分类号 O241.8 密 级
U D C 519.6 编 号 10486

武漢大學

硕 士 学 位 论 文

棱边元离散时谐麦克斯韦方程组的预条件快速迭代解法

研 究 生 姓 名: 张仕洋

学 号: 2014202010040

指导教师姓名、职称: 向华 教授

专 业 名 称: 计算数学

研 究 方 向: 数值代数

二○一七年五月

Preconditioners and Their Analyses for Edge Element Saddle-point Systems Arising from Time-harmonic Maxwell Equations

Candidate: Shiyang Zhang

Student Number: 2014202010040

Supervisor: Prof. Hua Xiang

Major: Computational Mathematics

Speciality: Numerical Algebra

School of Mathematics and Statistics

Wuhan University

May, 2017

论 文 原 创 性 声 明

本人郑重声明: 所呈交的学位论文, 是本人在导师指导下, 独立进行研究工作所取得的研究成果. 除文中已经标明引用的内容外, 本论文不包含任何其他个人或集体已经发表或撰写过的研究成果. 对本文的研究做出贡献的个人和集体, 均已在文中以明确方式标明. 本声明的法律结果由本人承担.

学位论文作者 (签名):

年　　月　　日

摘　　要

对于棱边元离散时谐麦克斯韦方程组产生的鞍点系统，我们提出并分析了新的预条件子。这些预条件子的提出源于一个求解该鞍点系统的系数矩阵的逆矩阵的公式。这些预条件子是文献 [9] 中的预条件子的推广。我们在理论上和数值上展示了结合我们提出的预条件子，当波数 (k) 不太大时 (粗略来说数值上 $k \leq 4$)，共轭梯度法 (CG) 能有效地求解该鞍点系统。对于新提出的和一些已经有的预条件子，我们分析并且比较了预条件系统的谱性质。数值实验方面，我们也详细展现了这些预条件迭代方法的有效性。

关键词: 时谐麦克斯韦方程, 预条件, 共轭梯度法, 计算电磁学

ABSTRACT

We propose and analyze new preconditioners for the saddle-point systems arising from the edge element discretization of the time-harmonic Maxwell equations. The preconditioners come from a formula giving the inverse of the coefficient matrix of the saddle-point system with vanishing and non-vanishing wave numbers, and are generalizations of the preconditioner in [9]. we show theoretically and numerically that the conjugate gradient method (CG) with these new preconditioners can be applied efficiently when the wave number (k) is not too large (roughly $k \leq 4$ numerically). The spectral behaviors of the resulting preconditioned systems for the new and some existing preconditioners are analyzed and compared, and numerical experiments are presented to demonstrate and compare the efficiencies of these preconditioners.

Key words: time harmonic Maxwell equations, preconditioner, computational electromagnetics

目　　录

1 绪论

时谐麦克斯韦方程组的求解一直面临着巨大的挑战。本文关注的是其中的一个问题：对于广泛使用的棱边元离散方程组后形成的大型稀疏鞍点系统的求解问题。对于 H(curl) 椭圆方程组，目前最引人注目的工作是由 Hitmair 和 Xu 提出的辅助空间预条件 [12]。不定时谐麦克斯韦方程组使问题的难度上升到另一个层次，尤其是面对高频应用（$\omega \sim 1GHz$），此时 $k^2 \sim O(10)$. 文献 [28] 提出了一个没有收敛性分析的快速多水平预条件方法。文献 [29] 和 [30] 基于摄动法，分别提出了两种重叠型施瓦兹方法和一种多水平方法，文献 [31] 则发展了一些两网格方法。其中一种的思路是用粗糙网格解原问题，用细网格解一个相应的对称正定问题。文献 [14] 对原鞍点系统进行增广，结合区域分解逼近，对新系统采用 Uzawa 型算法求解。文献 [10] 提出了一个对称正定的预条件子，对鞍点系统直接采用 MINRES 迭代。但是此文中处理的情况是波数 k 远远小于 1，也就是微波的情形。文献 [7,22,23] 对这一结果进行了推广，使之能够处理更大一点的波数。对于此预条件在三维空间间断系数环境下的并行计算的表现，文献 [16] 进行了分析。最近，对于波数 $k = 0$ 时的鞍点系统，文献 [9] 利用系数矩阵特殊的极大秩亏结构，提出并分析了一种新的预条件。在此预条件下，有趣的是，虽然系数矩阵与预条件本身都不是正定的，我们却可以定义特殊的内积，在此意义下仍然可以使用共轭梯度法。本文将在这个方向上更进一步地证明，这种新的方法对于非零波数（一般 $k < 4$）也有效，并且我们引入一个参数使整个方法更灵活。

迭代方法（配合预条件）包含了广阔的内容 [2]，从经典的雅克比迭代，高斯-塞德尔迭代和超松弛 SOR 迭代法，到一大类 Krylov 子空间迭代，发展到多重网格和多水平迭代，并在 90 年代初衍生出空间分解以及子空间校正的思想 [1]，这种思想将各种类科学计算问题的解决方案联系到了一起，并催生出该领域许多根本性的突破。

对于求解大型稀疏线性系统，预条件是一种非常有意思的技术。还有什么方法能让你解决问题时，仅仅需要找到系统本身的一种模糊或者不确定远近的易求解的近似逼近，配合合适的迭代格式，就可以开始去测试原来的复杂的问题的缺口或方向，更有吸引力呢？而复杂问题本身，或有特定结构特点，或基于某类特定知识，因而人们一般都能慢慢找到关于该系统的越来越多的先验经验和知识。预条件的发展是解决科学计算问题的一大阶梯。回到本文考虑的不定时谐麦克斯韦方程组，预条件子本身是几个对称正定系统的耦合，其中一个是 H(curl) 椭圆问题，我们将在数值实验部分直接用节点辅助空间预条件技术去解决它，另一个拉普拉斯问题，已有许多较成熟的方法，如代数多重网格方法。本文数值实验部分用不完全楚列斯基分解预条件共轭梯度法处理它。由于谱的聚集度很好地反映了本文的迭代格式的收敛速度 [3]，我们也将加入对预条件系统的谱的理论和数值分析。

本文的结构安排如下。在第二章我们介绍了时谐麦克斯韦方程组，它的混合形式及其棱边元的离散，还有已有的一些预条件技术。第三章中我们发展了两个重要的公式：一是求时谐麦克斯韦方程组离散得来的鞍点矩阵的逆的公式，二是给出计算矩阵 $\mathcal{K}$ 的逆矩阵的公式。第四章我们提出了新的预条件，并且分析了预条件系统的谱分布性质。作为对比，我们也考虑了已有的预条件系统的谱性质。第五节我们对于收敛性和谱的分布都进行了大量的数值实验，并加入了大量与已有预条件之间的的对比。最后，在第五章中对一些相关问题进行讨论，对可能存在的问题进行了思考，对以后的方向和可能的改进做出说明。

2 背景知识

2.1 时谐麦克斯韦方程组

麦克斯韦方程组是描述电磁场规律的最基本的偏微分方程组。关于它的数值离散和快速解法一直是计算数学和工程领域最为基本和重要的一部分。我们将以文献 [10, 16, 17] 为基础简要描述我们所使用的物理模型背景。描述电磁场的四个以位置 $\mathbf{x} \in \mathbb{R}^3$ 和时间 $t \in \mathbb{R}$ 为定义域的场函数分别为：E 和 H, 表示电场和磁场强度；D 和 B，表示电场分布和磁感应。经典电动力学理论中，他们满足如下方程：

$$\begin{aligned} \text{法拉第定律:} \quad & \frac{\partial B}{\partial t} + \nabla \times E = 0, \\ \text{库伦定律:} \quad & \nabla \cdot D = \tau, \\ \text{安培法则:} \quad & \frac{\partial D}{\partial t} - \nabla \times H = -\mathcal{F}, \\ \text{高斯定律:} \quad & \nabla \cdot B = 0. \end{aligned} \tag{2.1}$$

其中电荷密度 τ 与电流密度函数 $\mathcal{F}$ 满足 $\nabla \cdot \mathcal{F} + \frac{\partial \tau}{\partial t} = 0$. 本文考虑考虑一种特殊但普遍的情况，即电流密度和电荷密度为时谐的，此时 $\mathcal{F}(\mathbf{x}, t) = \mathcal{R}(exp(-i\omega t)\bar{\mathbf{J}}(\mathbf{x}))$, 且 $\tau(\mathbf{x}, t) = \mathcal{R}(exp(-i\omega t)\bar{\tau}(\mathbf{x}))$, 这里 i 是虚数单位，$\mathcal{R}$ 表示取式子的实部，ω 表示时间频率。相应地，考虑时谐形式的场函数：

$$\begin{aligned} E(\mathbf{x}, t) :=& \quad \mathcal{R}(exp(-i\omega t)\bar{\mathbf{E}}(\mathbf{x})), \\ D(\mathbf{x}, t) :=& \quad \mathcal{R}(exp(-i\omega t)\bar{\mathbf{D}}(\mathbf{x})), \\ H(\mathbf{x}, t) :=& \quad \mathcal{R}(exp(-i\omega t)\bar{\mathbf{H}}(\mathbf{x})), \\ B(\mathbf{x}, t) :=& \quad \mathcal{R}(exp(-i\omega t)\bar{\mathbf{B}}(\mathbf{x})). \end{aligned} \tag{2.2}$$

结合(2.1)与(2.2)不难得出如下的时谐形式的麦克斯韦方程组：

$$\begin{aligned} -i\omega\bar{\mathbf{B}} + \nabla \times \bar{\mathbf{E}} &= 0, \\ \nabla \cdot \bar{\mathbf{D}} &= \bar{\tau}, \\ -i\omega\bar{\mathbf{D}} - \nabla \times \bar{\mathbf{H}} &= -\bar{\mathbb{J}}, \\ \nabla \cdot \bar{\mathbf{B}} &= 0. \end{aligned} \tag{2.3}$$

我们进一步假设线性关系 $\bar{\mathbf{D}} = \epsilon\bar{\mathbf{E}}$，和 $\bar{\mathbf{B}} = \mu\bar{\mathbf{H}}$ 成立, 这里 ϵ 和 μ 分别表示电容率和磁导率。真空中他们的取值分别记为 ϵ_0 和 μ_0. 最后，我们假设有如下的欧姆定律成立：$\bar{\mathbb{J}} = \sigma\bar{\mathbf{E}} + \bar{\mathbb{J}}_a$, 这里 $\bar{\mathbb{J}}_a$

描述感应电流密度，而 σ 表示电导率。方程组(2.3)可化为如下常用的二阶麦克斯韦系统：

$$\nabla \times (\mu_r^{-1} \nabla \times \mathbf{E}) - k^2 \epsilon_r \mathbf{E} = f. \tag{2.4}$$

其中，$\epsilon_r = \frac{1}{\epsilon_0}(\epsilon + \frac{i\sigma}{\omega})$, $\mu_r = \mu/\mu_0$, $\mathbf{E} = \epsilon_0^{1/2}\bar{\mathbf{E}}$, $\mathbf{H} = \mu_0^{1/2}\bar{\mathbf{H}}$, $f = ik\mu_0^{1/2}\bar{\mathbb{J}}_a$, 波数 k 满足 $k = \omega\sqrt{\mu_0\epsilon_0}$.

2.2 混合形式时谐麦克斯韦方程组及其离散

为了简便考虑，我们不考虑变参数的情形，假设 $\epsilon_r = 1$ 且 $\mu_r = 1$。现在将模型重新总结并简化一下并重新安排下记号。根据赫姆霍兹分解 [17], 当 $k \neq 0$ 时，记 $\mathbf{E} = u + (-1/k^2)\nabla p$, 我们直接考虑真空中以下混合形式的时谐麦克斯韦方程组 [6,8,13,14,20]：

$$\begin{cases} \nabla \times \nabla \times u - k^2 u + \nabla p = J & \text{in} \quad \Omega, \\ \nabla \cdot u = \rho & \text{in} \quad \Omega, \\ u \times n = 0 & \text{on} \quad \partial\Omega, \\ p = 0 & \text{on} \quad \partial\Omega. \end{cases} \tag{2.5}$$

其中 u 是向量场，p 为标量乘子，J 为给定的源项，ρ 为电荷密度。Ω 是 $\mathbb{R}^3$ 中简单连通多面体区域，$\partial\Omega$ 为其连通边界，n 为边界的单位法向量。波数 k 满足 $k^2 = \omega^2\varepsilon\mu$ 其中 ω, ε 和 μ 分别为正的频率，介电常数和介质的磁导率。我们假设 k^2 不是内麦克斯韦特征值, 但是允许是零。小波数及大波数的情 k 况在实际中，如静磁场, 波传播和其他应用中广泛出现 [10]. 读者亦可参考文献文献 [5, 第 11 章]。

当 $k \neq 0$ 时，(2.5)中的拉格朗日乘子 p 并不是一定要引入的，因为此时散度约束并不需要显式强制满足。我们可直接通过(2.5)中第一个方程求出 u，而数学上 $p = 0$ [11], 虽然目前设计一个高效的解法求解不定系统仍然是充满挑战的。引入拉格朗日乘子 p 的鞍点公式(2.5)是稳定和适定的 [8]。特别的，它直接确保了当 k 很小时高斯律的成立，并且能够更好地处理解在边界上的奇异性 [6,8,20]。更加重要的是, 混合形式(2.5)在计算方面提供了额外的灵活性 [16], 数值上可获得更稳定高效的求解方法 [9,10]。这即是本文的主要的动机和关注点。

首先我们规定一些常用的记号。类似于通常的有限元方法采用的方式，我们将区域 Ω 划分为形状规则的, 一致的四面体单元集合 $\mathcal{T}_h = \{K\}$. 四面体 K 的直径记为 h_K; 定义 $h = max_{K\in\mathcal{T}_h} h_K$. 记 $\mathcal{P}_l(K)$ 为四面体 K 上度数是 l $(l = 1, 2, 3...)$ 的多项式，而 $\mathcal{N}_l(K)$ 代表第一类 Nédélec 棱边元空间，其度数满足 $\mathcal{P}_{l-1}(K)^3 \in \mathcal{N}_l(K) \in \mathcal{P}_l(K)^3$. 为了离散向量场 u, 考虑如下 Sobolev 空间：

$$H_0(curl) = \{v \in L^2(\Omega)^3 \,:\, \nabla \times v \in L^2(\Omega)^3,\ v \times n = 0 \text{ on } \partial\Omega\}$$

及其离散子空间

$$V_h = \{v_h \in H_0(curl) \mid v_{h|K} \in \mathcal{N}_l(K),\, K \in \mathcal{T}_h\}$$

同时，为了离散标量乘子 p, 考虑 $H_0^1(\Omega)$ 的如下离散子空间：

$$Q_h = \{q_h \in H_0^1(\Omega) \mid q_{h|K} \in \mathcal{P}_l(K), K \in \mathcal{T}_h\}$$

混合问题的变分公式为：求 $(u,p) \in H_0(curl) \times H_0^1(\Omega)$, 使得对于所有的 $(v,q) \in H_0(curl) \times H_0^1(\Omega)$, 满足

$$\begin{aligned} A(u,v) - k^2 M(u,v) + B(v,p) &= (J,v)_\Omega \\ B(u,q) &= (\rho,q), \end{aligned}$$

而有限元离散问题为：求 $(u_h,p_h) \in V_h \times Q_h$, 使得对于所有的 $(v_h,q_h) \in V_h \times Q_h$, 满足

$$\begin{aligned} A(u_h,v_h) - k^2 M(u_h,v_h) + B(v_h,p_h) &= (J,v)_\Omega \\ B(u_h,q_h) &= (\rho,q_h). \end{aligned}$$

变分型由以下公式给出：

$$\begin{aligned} A(u,v) &= \int_\Omega (\nabla \times u) \cdot (\nabla \times v)\, d\mathrm{x} \\ M(u,v) &= \int_\Omega u \cdot v\, d\mathrm{x} \\ B(u,q) &= \int_\Omega u \cdot \nabla q\, d\mathrm{x}. \end{aligned}$$

假设 $\langle \psi_j \rangle_{j=1}^n$ 为 V_h 的一组基，而 $\langle \phi_i \rangle_{i=1}^m$ 为 Q_h 的一组基, 即 $V_h = span\, \langle \psi_j \rangle_{j=1}^n$, $Q_h = span\, \langle \phi_i \rangle_{i=1}^m$, 那么以上有限元离散问题在此组基下可以表示为如下方程组：

$$\mathcal{K} \begin{pmatrix} u \\ p \end{pmatrix} \equiv \begin{pmatrix} A - k^2 M & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{2.6}$$

其中 $u \in \mathbb{R}^n$, $p \in \mathbb{R}^m$, A, $M \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{m \times n}$, $m < n$ 满足：

$$\begin{aligned} A_{i,j} &= \int_\Omega (\nabla \times \psi_j) \cdot (\nabla \times \psi_i)\, d\mathrm{x} \\ M_{i,j} &= \int_\Omega \psi_j \cdot \psi_i\, d\mathrm{x} \\ B_{i,j} &= \int_\Omega \psi_j \cdot \nabla \phi_i\, d\mathrm{x}. \end{aligned}$$

简单的说，我们用第一类 Nédélec 元 [17,18] 逼近向量场 u，用阶数相容的标准的节点有限元逼近乘子 p, 导出了我们感兴趣的系统(2.6).

我们假设方程(2.6)中的系数矩阵 $\mathcal{K}$ 和它的 $(1,\ 1)$ 块 $A - k^2 M$ $(k \neq 0)$ 都是非奇异的，事实上如果网格尺寸足够小，这总是对的 [10]。

相应地，在此组基下，我们还可以构造拉普拉斯矩阵 $L \in \mathbb{R}^{m \times m}$, 即

$$L_{i,j} = \int \nabla \phi_j \cdot \nabla \phi_i\, d\mathrm{x}$$

由于有 $\nabla Q_h \in V_h$, 我们可以构造从 Q_h 到 V_h 的梯度插值算子，我们记该算子在该组基下的矩阵表示为 $C \in \mathbb{R}^{n \times m}$. 由于 ∇Q_h 是无旋的，我们有 $AC = 0$. 我们构造了一系列矩阵，我们将在下一节总结他们的性质。

注意如果我们考虑线性基函数，即 $l=1$, 那么 n 等于网格内部（不含边界上的）的边的数目，而 m 等于网格内部的（不含边界上的）顶点的数目。

有意思的一点是，A 是对称半正定的，它的零空间的维数是 m, 达到了整个非奇异鞍点系统所能允许的最大值，也即，当 $k=0$ 时整个系统是极大秩亏的 [9,10]。这个性质在文献 [9] 被用于求解鞍点矩阵的逆矩阵。有极大秩亏的（1，1）块的鞍点系统出现在许多应用中，包括麦克斯韦方程组的求解 [10,11,20]，欠定范数极小化问题和和地球物理反问题 [9]。故我们在3.1节中将推广这一结果到文献尚未涉及到的非对称极大秩亏鞍点系统。

2.3 已有的部分预条件

在文献 [9] 中考虑了 $k=0$ 的情形，此时方程 (2.6) 简化为

$$\mathcal{A}\begin{pmatrix} u \\ p \end{pmatrix} \equiv \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}. \tag{2.7}$$

该文献中的结论基于以下命题的性质，他们对于我们的进一步分析也是基本的。为此，我们首先引用矩阵 A, B, M, L 和 C 的一些有用的性质。

性质 2.3.1 矩阵 A, B, M, L 和 C 满足以下关系 [9,10]:

(i) $\mathbb{R}^n = ker(A) \oplus ker(B)$.

(ii) $C = M^{-1}B^T$, 并且存在一个与网格大小无关的常数 $\bar{\alpha} > 0$ 使 $u^T A u \geq \bar{\alpha}\, u^T M u, \quad \forall\, u \in ker(B)$.

(iii) $L = BM^{-1}B^T$, 或 $L = BC$.

(iv) 矩阵 $\mathcal{A}$ 的逆可以表示为

$$\mathcal{A}^{-1} = \begin{pmatrix} V & CL^{-1} \\ L^{-1}C^T & 0 \end{pmatrix}, \tag{2.8}$$

其中的对角块 V 为

$$V = (A + B^T L^{-1} B)^{-1}(I - B^T L^{-1} C^T) = (A + B^T L^{-1} B)^{-1} - CL^{-1}C^T. \tag{2.9}$$

事实上命题(i)来自于离散亥姆霍兹分解，参见文献 [32, 7.2.1 节] 中的例子。命题(iii)直接根据它们的定义验证即可。命题(ii)涉及到椭圆性的性质，可参考文献 [11] 中的定理 4.7. 我们在(3.1)节中给出了(iv)的另外一个证明。

为了求解鞍点系统(2.7)，文献 [9] 中提出以下的预条件 $\mathcal{P}_0$:

$$\mathcal{P}_0^{-1} = \begin{pmatrix} (A+M)^{-1}(I - B^T L^{-1} C^T) & CL^{-1} \\ L^{-1}C^T & 0 \end{pmatrix}, \tag{2.10}$$

其中矩阵 $L \in \mathbb{R}^{m\times m}$ 是如上节所定义的离散拉普拉斯, 而 $C \in \mathbb{R}^{n\times m}$ 上节所定义的离散梯度插值算子, 它的列张成 $ker(A)$，利用用标准节点基的梯度，我们能够很轻易和显式地组装它 [9,10]。该文中证明了 $\mathcal{P}_0^{-1}\mathcal{A}$ 可以表示为对角阵

$$\mathcal{P}_0^{-1}\mathcal{A} = \begin{pmatrix} (A+M)^{-1}(A+B^TL^{-1}B) & 0 \\ 0 & I \end{pmatrix}.$$

矩阵 $A+M$ 和 $A+B^TL^{-1}B$ 都是对称正定的, 因此以 $\mathcal{P}_0^{-1}$ 为预条件，在非标准内积下，虽然矩阵 $\mathcal{A}$ 和 $\mathcal{P}_0$ 都不是正定的我们仍然可以使用共轭梯度法。

需要注意的是我们不一定非要求解鞍点系统 [10]，相反，我们可以直接求解分解后的系统。对简单的情形如 $k=0$ 时，以下已经分解的系统是广为人知的：

- 前处理: 首先计算乘子 p: $p = L^{-1}C^Tf$. 这里需要求解一个 L-系统。若 f 是无散的, 我们有 $p=0$，此时这一步一般会省略掉。
- 求解半正定 H(curl) 系统: $Au = f - B^Tp$. 这步的方法有很多，例如，可通过节点辅助空间预条件技术 [12,15]. 在高性能预条件子软件 Hypre [25] 中, 不仅有此预条件直接处理奇异系统 $Ax = b$ （$b \in \mathcal{R}(A)$）的方法, 也可以让用户选择求解一个等价的系统 $(A+\alpha CC')x = b$, 这里 α 是一个很小的正数.
- 后处理: 零空间纠正。这里需要另外一个准去确的 L-求解器。这一步我们需要确保解满足约束条件 $Bu = g$.

对于更一般的 $k \neq 0$ 的情形, 文献 [7,10,22,23] 中探讨了有两个可变松弛参数 $\eta > k^2$ 和 $\varepsilon \neq 0$ 的块三角预条件子：

$$\mathcal{M}_{\eta,\varepsilon} = \begin{bmatrix} A+(\eta-k^2)M & (1-\eta\varepsilon)B^T \\ 0 & \varepsilon L \end{bmatrix} \tag{2.11}$$

这篇文章中, 基于以上预条件子，我们为系统 (2.6)提供了一些新的预条件。正如文献 [9] 中所讲，(2.10)中的的预条件子 $\mathcal{P}_0^{-1}$ 对于稳态麦克斯韦方程非常有效。我们将说明类似的结果能够创造，推广到波数非零，频率更高的情形。并且，我们将分析谱的分布，且发现它们与使用(2.11)中块三角预条件产生的分布十分类似。但是在非标准内积下，新的方法能够使用共轭梯度法迭代，数值上我们发现新的方法比块三角预条件(2.11)迭代法收敛更快。

3 逆矩阵的计算公式

3.1 极大秩亏矩阵的逆矩阵

注意本节中的所有记号是独立的并且只在本节中使用。

在这一节中，我们借助一些文献中的已有的结果，推广(2.8)中计算对称鞍点矩阵 $\mathcal{A}$ 的逆的公式。考虑如下非对称广义鞍点系统：

$$\mathcal{S}\begin{pmatrix} u \\ p \end{pmatrix} \equiv \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{3.1}$$

其中所有的矩阵 A, B, C 和 D 可以被允许是非方阵, 且 $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times k}$, $C \in \mathbb{R}^{l\times n}$, $D \in \mathbb{R}^{l\times k}$. 但是注意整个块矩阵 $\mathcal{S}$ 本身是方的, 即 $m+l = n+k = t$. 进一步假设矩阵 A 的秩是 $m+n-t$. 接着记 $C_r \in \mathbb{R}^{n\times l}$ 为列满秩矩阵，且它的列张成 $ker(A)$，而 $C_l \in \mathbb{R}^{k\times m}$ 为行满秩矩阵，且它的行张成矩阵 A 的左零空间，也即 $C_lA = 0$, 且 $AC_r = 0$. 我们记 $L_l = C_lB$，$L_r = CC_r$, rank(A) 表示矩阵 A 的秩, 且将 A 的值域空间记为 $\mathcal{R}(A)$.

现在我们给出本节的主要结论：

定理 3.1.1 假设

$$\mathcal{R}(A)\cap\mathcal{R}(B) = 0\,, \quad \mathcal{R}(A^T)\cap\mathcal{R}(C^T) = 0\,, \tag{3.2}$$

$$rank(A) = m+n-t\,, \quad rank(B) = k\,, \quad rank(C) = l\,, \tag{3.3}$$

那么矩阵 $\mathcal{S}$ 不是奇异的，并且它的逆可以表示为

$$\mathcal{S}^{-1} = \begin{pmatrix} N & C_rL_r^{-1} \\ L_l^{-1}C_l & 0 \end{pmatrix}, \tag{3.4}$$

这里 N 满足

$$NA = I - C_rL_r^{-1}C\,, \quad AN = I - BL_l^{-1}C_l\,, \tag{3.5}$$

$$NB = -C_rL_r^{-1}D\,, \quad CN = -DL_l^{-1}C_l. \tag{3.6}$$

如果 $m = n$, 那么对于任意的 $X \in \mathbb{R}^{m\times l}$，只要矩阵 $A+XC$ 非奇异，就有

$$N = (A+XC)^{-1}(I - BL_l^{-1}C_l - XDL_l^{-1}C_l). \tag{3.7}$$

容易验证(2.7)中的系数矩阵满足假设条件(3.2) 和 (3.3).

为了给出该定理的证明，我们首先引入一些记号和辅助的结果。我们记 $E_A = I - AA^\dagger$，$F_A = I - A^\dagger A$, 其中 $A^\dagger$ 是矩阵 A 的广义逆（Moore-Penrose 逆），即, $A^\dagger$ 是满足以下方程的 X 的唯一解。

$$XAX = X, \quad AXA = A, \quad (AX)^T = AX, \quad (XA)^T = XA. \tag{3.8}$$

我们借用以下的结果来显式地得到我们的(3.4)中的 $\mathcal{S}^{-1}$ 的逆。

定理 3.1.2 ([24,27]) 假设定理 (3.1.1)中的假设条件成立, 那么矩阵 $\mathcal{S}$ 不是奇异的，并且它的逆可以表示为

$$\mathcal{S}^{-1} = \begin{pmatrix} A^\dagger - A^\dagger B B_0^\dagger - C_0^\dagger C A^\dagger - C_0^\dagger \left(D - CA^\dagger B\right) B_0^\dagger & C_0^\dagger \\ B_0^\dagger & 0 \end{pmatrix}, \tag{3.9}$$

其中 $B_0 = E_A B$，$C_0 = C F_A$.

首先，我们推导出(3.4)的 (1,2) 块和 (2,1) 块的公式。.

定理 3.1.3 以下结果成立：

$$B_0^\dagger = L_l^{-1} C_l, \quad C_0^\dagger = C_r L_r^{-1}. \tag{3.10}$$

证明 假设 $x \in \mathbb{R}^k$ 满足 $x^T L_l = x^T C_l B = 0$, 那么直接计算可得 $(xC_l \ \ 0)\,\mathcal{S} = 0$, 由于 $\mathcal{K}$ 非奇异，故 $x^T C_l = 0$. 这表明 $x = 0$. 因此方阵 L_l 是非奇异的。类似的，我们有 L_r 是非奇异的。

根据定理(3.1.2)中的定义，容易验证

$$E_A = I - AA^\dagger = C_l^\dagger C_l\,, \quad F_A = I - A^\dagger A = C_r C_r^\dagger. \tag{3.11}$$

所以 $B_0 = C_l^\dagger C_l B = C_l^\dagger L_l$. 由于矩阵 L_l^{-1} 列满秩，C_l 行满秩, 我们推得 $(L_l^{-1} C_l)^\dagger = C_l^\dagger (L_l^{-1})^\dagger = B_0$, 这就结束了第一个等式的证明。类似的，第二个等式也成立。 □

现在我们更加进一步地导出(3.9)的 (1,1) 块的显式的公式。首先，我们给出以下结果：

性质 3.1.1 记 $V = A^\dagger - A^\dagger B B_0^\dagger - C_0^\dagger C A^\dagger$, $T = V + C_0^\dagger C A^\dagger B B_0^\dagger$, 和 $N = T - C_0^\dagger D B_0^\dagger$, 那么以下结论成立

$$NA = TA = VA = I - C_r L_r^{-1} C, \quad AN = AT = AV = I - B L_l^{-1} C_l, \tag{3.12}$$

$$TB = CT = 0, \quad NB = -C_0^\dagger D, \quad CN = -D B_0^\dagger. \tag{3.13}$$

并且, 对于任意的 $X \in \mathbb{R}^{m\times l}$ 和 $Y \in \mathbb{R}^{k\times n}$, 以下等式成立：

$$N(A + BY) = I - C_r L_r^{-1} C - C_r L_r^{-1} D Y\,,$$

$$(A + XC)N = I - B L_l^{-1} C_l - X D L_l^{-1} C_l\,.$$

证明 (3.12)中的第一个命题可由以下关系

$$VA = A^\dagger A - A^\dagger BB_0^\dagger A - C_0^\dagger CA^\dagger A = (I - C_0^\dagger C)A^\dagger A = (I - C_r L_r^{-1} C)(I - C_r C_r^\dagger)$$
$$= I - C_r L_r^{-1} C - C_r C_r^\dagger + C_r L_r^{-1} C C_r C_r^\dagger = I - C_r L_r^{-1} C$$

和关系式 $(C_0^\dagger CA^\dagger BB_0^\dagger)A = 0$ 推出, 后者是定理中(3.1.3)结论 $B_0^\dagger A = 0$ 的直接推论。(3.12)中的第二个命题可类似推出。注意到由定理(3.1.3)可知: $B_0^\dagger B = I$ 并且 $CC_0^\dagger = I$. 由此可直接证明关系式(3.13)。剩下的等式直接验证即可。 □

3.2 计算矩阵鞍点系统矩阵的逆

我们在这一节推导出一些公式来计算(2.6)中矩阵 $\mathcal{K}$ 的逆。现在我们已经准备好推导出矩阵 $\mathcal{K}$ 的逆矩阵了。回忆一下矩阵 $\mathcal{K}$ 和 $A - k^2 M$ 都是可逆矩阵。我们记逆矩阵 $\mathcal{K}^{-1}$ 的 (1,1) 块为 T, 那么对于鞍点矩阵 $\mathcal{K}$ 的逆矩阵, 我们有如下的表示:

定理 3.2.1 $\mathcal{K}$ 的逆矩阵有如下表示

$$\mathcal{K}^{-1} = \begin{pmatrix} T & CL^{-1} \\ L^{-1}C^T & k^2 L^{-1} \end{pmatrix}, \tag{3.14}$$

其中 T 满足

$$(A - k^2 M)T = I - B^T L^{-1} C^T, \quad BT = 0. \tag{3.15}$$

证明 我们记 $\mathcal{K}^{-1}$ 为矩阵 $\mathcal{A}^{-1}$ 的如下形式的摄动:

$$\mathcal{K}^{-1} = \mathcal{A}^{-1} + \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}, \tag{3.16}$$

那么由于 $\mathcal{K}\mathcal{K}^{-1} = I$, 即

$$\left[\mathcal{A} + \begin{pmatrix} -k^2 M & 0 \\ 0 & 0 \end{pmatrix}\right] \cdot \left[\mathcal{A}^{-1} + \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}\right] = I,$$

我们通过直接计算可以得到

$$-k^2 M(V + X_1) + AX_1 + B^T X_3 = 0, \tag{3.17}$$

$$-k^2 (B^T L^{-1} + MX_2) + AX_2 + B^T X_4 = 0, \tag{3.18}$$

$$BX_1 = 0, \quad BX_2 = 0. \tag{3.19}$$

同样地，通过直接验证，有

$$AV = I - B^T L^{-1} C^T, \quad BV = 0\,. \tag{3.20}$$

从 (3.16)我们知道 $V + X_1$ 为矩阵 $\mathcal{K}^{-1}$ 的 (1,1) 块, 所以从 (3.19) 和 (3.20) 我们可以推出

$$BT = B(V + X_1) = 0.$$

用矩阵 C^T 乘以(3.17)的两边，我们得到如下表达式：

$$-k^2 B(V + X_1) + LX_3 = 0,$$

即

$$X_3 = k^2 L^{-1} B(V + X_1) = 0. \tag{3.21}$$

相似地, 用矩阵 C^T 乘以(3.18)的两边，我们得到如下表达式：

$$-k^2(I + BX_2) + LX_4 = 0.$$

结合这个等式以及 (3.19)中的第二个关系式，不难获得

$$X_4 = k^2 L^{-1}. \tag{3.22}$$

然后我们将 (3.22) 代入 (3.18)中，可得

$$(A - k^2 M)X_2 = 0\,, \tag{3.23}$$

意即 $X_2 = 0$.

注意到我们已经证明了 $X_3 = 0$, 那么等式 (3.17)可化为 $-k^2 M(V + X_1) + AX_1 = 0$, 或者说 $(A - k^2 M)(V + X_1) = AV$, 这样我们就完成了所有需要的证明。 □

以下的重要结果有助于我们理解(3.14)中的 $\mathcal{K}$ 的逆矩阵的 (1,1) 块 T。

定理 3.2.2 对于任何的参数 $\eta \neq k^2$，矩阵 $A + \eta B^T L^{-1} B - k^2 M$ 是非奇异的，并且当 $\eta = k^2$ 时，它的核空间与 A 的相同。

证明 根据(2.3.1)中的性质(i)，我们能将 $u \in \mathbb{R}^n$ 写作 $u = u_A + u_B$，其中 $u_A \in ker(A)$, $u_B \in ker(B)$. 假设 $(A+\eta B^T L^{-1} B - k^2 M)u = 0$, 那么有 $(A - k^2 M)u_B + \eta B^T L^{-1} B u_A - k^2 M u_A = 0$. 由于矩阵 C 的列张成 A 的零空间, 存在着 $t \in \mathbb{R}^m$ 使得 $u_A = Ct$. 所以有 $(A - k^2 M)u_B + (\eta - k^2)B^T t = 0$. 用 C^T 乘在该等式的两边, 我们有 $t = 0$, 因而 $(A - k^2 M)u_B = 0$, 也就意味着 $u_B = 0$. 故此我们证明了 $u = 0$, 也就证明了定理的前半部分。

接着, 我们考虑参数 $\eta = k^2$ 的情形。我们将证明矩阵 $A + k^2 B^T L^{-1} B - k^2 M$ 和 A 有同样的零空间。首先, 注意到 $(A + k^2 B^T L^{-1} B - k^2 M)C = 0$. 另外, 我们假设 u 属于矩阵 $A + k^2 B^T L^{-1} B - k^2 M$

的零空间。我们仍然用直和分解 $u = u_A + u_B$, 并沿着证明的前半部分的思路去证明此部分, 不同之处在于这时我们取参数 $\eta = k^2$ 以此方式易推得 $(A - k^2M)u_B = 0$, 也即 $u_B = 0$, 因而我们得到 $Au = 0$. □

以下结果为方程(3.15) 和定理 (3.2.2) 的直接推论。该结果给出了矩阵 $\mathcal{K}^{-1}$ 的 (1,1) 块的表达式 T, 并且同时引入了一个非常重要的参数 η。这个参数只要满足 $\eta \neq k^2$ 即可。

推论 3.2.1 对于任意的 $\eta \neq k^2$, 我们有

$$\begin{aligned} T &= (A + \eta B^T L^{-1} B - k^2 M)^{-1}(I - B^T L^{-1} C^T) \\ &= (A + \eta B^T L^{-1} B - k^2 M)^{-1} - \frac{1}{\eta - k^2} C L^{-1} C^T. \end{aligned} \tag{3.24}$$

总而言之, 从定理 (3.2.1)和推论 (3.2.1)中不难推出计算方程(2.6)中的矩阵 $\mathcal{K}$ $(\eta \neq k^2)$ 的逆的公式:

$$\mathcal{K}^{-1} = \begin{pmatrix} (A + \eta B^T L^{-1} B - k^2 M)^{-1}(I - B^T L^{-1} C^T) & CL^{-1} \\ L^{-1} C^T & k^2 L^{-1} \end{pmatrix}. \tag{3.25}$$

这个公式是我们下一节中构建新的预条件的基础。

4 新的预条件子及其谱性质

4.1 预条件及其分析

公式(3.25)给我们提供了一些自然的预处理器去处理方程(2.6)中的鞍点矩阵 $\mathcal{K}$。由于矩阵 $B^TL^{-1}B$ 是稠密矩阵，求解(3.25)的（1,1）块需要非常昂贵的计算。为了克服这个部分的困难，我们用与稠密矩阵 $A+\eta B^TL^{-1}B-k^2M$ 谱等价 [10] 的稀疏矩阵 $A+\eta M-k^2M$ 来逼近这个稠密矩阵。这样我们可以得到如下的经过简化的预条件子：

$$\mathcal{P}^{-1}\equiv\begin{pmatrix}(A+\eta M-k^2M)^{-1}(I-B^TL^{-1}C^T) & CL^{-1}\\ L^{-1}C^T & k^2L^{-1}\end{pmatrix}. \tag{4.1}$$

对于最简单的波数等于零 ($k=0$) 并且 $\eta=1$ 的情形，预条件子(4.1)简化为 (2.10)中已经存在的 $\mathcal{P}_0^{-1}$. 为了确保(4.1)中的矩阵 $A+\eta M-k^2M$ 的非奇异性，我们可以令参数 $\eta>k^2$，这样这个矩阵变为对称正定矩阵。选取参数满足这个条件还可以确保等式(4.1)的右端的非奇异性，正如下面的定理所示：

定理 4.1.1 对于任意的 $\eta>k^2$, (4.1)中等号右边的矩阵是非奇异的。

证明 我们可以直接验证预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 满足

$$\mathcal{P}^{-1}\mathcal{K}=\begin{pmatrix}(A+\eta M-k^2M)^{-1}(A+k^2B^TL^{-1}B-k^2M)+CL^{-1}B & 0\\ 0 & I\end{pmatrix}.$$

根据命题 (2.3.1) (i), 我们能进一步将如上矩阵的 (1, 1) 块写作:

$$\begin{aligned}&(A+\eta M-k^2M)^{-1}(A+k^2B^TL^{-1}B-k^2M)+CL^{-1}B\\ =&(A+\eta M-k^2M)^{-1}(A+k^2B^TL^{-1}B-k^2M)\\ &+(A+\eta M-k^2M)^{-1}(\eta MCL^{-1}B-k^2MCL^{-1}B)\\ =&(A+\eta M-k^2M)^{-1}(A+\eta B^TL^{-1}B-k^2M),\end{aligned}$$

因此预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 也可以写作

$$\mathcal{P}^{-1}\mathcal{K}=\begin{pmatrix}(A+\eta M-k^2M)^{-1}(A+\eta B^TL^{-1}B-k^2M) & 0\\ 0 & I\end{pmatrix}. \tag{4.2}$$

定理(3.2.2)告诉我们，方程(4.2)中矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的（1，1）块不是奇异的。这样我们推出了所需结论。 □

注意当 $\eta > k^2$ 时，$A + \eta M - k^2 M$ 和它的逆矩阵总是对称正定的。事实上，原矩阵 $A + \eta B^T L^{-1} B - k^2 M$ 也可以是对称正定的，如下所示:

定理 4.1.2 参数满足 $\eta > k^2$ 和 $k^2 < \bar{\alpha}$ 时, 矩阵 $A + \eta B^T L^{-1} B - k^2 M$ 是对称正定的.

证明 对于任意的 $u \in \mathbb{R}^n$, 我们将 u 写作 $u = u_A + u_B$，其中 $u_A \in ker(A)$，$u_B \in ker(B)$. 由命题 (2.3.1)知 $u_A^T M u_B = 0$ 并且 $u_A^T B^T L^{-1} B u_A = u_A^T M u_A$. 因此,

$$\begin{aligned} & u^T (A + \eta B^T L^{-1} B - k^2 M) u \\ =& u_A^T (A + \eta B^T L^{-1} B - k^2 M) u_A + u_B^T (A + \eta B^T L^{-1} B - k^2 M) u_B \\ =& u_A^T (\eta B^T L^{-1} B - k^2 M) u_A + u_B^T (A - k^2 M) u_B \\ =& u_B^T (A - k^2 M) u_B + (\eta - k^2) u_A^T M u_A \,. \end{aligned} \tag{4.3}$$

但是根据命题 (2.3.1)中的(ii)我们知道 $u_B^T A u_B \geq \bar{\alpha} u_B^T M u_B$. 这意味着

$$u^T (A + \eta B^T L^{-1} B - k^2 M) u \geq (\bar{\alpha} - k^2) u_B^T M u_B + (\eta - k^2) u_A^T M u_A > 0 \,,$$

由此我们证明了所需结论。 □

对于参数 $\eta > k^2$ 和 $k^2 < \bar{\alpha}$, 预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 关于如下内积是自伴并且正定的:

$$\langle x, y \rangle = x^T \begin{pmatrix} (A + \eta M - k^2 M) & 0 \\ 0 & I \end{pmatrix} y. \tag{4.4}$$

因此在此内积空间下我们可以使用 CG 迭代 [4] 去求解系数矩阵为 $\mathcal{P}^{-1}\mathcal{K}$ 的预条件系统。但是定理 (4.1.2) 并没有告诉我们 $k^2 \geq \bar{\alpha}$ 时原系统是否是正定的，因此理论上 CG 迭代可能失败。我们将在(5)节看到数值上迭代并不会失败 (参见图 (5.3))。

我们知道 Krylov 子空间迭代方法的收敛速度经常可以由预条件系统的谱聚集度所反应。因此，我们现在研究预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的谱性质。首先, 我们观察到矩阵 $A + \eta B^T L^{-1} B - k^2 M$ 的对称正定性由参数 k 决定。

定理 4.1.3 对于任意满足 $\eta_1, \eta_2 > k^2$ 的两个实数 η_1, η_2，矩阵 $A + \eta_1 B^T L^{-1} B - k^2 M$ 是对称正定的当且仅当 $A + \eta_2 B^T L^{-1} B - k^2 M$ 是对称正定的。

证明 对于任意的 $\eta_1 > k^2$, 假设 $A + \eta_1 B^T L^{-1} B - k^2 M$ 不是对称正定的。由于这个矩阵是非奇异的（见定理(3.2.2)），所以它也不是对称半正定的。因此，存在 $u \in \mathbb{R}^n$ 满足 $u^T (A + \eta_1 B^T L^{-1} B - k^2 M) u < 0$. 注意我们可以将 u 写作 $u = u_A + u_B$, 其中 $u_A \in ker(A)$, $u_B \in ker(B)$. 我们可以从(4.3)得到 $u_B \neq 0$，并且 $u_B^T (A - k^2 M) u_B < 0$. 这样对于任意的 $\eta_2 > k^2$, 我们可以验证如下关系: $u_B^T (A + \eta_2 B^T L^{-1} B - k^2 M) u_B = u_B^T (A - k^2 M) u_B < 0$, 因此 $A + \eta_2 B^T L^{-1} B - k^2 M$ 不是对称正定的。 □

4.2 谱的分布情况

接下来我们给出一些关于预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的特征值分布的结果。

引理 4.2.1 对于任意的 $\eta > k^2$, $\lambda = 1$ 是 $(A + \eta M - k^2 M)^{-1}(A + \eta B^T L^{-1} B - k^2 M)$ 的代数重数为 m 的特征值。其余的特征值有如下的上下界：

$$\frac{\bar{\alpha} - k^2}{\bar{\alpha} + \eta - k^2} < \lambda < 1. \tag{4.5}$$

证明 当 $\eta = 1$ 和 $k^2 < 1$ 时，该结论已经在文献 [10, 定理 5.1] 中被证明。我们的关于任意参数 η 的结论可以类似地被证明。 □

结合公式(4.2)，引理 (4.2.1)有如下的直接的推论：

定理 4.2.1 对于任意的 $\eta > k^2$, $\lambda = 1$ 是预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的代数重数为 $2m$ 的特征值。其余的特征值被(4.5)所限制。

对于我们的新的预条件 $\mathcal{P}$ 和文献中已经存在的块三角预条件 $\mathcal{M}_{\eta,\varepsilon}$（参见(2.11)），我们将比较下它们生成的预条件系统的的谱的分布情况。

首先，我们回忆 [23, 定理 2.6] 中的如下的结果：

定理 4.2.2 对于任意的 $\eta > k^2$, $\lambda_1 = 1$ 和 $\lambda_2 = -\frac{1}{\varepsilon(\eta - k^2)}$ 是矩阵 $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$, 的特征值，每一个的代数重数都是 m. 剩余的特征值满足 (4.5).

从定理(4.2.1) 和 (4.2.2) 中我们知道 $\mathcal{P}^{-1}\mathcal{K}$ 和 $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$ 的谱分布是很相似的，除了后者有一个额外的代数重数为 m 的特征值。我们也将在下一节数值验证这一结论。

另外，需要注意的是，如果我们令 $\varepsilon = \frac{1}{\eta}$，块三角预条件子 $\mathcal{M}_{\eta,\varepsilon}$ 将变为对称的：

$$\mathcal{M}_{\eta,1/\eta} = \begin{bmatrix} A + (\eta - k^2)M & 0 \\ 0 & \frac{1}{\eta}L \end{bmatrix}. \tag{4.6}$$

文献 [10, 22] 中分析了这个预条件在 MINRES 迭代下的表现。从定理 (4.2.1) 和 (4.2.2) 中可以推出，我们的预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的谱聚集度比 $\mathcal{M}_{\eta,1/\eta}^{-1}\mathcal{K}$ 的稍微好一点。这是因为特征值 λ_2 要比 $\frac{\bar{\alpha}-k^2}{\bar{\alpha}+\eta-k^2}$ 小。然而预条件 $\mathcal{P}$ 在 $k^2 < \bar{\alpha}$ 时可以结合 CG 方法使用，在 $k^2 \geq \bar{\alpha}$ 时可以结合 MINRES 使用。更重要的是，正如我们将在下一节的数值实验中看到的一样，即使当 $4 > k^2 \geq \bar{\alpha}$ 时，我们仍然可以使用 CG 方法，结合我们的预条件，并且数值上收敛是相当稳定的。相反地，CG 法结合(4.6)中的预条件 $\mathcal{M}_{\eta,1/\eta}$ 大多时候都会在收敛到一定精度之前迭代终止。

另外一方面，如果我们取 $\varepsilon \neq 1/\eta$, 预条件 $\mathcal{M}_{\eta,\varepsilon}$ 不是对称的，此时我们没法使用像 CG 或 MINRES 这样经济稳定的方法。如果我们取 $\varepsilon = -\frac{1}{\eta - k^2}$, 我们有 $\lambda_2 = \lambda_1$, 故 $\lambda = 1$ 是 $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$ 代数重数为 $2m$ 的特征值，与 $\mathcal{P}^{-1}\mathcal{K}$ 的相同。

现在我们来考虑和预条件 $\mathcal{P}$ 相对应的内迭代对于任意的两个向量 $x \in \mathbb{R}^n$ 和 $y \in \mathbb{R}^m$, 我们记

$$\begin{aligned}\mathcal{P}^{-1}\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} (A+\eta M-k^2M)^{-1}(x-B^TL^{-1}C^Tx)+CL^{-1}y \\ L^{-1}C^Tx+k^2L^{-1}y \end{pmatrix} \\ &= \begin{pmatrix} (A+\eta M-k^2M)^{-1}x-\frac{1}{\eta-k^2}CL^{-1}C^Tx+CL^{-1}y \\ L^{-1}C^Tx+k^2L^{-1}y \end{pmatrix} \\ &= \begin{pmatrix} (A+\eta M-k^2M)^{-1}x+CL^{-1}(y-\frac{1}{\eta-k^2}C^Tx) \\ L^{-1}(C^Tx+k^2y) \end{pmatrix}.\end{aligned}$$

在任意一种形式中，为了得到 $\mathcal{P}^{-1}$ 作用于向量的结果，我们都需要解两个线性系统，一个方程组的的系数矩阵是离散拉普拉斯算子 L，另一个的是 $A+(\eta-k^2)M$. 许多快速解法可以用来求解这两个对称正定系统 [12,16]。我们使用如上公式中的第一个来计算 $\mathcal{P}^{-1}$ 作用于向量的结果。

我们知道参数 $\bar{\alpha}$ 仅仅依赖于网格形状以及有限元的阶，而与网格大小无关（参见 [11, 定理 4.7]）。数值上我们预期对于参数 k^2 使得 $A+\eta B^TL^{-1}B-k^2M$ 对称正定的上界应该是与网格尺寸无关的。我们将在下一节的数值实验中测试此上界。

5 数值实验

5.1 实验说明

我们给出数值实验来分析和比较在不同预条件下，鞍点问题(2.6)的预条件系统的谱的分布情况，以及一些 Krylov 空间迭代方法的结果。我们已经将代码上传至网址 [26]. 我们在矩形区域 $(-1 \le x \le 1, -1 \le y \le 1)$ 和 L 形区域 (见图 (5.1) 和 (5.2)), 采用最低阶棱边元离散系统(2.5)。我们用非结构化单纯形网格划分工具 EasyMesh [19] 建立这些区域上的网格。对于网格 G1 到 G5, 包含区域顶点的所有三角形的边的目标长度设为一样, 相应产生的线性系统大小分别满足 $m+n=$ 187, 437, 1777, 7217, 23769. 对于网格 L1 到 L5, 包含原点的三角形的边的目标长度设为包含其他顶点的三角形的边的目标长度的十分之一，相应产生的线性系统大小分别满足 $m+n=185$, 409, 1177, 5325, 29277.

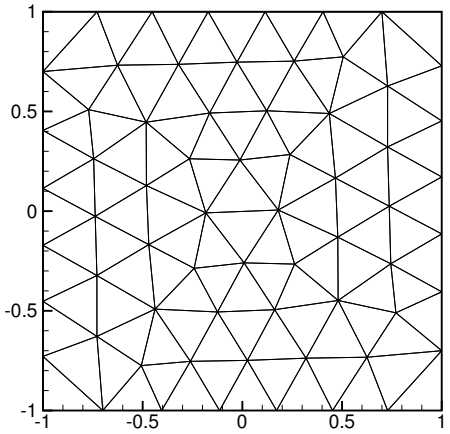
我们使用 MATLAB (inter(R) Core(TM) i7-4510U CPU @ 2.00 GHz 2.60 GHz, 4 GB RAM) 实现所有的数值迭代解法，同时，$A+(\eta-k^2)M$-求解器为 Hiptmair-Xu 预条件 [12]) CG 方法, 拉普拉斯系统求解使用的是通过不完全楚列斯基分解预条件 CG 方法。方程(2.6)的右端项, 记为 b, 设为元素全为一的向量, 对于所有迭代，零向量作为初始猜测解 $x^{(0)}$. 为了求解鞍点系统(2.6), 我们分别运行预条件 $\mathcal{P}$-CG 方法 ([21] 中算法 1 的修改版）和块预条件 $\mathcal{M}_{\eta,1/\eta}$-MINRES 方法, 记这两种方法为 $\mathcal{P}$-CG 和 $\mathcal{M}_{\eta,1/\eta}$-MINRES。外迭代终止条件是 $\|b-\mathcal{K}x^{(k)}\|_2 \le 10^{-6}\cdot\|b\|_2$, 其中 $x^{(k)}$ 是 k 次迭代解。我们设参数 $\eta=k^2+1$ 并将所有的拉普拉斯求解器 (包含外迭代的 L-求解器和 Hiptmair-Xu 预条件内部的拉普拉斯求解器) 的停止准则设为残量的相对 l_2 范数误差小于一个相同的误差界（除非另有说明)。迭代时间的单位为秒，两类拉普拉斯矩阵的不完全楚列斯基分解所花费的时间都加到了所有方法花费的迭代时间上。

5.2 迭代表现

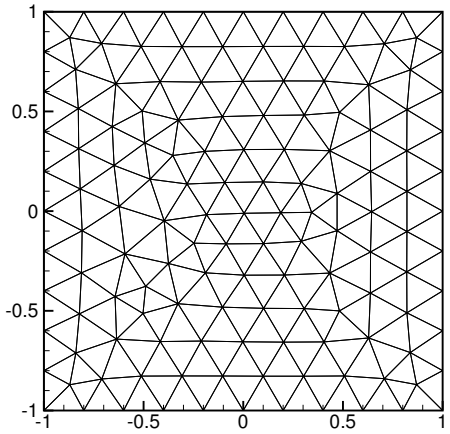
在(5.1)和(5.2) 中我们列出在不同网格和波数下，方法 $\mathcal{P}$-CG 和 $\mathcal{M}_{\eta,1/\eta}$-MINRES 的迭代次数和时间。其中 $A+(\eta-k^2)M$-求解器和拉普拉斯求解器采用相同的紧的内迭代限度 10^{-6}. 与我们(4)节中的的理论预期相同，新的方法 $\mathcal{P}$-CG 需要的迭代次数稍微比方法 $\mathcal{M}_{\eta,1/\eta}$-MINRES 所需要的少一点。

比值行给出了方法 $\mathcal{M}_{\eta,1/\eta}$-MINRES 和 $\mathcal{P}$-CG 所花费的时间之比。当 $k \le 2$ 时，$\mathcal{M}_{\eta,1/\eta}$-MINRES 所花费的时间大约比 $\mathcal{P}$-CG 的多 19% $\sim$ 55%, 而当我们取 $k=4$ 时这个差异稍微地变

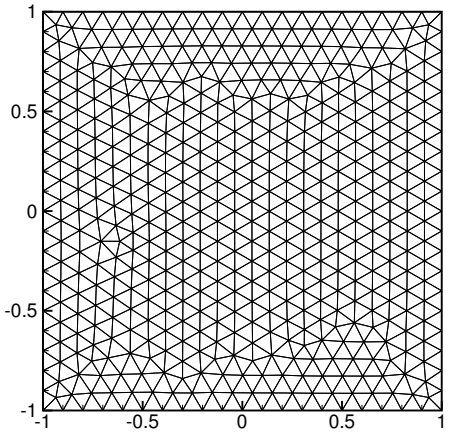
图 5.1: 网格 *G1* 到 *G4*



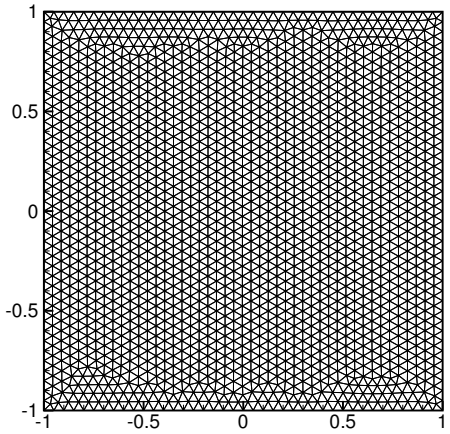
(a) 网格 G1: n+m=187



(b) 网格 G2: n+m=437



(c) 网格 G3: n+m=1777



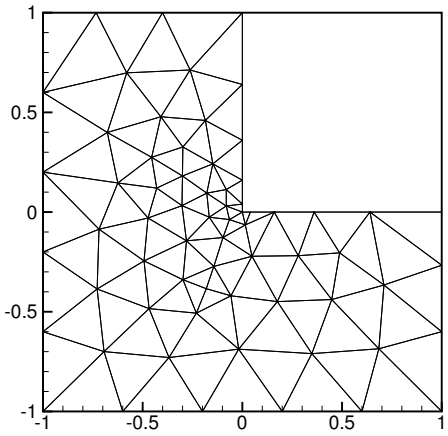
(d) 网格 G4: n+m=7217

小了一些。我们还可以观察到迭代次数基本上是与网格尺寸没有关系的。这些实验仅仅用作在这两种方法间做一个粗略的比较，实际应用中，我们需要更加有效的拉普拉斯求解器。

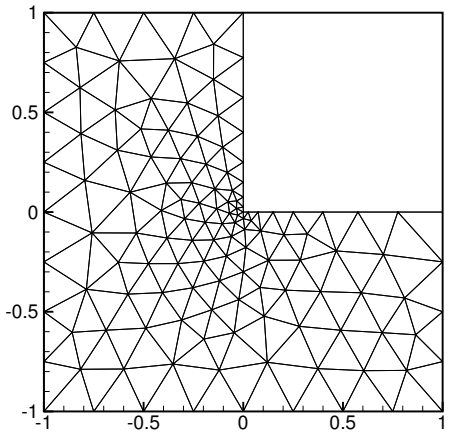
在表5.3和5.4中，我们检查了对于 $A+(\eta-k^2)M$-求解器 (在左边列出) 和 L-求解器 (在上面列出) 不同松弛程度的内迭代准则下，这两种方法的表现情况。此时, Hitmair-Xu 内部的拉普拉系求解器的残量相对误差界特别地被固定为 10^{-1}. 首先我们可以观察到下三角部分的时间比上三角部分的要小，这说明对 $A+(\eta-k^2)M$-求解器采用相对宽松的停止准则，对 L-求解器采用相对紧的停止准则是一个较好的组合策略。实际上 $A+(\eta-k^2)M$-求解器相对于 L-求解器来说要贵。可见这种组合策略是有意义的。表格 (5.5) 给出了表格(5.4) 和 (5.3)中列出的时间之间的比值。我们可以观察到方法 $\mathcal{P}$-CG 相对于 $\mathcal{M}_{\eta,1/\eta}$-MINRES 有比较明显的优势。我们接下来将对最优内迭代有个更加详细的观察和讨论。

在表格 (5.6)中我们报告参数 $\eta-k^2$ 取值不同时 $\mathcal{P}$-CG ($\mathcal{M}_{\eta,1/\eta}$-MINRES) 所需的迭代次数。

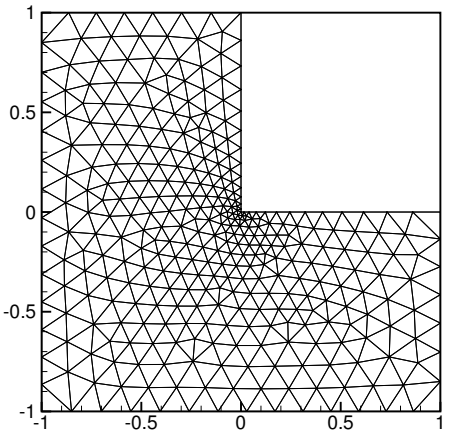
图 5.2: 网格 *L1* 到 *L4*



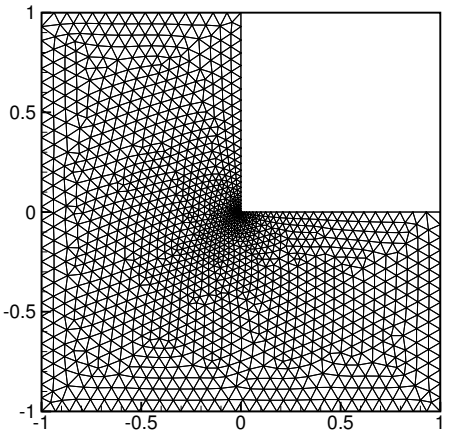
(a) 网格 L1：n+m=185



(b) 网格 L2：n+m=409



(c) 网格 L3：n+m=1177



(d) 网格 L4：n+m=5325

此时我们给定不那么准确的内迭代（L-求解器和 $A+(\eta-k^2)M$-求解器）精度。从数据可以看出，如果我们取参数 $k=0$ 或 1, 那么参数 $\eta-k^2=1$ 足够好；然而, 当 $k=2$ 或者 $k=4$ 时却不是这样. 此时, 我们需要 $\eta-k^2$ 的取值更加大一些, 例如, $\eta-k^2=1/2k^2$ 时迭代表现更佳。

从表格(5.3)和(5.4)中我们可以观察到在参数 $k=0,\eta=1$ 时, $\mathcal{P}$-CG 和 $\mathcal{M}_{\eta,1/\eta}$-MINRES 在内迭代精度为 (1e-1, 1e-5) (分别对应于 $A+(\eta-k^2)M$-求解器和 L-求解器) 时表现最佳。在表 (5.7)中我们做了一些额外的实验来调查这个现象. 此表显示当取 $k=0,1,1.2$ 或者 1.25 时, 内迭代的精度为 (1e-1,1e-5) 时外迭代表现最佳。然而, 当我们取参数 $k=4$ 时，我们需要稍微紧一点的内迭代精度，以防止出现不稳定情况。在表 (5.7)中我们还列出了 $\mathcal{M}_{\eta,1/\eta}$-MINRES 和 $\mathcal{P}$-CG 的迭代时间的比值。

表 5.1: $\mathcal{P}$-*CG* 和 $\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES* 行: 各种波数 k 下 $\mathcal{P}$-*CG* 方法和 $\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES* 方法的迭代次数（时间）; 比值行: $\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES* 方法和 $\mathcal{P}$-*CG* 方法花费的时间的比值。我们取参数 $\eta = k^2 + 1$, 网格为 *G1* 到 *G5*.

k	0	1.0	1.55	1.6	2	4
网格 G1						
$\mathcal{P}$-CG	5(0.6802)	6(0.7485)	11(1.2736)	11(1.268)	11(1.2619)	25(2.6965)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	7(0.8907)	9(1.1619)	15(1.6414)	15(1.6309)	14(1.5287)	31(3.2299)
比值	1.3095	1.5523	1.2888	1.2863	1.2114	1.1978
网格 G2						
$\mathcal{P}$-CG	5(0.949)	7(1.1789)	12(1.9092)	12(1.8898)	11(1.7518)	28(4.1755)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	7(1.2051)	9(1.4634)	15(2.3171)	15(2.261)	15(2.2492)	31(4.4716)
比值	1.2698	1.2414	1.2137	1.1965	1.2839	1.0709
网格 G3						
$\mathcal{P}$-CG	5(1.9158)	6(2.1352)	11(3.5818)	11(3.6182)	11(3.5805)	25(7.5431)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	8(2.7483)	9(2.9572)	15(4.6596)	15(4.6461)	14(4.337)	31(9.0741)
比值	1.4346	1.385	1.3009	1.2841	1.2113	1.203
网格 G4						
$\mathcal{P}$-CG	5(6.2827)	6(7.2006)	9(10.3151)	9(10.2293)	11(12.1859)	24(25.2278)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	7(8.1305)	9(10.0955)	12(13.2987)	12(13.0848)	14(15.0183)	29(29.4057)
比值	1.2941	1.402	1.2892	1.2791	1.2324	1.1656
网格 G5						
$\mathcal{P}$-CG	5(32.9871)	6(37.8760)	9(53.7789)	9(53.6557)	11(63.9640)	23(127.2520)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	7(43.0383)	8(48.4078)	12(69.4350)	12(69.3690)	14(79.6365)	29(156.8750)
比值	1.3047	1.2781	1.2911	1.2929	1.2450	1.2328

5.3 谱分析

在表(5.8)对应的数值实验中，为了测试矩阵 A_η 的正定性，我们计算了

$$A_\eta = \begin{pmatrix} A + \eta B^T L^{-1} B - k^2 M & 0 \\ 0 & I_m \end{pmatrix}$$

的最小特征值，记为 $\lambda_{\min}(A_\eta)$. 从方程(4.2)容易看出，如果 A_η 是对称正定的，我们能够结合我们的新的预条件在(4.4)定义的特殊内积下使用 CG 方法。首先可以观察到对于网格 G1 到 G4，如果 $k = 0, 1,$ 或 1.55，矩阵 A_η 是对称正定的，如果 $k = 1.6, 2$ 或 4，矩阵 A_η 不是对称正定的。相似地, 对于网格 L1 到 L4，如果 $k = 0, 1,$ 或 1.2，矩阵 A_η 是对称正定的，如果 $k = 1.25, 2$ 或 4，矩阵 A_η 不是对称正定的。正如定理(4.1.2)所预测, 矩阵 A_η 的正定性与网格尺寸无关。有很重要的一点需要注意到：当波数 k 很小时, A_η 是对称正定的, 因此 CG 法可以结合我们的预条件 $\mathcal{P}$ 使用，即使原系统 $\mathcal{K}$ 是不定的。而当 k 不是足够小时, 对应的预条件矩阵不再是正定的，即使仍然采用非标准的内积空间。此时理论上我们应该采用相应的非正定 Krylov 子空间迭代法如最小残差

表 5.2: $\mathcal{P}$-*CG* 和 $\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES* 行: 各种波数 k 下 $\mathcal{P}$-*CG* 方法和 $\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES* 方法的迭代次数（时间）; 比值行: $\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES* 方法和 $\mathcal{P}$-*CG* 方法花费的时间的比值。我们取参数 $\eta = k^2 + 1$, 网格为 *L1* 到 *L5*.

k	0	1.0	1.2	1.25	2	4
网格 L1						
$\mathcal{P}$-CG	5(0.8006)	7(1.0285)	9(1.2749)	8(1.1593)	10(1.3678)	25(3.3425)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	8(1.1821)	9(1.3191)	12(1.7786)	10(1.3660)	15(1.9633)	31(3.8042)
Ratios	1.4765	1.2826	1.3951	1.1783	1.4353	1.1382
网格 L2						
$\mathcal{P}$-CG	6(1.3000)	7(1.3541)	9(1.7064)	8(1.5272)	12(2.2195)	28(4.8256)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	8(1.5563)	9(1.6863)	12(2.1860)	10(1.8770)	15(2.7282)	32(5.4724)
Ratios	1.1972	1.2454	1.2810	1.2290	1.2292	1.1340
网格 L3						
$\mathcal{P}$-CG	5(1.8053)	7(2.1953)	9(2.7333)	8(2.5120)	12(3.5724)	25(7.0046)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	8(2.5005)	9(2.7071)	11(3.2302)	11(3.3899)	15(4.2672)	31(8.4044)
Ratios	1.3851	1.2331	1.1818	1.3495	1.1945	1.1998
网格 L4						
$\mathcal{P}$-CG	5(5.3455)	7(6.9149)	8(7.7637)	8(7.7648)	12(11.0665)	24(21.0020)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	8(7.7230)	9(8.5107)	12(11.0335)	12(11.0477)	15(13.4594)	31(26.6709)
Ratios	1.4448	1.2308	1.4212	1.4228	1.2162	1.2699
网格 L5						
$\mathcal{P}$-CG	5(53.0033)	7(69.6419)	8(78.3754)	8(78.5196)	10(95.8931)	26(231.7050)
$\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES	8(78.9650)	9(87.1453)	10(95.7683)	10(95.6296)	13(121.6840)	30(261.2080)
Ratios	1.4898	1.2513	1.2219	1.2179	1.2690	1.1273

表 5.3: 在网格 *L4* 上, 对于 $A + (\eta - k^2)M$-求解器 *(列于左边)* 和 L-求解器 *(列于顶部)* 采用不同的内迭代相对误差界时, $\mathcal{P}$-*CG* 方法的迭代次数（时间）。*Hitmair-Xu* 预条件内部的拉普拉斯求解器采用的终止限度特别的固定为 *1e-1* . 参数取为 $k = 0$, $\eta = 1$. 这里及后面的 *1e-1*（或其他指数）表示 10^{-1}.

	1e-5	1e-4	1e-3	1e-2	1e-1
1e-5	5(2.0808)	5(2.1045)	5(2.1666)	6(2.3585)	13(4.5092)
1e-4	5(1.6821)	5(1.6615)	5(1.6706)	6(2.0957)	13(3.9118)
1e-3	5(1.3027)	5(1.3421)	5(1.3220)	6(1.5743)	13(3.0280)
1e-2	7(1.1731)	6(1.1331)	6(1.1461)	6(1.1881)	13(2.3217)
1e-1	8(0.7564)	9(0.7813)	8(0.7640)	7(0.8225)	13(1.5585)

法。然而, (5.2)节中的数值结果表明此时 CG 方法仍然收敛得十分稳定和快速。产生这个重要现象的原因我们将在接下来的一段中说明。

我们做了更多的实验来深入比较新的和旧的预条件 $\mathcal{P}$ 和 $\mathcal{M}_{\eta,1/\eta}$ 迭代方法的稳定性和速度。从(5.2)节我们发现数值上预条件 $\mathcal{P}$-CG 法总能稳定地迭代和很好地收敛, 虽然理论上 CG 法此时不一定会收敛。但是对于预条件子 $\mathcal{M}_{\eta,1/\eta}$, 这种情况却不会发生, 也即 CG 法数值上也不能用。为了验证这一点, 我们重新跑了表格 (5.4)中的所有的实验, 但是这次使用 CG 迭代, 取代 MINRES

表 5.4: 在网格 *L4* 上，对于 $A+(\eta-k^2)M$-求解器 *(*列于左边*)* 和 L-求解器 *(*列于顶部*)* 采用不同的内迭代相对误差界时，$\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES* 方法的迭代次数（时间）。*Hitmair-Xu* 预条件内部的拉普拉斯求解器采用的终止限度特别的固定为 *1e-1* . 参数取为 $k=0$，$\eta=1$.

	1e-5	1e-4	1e-3	1e-2	1e-1
1e-5	8(2.9245)	8(3.0198)	9(3.1409)	10(3.5018)	21(7.5691)
1e-4	8(2.4044)	8(2.4364)	9(2.6340)	10(2.9974)	21(5.8159)
1e-3	8(1.7549)	8(1.7433)	9(2.0404)	10(2.3071)	21(4.5918)
1e-2	10(1.4443)	10(1.4053)	9(1.3736)	10(1.6640)	21(3.5088)
1e-1	18(1.2268)	18(1.2463)	15(0.9492)	17(1.1446)	21(2.0830)

表 5.5: 表格 (5.3) 和 (5.4)中列出的时间的比值。方法 $\mathcal{P}$-*CG* 相对于方法 $\mathcal{M}_{\eta,1/\eta}$-*MINRES* 在时间上有部分优势。

	1e-5	1e-4	1e-3	1e-2	1e-1
1e-5	1.4054	1.4349	1.4497	1.4848	1.6786
1e-4	1.4294	1.4664	1.5767	1.4303	1.4868
1e-3	1.3471	1.299	1.5435	1.4655	1.5164
1e-2	1.2311	1.2402	1.1985	1.4005	1.5113
1e-1	1.6219	1.5952	1.2425	1.3915	1.3365

表 5.6: 在网格 *L4* 上，取不同的参数 k 和 $\eta-k^2$ 时，方法 $\mathcal{P}$-*CG* ($\mathcal{M}_{\eta,\frac{1}{\eta}}$-*MINRES)* 所花费的时间。内迭代中 $A+(\eta-k^2)M$-求解器和 L-求解器的精度分别设为 *1e-2* 和 *1e-4*.

η	k^2+1	k^2+2	k^2+4	k^2+8	k^2+16
$k=0$	1.5594(1.9346)	1.5704(2.1003)	1.7223(2.2196)	2.0911(2.7166)	3.1828(3.7524)
$k=1$	2.2820(2.5947)	2.2033(2.6732)	2.2419(2.7203)	2.3651(3.7658)	3.4197(4.2749)
$k=2$	3.8534(4.5152)	3.4527(3.7501)	3.1498(3.5301)	3.8618(5.0507)	4.3371(6.1675)
$k=4$	9.3866(11.9675)	8.5725(8.8261)	6.4138(7.3868)	7.3863(8.9812)	7.9024(9.1783)

迭代。在所有 30 个数值实验中, 我们总是遇到一个除数变得太小，使得迭代过程终止的情况。这种现象背后的原因很简单：当我们使用预条件子 $\mathcal{M}_{\eta,1/\eta}$ 时, 我们在第 k 次 CG 迭代步时需要除以 $p_k^T\mathcal{K}p_k$ (这里 p_k 是 k 次搜索方向)，当我们使用预条件子 $\mathcal{P}$ 时, 由于特殊内积 (4.4)的存在，我们在第 k 次 CG 迭代步时需要除以 $p_k^T A_\eta p_k$ (这里 p_k 是 k 次搜索方向)。图 (5.3) 展示了 $k=4$ 时矩阵 $\mathcal{K}$ 和 A_η 的小于 0.3 的特征值的分布情况。这些小的和负的特征值是导致迭代失败的主要原因(大多数特征值大于 0.3，他们并没有在图中被表示出来)。从图中可以看出, 红色点号部分 $\mathcal{K}$ 的特征值的数量比蓝色加号部分 A_η 的特征值的数量要多得多。这很清楚地解释了预条件子 $\mathcal{M}_{\eta,1/\eta}$ 下 CG 方法的高度不稳定性和预条件子 $\mathcal{P}$ 下 CG 方法的高度稳定性。

图 (5.4) 展示了参数 $\eta-k^2$ 的值对绝对值最小的特征值的大小的影响。从中可以看出，当 $\eta-k^2$ 太小时，$\eta-k^2$ 增大会让 A_η 的绝对值最小的特征值摸变大。

为了检查当参数 $k\le 4$ 时, 对于任意大小的网格, 我们是否能一直用预条件子 $\mathcal{P}$CG 方法，我

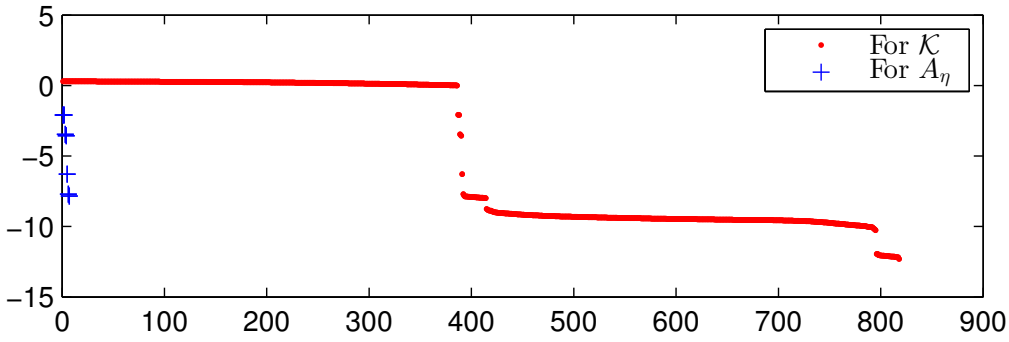
表 5.7: 第一列：$A+(\eta-k^2)M$-求解器/L-求解器的精度对；每个精度对的第一行：网格 $L4$ 上取不同波数 k 时方法 $\mathcal{P}$-CG ($\mathcal{M}_{\eta,\frac{1}{\eta}}$-$MINRES$) 所需时间；第二列：方法 $\mathcal{M}_{\eta,\frac{1}{\eta}}$-$MINRES$ 和 $\mathcal{P}$-CG 所花费的时间之间的比值。对于参数 η，如果 $k^2<1$，则 $\eta=1$，否则 $\eta=1.5k^2$。

k	0	1	1.2	1.25	2	4
1e-4/1e-5	3.4148(4.5445)	3.7516(5.6651)	4.6408(5.9792)	4.5615(5.9787)	5.9447(7.5844)	12.3785(13.2688)
	1.3308	1.5100	1.2884	1.3107	1.2758	1.0719
1e-3/1e-5	2.3986(3.1542)	3.2000(4.4446)	3.4226(4.7965)	3.7015(4.8802)	4.4020(5.4902)	10.1951(10.8177)
	1.3150	1.3889	1.4014	1.3185	1.2472	1.0611
1e-2/1e-5	1.5173(2.0689)	1.9599(2.6223)	2.6071(3.6012)	2.6189(3.2503)	3.0453(3.4254)	6.8830(8.1561)
	1.3635	1.3380	1.3813	1.2411	1.1248	1.1850
1e-1/1e-5	0.8272(1.3258)	1.2529(1.5368)	2.7211(4.9149)	1.8538(3.4134)	3.7716(8.3334)	7.5735(9.7269)
	1.6027	1.2266	1.8062	1.8413	2.2095	1.2843

表 5.8: 不同网格上矩阵 A_η 的最小特征值。参数满足 $\eta=k^2+1$. 虚线以上的数值值是正的（等价地，此时 A_η 是正定的)；虚线以下的数值值是负的（等价地，此时 A_η 不是正定的)。

k	G1	G2	G3	G4	k	L1	L2	L3	L4
0	0.4677	0.4738	0.4776	0.4769	0	0.4787	0.4582	0.4758	0.4654
1	0.4677	0.4738	0.4776	0.4769	1	0.2496	0.2575	0.2704	0.2753
1.55	0.0340	0.0360	0.0369	0.0373	1.2	0.0039	0.0128	0.0175	0.0200
1.6	-0.0544	-0.0543	-0.0536	-0.0533	1.25	-0.0646	-0.0556	-0.0530	-0.0512
2	-0.8719	-0.8988	-0.8875	-0.8823	2	-1.4349	-1.4249	-1.4580	-1.4674
4	-7.7031	-7.9434	-7.8425	-7.7907	4	-8.3127	-8.3664	-8.3974	-8.4450

图 5.3: 取参数 $k=4$, 并且 $\eta=k^2+1$ 时，在网格 $G3$ 上矩阵 $\mathcal{K}$（红色点号部分）和矩阵 A_η （蓝色加号部分）的最小特征值。相比于矩阵 A_η 的特征，$\mathcal{K}$ 的特征值中有更多的接近于零。



们在表(5.9)中调查了网格尺寸对矩阵 A_η 的模最小特征值的影响。我们可以观察到 A_η 的模最小特征值的模基本不受网格尺寸的影响。这个现象表明即使当参数 $k^2\geq\bar{\alpha}$ 时，我们的预条件仍然可以结合 CG 方法使用。总之，这些数值实验表明新的预条件子 $\mathcal{P}$ 下 CG 方法有很好的稳定性和收敛性。

图 5.4: 取参数 $k=4$, $\eta-k^2=1,2\cdots 20.$ 时，网格 $L4$ 上矩阵 A_η 的绝对值最小的特征值。这幅图说明参数 $\eta-k^2$ 不应该取得过小。

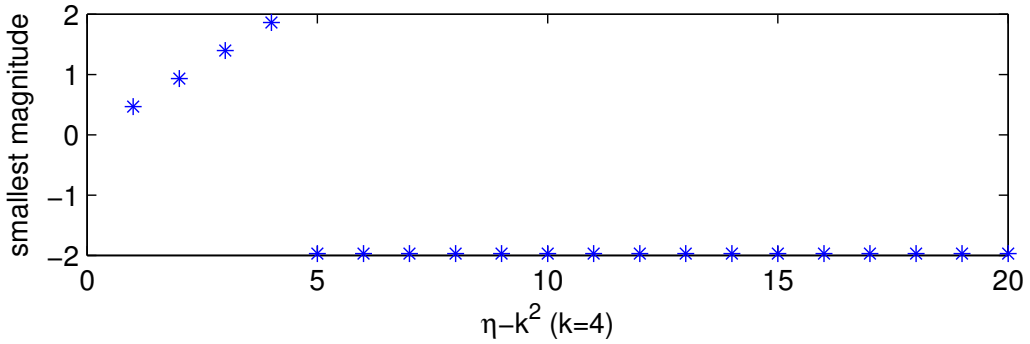
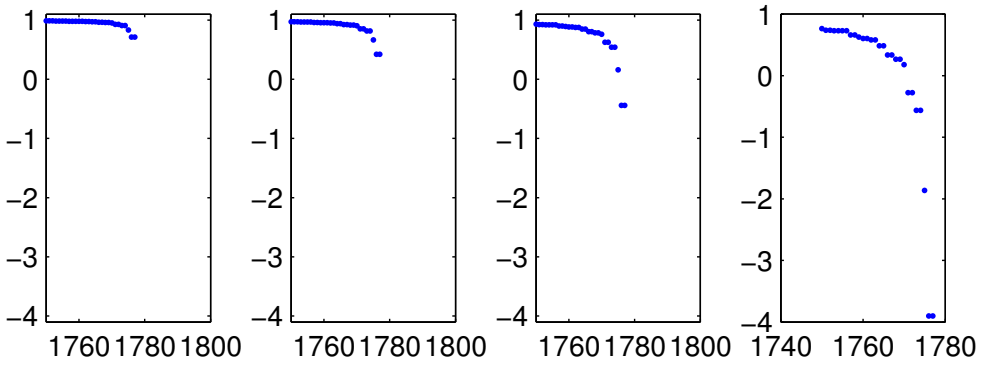


表 5.9: 不同网格上矩阵 A_η 的绝对值最小的特征值。这些特征值的模下有界。

网格	G1	G2	G3	G4	L1	L2	L3	L4
$k=4,\eta=17$	0.4677	0.4738	0.4776	0.4769	0.4787	0.4582	0.4758	0.4654
$k=4,\eta=24$	1.8963	-2.0601	-2.0846	-2.0966	-1.9134	-1.9650	-1.9311	-1.9705
$k=2,\eta=5$	0.4677	0.4738	0.4776	0.4769	-0.2541	-0.2640	-0.2705	-0.2692
$k=2,\eta=6$	0.5061	0.5310	0.5311	0.5338	-0.2540	-0.2640	-0.2705	-0.2692

图 5.5: 网格 $G3$ 上预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的最小的 27 个特征值的分布情况 (从左到右: $k=0,1,2,4$)。参数 η 满足 $\eta=k^2+1$. 参数 k 越大，分布情况越糟糕。



接下来我们说明在网格 G3 上预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的最小的 27 个特征值的分布情况。

图 (5.5)画出了网格 G3 上取不同特征值预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 的特征值的分布情况。从中可以看出 $k=0$ 和 $k=1$ 时，特征值有良好的边界，只有很少几个特征值在 0.22 和 0.8 之间，其余所有的特征值在 0.8 和 1 之间。这些数值结果与我们的理论预期相符 (定理 (4.2.1))；当取参数 $k=2$ 或 $k=4$ 时, 负的特征值会出现。波数越高，特征值的分布情况越糟糕。

图 5.6: 取参数 $k = 1.3$, $\varepsilon = -\frac{1}{\eta - k^2}$ 并且 $\eta = k^2 + 1$ 时，在网格 *G3* 上，预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ *(红色 + 号部分)* 和 $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$ *(蓝色 * 号部分)* 的特征值的分布情况。这两个分布完全重合。

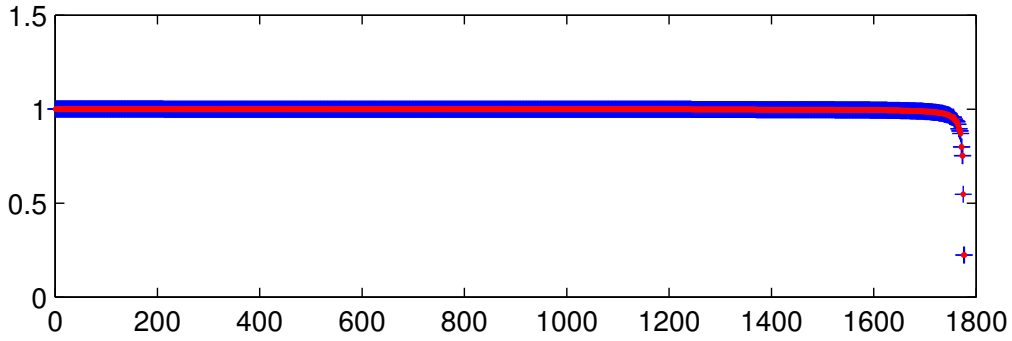


图 (5.6)表明, 在网格 G3 上, 当取参数 $k = 1.3$ 和 $\eta = k^2 + 1$ 时, 预条件矩阵 $\mathcal{P}^{-1}\mathcal{K}$ 和 $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$ 的特征值完全一样。

6 结果与展望

基于文献 [9] 和 [10] 中的结果，我们在这篇文章中提出了求解棱边元离散稳态麦克斯韦方程组和时谐麦克斯韦方程组后形成的鞍点问题的推广的预条件技术。新的参数 η 的引入使得整个方法适应性更强。这些预条件系统的谱的性质被简要地分析了, 并与块预条件系统做了对比。数值上我们发现新的方法从迭代时间上来说比原有的 MINRES 方法要优。我们在数值实验中也分析了新引进的参数 η 和非精确内迭代的精度对迭代的影响。

我们需要更多的数值实验来观察三维麦克斯韦问题中该方法的有效性，尤其是实际中涉及到间断系数, 复杂区域的情形。对于内迭代, 我们需要用高效的多重网格法取代不完全楚列斯基分解分解预条件子。另外，我们需要测试出这些情形下的最优参数 η 和内迭代精度。对于未来的可能的提高这些迭代法的效率的方法，我们有一些建议。首先，虽然所有 L-求解器使用相同精度已经很好了，不同的 L-求解器采用不同的收敛停止策略是值得考虑的。我们有不同的方式来施行 $\mathcal{P}^{-1}$ 作用于向量的动作，数值上他们需要比较一下效率，尤其是处理不同的右端项时。

最后需要指出的是我们所有的理论分析都基于准确的内迭代求解器。所以理论上为了用 CG 方法，我们可能需要精确的预条件内迭代。然而，数值实验表明，$A+(\eta-k^2)M$-求解器的精确不需要那么高也可以，而且此时非精确的内迭代能够显著地加速整个迭代过程。虽然我们需要更多的理论分析，并且可能含有不稳定的风险，我们认为未来对非精确内迭代下整个方法的稳定性进行更详细和创新的研究是有意义的。

我们的预条件推广到特定的非常数系数是容易的。实际上我们需要做的仅仅是验证(2.3.1)中的关系是否成立即可。而这部分在文献 [16] 中已经完成。

参考文献

[1] J. Xu *Iterative Methods by Space Decomposition and Subspace Correction,* SIAM Review, Vol. 34, No. 4 (Dec., 1992), pp. 581-613.

[2] Michele B. *Preconditioning Techniques for Large Linear Systems: A Survey,* Journal of Computational Physics, 182, 418–477 (2002)

[3] L. N. Trefethen 和 D. Bau, III, *Numerical Linear Algebra,* SIAM, Philadelphia, PA, 1997.

[4] S. F. Ashby, T. A. Manteuffel, and P. E. Saylor, *A taxonomy for conjugate gradient methods*, SIAM J. Numer. Anal., 27 (1990), pp. 1542–1568.

[5] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Vol. 44 of Springer Series in Computational Mathematics, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[6] Z. Chen, Q. Du, and J. Zou, *Finite element methods with matching and nonmatching meshes for Maxwell equations with discontinuous coefficients*, SIAM J. Numer. Anal., 37 (2000), pp. 1542–1570.

[7] G.-H. Cheng, T.-Z. Huang, and S.-Q. Shen, *Block triangular preconditioners for the discretized time-harmonic Maxwell equations in mixed form*, Comput. Phys. Commun., 180 (2009), pp. 192–196.

[8] L. Demkowicz and L. Vardapetyan, *Modeling of electromagnetic absorption/scattering problems using hp-adaptive finite elements*, Comput. Methods Appl. Mech. Engrg., 152 (1998), pp. 103–124.

[9] R. Estrin and C. Greif, *On nonsingular saddle-point systems with a maximally rank deficient leading block*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 367–384.

[10] C. Greif and D. Schötzau, *Preconditioners for the discretized time-harmonic Maxwell equations in mixed form*, Numer. Lin. Algebra Appl., 14 (2007), pp. 281–297.

[11] R. Hiptmair, *Finite elements in computational electromagnetism*, Acta Numerica, 11 (2002), pp. 237-339.

[12] R. Hiptmair and J. Xu, *Nodal Auxiliary Space Preconditioning in H(curl) and H(div) Spaces*, SIAM J. Numer. Anal., 45 (2007), pp. 2483–2509.

[13] P. Houston, I. Perugia, and D. Schötzau, *Mixed discontinuous Galerkin approximation*

of the Maxwell operator: non-stabilized formulation, J. Sci. Comput., 22-23 (2005), pp. 315–346.

[14] Q. Hu and J. Zou, *Substructuring preconditioners for saddle-point problems arising from Maxwell's equations in three dimensions*, Math. Comput., 73 (2004), pp. 35–61.

[15] T. Kolev and P. Vassilevski, *Some experience with a H^1-based auxiliary space AMG for H(curl) Problems*, Report UCRL-TR-221841, LLNL, Livermore, CA, 2006.

[16] D. Li, C. Greif, and D. Schötzau, *Parallel numerical solution of the time-harmonic Maxwell equations in mixed form*, Numer. Lin. Algebra Appl., 19 (2012), pp. 525–539.

[17] P. Monk, *Analysis of a finite element method for Maxwell's equations*, SIAM J. Numer. Anal., 29 (1992), pp. 714–729.

[18] J. C. Nédélec, *Mixed finite elements in* $\mathbb{R}^3$, Numer. Math., 35 (1980), pp. 315–341.

[19] B. Niceno, *EasyMesh*. `http://web.mit.edu/easymesh_v1.4/www/easymesh.html`.

[20] I. Perugia, D. Schötzau, and P. Monk, *Stabilized interior penalty methods for the time-harmonic Maxwell equations*, Comput. Methods Appl. Mech. Eng., 191 (2002), pp. 4675–4697.

[21] J. Pestana and A. J. Wathen, *Combination preconditioning of saddle point systems for positive definiteness*, Numer. Lin. Algebra Appl., 20 (2013), pp. 785–808.

[22] S. L. Wu, T. Z. Huang, and C. X. Li, *Modified block preconditioners for the discretized time-harmonic Maxwell equations in mixed form*, J. Comput. Appl. Math., 237 (2013), pp. 419–431.

[23] Y. Zeng and C. Li, *New preconditioners with two variable relaxation parameters for the discretized time-harmonic Maxwell wquations in mixed form*, Math. Comput. Probl. Eng., 2012 (2012), pp. 1–13.

[24] Tian Y. and Takane Y., *The inverse of any two-by-two nonsingular partitioned matrix and three matrix inverse completion problems*, Comp. Math. Appl., 2009, 57(8), pp. 1294–1304.

[25] *hypre* : High performance preconditioners. `http://www.llnl.gov/CASC/hypre/`.

[26] code: `https://github.com/ShiyangZhang/Preconditioners`.

[27] J. M. Miao, *General expressions for the Moore-Penrose inverse of a* 2×2 *block matrix,* Linear Algebra Appl., 151 (1991), pp. 1-15.

[28] Beck R, Hiptmair R., *Mulitlevel solution of the time-harmonic Maxwell' s equations based on edge elements,* International Journal for Numerical Methods in Engineering, 1999; 45:901–920.

[29] Gopalakrishnan J., Pasciak J., *Overlapping Schwarz preconditioners for indefinite time harmonic Maxwell equations,* Mathematics of Computation 2003; 72:1–15.

[30] Gopalakrishnan J., Pasciak J., *Demkowicz LF. Analysis of a multigrid algorithm for*

time harmonic Maxwell equations, SIAM Journal on Numerical Analysis 2004; 42(1):90–108.

[31] L.Zhong, S. Shu, J. Wang and J. Xu, *Two-grid methods for time-harmonic Maxwell equations*, Numer. Linear Algebra Appl. 2013; 20:93–111.

[32] P. Monk, *Finite element method for Maxwell's Equations,* Oxford University Press, New York, 2003.

致　　谢

这篇文章能够写完得益于许多人的帮助。首先，感谢武汉大学提供了宽松自由与友善友爱的环境，感谢校园的美好。同时，是导师向华老师指导我入门，从数值实验开始，一步步完成这份工作。生活中向老师也经常帮助学生。感谢向华老师。另外，也很感谢各位任课的老师，特别地，感谢邹军教授。最后，感谢同班同学和朋友们。这篇文章包含部分尚未发表的与邹军教授和向华教授的讨论结果。经过二零一四年到二零一七在武汉大学的四年间的准备，阅读，探讨，交流和写作，在许多人的帮助下，这篇硕士论文，尽管不完美，但是还是完成了。希望它能对其他人有所帮助。如果能这样，这份工作就是值得的。

武汉大学学位论文使用授权协议书

本学位论文作者愿意遵守武汉大学关于保存、使用学位论文的管理办法及规定, 即: 学校有权保存学位论文的印刷本和电子版, 并提供文献检索与阅览服务; 学校可以采用影印、缩印、数字化或其它复制手段保存论文; 在以教学与科研服务为目的前提下, 学校可以在校园网内公布部分及全部内容.

1、 在本论文提交当年, 同意在校园网内以及中国高等教育文献保障系统 (CALIS) 高校学位论文系统提供查询及前十六页浏览服务.

2、 在本论文提交 □ 当年/ □ 一年/ □ 两年 / □ 三年/ □ 五年以后, 同意在校园网内允许读者在线浏览并下载全文, 学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务.(保密论文解密后遵守此规定)

论文作者 (签名): _____

学　　号: _____

学　　院: _____

日期:　　年　　月　　日