & \text{in } \Omega, \\ v_i &= 0, & \text{on } \partial\Omega, \end{aligned} \quad (5.38)$$

Same as in Sec. 5.3, the spectral methods with fictitious domain techniques can be directly utilized when Ω is of complex geometry. Denote $v_{i,M}$ by the approximate solution of (5.38) solved by numerical methods with degree of freedom M , then the final solution u can be approximated by

$$u = (-\Delta)^{-s} f \approx u_{MN} := \frac{h \sin(s\pi)}{\pi} \sum_{i=-N_-}^{N_+} e^{(1-s)y_i} v_{i,M}. \quad (5.39)$$

5.4.2 Numerical examples

The same model problem as in Sec. 5.3.3 is solved by using the representation (5.34). First, we employ the sinc quadrature with choosing the step length h such that the number of quadrature nodes $N := N_+ + N_- = 4M$. The L^2 error versus N for different s are shown in Fig. 5.3, from which we can see the errors all decay exponentially for all s . This implies the robustness of sinc quadrature on Dunford-Taylor integrals.

On the other hand, we attempt to evaluate the integral (5.34) by global adaptive quadrature with given tolerance. In this example, the tolerance is set by $\varepsilon = 10^{-N/8}$ for each N . The L^2 error decay for different s is shown in Fig. 5.3 and the convergence rate is also observed to be exponential. Moreover, the error from $s = 0.5$ decays much faster than s of other values, which attains to the machine precision by $N = 50$. By comparison, it is also noticed that even for the cases of $s \neq 0.5$, the error obtained by global adaptive quadrature reaches to lower levels than that of using sinc quadrature for same N .

5.5 Conclusion

Totally two approaches to solve the fractional Laplacian problem with homogeneous Dirichlet condition in a complex domain have been presented.

The first is to extend the original d -dimensional fractional problem to a $(d+1)$ -dimensional non-fractional problem. In practice, we do not have to solve the extended

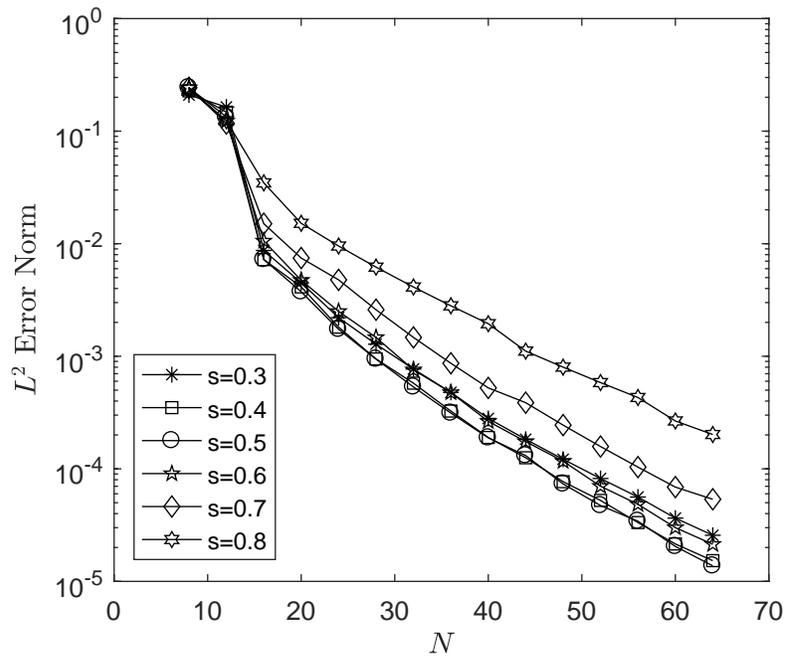


Fig. 5.2.: $\|u - u_{MN}\|_{L^2(\Omega)}$ versus N for the example solved by Dunford-Taylor representation using sinc quadrature

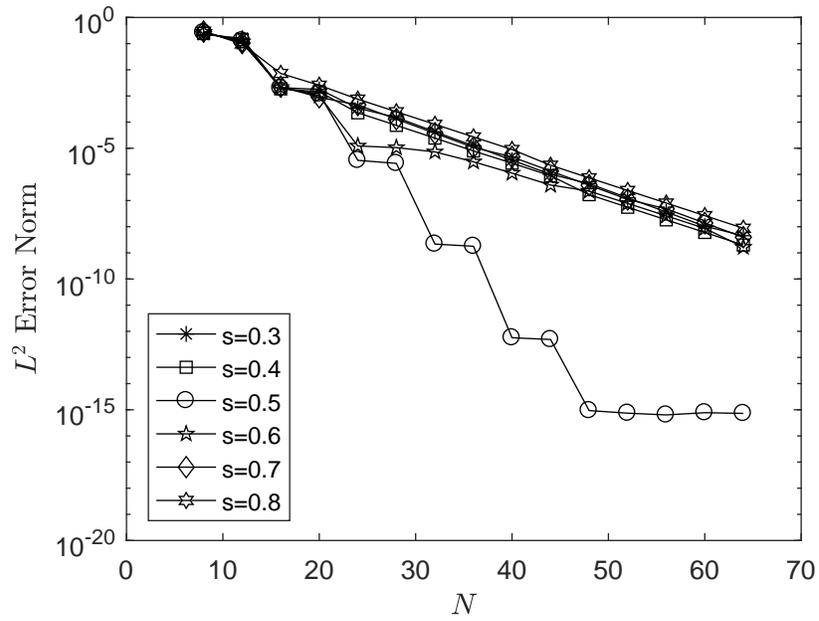


Fig. 5.3.: $\|u - u_{MN}\|_{L^2(\Omega)}$ versus N for the example solved by Dunford-Taylor representation using global adaptive quadrature

problem of high dimension since we only need to evaluate the solution at the initial slice $y = 0$. Instead, we employ the eigen-decomposition technique and turn to solve a sequence of Poisson-type problems, which can be simply dealt with by the spectral methods with fictitious domain embedding. However, the numerical result shows the exponential convergence rate can only be obtained when $s = 0.5$, because of the singularity in y direction.

The second is to rewrite the fractional solution as an one-dimensional integral of the solution of Poisson-type problems. Hence, by utilizing effective numerical quadrature, the solution can be computed straightforwardly. We only need to solve one such problem for each quadrature node. The sinc quadrature and global adaptive quadrature are both tested in numerical examples, and the convergence rates of both are observed to be exponential, even the latter appears faster.

6. APPLICATION II: HELMHOLTZ EQUATION

6.1 Introduction

As we know, time harmonic wave propagations appear in many applications such as wave scattering and transmission, noise reduction, fluid-solid interaction, etc. In this chapter, the Helmholtz equation arising from acoustic scattering is introduced and the specific numerical approach is presented. We start with giving a bounded obstacle in space that is described by the region $\Omega_1 \subset \mathbb{R}^2$. And the problem is considered in the exterior domain $\Omega_{\text{ext}} := \mathbb{R}^2 \setminus \Omega_1$.

However, the unboundedness of Ω_{ext} prevents us from using most of discretization techniques. One classical approach to reduce the problem from a unbounded domain to a bounded one is to use the Dirichlet-to-Neumann map. We first introduce a large domain D_1 (usually of simple geometry) which encloses Ω_1 , and reduce the problem in $\Omega := D_1 \cap (\mathbb{R}^2 \setminus \Omega_1)$. A boundary condition corresponding to the DtN map is posed on ∂D_1 . For convenience, D_1 is chosen by an open disk with radius $b > 0$, i.e. $D_1 : \{(r, \rho) : r \leq b\}$. It allows us to apply the spectral method for bounded domains to the newly reduced problem.

6.2 Dimension reduction

Let us consider the following Helmholtz

$$\begin{aligned} -\Delta u - k^2 u &= f && \text{in } \Omega, \\ u &= g && \text{on } \partial\Omega_1, \\ \frac{\partial u}{\partial r} + T(u) &= 0 && \text{on } \partial D_1, \end{aligned} \tag{6.1}$$

where k is the wave number and T is the DtN map. Note the problem domain Ω is exactly formed by removing an interior region from a larger disk, hence it is natural

to employ the spectral method with annular embedding scheme discussed in previous chapters. Specifically, let $D_2 := \{(r, \rho) : r \leq a\}$ be a small disk such that $D_2 \subset \Omega_1$. we extend f smoothly from Ω to $\tilde{\Omega} := D_1 \cap (\mathbb{R}^2 \setminus D_2)$ (still denoted by f), and then solve

$$\begin{aligned} -\Delta u - k^2 u &= f \quad \text{in } \tilde{\Omega}, \\ u &= g \quad \text{on } \partial\Omega_1, \\ \frac{\partial u}{\partial r} + T(u) &= 0 \quad \text{on } \partial D_1. \end{aligned} \tag{6.2}$$

Similar to the case of Poisson problems, we apply dimension reduction on the solution by expanding it by Fourier basis, that is, we expand

$$f = \sum_{m=-\infty}^{\infty} f^m(r) e^{im\theta} \tag{6.3}$$

and look for the solution u having the following form

$$u = \sum_{m=-\infty}^{\infty} u^m(r) e^{im\theta}. \tag{6.4}$$

By taking (6.3) and (6.4) in (6.2), it can derived the ordinary differential equations for u^m , which are

$$\begin{aligned} -\frac{1}{r} \partial_r (r \partial_r u^m) + \frac{m^2}{r^2} u^m - k^2 u^m &= f^m, \quad r \in (a, b), \\ \partial_r u^m(b) - k D_{m,k} u^m(b) &= 0, \end{aligned} \tag{6.5}$$

for all m . In (6.5), $D_{m,k} := \frac{\partial_z H_m^{(1)}(kb)}{H_m^{(1)}(kb)}$ where $H_m^{(1)}$ is the Hankel function of the first kind of order m .

To make the 1-D problems (6.5) to be well-defined, we add an artificial boundary condition for each m at the left ending point $r = a$, namely,

$$u^m(a) = t^m. \tag{6.6}$$

The values of $\{t^m\}$ are exactly that we need to evaluate subject to the original interior boundary condition $u = g$ on $\partial\Omega_1$.

To figure out the relation between u^m and $\{t^m\}$, we need the following result.

Lemma 6.2.1 *Let ϕ^m be the solution to*

$$\begin{aligned} -\frac{1}{r}\partial_r(r\partial_r\phi^m) + \frac{m^2}{r^2}\phi^m - k^2\phi^m &= 0, \quad r \in (a, b), \\ \phi^m(a) = 1, \quad \partial_r\phi^m(b) - kD_{m,k}\phi^m(b) &= 0, \end{aligned} \quad (6.7)$$

and ψ^m be the solution to

$$\begin{aligned} -\frac{1}{r}\partial_r(r\partial_r\psi^m) + \frac{m^2}{r^2}\psi^m - k^2\psi^m &= f^m, \quad r \in (a, b), \\ \psi^m(a) = 0, \quad \partial_r\psi^m(b) - kD_{m,k}\psi^m(b) &= 0. \end{aligned} \quad (6.8)$$

Then the solution to (6.5) satisfies

$$u^m = t^m\phi^m + \psi^m. \quad (6.9)$$

6.3 Approximation

By virtue of Lemma 6.2.1, we can construct an approximate solution u_{MN} as follows

$$u_{MN}(r, \theta) = \sum_{m=-M}^M u_N^m(r) e^{im\theta}, \quad (6.10)$$

where

$$u_N^m = t^m\phi_N^m + \psi_N^m, \quad (6.11)$$

and ϕ_N^m and ψ_N^m are solved from (6.7) and (6.8) by appropriate ODE schemes with degree of freedom N . An effective spectral-Galerkin method for (6.7) and (6.8) is put forward in [1, Chap. 9.1].

We discretize the original interior boundary condition $u = g$ on $\partial\Omega_1$ in some sense to help to determine $\{t^m\}_{m=-M}^M$. One straightforward way is to prescribe K collocation nodes $\{(\hat{r}_k, \hat{\theta}_k)\}_{k=1}^K$ on $\partial\Omega_1$, and enforce

$$u_{MN}(\hat{r}_k, \hat{\theta}_k) \approx g(\hat{\theta}_k), \quad k = 1, \dots, K, \quad (6.12)$$

or

$$\sum_{m=-M}^M (t^m\phi_N^m(\hat{r}_k) + \psi_N^m(\hat{r}_k)) e^{im\hat{\theta}_k} \approx g(\hat{\theta}_k), \quad k = 1, \dots, K. \quad (6.13)$$

For $K > M$, (6.13) corresponds to the following least square problem,

$$\min_{\hat{\theta}_k} \sum_{k=1}^K \left| \sum_{m=-M}^M (t^m \phi_N^m(\hat{r}_k) + \psi_N^m(\hat{r}_k)) e^{im\hat{\theta}_k} - g(\hat{\theta}_k) \right|^2, \quad (6.14)$$

In practical implementation, we find choosing K between $4M$ and $8M$ can keep the best conditioning-efficiency balance on (6.14). Another discretization is taking the projection of the residue onto the finite-dimensional Fourier subspace to be zero. Specifically, suppose Ω_1 is characterized by the curve $r = \rho(\theta)$, then we enforce

$$(u_{MN}(\rho(\theta), \theta) - g(\theta), e^{il\theta})_{[0,2\pi)} = 0, \quad (6.15)$$

or

$$\left(\sum_{m=-M}^M (t^m \phi^m(\rho(\theta)) + \psi^m(\rho(\theta))) e^{im\theta} - g(\theta), e^{il\theta} \right)_{[0,2\pi)} = 0, \quad (6.16)$$

for $l = -M, \dots, M$. Usually the linear system (6.16) is better-conditioned if Ω_1 is closer to a circle, namely, if $\rho(\theta) \approx \rho_0$ which is a constant.

To sum up, the following algorithm results from the above discussion.

Algorithm 6.3.1 *Solve the Helmholtz problem (6.1)*

1. perform the domain embedding $\Omega \subset \subset \tilde{\Omega}$ and extend f from Ω to $\tilde{\Omega}$ smoothly;
2. compute the truncated Fourier expansion of $f(r, \theta) := F(r \cos \theta, \sin \theta)$ with respect to θ , obtaining (6.3);
3. compute approximate solutions $\{\phi_N^m\}$ and $\{\psi_N^m\}$ to the 1-D problems (6.7) and (6.8);
4. solve the optimization (6.14) or the linear system (6.16), to obtain $\{t^m\}_{m=-M}^M$;
5. compute u_{MN} by (6.10) and (6.11);

Same as the case of circular embedding, the total complexity of this algorithm is no greater than $O(N^3)$.

6.4 Numerical examples

We take the numerical example of (6.1) with f set by

$$f = \hat{f}_{-5}^k(r, \theta) + \hat{f}_4^k(r, \theta) + \hat{f}_{10}^k(r, \theta), \quad (6.17)$$

where

$$\hat{f}_j^k(r, \theta) = \left(-k^2 H_j^{(1)}(kr) + \frac{2k(j+1)}{r} H_{j+1}^{(1)}(kr) - k^2 H_{j+2}^{(1)}(kr) \right) e^{ij\theta}, \quad (6.18)$$

and the exact solution

$$u = \hat{u}_{-5}^k(r, \theta) + \hat{u}_4^k(r, \theta) + \hat{u}_{10}^k(r, \theta) \quad (6.19)$$

where

$$\hat{u}_j^k(r, \theta) = H_j^{(1)}(kr) e^{ij\theta}, \quad (6.20)$$

First, the shape of the obstacle is chosen by a smoothly perturbed curve

$$\Omega = \{(r, \theta) : r < 1 + 0.2 \sin(4\theta)\}. \quad (6.21)$$

And the embedding annulus is chosen by

$$\tilde{\Omega} = \{(r, \theta) : 0.8 < r < 2\}. \quad (6.22)$$

We perform Algorithm 6.3.1 for $k = 10, 50, 100$ and $M = 10, 15, \dots, 120$. The degree of spatial discretization is set by $N = M$, and the number of collocation nodes on $\partial\Omega_1$ is set by $K = 4M$. The 1-D spectral Galerkin method described in [1, Chap. 9.1] is utilized to solve (6.7) and (6.8). The problem domains and collocation nodes for $M = 10$ are shown in Fig. 6.1. The L^2 and L^∞ errors versus M for each k are shown in Fig. 6.2 and Fig. 6.3

It can be observed that for smaller $k = 10$, the error decays directly to machine precision in exponential rate. As M becomes greater than 90, the system becomes ill-conditioned that hence deteriorates the accuracy (leading to a very large error). On the other hand, for larger $k = 50$ or 100 , the errors decay very slowly in the

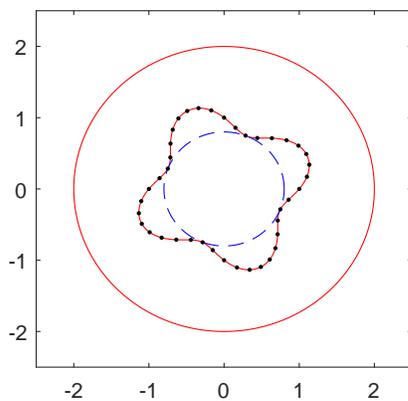


Fig. 6.1.: *the original domain, embedding annulus and collocation nodes for $M = 15$ in the first example*

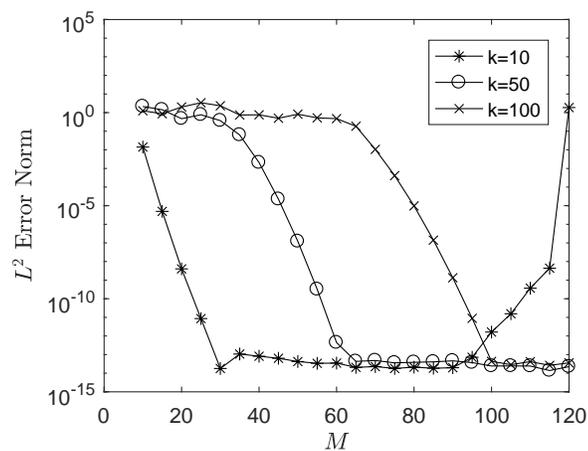


Fig. 6.2.: *L^2 errors versus M for $k = 10, 50, 100$ in the first example*

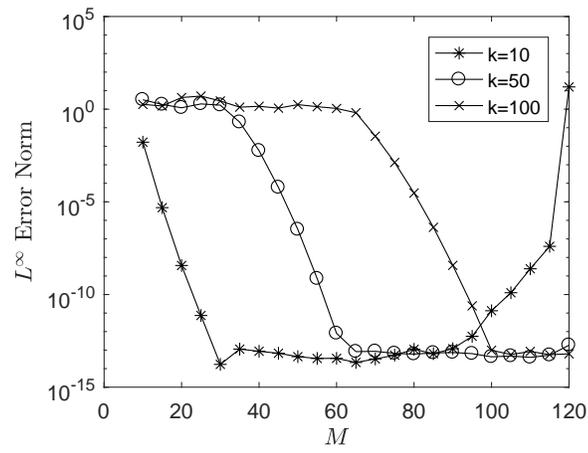


Fig. 6.3.: L^∞ errors versus M for $k = 10, 50, 100$ in the first example

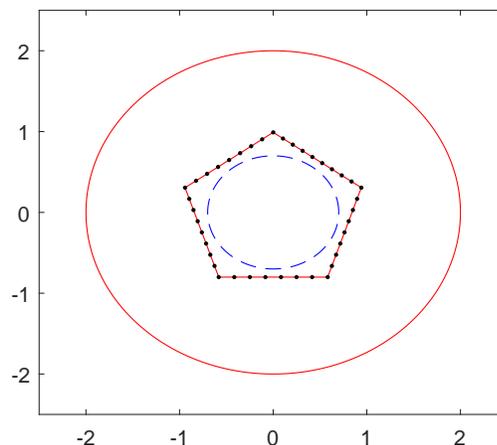


Fig. 6.4.: *the original domain, embedding annulus and collocation nodes for $M = 15$ in the second example*

beginning when M is small. They then decay exponentially to machine precision as M grows to certain levels. All phenomena is consistent with the case of circular obstacle (see [1, Chap. 9.1]).

Second, we choose the obstacle to be a equilateral pentagon with following vertices

$$\left(-r_0 \sin\left(\frac{2j\pi}{5}\right), r_0 \cos\left(\frac{2j\pi}{5} + \frac{\pi}{2}\right) \right), \quad j = 0, \dots, 4 \quad (6.23)$$

where $r_0 = 0.7\sqrt{2}$. And the embedding annulus is chosen by

$$\tilde{\Omega} = \{(r, \theta) : 0.7 < r < 2\}. \quad (6.24)$$

The results are shown in Fig. ???. The behaviors of L^2 and L^∞ error decays are exactly same as the preceding case.

6.5 Conclusion

In this chapter, we extend the spectral method with annular embedding scheme from Poisson-type problem to Helmholtz problem for acoustic scattering. We put forward the algorithm for the problems of obstacle of complex geometry, whose complexity is no greater than the classic spectral method for circular obstacles. Moreover,

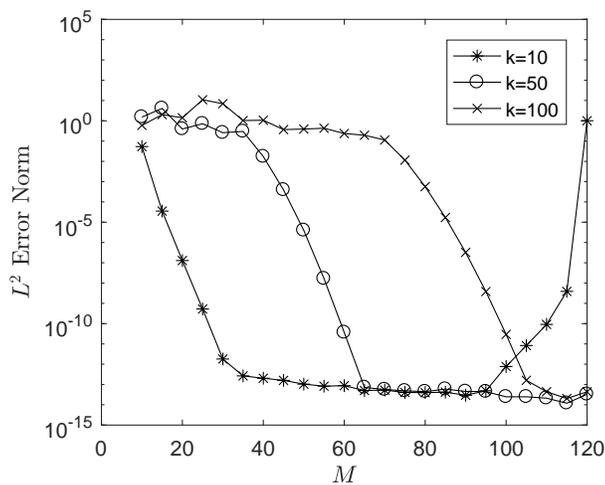


Fig. 6.5.: L^2 errors versus M for $k = 10, 50, 100$ in the second example

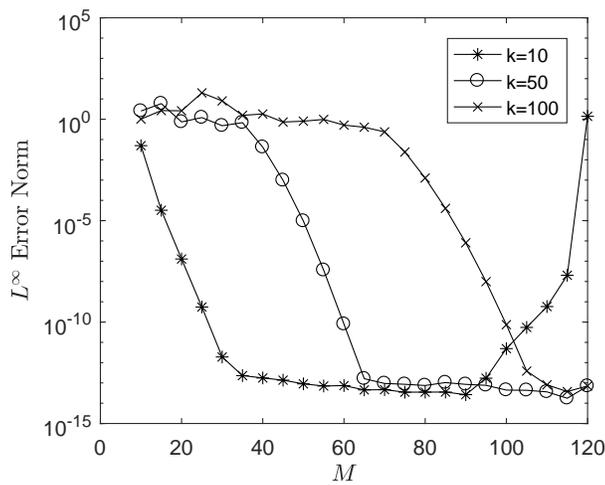


Fig. 6.6.: L^∞ errors versus M for $k = 10, 50, 100$ in the second example

the numerical examples show the behaviors of error decay is consistent with that from the case of circular obstacles.

REFERENCES

- [1] J. Shen, T. Tang, and L. Lian Wang, *Spectral Methods: Algorithms, Analysis and Applications*. Springer, 2011.
- [2] D. Gottlieb and S. Orszag, *Numerical Analysis of Spectral Methods*. Society for Industrial and Applied Mathematics, 1977.
- [3] A. Q. Claudio Canuto, M. Yousuff Hussaini and T. A. Z. Jr., *Spectral Methods in Fluid Dynamics*. Springer, Berlin, Heidelberg, 1988.
- [4] J. P. Boyd, *Chebyshev and Fourier spectral methods*. Dover Publications Inc., 2001.
- [5] L. N. Trefethen, *Spectral Methods in MATLAB*. Society for Industrial and Applied Mathematics SIAM, 2000.
- [6] P. F. M.O. Deville and E. Mund, *High-Order Methods for Incompressible Fluid Flow*. Cambridge : Cambridge Univ. Press, 2002.
- [7] G. E. Karniadakis and S. J. Sherwin, *Spectral/hp Element Methods for Computational Fluid Dynamics*. Oxford University Press, 1999.
- [8] R. Glowinski, T.-W. Pan, and J. Periaux, "A fictitious domain method for dirichlet problem and applications," *Computer Methods in Applied Mechanics and Engineering*, vol. 111, pp. 283–303, 1994.
- [9] P. Angot, C.-H. B. Pan, and P. Fabrie, "A penalization method to take into account obstacles in incompressible viscous flows," *Numerische Mathematik*, vol. 81, pp. 497–520, 1999.
- [10] K. Schneider, "Numerical simulation of the transient flow behaviour in chemical reactors using a penalisation method," *Computers & Fluids*, vol. 34, pp. 1223–1238, 2005.
- [11] R. N. van yen, D. Kolomenskiy, and K. Schneider, "Approximation of the laplace and stokes operators with dirichlet boundary conditions through volume penalization: a spectral viewpoint," *Numerische Mathematik*, vol. 128, pp. 301–338, 2014.
- [12] A. Bueno-Orovio, V. M. Perez-Garcia, and F. H. Fenton, "Spectral methods for partial differential equations in irregular domains: the spectral smoothed boundary method," *SIAM Journal on Scientific Computing*, vol. 28, pp. 886–900, 2006.
- [13] A. Bueno-Orovio and V. M. Perez-Garcia, "Spectral smoothed boundary methods: The role of external boundary conditions," *Numerical Methods for Partial Differential Equations*, vol. 22, pp. 435–448, 2005.

- [14] X. Li, J. Lowengrub, A. Ratz, and A. Voigt, “Solving pdes in complex geometries: a diffuse domain approach,” *Communications in Mathematical Sciences*, vol. 7, pp. 81–107, 2009.
- [15] A. Sarthou, S. Vincent, and P. Angot, “Eulerian-lagrangian grid coupling and penalty methods for the simulation of multiphase flows interacting with complex objects,” *International Journal for Numerical Methods in Fluids*, vol. 56, pp. 1093–1099, 2008.
- [16] A. Sarthou, S. Vincent, and J.-P. Caltagirone, “Consistent velocity-pressure coupling for second-order l2-penalty and direct-forcing methods,” 2011, hal-00592079, Version 1.
- [17] D. Shirokoff and J.-C. Nave, “A sharp-interface active penalty method for the incompressible navier-stokes equations,” *Journal of Scientific Computing*, vol. 62, pp. 53–77, 2015.
- [18] Q. Dinh, R. Glowinski, J. He, V. Kwock, T. Pan, and J. Periaux, “Lagrange multiplier approach to fictitious domain methods: application to fluid dynamics and electro-magnetics,” in *Domain Decomposition Methods for Partial Differential Equations*, D. Keyes, T. Chan, G. Meurant, J. Scroggs, and R. Voigt, Eds. Philadelphia: SIAM, 1992.
- [19] C. A. Brebbia, J. C. F. Telles, and L. C. Wrobel, *Boundary Element Techniques*. Springer-Verlag, 1984.
- [20] M. Elghaoui and R. Pasquetti, “A spectral embedding method applied to the advection-diffusion equation,” *Journal of Computational Physics*, vol. 125, pp. 464–476, 1996.
- [21] K. E. Atkinson, *The Numerical Solution of Integral Equations of the Second Kind*, ser. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 1997.
- [22] J. P. Boyd, “A comparison of numerical algorithms for fourier extension of the first, second, and third kinds,” *Journal of Computational Physics*, vol. 178, pp. 118–160, 2002.
- [23] M. Lyon, “A fast algorithm for fourier continuation,” *SIAM Journal on Scientific Computing*, vol. 33, pp. 3241–3260, 2011.
- [24] B. Adcock, D. Huybrechs, and J. Martin-Vaquero, “On the numerical stability of fourier extensions,” *Foundations of Computational Mathematics*, vol. 14, pp. 635–687, 2014.
- [25] O. P. Bruno and M. Lyon, “High-order unconditionally stable fc-ad solvers for general smooth domains i. basic elements,” *Journal of Computational Physics*, vol. 229, pp. 2009–2033, 2010.
- [26] M. Lyon and O. P. Bruno, “High-order unconditionally stable fc-ad solvers for general smooth domains ii. elliptic, parabolic and hyperbolic pdes; theoretical considerations,” *Journal of Computational Physics*, vol. 229, pp. 3358–3381, 2010.

- [27] N. Albin and O. P. Bruno, “A spectral fc solver for the compressible navier-stokes equations in general domains i: Explicit time-stepping,” *Journal of Computational Physics*, vol. 230, pp. 6248–6270, 2011.
- [28] N. Albin, O. P. Bruno, T. Y. Cheung, and R. O. Cleveland, “Fourier continuation methods for high-fidelity simulation of nonlinear acoustic beams,” *The Journal of the Acoustical Society of America*, vol. 132, pp. 2371–2387, 2012.
- [29] S. Lui, “Spectral domain embedding for elliptic pdes in complex domains,” *Journal of Computational and Applied Mathematics*, vol. 225, pp. 541–557, 2009.
- [30] M. Buffat and L. L. Penven, “A spectral fictitious domain method with internal forcing for solving elliptic pdes,” *Journal of Computational and Applied Mathematics*, vol. 230, pp. 2433–2450, 2011.
- [31] L. L. Penven and M. Buffat, “On the spectral accuracy of a fictitious domain method for elliptic operators in multi-dimensions,” *Journal of Computational Physics*, vol. 231, pp. 7893–7906, 2012.
- [32] H. Choi, “On several efficient algorithms for some partial differential equations,” Ph.D. dissertation, Purdue University, 2015.
- [33] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*. Springer, 2001.
- [34] I. Babuska and A. K. Aziz, “Survey lectures on the mathematical foundation of the finite element method,” in *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, A. K. Aziz, Ed. New York: Academic Press, 1972.
- [35] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*. Springer, 2004.
- [36] G. Strang, “Variational crimes in the finite element method,” in *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, A. K. Aziz, Ed. New York: Academic Press, 1972.
- [37] D. B. Haidvogel and T. Zang, “The accurate solution of poisson’s equation by expansion in chebyshev polynomials,” *Journal of Computational Physics*, vol. 20, pp. 167–180, 1979.
- [38] J. Shen, “Efficient spectral-galerkin method i. direct solvers of second- and fourth-order equations using legendre polynomials,” *SIAM Journal on Scientific Computing*, vol. 15, pp. 1489–1505, 1994.
- [39] X. Cheng and A. W. Shaikh, “Iterative methods for the poisson equation in l-shape domain based on domain decomposition method,” *Applied Mathematics and Computation*, vol. 180, pp. 393–400, 2006.
- [40] R. Zamolo, L. Parussini, and V. Pediroda, “The finite and spectral cell methods for smart structure applications: transient analysis,” *Journal of Scientific Computing*, vol. 74, p. 805825, 2018.
- [41] S. Duczek and U. Gabbert, “Efficient integration method for fictitious domain approaches,” *Computational Mechanics*, vol. 56, pp. 725–738, 2015.

- [42] S. Duczek, S. Liefold, and U. Gabbert, “The finite and spectral cell methods for smart structure applications: transient analysis,” *Acta Mechanica*, vol. 226, p. 845869, 2015.
- [43] D. Giraldo and D. Restrepo, “The spectral cell method in nonlinear earthquake modeling,” *Computational Mechanics*, vol. 60, p. 883903, 2017.
- [44] L. Caffarelli and L. Silvestre, “An extension problem related to the fractional laplacian,” *Communications in Partial Differential Equations*, vol. 32, no. 8, pp. 1245–1260, 2007.
- [45] L. A. Caffarelli and A. Vasseur, “Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation,” *Annals of Mathematics*, vol. 171, no. 3, pp. 1903–1930, 2010.
- [46] L. Silvestre, “Regularity of the obstacle problem for a fractional power of the laplace operator,” *Communications on Pure and Applied Mathematics*, vol. 60, no. 1, pp. 67–112, 2007.
- [47] C. Brndle, E. Colorado, A. de Pablo, and U. Snchez, “A concaveconvex elliptic problem involving the fractional laplacian,” *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, vol. 143, no. 1, p. 3971, 2013.
- [48] A. Capella, J. Dvila, L. Dupaigne, and Y. Sire, “Regularity of radial extremal solutions for some non-local semilinear equations,” *Communications in Partial Differential Equations*, vol. 36, no. 8, pp. 1353–1384, 2011.
- [49] P. R. Stinga and J. L. Torrea, “Extension problem and harnack’s inequality for some fractional operators,” *Communications in Partial Differential Equations*, vol. 35, no. 11, pp. 2092–2122, 2010.
- [50] R. H. Nochetto, E. Otárola, and A. J. Salgado, “A pde approach to fractional diffusion in general domains: A priori error analysis,” *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 733–791, Jun 2015. [Online]. Available: <https://doi.org/10.1007/s10208-014-9208-x>
- [51] G. P. Szegő, *Orthogonal polynomials*, 4th ed. Amer. Math. Soc. Colloq. Publ., 1975.
- [52] A. Lunardi, *Interpolation theory*. Springer, 2009.
- [53] K. Yosida, *Functional analysis*. Springer-Verlag New York Inc., 1968.
- [54] A. Bonito, W. Lei, and J. E. Pasciak, “Numerical approximation of the integral fractional laplacian,” *Numerische Mathematik*, vol. 142, no. 2, pp. 235–278, Jun 2019.

VITA

Yiqi Gu grew up in Zhejiang, China. In 2008, he was admitted to the department of mathematics at Zhejiang University and graduated in 2012. He also attended the department of applied mathematics at University of Washington and graduated with M.Sc. degree in 2014. Later in the fall 2014, he attended Purdue University to pursue Ph.D. degree in the department of mathematics. He is now a Ph.D. candidate in Mathematics.