OPTIMAL ORDER PRECONDITIONERS FOR MIXED AND NONCONFORMING FINITE ELEMENT APPROXIMATIONS OF ELLIPTIC PROBLEMS WITH ANISOTROPY

A Dissertation
by
SERGUEI MALIASSOV

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 1996

Major Subject: Mathematics

OPTIMAL ORDER PRECONDITIONERS FOR MIXED AND NONCONFORMING FINITE ELEMENT APPROXIMATIONS OF ELLIPTIC PROBLEMS WITH ANISOTROPY

A Dissertation
by
SERGUEI MALIASSOV

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

R.D. Lazarov
(Chair of Committee)

R.E. Ewing
(Member)

M. Pilant
(Member)

T. Strouboulis
(Member)

W. Rundell
(Head of Department)

May 1996

Major Subject: Mathematics

ABSTRACT

Optimal Order Preconditioners for Mixed and Nonconforming Finite Element
Approximations of Elliptic Problems with Anisotropy. (May 1996)
Serguei Maliassov, M.S., Moscow Institute of Physics and Technology;
Ph.D., Institute of Numerical Mathematics, Russia
Chair of Advisory Committee: Dr. Raytcho Lazarov

The general area of this thesis is preconditioning techniques for mixed and nonconforming finite element approximations of elliptic boundary value problems. A special emphasis is placed on problems in three dimensions with possibly large anisotropy in the coefficients of the PDE's along with large jumps in the coefficients across the interfaces separating subregions. The optimal preconditioners developed exploit the techniques of domain decomposition methods, algebraic substructuring, and multigrid methods. As a result, the proposed iterative processes converge with rates independent of the mesh size, the jumps of the coefficients, and the ratio of anisotropy.

Using an equivalence between nonconforming finite element methods and hybrid-mixed methods the iterative methods constructed for algebraic systems with symmetric positive definite matrices are extended to saddle-point problems which arise from mixed finite element approximations.

A new construction of iterative methods for nonconforming approximations of elliptic PDE's on nonmatching grids is proposed. The computational domain is considered as a union of nonintersecting subdomains. In each subdomain the grid is constructed in accordance with its own coordinate system using the main directions of anisotropy. The original elliptic problem is posed as a problem with Lagrange multipliers at the interface between the subdomains, which ensure the continuity conditions of the solution. A mortar finite element subspace is constructed in the space of Lagrange multipliers, which results in algebraic systems of a saddle-point type.

Based on the technique of domain decomposition and fictitious components methods a construction of block diagonal preconditioners for the algebraic systems arising in the mortar finite element method is developed. The fictitious components method is used to precondition subdomain problems, while the interface problems are preconditioned by an inner Chebyshev iterative procedure. It is shown that the developed preconditioner is spectrally equivalent to the original saddle-point matrix.

Applications of the newly developed iterative methods and preconditioning techniques are considered. In particularly, these methods are applied in the simulator of fluid flow in porous media.

DEDICATION

To my Parents Yurij and Valentina
my Wife Olga
and my Son Aleksander
for their Patience and Understanding

ACKNOWLEDGMENTS

My deepest gratitude is to my friend and advisor Raytcho Lazarov for the support and guidance he has provided throughout the years of my graduate study. His generous attitude, care, and devotion to his students helped me enormously. I wish also to thank Raytcho for his attention to detail and the valuable suggestions which helped to improve my dissertation considerably. His unlimited energy and availability, along with his wisdom, will be warmly remembered.

Also I am deeply grateful to Richard Ewing for his support and guidance. Dr. Ewing has always been generous to me by providing excellent opportunities for research and an environment of very high standards. Under his supervision I enjoyed a research assistantship for the entire period of my studies at Texas A&M University. Because of him I was involved in some of the most exciting projects in numerical analysis and scientific computing which were a great source of knowledge and experience.

I would like to thank Michael Pilant and Theofanis Strouboulis for serving on my graduate committee and being helpful to me in this regard.

I am deeply indebted to my former advisor Yurij Kuznetsov for first inspiring me to study numerical analysis. This endeavor would not have started without his help and assistance in becoming a graduate student at Texas A&M University. I am very pleased to thank him for his attention to my work and for a very fruitful collaboration during the years of my study. The opportunity to do research with a mathematician of his class was an exceptional inspiration for me. I wish also to thank him for all the care and consideration he provided for me.

In addition, I am very grateful to Jim Bramble, Serguei Nepomnyaschikh, Zhangxin Chen, and Panayot Vassilevski for their attention to my work. Discussions with them and their valuable advices helped me in my research and allowed me to improve the presentation of the results.

I also would like to acknowledge the following scientists: Peter Oswald, Joe Pasciak, and Yurij Vassilevski for helpful discussions.

From the people in the Institute for Scientific Computations, I would like to thank Michael Mann for his valuable comments on the style and presentation of the material in this dissertation. Also, I am very grateful to Sidney Outlaw and all the faculty and staff for creating a warm atmosphere for my education.

My final personal acknowledgement goes to my parents, my wife, and my son for everything they did for me in order to make this dissertation possible. I am deeply indebted to them for all of their love, sacrifice, help, and encouragement.

My graduate research was supported in part by the US Department of Energy under grant #DE–FG05–92ER25143. I am thankful for the valuable funding provided through this grant. However, no part of the research described in this dissertation was subject to an official review by the Department of Energy and therefore does not necessarily reflect its view; no official endorsement should be inferred.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

> [7] *All streams flow into the sea yet the sea is never full.*
> *To the place the streams come from, there they return again.*
> [8] *All things are wearisome, more than one can say.*
> *The eye never has enough of seeing, nor the ear its fill of hearing.*
>
> THE HOLY BIBLE. ECCLESIASTES. CHAPTER 1.

The numerical modeling of physical processes has played an increasing role in solving scientific and engineering problems in recent decades. Numerical methods for large algebraic systems play an essential role in the construction of efficient codes for modeling in computational fluid dynamics, elasticity, and other core areas of continuum mechanics. For example, even moderate resolution requirements for a relatively simple three-dimensional model of groundwater flow result in algebraic systems with millions of unknowns. Many important problems in science and engineering will require in the future much higher resolution than available at the present time. The importance of the algebraic solvers has increased dramatically with the arrival of powerful computing systems. Also, they have become a cornerstone in high performance computing. Therefore, the development of numerical methods for solving the resulting very large algebraic systems efficiently and with affordable computational cost is a challenging problem in numerical mathematics.

Many physical processes are described by elliptic partial differential equations (PDE). A classical example of such PDE in a bounded domain Ω is: $-\text{div}\,(K\nabla u) + c \cdot u = f$, where c, f are given functions, $K(x) = \{k_{ij}\}$ is a symmetric and positive definite coefficient matrix, and the unknown function u is subject to certain boundary conditions.

This problem can be discretized in various ways. Among the most popular and frequently used methods of approximation are the finite volume method, the Galerkin finite element method and the mixed finite element method. Each of these methods has its advantages and disadvantages when applied to particular engineering problems. For example, for petroleum reservoir simulations in geometrically simple domains and heterogeneous media, the finite volume method has proven to be reliable, accurate, and mass conserving cell-by-cell. Many engineering problems, e.g., petroleum recovery, ground-water contamination, seismic exploration, etc., need an accurate flux $\mathbf{q} = -K\nabla u$ calculation in the presence of heterogeneities, anisotropy and large jumps in the coefficient matrix $K(x)$. The mixed finite element method proposed by Raviart and Thomas in [101] is known for its accurate flux approximation.

The mixed methods for second order elliptic equations have been extensively studied in the last two decades. The popularity of these methods is due to the advantages they offer in solving problems from elasticity and fluid flow. Namely, along with the direct approximation of the flux $\mathbf{q}$ the mixed methods conserve mass locally element by element. A large variety of mixed finite element spaces on triangles, rectangles, prisms, and tetrahedrons has been proposed [101, 92, 27, 26] and their convergence and superconvergence properties have been studied [113, 67, 90, 43, 54].

As shown by Russell and Wheeler in [104], mixed finite element approximations on rectangular grids with special quadratures are equivalent to finite volume methods. The superconvergent velocity calculations for smooth solutions have been established by Weiser and Wheeler in [118]. Based on that equivalence, Bramble et al. in [21] have developed efficient multigrid solution procedures for mixed approximations on structured grids. However, in general the technique of the mixed finite element method leads to an algebraic saddle point problem that is more difficult and more expensive to solve compared to the problem with a symmetric and positive definite operator. Although some reliable iterative algorithms for the saddle point problems have been proposed and studied (see, e.g., [16, 19, 56, 105, 115]), their efficiency depends strongly on the geometry of the domain, the coefficient matrix $K(x)$, and the type of finite elements used.

An alternative approach can be taken by developing hybrid-mixed methods. This approach has been studied in the pioneering work of Arnold and Brezzi [5] where the continuity of the normal component of the flux vector to the boundary of each element is enforced by Lagrange multipliers. In general, the Lagrange multipliers on the element boundaries turn out to be none other than the trace of the primary unknown $u(x)$.

The important discovery of Arnold and Brezzi is that the hybrid-mixed method is equivalent to an approximation of the initial equation by the Galerkin method with nonconforming finite elements. Namely in [5] it is shown that the lowest-order Raviart-Thomas mixed element approximations are equivalent to the usual P_1-nonconforming finite element approximations when the classical P_1-nonconforming space is augmented with P_3-bubbles. Such a relationship has been studied recently for a large variety of mixed finite element spaces by Brenner, Chen, Cowsar, Arbogast, and many others (see, e.g., [4, 23, 32]). The importance of this study is that the algebraic system for the Lagrange multipliers has a symmetric and positive definite matrix.

Once the discretization method has been chosen, the next problem to address is solving the corresponding system of linear equations, which in general is very large. Today, in computer simulation of real-life processes, systems with hundreds of thousands (and even millions) of unknowns are usual and often encountered. Obviously, direct methods of solving such systems are not practical even on the most powerful computers. As an alternative to direct methods one should consider iterative methods.

The term "iterative method" refers to a wide range of techniques that use successive approximations to obtain more accurate solutions to a linear system at each step. The development of efficient iterative methods for systems arising from finite element discretizations of second-order partial differential equations has been a very active area of research over the last few decades. The rate at which an iterative method converges depends strongly on the spectrum of the coefficient matrix. At present, iterative methods usually involve a second matrix that transforms the coefficient matrix into one with a more favorable spectrum [44, 45, 46, 39, 117]. This transformation matrix is called a *preconditioner*. The use of a good preconditioner improves the convergence of the iterative method sufficiently to overcome the extra cost of constructing and applying the preconditioner. Today, the success of finite element methods to a large extent is due to the existence of fast and robust techniques for preconditioning and solving the corresponding discrete problems. Such efficient techniques for symmetric and positive definite matrices are well developed and studied [8, 9, 17, 45, 59, 57, 62, 68, 70, 80, 114, 120].

The equivalence between the hybrid-mixed and the nonconforming finite element methods establishes a framework for preconditioning and/or solving the algebraic problem arising from the mixed finite element method and for postprocessing the finite element solution. Schematically this framework includes the following three steps:

(a) forming the reduced algebraic problem for the Lagrange multipliers, which is equivalent to a nonconforming approximation;

(b) construction and study of efficient methods, based on multigrid, multilevel or domain decomposition for solving or preconditioning the reduced system;

(c) recovery of the solution $u(x)$ and the flux $\mathbf{q}$ from the Lagrange multipliers, which were already found, by using only element-by-element computations.

The recent progress in each of the steps described above (see, e.g., [5, 116, 107, 22, 37]) gives us an indication that the mixed finite element method can be used as an accurate and efficient tool for solving general elliptic problems of second order in domains with complicated geometry.

It follows that the most expensive part in solving the mixed problem numerically is to find the solution of a P_1-nonconforming problem. Thus, **the first main objective** of this dissertation is the development and study of efficient iterative techniques for nonconforming finite element approximations to boundary value problems of second-order self-adjoint linear elliptic PDE's. A special emphasis is placed on problems in three dimensions with a possibly large anisotropy in the coefficients. There is a variety of engineering applications where these methods can be very useful. Among them are problems that arise in contaminant transport, groundwater flow, and oil reservoir simulation. Other important applications are composite materials, phase transitions, polycrystalline dielectrics, and polyphased fluids.

Although the methods of solving the algebraic systems resulting from the nonconforming approximations have been extensively studied in the past few years (see, e.g., [5, 15, 20, 22, 35, 107]), their efficiency depends on the coefficient matrix $K(x)$. In the case of strong anisotropy in the coefficients the question of constructing effective solution techniques is still open.

In this dissertation we propose several preconditioners for P_1-nonconforming finite element approximations of anisotropic problems using ideas of substructuring proposed by Kuznetsov in [71]. These ideas make it possible to construct very efficient preconditioning techniques for problems with a high anisotropic ratio in the coefficients in domains of simple form such as a parallelepiped or a topological parallelepiped.

To construct the iterative methods for solving the anisotropic problems in domains of complex geometric shape we consider numerical methods which involve a solution of analogous problems in domains of relatively simple form. The known methods of such type are the Schwarz alternating subdomain methods [42, 88, 95, 76], the fictitious components method [6, 82, 85, 86], and methods based on matrix bordering [48, 40, 89, 94, 93, 100]. The methods which are based on the partitioning of the initial domain into subdomains are called domain decomposition methods (DD).

It is believed that the first DD method was proposed by Hermann Schwarz [108]. It was originally used to show the existence of the solution of an elliptic boundary value problem on domains that consist of the union of simple overlapping subdomains.

Recently, DD algorithms have become increasingly popular because they take full advantage of modern parallel computing technology. DD methods make it possible to solve the subdomain problems independently on different processors while exchanging information between them only from time to time. DD methods have an advantage of "natural parallelization" in comparison with any other effective method of solving elliptic boundary value problems. Exhausting results of the development of DD algorithms in the last decade can be found in the Proceedings of International Conferences on Domain Decomposition methods, and also in numerous papers (see, e.g., [14, 29, 49, 82, 86, 93, 111, 120]).

In general, the DD algorithms are based on variational methods for decomposing and solving elliptic problems. Most of the applications use discretization grids which are defined globally over the whole domain and then split into subdomains. In mechanics, this results in a conforming approximation of the primary variable. However, it might be more convenient and efficient to use approximations which are defined independently on each subdomain. This allows the user to make local and adaptive changes to the models, the approximation strategies, or the grids in one subdomain without modifying the other ones. This, off course, is possible if there is an adequate way of imposing the continuity (possibly in a weak sense) of both the fluxes and primary variables across such nonconforming interfaces.

In the presence of discontinuities and anisotropies in various subdomains of the computational domain, from a practical point of view it is attractive to have a local coordinate system in each particular subdomain with its axes aligned with the main directions of the anisotropy. However, this will require using different grids in different subdomains which are defined independently and which do not match at the interfaces. This concept has been considered in the mortar finite element method (see, e.g., [12, 110]). At present, there are already several approaches to the iterative solution of finite element systems on nonmatching grids, presented, for example, in [1, 2, 12, 72, 74]. Thus, **the second objective** of the dissertation is the construction of an iterative method for algebraic systems that occur when using nonmatching grids. The developed technique combines the ideas of the domain decomposition method [58, 111, 120] with the algorithms of multilevel and algebraic multigrid methods [8, 20, 60, 70].

The third objective of this dissertation is conducting numerical experiments to establish experimentally the conclusions from the theoretical analysis of the algorithms considered and to assess their effectiveness in terms of error reduction after a fixed number of iterations. Also the goal is to apply the newly developed iterative methods and preconditioning techniques to real-life problems such as the simulation of fluid flow in porous media [30].

The dissertation is organized as follows.

In Chapter II we consider a second-order elliptic boundary value problem and its discretization. It contains basic theoretical results concerning the differential problem and its finite element approximations. We begin in Section 2.1 with some definitions and results about Sobolev spaces for scalar and vector functions. In Section 2.2 we introduce a model elliptic problem and its discretization by the standard Galerkin finite element method. Then, in Section 2.3 we discuss the lowest order Raviart-Thomas mixed method and provide classical results concerning error estimates and properties of the discrete operators. Next, in Section 2.4, we review the Arnold-Brezzi theory which takes advantage of an equivalent hybrid formulation of the discrete mixed problem to reduce a symmetric indefinite problem to a positive definite one. The resulting problem is directly related to the P_1-nonconforming finite element problem. Finally, in Section 2.5 we outline the classes of problems for which we develop preconditioned iterative methods in the subsequent chapters.

In Chapter III we outline iterative techniques for solving systems of linear algebraic equations with both symmetric positive definite and indefinite matrices. For both kinds of systems in the next two chapters we develop efficient preconditioners. First, we consider some basic facts of the theory of iterative methods. Next, in Section 3.2 we give the formulae for the preconditioned Lanczos method [81] as applied to the solution of systems with symmetric indefinite matrices and discuss the choice of preconditioners for saddle-point matrices. Then, in Section 3.3 we discuss conjugate gradient type methods. Finally, in Section 3.4 we sketch the theory of the Chebyshev [57] methods which we use in Chapter V.

The main results of the dissertation are contained in Chapters IV and V.

In Chapter IV we consider two- and three-dimensional anisotropic problems with both constant and almost constant matrix coefficient $K(x)$ in the domains of a simple shape. Most of the theory developed in this chapter is based on the results published by the author in [77], and in joint works with R. Ewing, R. Lazarov, Yu. Kuznetsov, and Z. Chen in [52, 55, 33, 51, 73].

In the first section 4.1 we describe the idea of algebraic substructuring which we use to construct preconditioners. In Section 4.2 we consider a two-dimensional problem with a diagonal matrix coefficient $K(x)$. A detailed description of the algebraic substructuring preconditioners for three-dimensional problems is given in Sections 4.3 and 4.4. We formulate the model problem with a diagonal tensor, develop an algebraic substructuring preconditioner for the resulting linear system, and give an implementation algorithm. In Section 4.3 we define the partition of the whole domain, subdividing it into topological parallelepipeds and splitting each parallelepiped into **six** tetrahedra. The case of splitting each topological parallelepiped into **five** tetrahedra when $K(x)$ is a diagonal tensor is considered in Section 4.4. In Section 4.5 we consider the case of full tensor function $K(x)$ and domain Ω being a topological parallelepiped and develop a variant of the fictitious components method for anisotropic problems.

In Chapter V we present a construction of a domain decomposition method for solving systems of grid equations approximating boundary value problems for second order elliptic problems with anisotropic coefficients. We consider problems for which the computational domain Ω is a union of nonoverlapping subdomains $\Omega = \overset{m}{\underset{i=1}{\cup}} \Omega_i$ such that inside each Ω_i the equation coefficients vary insignificantly. We develop two different methods for the nonconforming approximations of anisotropic problems. This chapter is based on the results published in [78, 79].

In Section 5.2 we consider a variant of the block bordering method [89, 94] for the anisotropic problem. This algorithm uses the preconditioner developed in Chapter IV for the problems in subdomains. For the interface problems we construct a preconditioner in the form of an inner Chebyshev iterative procedure. More precisely, this is a preconditioner for the Schur complement of the original symmetric positive definite matrix, which is obtained after eliminating the block corresponding to the unknowns in the subdomains.

This approach combines the ideas of domain decomposition methods [14, 18, 29, 111, 120] and the algorithms of multilevel and algebraic multigrid methods [8, 20, 60, 70], with the bordering method for solving systems of mesh equations.

In Section 5.3 we propose iterative methods for solving systems of linear equations which arise in nonconforming finite element approximations of elliptic PDE's on nonmatching grids. More precisely, we use the technique of mortar finite elements (see, e.g., [1, 2, 12, 72, 109, 110]). The mortar element method is an optimal nonconforming domain decomposition method for the discretization of partial differential equations which provides for a maximum of mesh, refinement, and resolution flexibility while simultaneously preserving locality and elemental structure.

Using the results of Section 4.5, in each subdomain we construct its own coordinate system and grid (triangular in 2-D and tetrahedral in 3-D) in accordance with the main directions of anisotropy, so that the coefficient matrix is diagonal in the local coordinates. The original elliptic problem is posed as a problem with Lagrange multipliers at the interfaces between subdomains. A mortar finite element subspace is constructed in the space of Lagrange multipliers. The resulting algebraic systems have the form of a saddle-point problem. The iterative method involves a block diagonal preconditioner with the inner Chebyshev iterative procedure and the preconditioned Lanczos method as an outer iterative procedure.

Chapter VI is devoted to the applications of the theory developed in Chapters IV and V. We present the results from numerical experiments that illustrate this theory and apply it to the real-life problem of modeling fluid flow in porous media. In Section 6.1 we provide experiments for substructuring methods and the fictitious component method developed in Chapter IV. In Section 6.2 we consider experiments with the domain decomposition method on matching and nonmatching grids illustrating the theory of Chapter V. Finally, in Section 6.3 we consider an application of the Lagrange multiplier approach described in Chapter II to modeling the fluid flow in porous media.

In Chapter VII we summarize the results of the research presented for the defense.

CHAPTER II

DIFFERENTIAL AND FINITE ELEMENT PROBLEMS

In this chapter we consider a second-order elliptic boundary value problem and its discretization. It contains basic theoretical results concerning differential problems and finite element approximations.

The chapter is organized as follows. We begin in Section 2.1 with some basic definitions and useful results about Sobolev spaces for scalar and vector functions. In Section 2.2 we introduce the elliptic model problem and the corresponding discrete system using the standard Galerkin finite element method. Then, in Section 2.3 we discuss the lowest-order Raviart-Thomas mixed method and provide classical results concerning error estimates and properties of discrete operators. This formulation is useful in applications for which accurate approximation to the flux variable of the elliptic problem is required and where the solution (of the elliptic problem) is not sufficiently smooth. This is the case where there are highly discontinuous or anisotropic coefficients. Next, in Section 2.4, we review the Arnold-Brezzi theory which takes advantage of an equivalent hybrid formulation of the discrete mixed problem to reduce a symmetric indefinite problem to a positive definite one. The resulting problem is directly related to the nonconforming P_1 finite element problem. Finally, in Section 2.5 we formulate algebraic problems for which we develop preconditioned iterative methods in the subsequent chapters.

2.1 Sobolev spaces

Sobolev spaces play a fundamental role in studying partial differential equations. These spaces are very natural and often are used in analysis of scientific and engineering problems because norms in some such spaces have a sense of the energy of considered systems. Also, the existence of generalized solutions for many elliptic boundary problems is easily established using variational principles in some Sobolev spaces. The existence of classical solutions is accordingly transformed into the question of regularity of generalized solutions under appropriate boundary conditions. In the last decades Sobolev spaces have become very important in numerical analysis. They are used to answer questions related to the well-posedness of discrete systems and approximation properties of discrete solutions, or how close the discrete solution is to that of the continuous problem. In this dissertation the main use of Sobolev space theory is to analyze preconditioners for discrete systems.

Many elliptic boundary value problems arising in practice are formulated in domains, which are simple but not smooth. In finite element analysis, for instance, domains are very often composed of polyhedra. In domain decomposition methods we also encounter polyhedral substructures. Thus, it is natural to introduce Sobolev spaces for the class of *Lipschitz domains.*

Let $\mathcal{D}$ denote a domain, i.e. an open connected set, in $\mathbb{R}^d$, $d = 2, 3$. In later chapters $\mathcal{D}$ can be the whole region Ω, a substructure Ω_i, or a finite element τ.

Definition 2.1 (Lipschitz domain) *Let $\mathcal{D}$ be an open subset of $\mathrm{I\!R}^d$. $\mathcal{D}$ is a Lipschitz domain if for every $x \in \partial\mathcal{D}$ there exist a neighborhood $\mathcal{O} \subset \mathrm{I\!R}^d$ of x and a map ψ from $\mathcal{O}$ onto an open unit cube $\mathcal{C} = \{|\xi_i| < 1, i = 1, \ldots, d\}$ such that*

(A) ψ is injective;

(B) ψ together with ψ^{-1} (defined on $\mathcal{C}$) is Lipschitz continuous;

(C) $\mathcal{D} \cap \mathcal{O} = \{y \in \mathcal{D} : \xi_d = (\psi(y))_d < 0\}$, where $(\psi(y))_d$ denotes the d-th component of $\psi(y)$.

A consequence of (C) is that the boundary $\partial\mathcal{D}$ is defined locally by the equation $(\psi(y))_d = 0$.

We remark that if $\mathcal{D}$ is a bounded Lipschitz domain, we can select a finite number of pairs $(\mathcal{O}_i, \psi_i)$, $i = 1, \ldots, M$, to cover a neighborhood of $\partial\mathcal{D}$.

Definition 2.2 ($L^2(\mathcal{D})$) *Let u be a Lebesque measurable function and let $\mathcal{D}$ be a domain in $\mathrm{I\!R}^d$. The Hilbert space $L^2(\mathcal{D})$ is defined by the norm*

$$\|u\|^2_{L^2(\mathcal{D})} = \int_{\mathcal{D}} u^2 \, dx.$$

Definition 2.3 ($H^1(\mathcal{D})$) *Let $\mathcal{D}$ be a domain in $\mathrm{I\!R}^d$. The Sobolev space $H^1(\mathcal{D})$ is defined by the seminorm*

$$|u|^2_{H^1(\mathcal{D})} = \int_{\mathcal{D}} \nabla u \cdot \nabla u \, dx,$$

and the norm

$$\|u\|^2_{H^1(\mathcal{D})} = |u|^2_{H^1(\mathcal{D})} + \frac{1}{R^2_{\mathcal{D}}} \|u\|^2_{L^2(\mathcal{D})},$$

where $R_{\mathcal{D}}$ is a diameter of $\mathcal{D}$. Here ∇u has to be understood in the distributional sense.

The scale factor $R_{\mathcal{D}}$ is obtained by a change of variables beginning with the standard definition of the norm for a domain with unit diameter.

Definition 2.4 ($H^s(\mathcal{D})$, $0 < s < 1$) *Let $\mathcal{D}$ be a domain in $\mathrm{I\!R}^d$. The fractional order Sobolev space $H^s(\mathcal{D})$, $0 < s < 1$, is defined as the space of all $u \in L^2(\mathcal{D})$ such that*

$$|u|^2_{H^s(\mathcal{D})} = \int_{\mathcal{D}} \int_{\mathcal{D}} \frac{|u(x) - u(y)|^2}{|x - y|^{d+2s}} \, dx \, dy < \infty \tag{2.1}$$

with the norm

$$\|u\|^2_{H^s(\mathcal{D})} = |u|^2_{H^s(\mathcal{D})} + \frac{1}{R^{2s}_{\mathcal{D}}} \|u\|^2_{L^2(\mathcal{D})}.$$

For a bounded Lipschitz domain $\mathcal{D}$ it can be shown [3, 61, 91] that the space $H^s(\mathcal{D})$ is a completion of the space $C^\infty(\mathcal{D})$ (or $C^\infty(\bar{\mathcal{D}})$) with respect to the norm $\|\cdot\|_{H^s(\mathcal{D})}$. The space $C^\infty(\mathcal{D})$ consists of the infinitely continuously differentiable functions defined in $\mathcal{D}$. The space $C^\infty(\bar{\mathcal{D}}) \in C^\infty(\mathcal{D})$ is the restriction of $C^\infty(\mathrm{I\!R}^d)$ into $\bar{\mathcal{D}}$.

For a domain $\mathcal{D}$ the space $H^s_0(\mathcal{D}) \in H^s(\mathcal{D})$ is defined as the completion of $C^\infty_0(\mathcal{D})$ with respect to $\|\cdot\|_{H^s(\mathcal{D})}$. Here the space $C^\infty_0(\mathcal{D})$ is the subspace of $C^\infty(\mathcal{D})$ of functions with compact support in $\mathcal{D}$. For a bounded Lipschitz domain it can be shown [61] that $H^s(\mathcal{D}) = H_0^s(\mathcal{D})$, for $0 \leq s < 1/2$.

Next, we state a lemma which makes it possible to extend results from a d-dimensional cube or a smooth domain to a bounded Lipschitz domain.

Lemma 2.1 *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be bounded domains, and let ψ be a bi-Lipschitz map from $\bar{\mathcal{D}}_1$ to $\bar{\mathcal{D}}_2$. Then, for $u \in H^s(\mathcal{D}_2)$, $0 \leq s \leq 1$,*

$$c_0 |u \circ \psi|_{H^s(\mathcal{D}_1)} \leq |u|_{H^s(\mathcal{D}_2)} \leq c_1 |u \circ \psi|_{H^s(\mathcal{D}_1)}.$$

The proof of this lemma for $s = 0, 1$ can be found in [91]. For an intermediate s it can be proved using explicit presentation (2.1) (see, e.g., [61]).

We shall need Sobolev spaces on manifolds such as $\partial\mathcal{D}$ or an open subset $\Gamma \in \partial\mathcal{D}$. Let $\mathcal{D}$ be a bounded Lipschitz domain in $\mathbb{R}^d$. Then an outward vector normal to $\partial\mathcal{D}$ is defined almost everywhere with respect to a hypersurface measure dS [61]. This measure is uniquely defined in terms of the d-dimensional Lebesque measure dx and $\partial\mathcal{D}$. For domains with a piecewise smooth boundary dS coincides with the standard notion of surface area.

Definition 2.5 **($\|u\|_{L^2(\Gamma)}$)** *Let $\mathcal{D}$ be a bounded Lipschitz domain and Γ be an open subset of $\partial\mathcal{D}$. Let u be a measurable function with respect to the hypersurface measure dS. The Hilbert space $L^2(\Gamma)$ is defined by the norm*

$$\|u\|^2_{L^2(\Gamma)} = \int\limits_{\Gamma} u^2 \; dS. \tag{2.2}$$

We introduce now the concept of trace maps. We have an obvious definition of boundary values, or trace on $\partial\mathcal{D}$, for functions in $C^\infty(\bar{\mathcal{D}})$. These maps can be generalized to functions in $H^1(\mathcal{D})$ for a bounded Lipschitz region $\mathcal{D}$ [75, 91].

Lemma 2.2 (Trace and Extension theorem) *Let $\mathcal{D}$ be a bounded Lipschitz domain. The trace map $\gamma : u \to u|_{\partial\mathcal{D}}$, defined for $C^\infty(\bar{\mathcal{D}})$, has a unique continuous extension from $H^1(\mathcal{D})$ onto $H^{1/2}(\partial\mathcal{D})$. This operator has a right continuous inverse.*

Using this definition of the map γ we can introduce a seminorm for the space $H^{1/2}(\partial\mathcal{D})$.

Definition 2.6 **($H^{1/2}(\Gamma)$)** *Let $\mathcal{D}$ be a bounded Lipschitz domain in $\mathbb{R}^d$. Let u be a square integrable function with respect to the hypersurface measure dS. We define the norm and seminorm for the space $H^{1/2}(\partial\mathcal{D})$ by*

$$|u|_{H^{1/2}(\partial\mathcal{D})} = \inf_{v \in H^1(\mathcal{D}),\; \gamma v = u} |v|_{H^1(\mathcal{D})}, \tag{2.3}$$

and

$$\|u\|^2_{H^{1/2}(\partial\mathcal{D})} = |u|^2_{H^{1/2}(\partial\mathcal{D})} + \frac{1}{R_{\mathcal{D}}} \|u\|^2_{L^2(\partial\mathcal{D})}, \tag{2.4}$$

respectively.

We now introduce spaces that are used in the mixed formulation of elliptic problems.

Definition 2.7 ($H^{-1/2}(\partial\mathcal{D})$) *The dual space of $H^{1/2}(\partial\mathcal{D})$ is denoted by $H^{-1/2}(\partial\mathcal{D})$ and is a Hilbert space with the norm*

$$\|u\|_{H^{-1/2}(\partial\mathcal{D})} = \sup_{v\in H^{1/2}(\partial\mathcal{D})} \frac{< u,\, v >_{\partial\mathcal{D}}}{\|v\|_{H^{1/2}(\partial\mathcal{D})}},$$

where $< \cdot, \cdot >_{\partial\mathcal{D}}$ is the duality pairing between $H^{1/2}(\partial\mathcal{D})$ and $H^{-1/2}(\partial\mathcal{D})$.

Definition 2.8 ($H(\mathrm{div}\,;\mathcal{D})$) *The space $H(\mathrm{div}\,;\mathcal{D})$ is defined by*

$$H(\mathrm{div}\,;\mathcal{D}) = \left\{ \mathbf{p} = \{p_i\}_{i=1}^d \in \left(L^2(\mathcal{D})\right)^d : \mathrm{div}\,\mathbf{p} \in L^2(\mathcal{D}) \right\}$$

and is a Hilbert space with the usual graph norm

$$\|\mathbf{p}\|^2_{H(\mathrm{div};\mathcal{D})} = \sum_{i=1}^d \|p_i\|^2_{L^2(\mathcal{D})} + \|\mathrm{div}\ \mathbf{p}\|^2_{L^2(\mathcal{D})}.$$

Since for a Lipschitz domain the unit normal $\mathbf{n}$ to the boundary $\partial\mathcal{D}$ is defined almost everywhere, for a smooth vector function $\mathbf{p}$ in $\mathcal{D}$ the normal component $\mathbf{p}\cdot\mathbf{n}$ is defined almost everywhere on $\partial\mathcal{D}$. The following lemma extends the notion of the normal component to functions in $H(\mathrm{div}\,;\mathcal{D})$.

Lemma 2.3 (Trace and Extension theorem for $H(\mathrm{div}\,;\mathcal{D})$) *Let $\mathcal{D}$ be a bounded Lipschitz domain. The trace map $\gamma_n : \mathbf{p} \to \mathbf{p}|_{\partial\mathcal{D}}$, defined a priori from $\left(H^1(\Omega)\right)^d$ into $L^2(\partial\mathcal{D})$, has a unique continuous extension from $H(\mathrm{div}\,;\mathcal{D})$ onto $H^{-1/2}(\partial\mathcal{D})$. As a consequence, the following characterization of the norm for functions $\mu \in H^{-1/2}(\partial\mathcal{D})$ is valid:*

$$\|\mu\|_{H^{-1/2}(\partial\mathcal{D})} = \inf_{\mathbf{p}\in H(\mathrm{div};\mathcal{D}),\ \mathbf{p}\cdot\mathbf{n}=\mu} \|\mathbf{p}\|_{H(\mathrm{div};\mathcal{D})}. \tag{2.5}$$

A demonstration of the first part of this theorem can be found in [112]. The characterization of the norm was given in [113]. To avoid, whenever possible, working in this space $H^{-1/2}(\partial\mathcal{D})$ which contains all the functions of $L^2(\partial\mathcal{D})$, we define, with the aid of Lemma 2.3, the space

$$\mathcal{H}(\mathrm{div}\,;\mathcal{D}) = \left\{ \mathbf{p} \in H(\mathrm{div}\,;\mathcal{D}) : \mathbf{p}\cdot\mathbf{n} \in L^2(\partial\mathcal{D}) \right\}, \tag{2.6}$$

which is a Hilbert space with norm

$$\|\mathbf{p}\|^2_{\mathcal{H}(\mathrm{div};\mathcal{D})} = \|\mathbf{p}\|^2_{H(\mathrm{div};\mathcal{D})} + \|\mathbf{p}\cdot\mathbf{n}\|^2_{L^2(\partial\mathcal{D})}.$$

Then, we have Green's formula.

Lemma 2.4 (Green's Formula) *Let $\mathcal{D}$ be a bounded Lipschitz domain. Let $\mathbf{p} \in \mathcal{H}(\mathrm{div}\,;\mathcal{D})$. Then,*

$$\int_{\mathcal{D}} (\nabla v \cdot \mathbf{p} + v\ \mathrm{div}\,\mathbf{p})\ dx = \int_{\partial\mathcal{D}} v\ (\mathbf{p}\cdot\mathbf{n})\ dS, \qquad \forall v \in H^1(\mathcal{D}). \tag{2.7}$$

The following abstract lemma [36, 61] allows us to show well-posedness of certain elliptic problems.

Lemma 2.5 (Lax-Milgram Lemma) *Let B be a bilinear form on a Hilbert space $\mathcal{H}$. Assume that B is bounded*

$$|B(w,v)| \leq C \cdot \|w\|_{\mathcal{H}} \cdot \|v\|_{\mathcal{H}}, \qquad \forall w, v \in \mathcal{H}$$

and coercive, i.e. there exists a $\nu > 0$ such that

$$B(v,v) \geq \nu \|v\|_{\mathcal{H}}^2, \qquad \forall v \in \mathcal{H}.$$

Then, for every bounded functional $f \in \mathcal{H}^*$ there exists a unique element $u_f \in \mathcal{H}$ such that

$$B(u_f, v) = f(v), \qquad \forall v \in \mathcal{H},$$

and

$$\|u_f\|_{\mathcal{H}} \leq \frac{1}{\nu} \|f\|_{\mathcal{H}^*}.$$

The counterpart of the Lax-Milgram Lemma for certain saddle-point problems is given by (cf. Brezzi and Fortin [25])

Lemma 2.6 (Babuska-Brezzi Lemma) *Let V and Q be Hilbert spaces with the norms $\|\cdot\|_V$ and $\|\cdot\|_Q$, respectively. Let $a(\cdot,\cdot)$ be a continuous bilinear form on $V \times V$ and $b(\cdot,\cdot)$ be a continuous bilinear form on $V \times Q$. Let us define an operator $B : V \to Q'$ by $(B\mathbf{p}, v) = b(\mathbf{p}, v)$ for all $\mathbf{p} \in V$ and $v \in Q$, and assume that the range of B is closed in Q', i.e. there exists a constant $\alpha_0 > 0$ such that*

$$\sup_{\mathbf{p} \in V} \frac{b(\mathbf{p}, v)}{\|\mathbf{p}\|_V} \geq \alpha_0 \|v\|_{Q \setminus \mathrm{Ker}\ B^T} = \alpha_0 \left(\inf_{v_0 \in \mathrm{Ker}\ B^T} \|v + v_0\|_Q \right), \forall v \in Q. \tag{2.8}$$

Let us also suppose that the operator defined by the bilinear form $a(\cdot,\cdot)$ is elliptic on Ker B, *i.e. there exists a constant $\alpha_1 > 0$ such that*

$$\begin{cases} \inf\limits_{\mathbf{q}_0 \in \mathrm{Ker}\ B} \sup\limits_{\mathbf{p}_0 \in \mathrm{Ker}\ B} \dfrac{a(\mathbf{q}_0, \mathbf{p}_0)}{\|\mathbf{q}_0\|_V \cdot \|\mathbf{p}_0\|_V} \geq \alpha_1, \\ \inf\limits_{\mathbf{p}_0 \in \mathrm{Ker}\ B} \sup\limits_{\mathbf{q}_0 \in \mathrm{Ker}\ B} \dfrac{a(\mathbf{q}_0, \mathbf{p}_0)}{\|\mathbf{q}_0\|_V \cdot \|\mathbf{p}_0\|_V} \geq \alpha_1. \end{cases} \tag{2.9}$$

Then the problem: find $(\mathbf{p}, u) \in V \times Q$ such that

$$\begin{cases} a(\mathbf{p}, \mathbf{q}) + b(\mathbf{q}, u) &= g(\mathbf{q}), \qquad \forall \mathbf{q} \in V, \\ b(\mathbf{p}, v) &= f(v), \qquad \forall v \in Q, \end{cases} \tag{2.10}$$

has a solution $(\mathbf{p}, u)$ for any $g \in V'$ and for any $f \in$ Im B. *The first component $\mathbf{p}$ is unique and the second component u is defined up to an element of* Ker B^T. *Furthermore,*

$$\|\mathbf{p}\|_V \leq \frac{1}{\alpha_1} \|g\|_{V'} + \frac{1}{\alpha_0} \left(1 + \frac{\|a\|}{\alpha_1}\right) \|f\|_{Q'}, \tag{2.11}$$

and

$$\|u\|_{Q \setminus \mathrm{Ker}\ B^T} \leq \frac{1}{\alpha_0} \left(1 + \frac{\|a\|}{\alpha_1}\right) \|g\|_{V'} + \frac{\|a\|}{\alpha_0^2} \left(1 + \frac{\|a\|}{\alpha_1}\right) \|f\|_{Q'}. \tag{2.12}$$

We note that the conditions (2.8) and (2.9) of this Lemma are not only sufficient but also necessary for the existence of a solution (cf. Brezzi [24]).

2.2 Elliptic problem

In this section we formulate a model elliptic problem, give its weak formulation, and introduce the standard Galerkin finite element discretization.

2.2.1 Formulation of the problem

Let Ω be a convex bounded domain in $\mathbb{R}^d$, $d = 2, 3$, with boundary $\partial\Omega$. Consider an elliptic problem

$$\begin{aligned} -\mathrm{div}\,(K\nabla u) &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma_0, \\ (K\nabla u, \mathbf{n}) &= 0 && \text{on } \Gamma_1, \end{aligned} \tag{2.13}$$

where $f(\mathbf{x}) \in L^2(\Omega)$, $\overline{\Gamma_0 \cup \Gamma_1} = \partial\Omega$, $\Gamma_0 \cap \Gamma_1 = \emptyset$, and Γ_0 is a nonempty closed subset of $\partial\Omega$, i.e. $\Gamma_0 \equiv \overline{\Gamma_0} \neq \emptyset$. We assume that $K(\mathbf{x})$ is a $d \times d$ symmetric, uniformly positive definite matrix-valued Lebesque measurable function bounded above on Ω such that

$$0 < \lambda_{\min}|\boldsymbol{\xi}|^2 \leq \boldsymbol{\xi}^T K(\mathbf{x})\boldsymbol{\xi} \leq \lambda_{\max}|\boldsymbol{\xi}|^2 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^d \text{ a.e. } \mathbf{x} \in \Omega. \tag{2.14}$$

Note that approaches considered in the dissertation are valid also for the case of the Neumann problem, i.e. $\Gamma_0 = \emptyset$ but for the sake of simplicity are not described here.

Let us consider a space $V_0(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_0\}$, and define a bilinear form $a(\cdot,\cdot)$ by

$$a(u,v) = (K\ \nabla u, \nabla v), \qquad u, v \in V_0(\Omega), \tag{2.15}$$

where $(\cdot,\cdot)$ denotes the $L^2(\Omega)$ inner product. Then the usual weak formulation of (2.13) in $V_0(\Omega)$ is: *given $f \in L^2(\Omega)$ find $u \in V_0(\Omega)$ such that*

$$a(u,v) = (f,v), \qquad \forall v \in V_0(\Omega). \tag{2.16}$$

This problem is uniquely solvable. It can be shown by checking the hypothesis of the Lax-Milgram Lemma with $\mathcal{H} = V_0(\Omega)$ and $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_{H^1(\Omega)}$.

The boundedness is obtained by

$$a(w,v) \leq \lambda_{\max}|w|_{H^1(\Omega)} \cdot |v|_{H^1(\Omega)} \leq \lambda_{\max}\|w\|_{H^1(\Omega)} \cdot \|v\|_{H^1(\Omega)}.$$

The coercivity is obtained by using Friedrich's inequality for the space $V_0(\Omega)$.

$$a(v,v) \geq \lambda_{\min}|v|^2_{H^1(\Omega)} \geq c(\Omega)\lambda_{\min}\|v\|^2_{H^1(\Omega)} \qquad \forall v \in V_0(\Omega).$$

Hence, the solution u of (2.16) satisfies

$$\|u\|_{H^1(\Omega)} \leq C(\Omega)\frac{\|f\|_{H^{-1}(\Omega)}}{\lambda_{\min}} \leq C(\Omega)\frac{\|f\|_{L^2(\Omega)}}{\lambda_{\min}}. \tag{2.17}$$

2.2.2 Galerkin finite element method

Now we outline the Galerkin finite element method for problem (2.16). First, we introduce a conforming quasi-uniform triangulation of Ω [36] by dividing the domain into simplices $\left\{\tau_i^h\right\}_{i=1}^M$, with diameters of order h. In this dissertation we will consider as simplices either triangles in $\mathbb{R}^2$ or tetrahedra in $\mathbb{R}^3$. We denote this triangulation by $\mathcal{T}_h$. The collection of simplex vertices belonging to $\bar{\Omega} \setminus \Gamma_0$ we denote by $\{\mathbf{x}_i\}_{i=1}^N$.

Let $V^h(\Omega)$ be the finite element space of continuous piecewise linear functions defined on the triangulation $\mathcal{T}_h$, and let $V_0^h(\Omega)$ be the subspace of $V^h(\Omega)$ of functions which vanish on Γ_0.

The finite element approximation to the solution u of problem (2.16) is given by: *find* $u^h \in V_0^h(\Omega)$ *such that*

$$a(u^h, v) = (f, v), \qquad \forall v \in V_0^h(\Omega). \tag{2.18}$$

The well-posedness of problem (2.18) follows directly from the Lax-Milgram Lemma. We also have a stability result as in (2.17) with a constant independent of the mesh size h.

Let a set of functions $\{\varphi_i(\mathbf{x})\}_{i=1}^N$ be a nodal basis for $V_0^h(\Omega)$. Then, every function $v \in V_0^h(\Omega)$ is represented by $v(\mathbf{x}) = \sum_{i=1}^N v_i \varphi_i(\mathbf{x})$. The choice of the basis in $V_0^h(\Omega)$ induces a one-to-one correspondence between the functions from $V_0^h(\Omega)$ and the vectors of the linear space $\mathbb{R}^N$. Thus, (2.18) becomes

$$\sum_{i=1}^N u_i a(\varphi_i, \varphi_j) = (f, \varphi_j), \qquad j = 1, \ldots, N,$$

or in matrix presentation

$$A\mathbf{u} = \mathbf{f}, \tag{2.19}$$

where $A_{ji} = a(\varphi_i, \varphi_j)$, $f_j = (f, \varphi_j)$, $i, j = 1, \ldots, N$.

Note, that the symmetric and positive definite matrix A can also be defined by an expression:

$$(A\mathbf{u}, \mathbf{v}) = a(u, v), \qquad \forall u, v \in V_0^h(\Omega), \tag{2.20}$$

where the vectors $\mathbf{u}$ and $\mathbf{v}$ correspond to the finite element functions u and v, respectively.

2.3 Saddle-point problem

2.3.1 Motivation

Many engineering problems, e.g., petroleum recovery, ground-water contamination, seismic exploration, etc., need very accurate flux $\mathbf{q} = -K\nabla u$ calculation in the presence of heterogeneities, anisotropy and large jumps in the coefficient matrix $K(\mathbf{x})$. Here u is the solution of an elliptic problem with the coefficient K. More accurate and direct approximation of the velocity can be achieved through the use of the mixed finite element method [101, 25].

First, we introduce a new independent variable $\mathbf{q}$ and rewrite problem (2.13) in the form:

$$\begin{array}{rcll} \mathbf{q} + K\nabla u & = & \mathbf{0} & \text{in } \Omega, \\ \operatorname{div} \mathbf{q} & = & f & \text{in } \Omega, \\ u & = & 0 & \text{on } \Gamma_0, \\ -(\mathbf{q}, \mathbf{n}) & = & 0 & \text{on } \Gamma_1. \end{array} \tag{2.21}$$

The mixed finite element method is based on the approximation of the weak form of this first-order system. Along with the direct approximation of the flux $\mathbf{q}$, mixed methods have another advantage in comparison with other methods — they conserve mass locally element by element.

2.3.2 Mixed method

Assume that $f \in L^2(\Omega)$. It is easy to see that if u is the solution of (2.16) then

$$\mathbf{q} = -K\nabla u \quad \in \quad \mathcal{H}(\mathrm{div}\,;\Omega). \tag{2.22}$$

Now we use Green's formula (2.7) and the argument that $\mathcal{H}(\mathrm{div}\,;\Omega)$ is dense in $H(\mathrm{div}\,;\Omega)$ to see that $(\mathbf{q}, u)$ is a solution of the following mixed formulation of (2.16):

find $(\mathbf{q}, u) \in H(\mathrm{div}\,;\Omega) \times L^2(\Omega)$ *such that*

$$\begin{aligned} \int_\Omega K^{-1}\mathbf{q}\cdot\mathbf{p}\,dx - \int_\Omega u\,\mathrm{div}\,\mathbf{p}\,dx &= 0, && \forall \mathbf{p} \in H(\mathrm{div}\,;\Omega), \\ -\int_\Omega v\,\mathrm{div}\,\mathbf{q}\,dx &= -\int_\Omega f\,v\,dx, && \forall v \in L^2(\Omega). \end{aligned} \tag{2.23}$$

We show the well-posedness for problem (2.23) by checking the hypothesis of Lemma 2.6 with $Q = L^2(\Omega)$, $V = H(\mathrm{div}\,;\Omega)$, and

$$a(\mathbf{q},\mathbf{p}) = \int_\Omega K^{-1}\mathbf{q}\cdot\mathbf{p}\,dx, \qquad b(\mathbf{q}, v) = -\int_\Omega v\,\mathrm{div}\,\mathbf{q}\,dx. \tag{2.24}$$

Using the definition of B we have

$$\mathrm{Ker}\ B = \{\mathbf{p}_0 \in H(\mathrm{div}\,;\Omega) : \mathrm{div}\,\mathbf{p}_0 = 0\}. \tag{2.25}$$

We also have

$$\mathrm{Im}\ B = L^2(\Omega). \tag{2.26}$$

To show (2.26) we have to show that for every $f \in L^2(\Omega)$ there exists a $\mathbf{p} \in H(\mathrm{div}\,;\Omega)$ such that $B\mathbf{p} = f$. Indeed, in the first step we find $w \in V_0(\Omega)$ such that

$$\int_\Omega \nabla w \cdot \nabla\psi\,dx = \int_\Omega f\psi\,dx, \qquad \forall\psi \in V_0(\Omega),$$

and then, set $\mathbf{p} = \nabla w$.

Using (2.26) we easily obtain that Ker $B^T = 0$.

We also obtain $\|a\| \le 1/\lambda_{\min}$ since for any $\mathbf{p},\mathbf{q} \in \mathcal{H}(\mathrm{div}\,;\Omega)$ we have

$$a(\mathbf{p},\mathbf{q}) \le \frac{1}{\lambda_{\min}}\|\mathbf{p}\|_{L^2(\Omega)}\|\mathbf{q}\|_{L^2(\Omega)} \le \frac{1}{\lambda_{\min}}\|\mathbf{p}\|_{H(\mathrm{div};\Omega)}\|\mathbf{q}\|_{H(\mathrm{div};\Omega)}.$$

We obtain an estimation $\alpha_1 \ge 1/\lambda_{\max}$ since

$$\begin{aligned} \inf_{\mathbf{q}_0\in\mathrm{Ker}\ B}\ \sup_{\mathbf{p}_0\in\mathrm{Ker}\ B} \frac{a(\mathbf{q}_0,\mathbf{p}_0)}{\|\mathbf{q}_0\|_V\|\mathbf{p}_0\|_V} &\ge \inf_{\mathbf{q}_0\in\mathrm{Ker}\ B} \frac{a(\mathbf{q}_0,\mathbf{q}_0)}{\|\mathbf{q}_0\|_V\|\mathbf{q}_0\|_V} \\ &= \inf_{\mathbf{q}_0\in\mathrm{Ker}\ B} \frac{a(\mathbf{q}_0,\mathbf{q}_0)}{\|\mathbf{q}_0\|_{L^2(\Omega)}\|\mathbf{q}_0\|_{L^2(\Omega)}} \ge \frac{1}{\lambda_{\max}}. \end{aligned}$$

To show that $\alpha_0 \ge C_1(\Omega) > 0$, we first note that

$$\sup_{\mathbf{p}\in V}\frac{b(\mathbf{p},v)}{\|\mathbf{p}\|_V} \ge \frac{b(\boldsymbol{\xi},v)}{\|\boldsymbol{\xi}\|_V},$$

where $\boldsymbol{\xi} = \nabla w$, and w is the solution of the problem

$$\int\limits_{\Omega} \nabla w \cdot \nabla \psi \, dx = \int\limits_{\Omega} - v\psi \, dx, \qquad \forall \psi \in V_0(\Omega).$$

From Schwarz and Poincaré inequalities we have

$$\|\nabla w\|^2_{L^2(\Omega)} \leq \|w\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \leq C_2(\Omega) \|\nabla w\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)}.$$

So, using the estimate $\|\nabla w\|_{L^2(\Omega)} \leq C_2(\Omega)\|v\|_{L^2(\Omega)}$, we obtain

$$\begin{aligned} \frac{b(\boldsymbol{\xi}, v)}{\|\boldsymbol{\xi}\|_V} &= \frac{-\int_\Omega v \operatorname{div} \boldsymbol{\xi} \, dx}{\|\boldsymbol{\xi}\|_V} = \frac{\int_\Omega v^2 \, dx}{\|\nabla w\|_{H(\mathrm{div};\Omega)}} = \frac{\int_\Omega v^2 \, dx}{(\|\Delta w\|^2_{L^2(\Omega)} + \|\nabla w\|^2_{L^2(\Omega)})^{1/2}} \\ &= \frac{\int_\Omega v^2 \, dx}{(\|\mathbf{v}\|^2_{L^2(\Omega)} + \|\nabla w\|^2_{L^2(\Omega)})^{1/2}} \geq C_1(\Omega) \frac{\int_\Omega v^2 \, dx}{\|v\|_{L^2(\Omega)}} = C_1(\Omega)\|v\|_{L^2(\Omega)}. \end{aligned}$$

Using the bounds for $\|a\|$, α_0, and α_1 in (2.11) and (2.12), we obtain

$$\|\mathbf{q}\|_{H(\mathrm{div};\Omega)} \leq C_3(\Omega) \frac{\lambda_{\max}}{\lambda_{\min}} \|f\|_{L^2(\Omega)}, \tag{2.27}$$

and

$$\|u\|_{L^2(\Omega)} \leq C_4(\Omega) \frac{\lambda_{\max}}{\lambda^2_{\min}} \|f\|_{L^2(\Omega)}, \tag{2.28}$$

Remark 2.1 Estimate (2.27) can be recovered from (2.17). We only use the fact that the solution $(\mathbf{q}, u)$ of (2.23) is unique and satisfies relation (2.22). We note, however, that we must use the technique presented above to derive the estimates analogous to (2.27) and (2.28) with constants independent of h for the discrete mixed problem.

Remark 2.2 Problem (2.23) also can be viewed as the Euler-Lagrange equation of the following saddle-point problem:

$$\inf_{\mathbf{p} \in H(\mathrm{div};\Omega)} \sup_{v \in L^2(\Omega)} \left(\frac{1}{2} \int\limits_\Omega K^{-1} \mathbf{p} \cdot \mathbf{p} \, dx + \int\limits_\Omega f v \, dx + \int\limits_\Omega v \operatorname{div} \mathbf{p} \, dx \right). \tag{2.29}$$

2.3.3 Discrete mixed problem

Here we consider a discretization of the mixed problem (2.23) in finite dimensional subspaces of $H(\mathrm{div}\,;\Omega) \times L^2(\Omega)$ belonging to the Raviart-Thomas family of spaces [101]. In this dissertation we consider only the *lowest-order* Raviart-Thomas space.

Using the simplicial (triangular or tetrahedral) triangulation introduced in Section 2.2.2, first, we define the lowest-order Raviart-Thomas velocity space on a single simplex. For simplicity we consider only the case of a 3-dimensional domain. The definition of a velocity space in the 2-dimensional case is analogous and much simpler. We use these spaces on each simplex to define the lowest-order Raviart-Thomas space on the whole domain.

Let $\hat{\tau}$ be the unit reference tetrahedron with vertices

$$\hat{a}_1 = (0,0,0), \qquad \hat{a}_2 = (1,0,0), \qquad \hat{a}_3 = (0,1,0), \qquad \hat{a}_4 = (0,0,1).$$

The lowest-order Raviart-Thomas velocity space on $\hat{\tau}$ is defined by

$$RT^0_{-1}(\hat{\tau}) = \left\{ \mathbf{p} : \mathbf{p} = \begin{bmatrix} a_{\hat{\tau}} \\ b_{\hat{\tau}} \\ c_{\hat{\tau}} \end{bmatrix} + d_{\hat{\tau}} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} \right\}.$$

Let $\mathcal{T}_h$ be a triangulation, as before, of the 3-dimensional domain Ω. For a tetrahedron $\tau \in \mathcal{T}_h$ with vertices a_1, a_2, a_3, and a_4, we define an invertible, affine linear map $F_\tau : \tau \to \hat{\tau}$, such that $F_\tau(a_i) = \hat{a}_i$, $i = 1, \ldots, 4$. Here, $F_\tau = B_\tau \hat{x} + b_\tau$, where B_τ is a 3×3 invertible matrix and $b_\tau \in \mathbb{R}^3$. For any scalar function $\hat{v}$ defined on $\hat{\tau}$ (respectively, on $\partial\hat{\tau}$), we associate the function v defined on τ (respectively, on $\partial\tau$) by

$$v = \hat{v} \circ F_\tau^{-1}, \qquad \hat{v} = v \circ F_\tau,$$

and for any vector-valued function $\hat{\mathbf{p}}$ defined on $\hat{\tau}$, we associate the function $\mathbf{p}$ on τ by

$$\mathbf{p} = \frac{1}{\det(B_\tau)} B_\tau \hat{\mathbf{p}} \circ F_\tau^{-1}, \qquad \hat{\mathbf{p}} = \det(B_\tau) B_\tau^{-1} \mathbf{p} \circ F_\tau. \tag{2.30}$$

The choice of transformation (2.30) is based on the following identities:

$$\int_{\hat{\tau}} \hat{v} \operatorname{div} \hat{\mathbf{p}} \, d\hat{x} = \int_{\tau} v \operatorname{div} \mathbf{p} \, dx, \qquad \forall \hat{v} \in L^2(\hat{\tau}), \quad \forall \hat{\mathbf{p}} \in (H^1(\hat{\tau}))^3, \tag{2.31}$$

and

$$\int_{\partial\hat{\tau}} \hat{v} \; \hat{\mathbf{p}} \cdot \hat{\mathbf{n}} \, d\hat{S} = \int_{\partial\tau} v \; \mathbf{p} \cdot \mathbf{n} \, dS, \qquad \forall \hat{v} \in L^2(\hat{\tau}), \quad \forall \hat{\mathbf{p}} \in (H^1(\hat{\tau}))^3. \tag{2.32}$$

The space $RT^0_{-1}(\tau)$ is defined by

$$RT^0_{-1}(\tau) = \frac{1}{\det(B_\tau)} B_\tau RT^0_{-1}(\hat{\tau}) \circ F_\tau^{-1}. \tag{2.33}$$

It is easy to show that $RT^0_{-1}(\tau)$ consists of linear vector functions which have a constant normal component on the faces of τ.

We introduce the following Raviart-Thomas spaces

$$\begin{aligned} RT^0_{-1}(\mathcal{T}_h) &= \Big\{ \mathbf{p} : \mathbf{p} \in (L^2(\Omega))^3, \mathbf{p}|_\tau \in RT^0_{-1}(\tau) \ \forall \tau \in \mathcal{T}_h \Big\}, \\ \mathbf{V}_h \equiv RT^0_0(\mathcal{T}_h) &= \Big\{ \mathbf{p} : \mathbf{p} \in RT^0_{-1}(\mathcal{T}_h), \text{ the normal component of } \mathbf{p} \\ &\qquad \text{is continuous across the interelement boundaries} \Big\} \\ W_h \equiv M^0_{-1}(\mathcal{T}_h) &= \Big\{ v : v \in L^2(\Omega), v|_\tau = c_\tau \ \forall \tau \in \mathcal{T}_h \Big\}. \end{aligned} \tag{2.34}$$

Here, c_τ is a constant that depends only on the element τ. It is easy to check that $\mathbf{V}_h = RT^0_{-1}(\mathcal{T}_h) \cap H(\operatorname{div}; \Omega)$ and, consequently, $\mathbf{V}_h \in H(\operatorname{div}; \Omega)$.

The lowest-order Raviart-Thomas mixed element method is given by:

find $(\mathbf{q}_h, u_h) \in \mathbf{V}_h \times W_h$ *such that*

$$\left\{\begin{array}{lcll} \int\limits_\Omega K^{-1}\mathbf{q}_h \cdot \mathbf{p}_h \, dx - \int\limits_\Omega u_h \operatorname{div} \mathbf{p}_h \, dx & = & 0, & \forall \mathbf{p}_h \in \mathbf{V}_h, \\ -\int\limits_\Omega v_h \operatorname{div} \mathbf{q}_h \, dx & = & -\int\limits_\Omega f \; v_h \, dx, & \forall v \in W_h. \end{array}\right. \tag{2.35}$$

Let $\{\mathbf{\Phi}_i\}_{i=1}^n$ and $\{\chi\}_{j=1}^m$ be nodal bases of finite element spaces $\mathbf{V}_h$ and W_h, respectively. Then every pair $(\mathbf{q}_h, u_h) \in \mathbf{V}_h \times W_h$ can be represented in a form: $\mathbf{q}_h = \sum_{i=1}^n q_i \, \mathbf{\Phi}_i$, and $u_h = \sum_{j=1}^m u_j \, \chi_j$. The choice of bases in these spaces induces a one-to-one correspondence between the functions $\mathbf{q}_h \in \mathbf{V}_h$, $u_h \in W_h$, and the vectors of linear spaces $\mathbf{q} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^m$, respectively. In this basis the mixed problem (2.35) is represented in the matrix form:

$$\begin{bmatrix} A_h & B_h^T \\ B_h & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f} \end{bmatrix}, \tag{2.36}$$

where A_h is a symmetric positive definite matrix with $(A_h)_{ij} = \int_\Omega K^{-1}\mathbf{\Phi}_i \cdot \mathbf{\Phi}_j \, dx$ and B_h is an approximation of the divergence map which is given by $(B_h)_{ij} = -\int_\Omega \chi_j \operatorname{div} \mathbf{\Phi}_i \, dx$.

System (2.36) is a saddle-point problem. The matrix of this system is symmetric but indefinite. For this reason many efficient, robust, and fast iterative methods (e.g., the conjugate gradient method) cannot be used to solve problem (2.36).

We again use Babuska-Brezzi Lemma 2.6 to show well-posedness for the discrete problem (2.35). We show that stability results (2.11) and (2.12) are uniform in h. The spaces Q and V are given by $Q = W_h$, $V = \mathbf{V}_h$, and the bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ are given by (2.24).

We first note that the *discrete divergence-free* space Ker B_h is *divergence-free* in the sense of $L^2(\Omega)$, i.e.

$$\text{Ker } B_h \in \text{Ker } B = \{\mathbf{p}_0 \in H(\operatorname{div}; \Omega) : \operatorname{div} \mathbf{p}_0 = 0\}. \tag{2.37}$$

To show (2.37) we reduce the problem to one on the reference element. Using the structure of the space W_h it is easy to see that

$$\begin{array}{ll} \int\limits_\Omega v_h \, (\operatorname{div} \mathbf{p}_h) \, dx = 0, & \forall v_h \in W_h \equiv M_{-1}^0(\mathcal{T}_h), \\ & \Updownarrow \\ \int\limits_\tau v_h \, (\operatorname{div} \mathbf{p}_h) \, dx = 0, & \forall v_h \in M_{-1}^0(\tau) \quad \forall \tau \in \mathcal{T}_h. \end{array}$$

Using (2.31) we have for any $\tau \in \mathcal{T}_h$:

$$\int\limits_\tau v_h \, (\operatorname{div} \mathbf{p}_h) \, dx = 0 \quad \Longleftrightarrow \int\limits_{\hat\tau} \hat v_h \, (\operatorname{div} \hat{\mathbf{p}}_h) \, d\hat x = 0.$$

We now set $\hat v_h = \operatorname{div} \hat{\mathbf{p}}_h$ to obtain $\operatorname{div} \hat{\mathbf{p}}_h = 0$ in $\hat\tau$, which implies that $\operatorname{div} \mathbf{p}_h = 0$ in Ω.

Using the same ideas as in the continuous case, we have $\|a\| \leq 1/\lambda_{\min}$.

Using (2.37) and the same ideas as in the continuous case, we obtain $\alpha_1 \geq 1/\lambda_{\max}$.

To show that $\alpha_0 \geq c(\Omega) > 0$, with $c(\Omega)$ independent of h, we use the same ideas as the continuous case and the following lemma:

Lemma 2.7 *For any function $v_h \in W_h$ there exists a function $\mathbf{p}_h \in \mathbf{V}_h$ such that* $\operatorname{div} \mathbf{p}_h = v_h$ *in Ω, and $\|\mathbf{p}_h\|_{H(\mathrm{div};\Omega)} \leq C(\Omega)\|v_h\|_{L^2(\Omega)}$.*

The proof of this lemma is given in Raviart and Thomas [101, 102]. The arguments in [101] are for the two-dimensional case and they can be extended straightforwardly to the three-dimensional case.

2.4 Nonconforming formulation

2.4.1 Motivation

The mixed methods for second-order elliptic equations have been extensively studied in the last two decades. A large variety of mixed finite element spaces on triangles, rectangles, prisms, and tetrahedrons have been developed [101, 92, 27, 26] and their convergence and superconvergence properties have been investigated [113, 67, 90, 43, 54]. As shown by Russell and Wheeler in [104], the mixed finite element approximations with special quadratures on rectangular grids are equivalent to the finite volume methods. The superconvergent velocity calculations for smooth solutions has been established by Weiser and Wheeler in [118]. Based on that equivalence, Bramble et al. in [21] have developed efficient multigrid solution procedures for mixed approximations on structured grids. However, in general the technique of the mixed finite element method leads to an algebraic saddle-point problem that is more difficult and more expensive to solve than the problem with a symmetric and positive definite operator. Although some reliable preconditioning algorithms for these saddle-point problems have been proposed and studied (see, e.g., [16, 19, 56, 105, 115]), their efficiency depends strongly on the geometry of the domain, the coefficient matrix $K(\mathbf{x})$, and the type of the finite elements used.

An alternative approach can be taken by developing hybrid-mixed methods. This approach has been studied in the pioneering work of Arnold and Brezzi [5] where the continuity of the normal component of the flux vector to the boundary of each element is enforced by Lagrange multipliers. In general, the Lagrange multipliers on the element boundaries turn out to be none other than the trace of the primary unknown $u(\mathbf{x})$.

The important discovery of Arnold and Brezzi is that the hybrid-mixed method is equivalent in application to (2.13), the Galerkin method with nonconforming finite elements. Namely, in [5] it is shown that the lowest-order Raviart-Thomas mixed finite element approximations are equivalent to the usual P_1-nonconforming finite element approximations when the classical P_1-nonconforming space is augmented with P_3-bubbles. Such a relationship has been studied recently for a large variety of mixed finite element spaces (see, e.g., [4, 23, 32]).

The equivalence between the hybrid-mixed and the nonconforming finite element methods establishes a framework for preconditioning and/or solving the algebraic problem and for postprocessing the finite element solution. Schematically this framework includes the following three steps:

(a) forming the reduced algebraic problem for the Lagrange multipliers, which is equivalent to a nonconforming approximation;

(b) construction and study of efficient methods, based on multigrid, multilevel or domain decomposition for solving or preconditioning the reduced system;

(c) recovery of the solution $u(\mathbf{x})$ and the flux $\mathbf{q}$ from the Lagrange multipliers, which were already found, by using only element-by-element computations.

The recent progress in each of the steps described above (see, e.g., [116, 107, 37]) gives us an indication that the mixed finite element method can be used as an accurate and efficient tool for solving general elliptic problems of second-order in domains with complicated geometry.

In this dissertation we use the Arnold-Brezzi [5] theory and construct several preconditioners for nonconforming P_1 finite element approximations [52, 55, 33, 73, 77, 78] (see also Chapters IV and V of this dissertation).

2.4.2 Hybrid-mixed formulation

Let $\mathcal{F}_h$ be the set of faces e of simplices $\tau \in \mathcal{T}_h$, and let $\mathcal{F}_h^\partial$ be the set of those faces e which are on the Dirichlet boundary Γ_0, $\mathcal{F}_h^\partial = \{e \in \mathcal{F}_h : e \subseteq \Gamma_0\}$, and $\mathcal{F}_h^0 = \mathcal{F}_h \setminus \mathcal{F}_h^\partial$.

The property $\mathbf{V}_h \subset H(\mathrm{div}\,;\Omega)$ says that the normal components of the members of $\mathbf{V}_h$ are continuous across the interior boundaries in $\mathcal{F}_h^0$. Following [5], we skip this requirement on $\mathbf{V}_h$ by defining $\tilde{\mathbf{V}}_h \equiv RT_{-1}^0(\mathcal{T}_h)$. Then, to enforce the continuity on the normal components in $\tilde{\mathbf{V}}_h$, we introduce Lagrange multipliers. We define the space of Lagrange multipliers $\mathcal{L}_h \equiv M_{-1}^0(\mathcal{F}_h^0)$ as the set of all functions on the union of faces $\mathcal{F}_h$ that are constant on each face $e \in \mathcal{F}_h^0$ and vanish on $\mathcal{F}_h^\partial$:

$$\mathcal{L}_h = \left\{\mu \in L^2(\mathcal{F}_h) : \mu|_e \in \mathbf{V}_h \cdot \mathbf{n}|_e \text{ for each } e \in \mathcal{F}_h^0\right\},$$

where $\mathbf{n}$ is the unit normal vector to the face e.

To establish the relationship between the mixed method and the nonconforming Galerkin method and to construct efficient preconditioners, following [32], we introduce the projection of the coefficient matrix K, i.e. $C_h = P_h K^{-1}$, where P_h is the L^2-projection into W_h.

Then the *hybrid-mixed discrete* formulation is given by:

find $(\mathbf{q}_h^*, u_h^*, \lambda_h) \in \tilde{\mathbf{V}}_h \times W_h \times \mathcal{L}_h$ *such that*

$$\begin{aligned}
\int_\Omega C_h \mathbf{q}_h^* \cdot \mathbf{p}_h \, dx - \sum_{\tau\in\mathcal{T}_h} \left(\int_\tau u_h^* \mathrm{div}\, \mathbf{p}_h \, dx - \int_{\partial\tau} \lambda_h \left(\mathbf{p}_h \cdot \mathbf{n}_\tau\right) ds \right) &= 0, && \forall \mathbf{p}_h \in \tilde{\mathbf{V}}_h, \\
-\sum_{\tau\in\mathcal{T}_h} \int_\tau v_h \,\mathrm{div}\, \mathbf{q}_h^* \, dx &= -\int_\Omega f v_h \, dx, && \forall v_h \in W_h, \\
\sum_{\tau\in\mathcal{T}_h} \int_{\partial\tau} \mu_h \left(\mathbf{q}_h^* \cdot \mathbf{n}_\tau\right) ds &= 0, && \forall \mu_h \in \mathcal{L}_h.
\end{aligned} \tag{2.38}$$

Note that the last equation in (2.38) enforces the continuity requirement mentioned above, so in fact $\mathbf{q}_h^* \in \mathbf{V}_h$. Also, note that any vector function $\mathbf{p}_h \in \tilde{\mathbf{V}}_h$ belongs to the space $\mathbf{V}_h$ if and only if

$$\sum_{\tau\in\mathcal{T}_h} \int_{\partial\tau} \mu_h \left(\mathbf{p}_h \cdot \mathbf{n}_\tau\right) ds = 0, \qquad \forall \mu_h \in \mathcal{L}_h.$$

Therefore, using element-by-element arguments, it is easy to check for piecewise constant tensor K that system (2.38) has a unique solution with $\mathbf{q}_h^* = \mathbf{q}_h$ and $u_h^* = u_h$, where $(\mathbf{q}_h, u_h)$ is the solution of (2.35). The function λ_h is uniquely determined from the first equation of (2.38). Hence, systems (2.35) and (2.38) are equivalent, and we can therefore drop the superscript '$*$' in (2.38).

In matrix notation system (2.38) has the form

$$\begin{bmatrix} \bar{A}_h & \bar{B}_h^T & \bar{C}_h^T \\ \bar{B}_h & \mathbf{0} & \mathbf{0} \\ \bar{C}_h & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{q}_h \\ u_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ f_h \\ \mathbf{0} \end{bmatrix}. \tag{2.39}$$

Remark 2.3 An advantage of the hybrid-mixed formulation is that matrix $\bar{A}_h$ is block diagonal, with each block corresponding to a single element. Hence, $\bar{A}_h$ can be inverted easily and in parallel. After eliminating the flux in (2.39), we obtain a symmetric positive definite system

$$\begin{bmatrix} \bar{B}_h\bar{A}_h^{-1}\bar{B}_h^T & \bar{B}_h\bar{A}_h^{-1}\bar{C}_h^T \\ \bar{C}_h\bar{A}_h^{-1}\bar{B}_h^T & \bar{C}_h\bar{A}_h^{-1}\bar{C}_h^T \end{bmatrix} \begin{bmatrix} u_h \\ \lambda_h \end{bmatrix} = \begin{bmatrix} -f_h \\ \mathbf{0} \end{bmatrix}. \tag{2.40}$$

Remark 2.4 We also note that the weak formulation for $\mathbf{q} = K\nabla u$ on a single element $\tau \in \mathcal{T}_h$ is given by:

$$\int_\tau K^{-1}\mathbf{q}\cdot\mathbf{p}\,dx - \int_\tau u\,\mathrm{div}\,\mathbf{p}\,dx + \int_{\partial\tau} u\,(\mathbf{p}\cdot\mathbf{n}_\tau)\,ds = 0, \qquad \forall\mathbf{p}\in\mathcal{H}(\mathrm{div}\,;\Omega). \tag{2.41}$$

Hence, comparing (2.41) with the first equation of (2.38), it is easy to see that the Lagrange multiplier λ_h can be interpreted as an approximation of the trace of u on the boundaries of the elements.

2.4.3 Arnold-Brezzi theory

As shown in [5, 4, 32, 83], the solution to (2.38) can be obtained from a certain modified nonconforming Galerkin method by means of augmenting the latter with bubble functions. In this subsection, following [32, 33, 34], we show that the linear system generated by (2.38) can be algebraically condensed to a symmetric, positive definite system for the Lagrange multiplier λ_h. Next, we show that this linear system can be obtained from the standard nonconforming Galerkin method without using any bubbles.

As in the previous subsection the definition and computation are done locally, element-by-element. The lowest-order Raviart-Thomas space [101, 92] defined over $\tau \in \mathcal{T}_h$ is given by

$$\begin{aligned} \mathbf{V}_h(\tau) &\equiv RT_{-1}^0(\tau) = \left\{\mathbf{p} : \mathbf{p} = (P_0(\tau))^3 \oplus \left((x,y,z)^T P_0(\tau)\right)\right\}, \\ W_h(\tau) &= P_0(\tau), \\ \mathcal{L}_h(e) &= P_0(e), \end{aligned}$$

where $P_i(\tau)$ is the restriction of the set of all polynomials of total degree not higher than $i \geq 0$ to the set $\tau \in \mathcal{T}_h$.

For each τ in $\mathcal{T}_h$, let

$$\bar{f}_\tau = \frac{1}{|\tau|}(f,1)_\tau,$$

where $|\tau|$ denotes the volume of τ and $(\cdot,\cdot)_\tau$ means L^2-inner product over τ. Also, set $C_h = (C_{ij})$ and $\mathbf{q}_h|_\tau = (q_{\tau 1}, q_{\tau 2}, q_{\tau 3})^T = (g_{\tau,1} + t_\tau x,\ g_{\tau,2} + t_\tau y,\ g_{\tau,3} + t_\tau z)^T$. Then, by the second equation of (2.38), it follows that

$$t_\tau = \bar{f}_\tau / 3. \tag{2.42}$$

Now, take $\mathbf{p} = (1,0,0)^T$ in τ and $\mathbf{p} = \mathbf{0}$ elsewhere, $\mathbf{p} = (0,1,0)^T$ in τ and $\mathbf{p} = \mathbf{0}$ elsewhere, and $\mathbf{p} = (0,0,1)^T$ in τ and $\mathbf{p} = \mathbf{0}$ elsewhere, respectively, in the first equation of (2.38) to obtain

$$\left(\sum_{i=1}^{3} C_{h,ji}\, q_{\tau i}, 1 \right)_\tau + \sum_{i=1}^{4} |e_{\tau,i}|\, n^{(j)}_{\tau,i} \cdot \lambda_h|_{e_{\tau,i}} = 0, \qquad j = 1,2,3, \tag{2.43}$$

where $|e_{\tau,i}|$ is the area of the face $e_{\tau,i}$, and $\mathbf{n}_{\tau,i} = (n^{(1)}_{\tau,i}, n^{(2)}_{\tau,i}, n^{(3)}_{\tau,i})$. Let $\Psi^\tau = (\Psi^\tau_{ij}) = ((C_{h,ij}, 1)_\tau)^{-1}$. Then (2.43) can be solved for $g_{\tau,j}$:

$$\begin{aligned} g_{\tau,j} \ &= \ -\sum_{i=1}^{4} |e_{\tau,i}| \left(\Psi^\tau_{j1} n^{(1)}_{\tau,i} + \Psi^\tau_{j2} n^{(2)}_{\tau,i} + \Psi^\tau_{j3} n^{(3)}_{\tau,i} \right) \cdot \lambda_h|_{e_{\tau,i}} - \\ & \quad -\frac{\bar{f}_\tau}{3} \left(\sum_{i=1}^{3} \Psi^\tau_{ji}\ (C_{h,i1}x + C_{h,i2}y + C_{h,i3}z)\ ,\ 1 \right)_\tau, \qquad j = 1,2,3. \end{aligned} \tag{2.44}$$

A basis function in $\mathcal{L}_h$ is defined by taking $\mu = 1$ on one face between two elements and $\mu = 0$ elsewhere. Then, applying (2.42) and (2.43) we see that the contributions of the tetrahedron τ to the stiffness matrix and the right-hand side are

$$A^\tau_{ij} = \overline{\mathbf{n}}_{\tau,i}\ \Psi^\tau\ \overline{\mathbf{n}}_{\tau,j}, \qquad F^\tau_i = -\frac{\left(J^f_\tau,\ \overline{\mathbf{n}}_{\tau,i} \right)_\tau}{|\tau|} + \left(J^f_\tau,\ \mathbf{n}_{\tau,i} \right)_{e_{\tau,i}}, \qquad \tau \in \mathcal{T}_h, \tag{2.45}$$

where $\overline{\mathbf{n}}_{\tau,i} = |e_{\tau,i}| \cdot \mathbf{n}_{\tau,i}$ and $J^f_\tau = \bar{f}_\tau (x,y,z)^T/3$. Hence we obtain the system for λ_h:

$$A\lambda = \mathbf{F}. \tag{2.46}$$

After the computation of λ_h, we can recover $\mathbf{q}_h$ via (2.42) and (2.44). Also, if u_h is required, it follows from the first equation of (2.38) that

$$u_\tau = \frac{1}{3|\tau|} \left(\left(C_h \mathbf{q}_h,\ (x,y,z)^T \right)_\tau + \sum_{i=1}^{4} \lambda_h|_{e_{\tau,i}} \cdot \left((x,y,z)^T,\ \mathbf{n}_{\tau,i} \right)_{e_{\tau,i}} \right), \qquad \tau \in \mathcal{T}_h. \tag{2.47}$$

The above result is summarized in the following lemma (see also [33]).

Lemma 2.8 *Let a bilinear form $c_h(\cdot,\cdot)$ and a functional $F_h(\cdot)$ be defined as follows:*

$$\begin{aligned} c_h(\chi_h, \mu_h) \ &= \ \sum_{\tau \in \mathcal{T}_h} (\chi_h, \mathbf{n}_\tau)_{\partial\tau} \Psi^\tau (\mu_h, \mathbf{n}_\tau)_{\partial\tau}, && \chi_h,\ \mu_h \in \mathcal{L}_h, \\ F_h(\mu_h) \ &= \ -\sum_{\tau \in \mathcal{T}_h} \tfrac{1}{|\tau|} (J^f, 1)_\tau \cdot (\mu_h, \mathbf{n}_\tau)_{\partial\tau} + \sum_{\tau \in \mathcal{T}_h} (\mu_h J^f, \mathbf{n}_\tau)_{\partial\tau}, && \mu_h \in \mathcal{L}_h, \end{aligned}$$

where J^f is such that $J^f|_\tau = J^f_\tau$. Then $\lambda_h \in \mathcal{L}_h$ satisfies

$$c_h(\lambda_h, \mu_h) = F_h(\mu_h), \qquad \forall \mu_h \in \mathcal{L}_h. \tag{2.48}$$

Remark 2.5 Obviously, the algebraic system (2.46) is the matrix representation of equation (2.48), i.e.

$$(A\lambda, \mu) = c_h(\lambda_h, \mu_h), \qquad \forall \mu_h \in \mathcal{L}_h,$$

where λ and μ are vector representations of the functions λ_h and μ_h, respectively.

Note that there are at most seven nonzero entries per row in the stiffness matrix A. Also, it is easy to see that matrix A is a symmetric and positive definite matrix; moreover, if the angles of every τ in $\mathcal{T}_h$ are not bigger than $\pi/2$, then it is an M-matrix. Finally, (2.46) corresponds to the P_1-nonconforming finite element method system, as described below. This equivalence is used to construct preconditioners for A in Chapters IV and V.

Let

$$V_h = \Big\{ v \in L^2(\Omega) : \ \ v|_\tau \in P_1(\tau), \ \forall \tau \in \mathcal{T}_h; \ v \text{ is continuous at the barycenters of faces from } \mathcal{F}_h^0 \text{ and vanishes at the barycenters of faces on } \Gamma_0 \Big\}. \tag{2.49}$$

Proposition 2.1 ([33]) *Let $f_h = P_h f$ be L^2-projection of f into W_h. Then* (2.46) *coincides with the linear system corresponding to the problem: find $\psi_h \in V_h$ such that*

$$a_h(\psi_h, \varphi) = (f_h, \varphi), \qquad \forall \varphi \in V_h, \tag{2.50}$$

where $a_h(\psi_h, \varphi) = \sum\limits_{\tau \in \mathcal{T}_h} (C_h^{-1} \nabla \psi_h, \nabla \varphi)_\tau$.

Proof: From the definition of the nodal basis $\{\psi_i^h\}$ of V_h, for each $\tau \in \mathcal{T}_h$ we have

$$\psi_i^h|_\tau = \frac{1}{|\tau|} \overline{\mathbf{n}}_{\tau,i} \cdot \left((x,y,z)^T - p_l \right), \qquad i \neq l,$$

for some barycenter p_l. Then, we see that

$$(C_h^{-1} \nabla \psi_i^h, \nabla \psi_j^h)_\tau = \overline{\mathbf{n}}_{\tau,i} \Psi^\tau \overline{\mathbf{n}}_{\tau,j},$$

which is A_{ij}^τ. Also, note that for any linear functions ψ and ϕ on a tetrahedron $\tau \in \mathcal{T}_h$

$$(\psi, \phi)_\tau = \frac{1}{4} |\tau| \sum_{i=1}^{4} \psi(p_i) \phi(p_i), \tag{2.51}$$

where the p_i's are the barycenters of the faces of τ, so that

$$F_i^\tau = -\frac{1}{|\tau|} \left(J_\tau^f, \ \overline{\mathbf{n}}_{\tau,i} \right)_\tau + \left(J_\tau^f, \ \mathbf{n}_{\tau,i} \right)_{e_{\tau,i}} = -\frac{\bar{f}_\tau}{3} (1, \psi_i^h)_\tau + \frac{|\tau| \bar{f}_\tau}{3 |e_{\tau,i}|} (\psi_i^h, 1)_{e_{\tau,i}} = \bar{f}_\tau \left(1, \psi_i^h \right)_\tau,$$

which is $(f_h, \psi_i^h)_\tau$. □

Corollary 2.1 *The values of the degrees of freedom of the solution $\psi^h \in V_h$ of problem* (2.50) *coincide with the corresponding values of the solution $\lambda_h \in \mathcal{L}_h$ of problem* (2.48).

Using this corollary we can define a projection operator $P_h : V_h \to \mathcal{L}_h$ as follows: for any $u_h \in V_h$

$$\lambda_h|_e = P_h u_h|_e \equiv u_h(\mathbf{x}_e), \qquad \forall e \in \mathcal{F}_h^0, \tag{2.52}$$

where $\mathbf{x}_e$ is the barycenter of the face e. That is, the value of the Lagrange multiplier function $\lambda_h = P_h u_h$ on a given face e is equal to the value of the function u_h in the barycenter of this face.

Remark 2.6 If the matrix function $K(\mathbf{x})$ is piecewise constant in the domain and the partition $\mathcal{T}_h$ is chosen in such a way that $K(\mathbf{x})$ is constant in each element $\tau \in \mathcal{T}_h$, then the bilinear form in (2.50) coincides with (2.15), i.e.

$$a_h(\psi, \varphi) = \sum_{\tau \in \mathcal{T}_h} (C_h^{-1} \nabla \psi, \nabla \varphi)_\tau \equiv \sum_{\tau \in \mathcal{T}_h} (K \nabla \psi, \nabla \varphi)_\tau, \qquad \forall \psi, \varphi \in V_h.$$

Remark 2.7 It is well known [36, 121] that for problem (2.13) with $K(x) = I$ the bilinear form (2.50) satisfies

$$c\, h^2 \cdot (\varphi, \varphi) \le a_h(\varphi, \varphi) \le C \ \cdot (\varphi, \varphi), \qquad \forall \varphi \in V_h.$$

2.5 Nonconforming approximation of elliptic problems with anisotropy

We conclude this chapter by outlining the problems we are going to consider below. As was mentioned in the previous section, we consider methods of constructing the preconditioners for nonconforming P_1 finite element approximations of (2.13).

Until the end of this chapter we define $\mathcal{T}_h$ as a regular partitioning of $\Omega \in \mathbb{R}^d$, $d = 2, 3$, into simplices τ with a mesh size h, the P_1–nonconforming finite element space $V_h(\Omega)$ by (2.49), and a bilinear form $a_h(\cdot, \cdot)$ by (2.50). Once a nodal basis $\{\varphi_i(\mathbf{x})\}_{i=1}^N$ for $V_h(\Omega)$ is chosen, (2.50) leads to a system of linear algebraic equations

$$A\mathbf{u} = \mathbf{f}, \tag{2.53}$$

where A is sparse symmetric positive definite matrix and $\mathbf{u}, \mathbf{f} \in \mathbb{R}^N$.

Although the methods of solving (2.53) have been extensively studied in the past few years (see, e.g., [5, 15, 20, 22, 35, 107]), their efficiency depends on the coefficient matrix $K(\mathbf{x})$. In the case of strong anisotropy in the coefficients the question of constructing effective solution techniques is still open.

In this dissertation we describe and analyze **methods of constructing preconditioners** for (2.53) when the tensor $K(\mathbf{x})$ from (2.13) is an anisotropic matrix coefficient. Below we outline the **classes of problems** for which we construct the preconditioners.

2.5.1 Method of algebraic substructuring

This method is applied for two types of problems.

(1) The computational domain Ω is a union of parallelepipeds (rectangles if we consider the problem in $\mathbb{R}^2$). The tensor $K(\mathbf{x})$ is a smooth matrix function which is a small

perturbation of a diagonal constant matrix in the entire domain. It means that there exist a diagonal constant matrix $\tilde{K} = \text{diag}\,\{k_1, \ldots, k_d\}$, $d = 2, 3$, and some positive constants $\check{c}$, $\hat{c}$, such that

$$\check{c}\,\tilde{K} \leq K(\mathbf{x}) \leq \hat{c}\,\tilde{K}, \qquad \forall \mathbf{x} \in \Omega. \tag{2.54}$$

We construct the mesh $\mathcal{T}_h$, first, by partitioning the domain Ω into small cubes (squares in $\mathbb{R}^2$) and, then, by subdividing each cube (square) into tetrahedra (triangles) in a regular way.

Then we define the auxiliary bilinear form

$$\tilde{a}_h(u, v) = \sum_{\tau \in \mathcal{T}_h} (\tilde{K} \nabla u, \nabla v)_\tau, \qquad \forall\; u, v \in V_h(\Omega). \tag{2.55}$$

We construct preconditioners for this auxiliary bilinear form using substructuring in Chapter IV. Due to inequality (2.54) these preconditioners can be used for initial problem (2.53).

(2) The tensor $K(\mathbf{x})$ which is a full symmetric matrix and the domain Ω satisfy the following assumptions:

(a) There is an orientation-preserving smooth map $\mathcal{L}$ of the unit cube (or square if we consider the problem in $\mathbb{R}^2$) $\hat{\Omega}$ onto Ω and there are positive constants r and C (see [47]) such that

$$\begin{array}{rcll} r^{-1} \|\mathcal{J}(\mathbf{x})\| & \leq & C, & \qquad \forall \mathbf{x} \in \hat{\Omega}, \\ r\, \|\mathcal{J}^{-1}(\mathbf{x})\| & \leq & C, & \qquad \forall \mathbf{x} \in \Omega, \end{array} \tag{2.56}$$

where $\mathcal{J}(\mathbf{x})$ is the Jacobian matrix of $\mathcal{L}$ at $\mathbf{x}$ and $\|\cdot\|$ denotes a matrix norm.

(b) The transformed tensor $\hat{K}(\mathbf{x}) = \frac{1}{|\det(\mathcal{J})|} \mathcal{J}^T K(\mathbf{x}) \mathcal{J}$, $\mathbf{x} \in \hat{\Omega}$ falls into the class of problems described in item (1).

The definition of the nonconforming finite element space for the domains satisfying (2.56) is given below. Let $\mathcal{C}_{\hat{h}}$ and $\mathcal{T}_{\hat{h}}$ be the partitions of $\hat{\Omega}$ into cubes and tetrahedra, respectively, which are associated with the mesh-size $\hat{h} = 1/n$, and let $V_{\hat{h}}$ be the P_1-nonconforming space associated with $\mathcal{T}_{\hat{h}}$, as given in (2.50). Set $h = r \cdot \hat{h}$ and define

$$V_h(\Omega) = \left\{ \varphi = \psi \circ \mathcal{L}^{-1} \;: \psi \in V_{\hat{h}}(\hat{\Omega}) \right\}.$$

We also introduce the map $\mathcal{I} : V_h(\Omega) \to V_{\hat{h}}(\hat{\Omega})$ defined by $\mathcal{I}v = v \circ \mathcal{L}$.

Now we define the stiffness matrix A on domain Ω by

$$(A\mathbf{u}, \mathbf{v})_N = a_h(u, v), \qquad \forall u, v \in V_h(\Omega), \tag{2.57}$$

where

$$\begin{array}{rcl} a_h(u, v) & = & \sum\limits_{\tau \in \mathcal{T}_h} \int\limits_\tau K(\mathbf{x})\, \nabla u \cdot \nabla v \; d\mathbf{x} =, \\ & = & \sum\limits_{\hat{\tau} \in \mathcal{T}_{\hat{h}}} \int\limits_{\hat{\tau}} \frac{1}{|\det(\mathcal{J})|} \mathcal{J}^T K(\mathbf{x}) \mathcal{J}\; \nabla \mathcal{I}u \cdot \nabla \mathcal{I}v \; d\mathbf{x}, \end{array} \tag{2.58}$$

and $|\det(\mathcal{J})|$ is the determinant of the Jacobian $\mathcal{J}(\mathbf{x})$.

Note that taking into account (2.58), we can treat the bilinear form (2.57) as a form generated by some elliptic positive definite operator with piecewise smooth 3×3 symmetric matrix-valued function $K(\mathbf{x})$ on the cube $\hat{\Omega}$. This function satisfies the uniform positive definiteness condition. For this reason, below without loss of generality we suppose that $\Omega \equiv \hat{\Omega}$ is a parallelepiped with the partition into cubes $\mathcal{C}_h$ and into tetrahedra $\mathcal{T}_h$.

For each cube $C \in \mathcal{C}_{\hat{h}}$, we introduce the diagonal matrix $\mathcal{K}_C$ =diag$\{k_{1,C}, k_{2,C}, k_{3,C}\}$ with some as yet unspecified constants $k_{i,C}$, $i = 1,2,3$. Then we define on the reference parallelepiped $\hat{\Omega}$ a bilinear form

$$b_h(u,v) = \sum_{C \in \mathcal{C}_h} \delta_C \left(\sum_{\tau \in C} \int_\tau \mathcal{K}_C \nabla u \cdot \nabla v \; d\mathbf{x} \right), \qquad \forall u,v \in V_{\hat{h}}, \tag{2.59}$$

where the constants δ_C are scaling factors. One reasonable choice is to take $\delta_C = (\lambda_{1,C} + \lambda_{0,C})/2$, where $\lambda_{1,C}$ and $\lambda_{0,C}$ are the largest and smallest eigenvalues of the eigenvalue problem $\hat{K}(\mathbf{x}_0)\mathbf{v} = \lambda_C \mathcal{K}_C \mathbf{v}$, $\mathbf{v} \in \mathrm{I\!R}^3$, where $\hat{K}(\mathbf{x}) = \frac{1}{|\det(\mathcal{J})|} \mathcal{J}^T K(\mathbf{x}) \mathcal{J}$ and $\mathbf{x}_0 \in \mathcal{L}(C) \subset \Omega$ is some point.

We assume that the matrix function defined above, $\delta_C \mathcal{K}_C$, is a small perturbation of a diagonal constant matrix in the entire cube $\hat{\Omega}$.

Note that assumptions (2.56) imply that there are two constants c_0 and c_1 independent of r and $\hat{h}$ such that

$$c_0 a_h(u,u) \leq r \cdot b_h(\mathcal{I}u, \mathcal{I}u) \leq c_1 a_h(u,u), \qquad \forall u \in V_h. \tag{2.60}$$

We choose matrices $\mathcal{K}_C$ in the form: $\mathcal{K}_C$ =diag$\{\hat{K}(\mathbf{x}_0)\}$, $\forall C \in \mathcal{C}_h$, i.e. the matrix $\mathcal{K}_C$ is the diagonal part of $\hat{K}(\mathbf{x}_0)$ at some point $\mathbf{x}_0 \in \mathcal{L}(C)$. In this case constants c_0 and c_1 in (2.60) depend only on the local variation of the coefficients $\left\{ \left(\hat{K} \right)_{kl} \right\}$. Hence the problem of defining a preconditioner for $a_h(\cdot,\cdot)$ is reduced to the problem of finding a preconditioner for $r \cdot b_h(\cdot,\cdot)$, which has a diagonal coefficient tensor and is defined on the unit cube $\hat{\Omega}$. Therefore, all the analysis of the item (1) can be carried out here.

2.5.2 Fictitious components method

For the problem with symmetric full tensor $K(\mathbf{x})$ (like in item (2) of Section 2.5.1) in the domain Ω of complex geometric shape we also consider a variant of the fictitious components method, which can be outlined for the problem in $\mathrm{I\!R}^2$ as follows.

Let

$$K(\mathbf{x}) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

be a constant symmetric matrix which has eigenpairs $(k_1, \mathbf{u}_1)$ and $(k_2, \mathbf{u}_2)$, where $\mathbf{u}_1 = (\alpha, \beta)$, $\mathbf{u}_2 = (-\beta, \alpha)$, $\alpha^2 + \beta^2 = 1$. Let us consider a transformation of the coordinates $(\xi, \nu) = F(x,y)$: $\xi = \alpha \cdot x + \beta \cdot y$, $\nu = -\beta \cdot x + \alpha \cdot y$. In the coordinates (ξ, ν) problem (2.50) has the diagonal matrix coefficient $\tilde{K} = \text{diag}\ \{k_1, k_2\}$.

Now we construct a rectangle Π in the (ξ, ν) plane which contains $\tilde{\Omega} \equiv F(\Omega)$ and a uniform triangular mesh in Π. Mapping this mesh to the real domain Ω by the transformation $(x,y) = F^{-1}(\xi, \nu)$, we define a triangulation of Ω.

In the fictitious components method, instead of problem (2.53) we consider a problem

$$\tilde{A}\tilde{\mathbf{u}} = \tilde{\mathbf{f}}, \tag{2.61}$$

where a square matrix $\tilde{A}$ of an order $M > N$ is an approximation of the problem on the extended domain Π. We assume that there exists a permutation matrix $\tilde{P}$ such that

$$\tilde{P}\tilde{A}\tilde{P}^T = \begin{bmatrix} A & A_{12} \\ \mathbf{0} & A_{22} \end{bmatrix}, \qquad \tilde{P}\tilde{\mathbf{f}} = \begin{bmatrix} f \\ \mathbf{0} \end{bmatrix}, \tag{2.62}$$

with some matrices A_{12} and A_{22}. It is obvious that for any solution $\tilde{\mathbf{u}}$ of problem (2.61) the solution $\mathbf{u}$ of problem (2.53) can be found by the formula $\mathbf{u} = Q\tilde{P}\tilde{\mathbf{u}}$, where $Q = [I_N\mathbf{0}]$ is an $N \times M$ projection matrix.

So, instead of initial problem (2.53) we need to solve an algebraic problem corresponding to (2.50) with diagonal coefficient matrix $\tilde{K}$. Again, we can use all of the analysis of item (1) from Section 2.5.1. The analysis of the fictitious components method for an anisotropic problem is given in Chapter IV.

2.5.3 Domain decomposition method

Here we consider problems for which the computational domain Ω can be represented as a union of substructures $\Omega = \bigcup\limits_{i=1}^{m} \Omega_i$ in such a way that the tensor $K(\mathbf{x})$ is an almost diagonal constant matrix (in the sense of (2.54)) in each substructure. Then we can use the domain decomposition method to solve these problems.

The main idea is to use methods described in Section 2.5.1 to solve or precondition the problems in subdomains. Then, for the problem at the interfaces we construct a preconditioner in the form of an inner Chebyshev iterative procedure. More precisely, we construct a preconditioner for the Schur complement of the original symmetric positive definite matrix, which results after eliminating the blocks corresponding to the unknowns in the subdomains.

This analysis is given in Chapter V.

2.5.4 Domain decomposition method on nonmatching grids

This is a generalization of the method considered in Section 2.5.3. We assume that the computational domain Ω is represented as a union of substructures $\Omega = \bigcup\limits_{i=1}^{m} \Omega_i$ and the matrix $K(\mathbf{x})$ is a full symmetric matrix in each substructure. To solve this problem we use the domain decomposition method on nonmatching grids (see an example of nonmatching grids in Figure 2.1).

The computational domain is considered as a union of nonintersecting subdomains. In each subdomain we construct its own coordinate system and a grid (a triangular one for two-dimensional equations and a tetrahedral one for three-dimensional equations) in accordance with the main directions of anisotropy or, in other words, we define local coordinate systems on eigenvectors of the coefficient matrix $K(\mathbf{x})$. It is easy to see that this matrix is diagonal in such a local coordinate system. The original elliptic problem is posed as a problem with Lagrange multipliers at interfaces between subdomains and with the continuity conditions of the solution (in a weak form) at the same interfaces. A mortar finite element subspace is constructed in the space of Lagrange multipliers. The resulting algebraic systems have the form of a saddle-point problem.

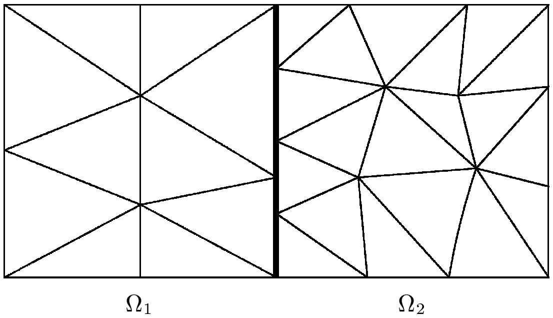


Figure 2.1: *Subdomains with nonmatching grids.*

In Chapter V we propose a new construction of block diagonal preconditioners for the algebraic systems that occur in using the mortar finite element method. This approach combines the ideas of the domain decomposition method (the substructure method) with the algorithms of multilevel and algebraic multigrid methods.

Assuming that the grid in each subdomain is the trace of a hierarchical grid we can use the results described in Sections 2.5.1 and 2.5.2 to construct the preconditioners for problems in subdomains. Then, for the problem at the interfaces we construct a preconditioner in the form of an inner Chebyshev iterative procedure. More precisely we construct a preconditioner for the Schur complement of the original saddle-point matrix. It can be shown that the constructed preconditioner is spectrally equivalent to the original saddle-point matrix with equivalence constants independent of the mesh size, the subdomain diameters, and anisotropy in the coefficients.

CHAPTER III

ITERATIVE METHODS

The term "iterative method" refers to a wide range of techniques that use successive approximations to obtain more accurate solutions to a linear system at each step. The development of efficient iterative methods for solving systems arising from finite element discretizations of second-order partial differential equations has been a very active area of research over the last few decades. The rate at which an iterative method converges depends strongly on the spectrum of the coefficient matrix. At present, iterative methods usually involve a second matrix that transforms the coefficient matrix into one with a more favorable spectrum [44, 45, 46, 39, 117]. The transformation matrix is called a *preconditioner.* The use of a good preconditioner improves the convergence of the iterative method, sufficiently to overcome the extra cost of constructing and applying the preconditioner. Today, the success of finite element methods is based to a large extent on the existence of fast and robust techniques for preconditioning and solving the corresponding discrete problems.

The main goal of this dissertation is the construction of preconditioners for nonconforming finite element approximations of a second-order elliptic problem. As an example of an iterative method for this type of problems we choose the conjugate gradient method [65, 81, 80]. In Chapter V we consider nonconforming approximations of equation (2.13) on nonmatching grids. The resulting algebraic systems have the form of an algebraic saddle-point problem. At present there are several approaches to the iterative solution of finite element systems on nonmatching grids, presented, for example, in [1, 109, 72]. At the same time there is a great number of papers on iterative methods for solving algebraic systems in the saddle-point form (see, e.g., [11, 16, 19, 74, 50, 105]). It is obvious that these methods, when appropriately modified, may be employed for solving the corresponding finite element systems.

In this chapter we outline iterative techniques for solving systems of linear algebraic equations with both symmetric positive definite and indefinite matrices. In the next chapters we develop efficient preconditioners for both kinds of systems. The rest of the chapter is organized as follows. First, we consider some basic facts from the theory of iterative methods. Next, in Section 3.2 we give formulae for the preconditioned Lanczos method [81] as applied to the solution of systems with symmetric indefinite matrices as well as the reasons for the choice of block diagonal preconditioners for saddle-point matrices. Then, in Section 3.3 we discuss the conjugate gradient type methods. Finally, in Section 3.4 we sketch the theory of the Chebyshev [57] methods which we use in Chapter V.

3.1 Preconditioned iterative methods

Let $A : X \to X$ be a linear symmetric invertible operator on a finite dimensional real space $X \equiv \mathbb{R}^N$ with inner product $(\cdot, \cdot)$. Consider an equation

$$Au = f, \tag{3.1}$$

where $u, f \in X$. We consider this problem in the context of the operators A induced by bilinear forms defined on finite element spaces.

To make the main idea of preconditioning clear we consider in this section a modified method of the simple iteration. Let $B : X \to X$ be another linear symmetric invertible operator on X. Given initial guess $u_0 \in X$, we define the basic linear iterative method for solving (3.1) by

$$B\,u_{k+1} = B\,u_k - \alpha\,(A\,u_k - f), \qquad k = 0, 1, \ldots, \tag{3.2}$$

or in other form

$$u_{k+1} = T\,u_k + \psi,$$

where $T = (I - B^{-1}A)$ and $\psi = B^{-1}f$.

Let u be the solution of (3.1). Then an error $e_k = u - u_k$ satisfies the equations

$$e_{k+1} = T\,e_k = T^2\,e_{k-1} = \ldots = T^k\,e_0. \tag{3.3}$$

A convergence of method (3.2) for a given initial guess depends on the behavior of the operator T^k. Iterations (3.2) converge when $T^k\,e_0 \to 0$ as $k \to \infty$.

From (3.1), (3.2), and (3.3) it follows that

$$u_{k+1} = T^k\,u_0 + (I - T^k)\,A^{-1}f. \tag{3.4}$$

In general, e_0 is unknown since it involves the unknown exact solution. Hence, it is better to study the computable quantity — a residue $r_k = f - A\,u_k$. From (3.3) we have

$$r_{k+1} = A\,T\,A^{-1}\,r_k = \ldots = A\,T^k\,A^{-1}\,r_0. \tag{3.5}$$

The notion of the *spectral radius* of the operator plays a very important role in the investigation of iterative methods.

Definition 3.1 *The value* $\mu(T) = \lim\limits_{k\to\infty} \|T^k\|^{1/k}$ *is called the spectral radius of the operator* T.

If $\mu(T) \neq 0$ then we have $\|T^k\|^{1/k} = \mu(T) \cdot b(k)$, where $b(k) \to 1$ as $k \to \infty$. From (3.3) we get

$$\|e_k\| \leq \|T^k\|\;\|e_0\| = \mu^k(T)\;b^k(k)\;\|e_0\|. \tag{3.6}$$

So, if $0 \neq \mu(T) < 1$ then to reduce the norm of error $\|e_0\|$ by a factor of $1/\varepsilon$ times for small ε it is sufficient to make

$$k(\varepsilon) \approx \frac{\ln \varepsilon}{\ln \mu(T)} \tag{3.7}$$

iterations.

The value of $(-\ln \mu(T))$ is known as the asymptotic rate of convergence of the iterative method.

One of the main problems of iterative methods is in some sense an optimal choice of operator B. We give the strong mathematical definition of the term *optimal preconditioner* later on page 31 after we provide some basic facts from numerical analysis. It is easy to see that taking $B = A$ in method (3.2) gives us the exact solution for one iteration. Obviously, it is not an optimal method since it includes the inversion of operator A. Assume that we have found some operator B such that $0 \neq \mu(T) < 1$. Let $W(B^{-1})$ be its implementation cost, i.e. a number of arithmetic operations required to implement multiplication of a vector by operator B^{-1}. Since operator A is the finite element approximation of a second-order differential operator using a nodal basis, the corresponding matrix is sparse and the computational work of multiplying a vector by A is on the order of N: $W(A) = O(N)$. Then method (3.2) gives an ε-approximation to the solution of (3.1) for $k(\varepsilon)$ iterations at a cost of

$$W_{it} = \frac{W(T)}{|\ln \mu(T)|} \cdot |\ln \varepsilon|, \tag{3.8}$$

where $W(T) = W(B^{-1}) + O(N)$.

Assume now that operators A and B are positive definite. It is well known from linear algebra [57] that a generalized eigenvalue problem

$$A\varphi = \lambda B\varphi \tag{3.9}$$

has a real and positive spectrum Sp $(B^{-1}A) \equiv \{\lambda_i\}_{i=1}^{N}$ and a corresponding complete orthonormal set of eigenvectors $\{\varphi_i\}_{i=1}^{N}$. Obviously, φ_i, $i = 1, \ldots, N$, are eigenvectors of the operator T:

$$T\ \varphi_i = \nu_i(\alpha)\ \varphi_i, \qquad i = 1, \ldots, N, \tag{3.10}$$

where $\nu_i(\alpha) = 1 - \alpha\lambda_i$. Method (3.2) is convergent if and only if

$$\sup_{\lambda \in \text{Sp }(B^{-1}A)} |1 - \alpha\lambda| \leq q < 1. \tag{3.11}$$

With the well known best choice for $\alpha = 2/(\lambda_{\min} + \lambda_{\max})$ the spectral radius of operator T is given by

$$\mu(T) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} < 1,$$

or

$$\mu(T) = \frac{\lambda_{\max}/\lambda_{\min} - 1}{\lambda_{\max}/\lambda_{\min} + 1}. \tag{3.12}$$

Definition 3.2 *The ratio $\nu = \lambda_{\max}/\lambda_{\min}$ of the extremal eigenvalues of problem* (3.9) *is called the condition number of matrix $B^{-1}A$. We denote this number by Cond $(B^{-1}A)$.*

Obviously, Cond $(B^{-1}A) \geq 1$. From (3.12) it is easy to see that $\mu(T) \to 1$ as Cond $(B^{-1}A) \to \infty$. That is, the bigger the condition number of matrix $B^{-1}A$ the slower the convergence rate of method (3.2).

Operator B is often referred to as a preconditioner. Taking into account (3.8) and (3.12), we would like B to satisfy two properties. First, the solution of the problem

$$B\,\mathbf{v} = \mathbf{g} \tag{3.13}$$

for a given $\mathbf{g} \in X$ should be easy to obtain. And second, B should be spectrally equivalent to A.

Recall that two $N \times N$ symmetric positive definite matrices A and B that result from the discretization of PDE's are called spectrally equivalent matrices [59] if there exist positive constants c_0 and c_1 independent of the grids such that the inequalities

$$c_0(B\varphi, \varphi) \leq (A\varphi, \varphi) \leq c_1(B\varphi, \varphi)$$

hold for any vector $\varphi \in \mathrm{I\!R}^N$.

These two properties will guarantee, firstly, that the work per iteration step in applying the preconditioned method will be small, and secondly, that the number of steps to reduce the error to a given size will also be small and will not depend on mesh size parameter h, so that an efficient algorithm will result. Now we can give the definition of an optimal preconditioner as was done by D'yakonov in [49].

Definition 3.3 *Operator B is an optimal preconditioner if it is spectrally equivalent to operator A and algorithms with estimate $W(B^{-1}) = O(N)$ are established for solving problem* (3.13).

With B being an optimal preconditioner, the computational work (3.8) becomes

$$W_{it} = O(N) \cdot |\ln \varepsilon|. \tag{3.14}$$

3.2 Iterative method for saddle-point problem

As is shown in Chapter V, the use of a nonconforming finite element method on nonmatching grids to problem (2.13) results in an algebraic saddle-point problem with nonsingular matrix

$$\mathcal{A} = \begin{bmatrix} A & C^T \\ C & \mathbf{0} \end{bmatrix}, \tag{3.15}$$

where block A is a symmetric and at least positive semidefinite matrix and block C is a full rank matrix.

In Chapter V we propose a symmetric and positive definite preconditioner $\mathcal{B}$ that is spectrally equivalent to matrix $\mathcal{A}$. The symmetric matrix $\mathcal{A}$ and the symmetric positive definite matrix $\mathcal{B}$ are said to be spectrally equivalent if the spectrum of matrix $\mathcal{B}^{-1}\mathcal{A}$ belongs to the set $[d_1, d_2] \cup [d_3, d_4]$, $d_1 \le d_2 < 0 < d_3 \le d_4$, with the boundaries of the segments independent of the mesh size parameter h.

We propose the preconditioner in the block diagonal form:

$$\mathcal{B} = \begin{bmatrix} B_A & \mathbf{0} \\ \mathbf{0} & B_C \end{bmatrix}. \tag{3.16}$$

In order to justify this choice of $\mathcal{B}$ we consider the eigenvalue problem

$$\mathcal{A} \begin{bmatrix} \mathbf{u}_A \\ \mathbf{u}_C \end{bmatrix} = \nu \hat{R} \begin{bmatrix} \mathbf{u}_A \\ \mathbf{u}_C \end{bmatrix} \tag{3.17}$$

with a symmetric positive definite matrix

$$\hat{R} = \begin{bmatrix} R_A & \mathbf{0} \\ \mathbf{0} & R_C \end{bmatrix}, \tag{3.18}$$

where $R_A = A$, $R_C = C\,A^{-1}\,C^T$, and $(\mathbf{u}_A^T, \mathbf{u}_C^T)^T \in \mathrm{I\!R}^M$. Obviously, matrices R_C and A are positive definite.

Assume in (3.17) that $\nu \ne 1$. Then eliminating the subvector $\mathbf{u}_A$ from the first equation, we obtain

$$\nu R_C\,\mathbf{u}_C = \frac{1}{\nu - 1} R_C\,\mathbf{u}_C. \tag{3.19}$$

It leads to the equation $\nu^2 - \nu - 1 = 0$, which has two roots: $\nu_{1,2} = (1 \pm \sqrt{5})/2$. Thus, the eigenvalues of problem (3.17) belong to the set:

$$\nu \in \left\{ \frac{1-\sqrt{5}}{2};\ \frac{1-\sqrt{5}}{2};\ 1 \right\}. \tag{3.20}$$

Now let the symmetric positive definite matrix $\mathcal{B}$ be spectrally equivalent to matrix $\hat{R}$ with the positive constants c_0 and c_1:

$$c_0(\mathcal{B}\mathbf{u}, \mathbf{u}) \le (\hat{R}\mathbf{u}, \mathbf{u}) \le c_1(\mathcal{B}\mathbf{u}, \mathbf{u}), \qquad \forall \mathbf{u} \in \mathbb{R}^M.$$

Then matrix $\mathcal{B}$ is spectrally equivalent to matrix $\mathcal{A}$ and the spectrum of matrix $\mathcal{B}^{-1}\mathcal{A}$ belongs to the set

$$\mathrm{Sp}\ (\mathcal{B}^{-1}\mathcal{A}) \in [d1, d2] \cup [d3, d4], \tag{3.21}$$

where the constants $d_1 \le d_2 < 0 < d_3 \le d_4$ depend only on the values of c_0 and c_1.

Now consider the system of linear algebraic equations in the saddle-point form:

$$\mathcal{A}\ \mathbf{u} = \mathbf{g}, \tag{3.22}$$

where

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_A \\ \mathbf{u}_C \end{bmatrix}, \qquad \mathbf{g} = \begin{bmatrix} \mathbf{g}_A \\ \mathbf{g}_C \end{bmatrix}, \tag{3.23}$$

and its preconditioned form

$$\mathcal{B}^{-1}\mathcal{A}\ \mathbf{u} = \mathcal{B}^{-1}\ \mathbf{g}. \tag{3.24}$$

In order to solve system (3.22) we can use the generalized Lanczos method of minimal iterations [81]. When applied to system (3.24) the formulae for implementing this method have the form:

$$\begin{gathered} \mathbf{p}_k = \begin{cases} \mathcal{B}^{-1}\boldsymbol{\xi}_0, & k = 1 \\ \mathcal{B}^{-1}\mathcal{A}\mathbf{p}_1 - \alpha_2 \mathbf{p}_1, & k = 2 \\ \mathcal{B}^{-1}\mathcal{A}\mathbf{p}_{k-1} - \alpha_k \mathbf{p}_{k-1} - \beta_k \mathbf{p}_{k-2}, & k \ge 3 \end{cases} \\ \mathbf{u}_0 \in \mathbb{R}^M, \qquad \mathbf{u}_k = \mathbf{u}_{k-1} - \gamma_k \mathbf{p}_k, \quad k \ge 1, \end{gathered} \tag{3.25}$$

where

$$\begin{aligned} \alpha_k &= \frac{(\mathcal{A}\mathcal{B}^{-1}\mathcal{A}\mathbf{p}_{k-1}, \mathcal{A}\mathbf{p}_{k-1})}{(\mathcal{B}^{-1}\mathcal{A}\mathbf{p}_{k-1}, \mathcal{A}\mathbf{p}_{k-1})}, \quad & k \ge 2, \\ \beta_k &= \frac{(\mathcal{B}^{-1}\mathcal{A}\mathbf{p}_{k-1}, \mathcal{A}\mathbf{p}_{k-1})}{(\mathcal{B}^{-1}\mathcal{A}\mathbf{p}_{k-2}, \mathcal{A}\mathbf{p}_{k-2})}, \quad & k \ge 3, \\ \gamma_k &= \frac{(\mathcal{B}^{-1}\boldsymbol{\xi}_{k-1}, \mathcal{A}\mathbf{p}_k)}{(\mathcal{B}^{-1}\mathcal{A}\mathbf{p}_k, \mathcal{A}\mathbf{p}_k)}, \quad & k \ge 1, \end{aligned} \tag{3.26}$$

and $\boldsymbol{\xi}_k = \mathcal{A}\mathbf{u}_k - \mathbf{g}$, $k \ge 0$.

The following expressions are introduced:

$$\kappa = \frac{\max\{d_4, |d_1|\}}{\min\{d_3, |d_2|\}}, \qquad q = \frac{1-\kappa}{1+\kappa}.$$

Then for method (3.25) with (3.26) the following estimate holds true [66, 81]:

$$\|\boldsymbol{\xi}_{2k}\|_{\mathcal{B}^{-1}} < 2q^k\ \|\boldsymbol{\xi}_0\|_{\mathcal{B}^{-1}}, \tag{3.27}$$

where $\|\cdot\|_{\mathcal{B}^{-1}}$ is the norm generated by the symmetric positive definite matrix $\mathcal{B}^{-1}$.

Remark 3.1 Due to property (3.20) the preconditioned Lanczos method with a preconditioner R defined by (3.18) gives the exact solution of system (3.22) in at most six iteration steps.

This statement has only a theoretical meaning since such convergence can never be reached practically because of an overly complicated and expensive preconditioner (3.18). But it gives an idea of how efficient preconditioners can be constructed.

3.3 Preconditioned conjugate gradient method

Now consider problem (3.1)

$$A\mathbf{u} = \mathbf{f}$$

with symmetric and positive definite matrix A and the given vector $\mathbf{f} \in \mathrm{I\!R}^N$. Assume that we have constructed a spectrally equivalent preconditioner B such that Cond $(B^{-1}A) \leq \nu$.

Then we can solve system (3.1) by the preconditioned conjugate gradient (PCG) method in the following form:

$$\begin{array}{c} \mathbf{p}_k = \begin{cases} B^{-1}\boldsymbol{\xi}_0, & k = 0 \\ B^{-1}\boldsymbol{\xi}_k + \beta_k \mathbf{p}_{k-1}, & k > 0 \end{cases} \\ \mathbf{u}_0 \in \mathrm{I\!R}^M, \qquad \mathbf{u}_{k+1} = \mathbf{u}_k + \alpha_k \mathbf{p}_k, \quad k \geq 0, \end{array} \tag{3.28}$$

where

$$\alpha_k = \frac{(B^{-1}\boldsymbol{\xi}_k, \boldsymbol{\xi}_k)}{(A\mathbf{p}_k, \mathbf{p}_k)}, \qquad \beta_k = \frac{(B^{-1}\boldsymbol{\xi}_k, \boldsymbol{\xi}_k)}{(B^{-1}\boldsymbol{\xi}_{k-1}, \boldsymbol{\xi}_{k-1})}, \tag{3.29}$$

and $\boldsymbol{\xi}_k = \mathcal{A}\mathbf{u}_k - \mathbf{g}$, $k \geq 0$.

It is well known that for a given accuracy ε ($\varepsilon \ll 1$) and $k_\varepsilon > \ln(\varepsilon/2) / \ln\ q$, with $q = \frac{\sqrt{\nu}-1}{\sqrt{\nu}+1}$, the following inequality is valid:

$$\|\mathbf{u}_{k_\varepsilon+1} - \mathbf{u}_*\|_A \leq \varepsilon \|\mathbf{u}_0 - \mathbf{u}_*\|_A, \tag{3.30}$$

where $\mathbf{u}_* = A^{-1}\mathbf{f}$.

An essential feature of PCG is that an explicit representation of A and B^{-1} are not needed. In fact, we only need their actions on a given vector.

3.3.1 Estimate for the extremal eigenvalues

In this dissertation we shall study substructuring multilevel and domain decomposition preconditioners. Using this framework, we shall be able to analyze and establish upper bounds for the condition number of our preconditioned matrices. To see how sharp these upper bounds are, we may compute Cond $(B^{-1}A)$ approximately by using a generalized Lancsoz procedure for eigenvalue problems ([59]). We note that the Lanczos algorithm is closely related to the conjugate gradient method. Both algorithms use Krylov subspaces and three-term recurrent formulae [59, 99, 98].

We first define the matrix of normalized residual vectors $R_m \in \mathrm{I\!R}^{N \times m}$ by

$$R_m = \left[\frac{\boldsymbol{\xi}_0}{\|\boldsymbol{\xi}_0\|} \ , \ \ldots \ , \ \frac{\boldsymbol{\xi}_{m-1}}{\|\boldsymbol{\xi}_{m-1}\|} \right],$$

where the vector $\boldsymbol{\xi}_k$ is the residual vector obtained on the k-th iteration of PCG method (3.28), (3.29).

It can be shown [59, 99] that a matrix $T_m = R_m^T \, B^{-1}A \, R_m$ is a $\mathrm{I\!R}^{m\times m}$ tridiagonal matrix:

$$T_m = \begin{bmatrix} \frac{1}{\alpha_1} & -\frac{\sqrt{\beta_1}}{\alpha_1} & & & \\ -\frac{\sqrt{\beta_1}}{\alpha_1} & \frac{\beta_1}{\alpha_1}+\frac{1}{\alpha_2} & -\frac{\sqrt{\beta_2}}{\alpha_2} & & \\ & -\frac{\sqrt{\beta_2}}{\alpha_2} & \frac{\beta_2}{\alpha_2}+\frac{1}{\alpha_3} & \ddots & \\ & & \ddots & \ddots & -\frac{\sqrt{\beta_{m-1}}}{\alpha_{m-1}} \\ & & & -\frac{\sqrt{\beta_{m-1}}}{\alpha_{m-1}} & \frac{\beta_{m-1}}{\alpha_{m-1}}+\frac{1}{\alpha_m} \end{bmatrix}. \tag{3.31}$$

Here α_k and β_k are the parameters of the PCG algorithm. From the theory of Lanczos methods [59, 99] it follows that extremal eigenvalues of matrices T_m provide a good approximation of the extremal eigenvalues of $B^{-1}A$. Thus, to compute an estimation of the condition number of matrix $(B^{-1}A)$ it is sufficient to find the condition number of the tridiagonal and relatively small matrix T_m. Questions related to convergence of the extremal eigenvalues of T_m to those of $(B^{-1}A)$ are considered in [98].

3.4 Chebyshev iterative method

In Chapter V we shall use some preconditioners in the form of the inner Chebyshev iterative procedure [8, 18, 70]. We present here the relevant results and constructions for the sake of completeness.

Again, we assume that B is a symmetric positive definite matrix and that the eigenvalues of matrix $B^{-1}A$ belong to the segment $[a, b]$, where $0 < a \leq b$. Let $P_m(t)$ be a polynomial of degree $m \geq 1$ of least deviation from zero on the segment $[a, b]$ and be normalized by condition $P_m(0) = 1$:

$$P_m(t) = \prod_{i=1}^{m} (1 - \beta_i t). \tag{3.32}$$

The polynomial $P_m(t)$ with these properties is defined in terms of the Chebyshev polynomials [114]:

$$P_m(t) = \frac{1}{T_m(\Theta)} \cdot T_m\left(\frac{b+a-2t}{b-a}\right), \tag{3.33}$$

where $\Theta = (b+a)/(b-a) > 1$ and the Chebyshev polynomial of the 1-st kind of degree m is given by

$$T_m(t) = \cos(m \cdot \arccos(t)) = \frac{1}{2}\left(\left(t+\sqrt{t^2-1}\right)^m + \left(t+\sqrt{t^2-1}\right)^{-m}\right). \tag{3.34}$$

From (3.32) it follows that β_i^{-1}, $i = 1, \ldots, m$, are the roots of polynomial $P_m(t)$. They are easily defined through the roots of the Chebyshev polynomial $T_m(t)$:

$$\beta_i = 2 \cdot \left(b + a - (b-a)\cos \pi \frac{2i-1}{2m}\right)^{-1}, \qquad i = 1, \ldots, m. \tag{3.35}$$

Then preconditioner $\hat{B}$ for matrix A is determined by the formula:

$$\hat{B}^{-1} = \left\{ I - \prod_{i=1}^{m} (I - \beta_i B^{-1} A) \right\} \, A^{-1}. \tag{3.36}$$

According to (3.3) after using one step of the modified method of simple iteration the error of the computed solution $e_k = u - u_k$ is decreased as follows:

$$\|e_{k+1}\| \leq \frac{2\, q^m}{1 + q^{2m}} \, \|e_k\|, \tag{3.37}$$

where $q = (\nu - 1)/(\nu + 1)$ and $\nu = b/a$.

The formulae for calculating the vector $\mathbf{w} = \hat{B}^{-1}\boldsymbol{\xi}$ for a given $\boldsymbol{\xi} \in \mathrm{I\!R}^N$ have the form:

$$\begin{aligned} &\mathbf{w}_0 = \mathbf{0}, \\ &\mathbf{w}_i = \mathbf{w}_{i-1} - \beta_i \, B^{-1} \, (A\mathbf{w}_{i-1} - \boldsymbol{\xi}), \qquad i = 1, \ldots, m, \\ &\mathbf{w} = \mathbf{w}_m. \end{aligned} \tag{3.38}$$

For computational stability, instead of (3.38), we can use the three-term formula [114].

CHAPTER IV

SUBSTRUCTURING PRECONDITIONERS FOR NONCONFORMING FINITE ELEMENT METHOD

4.1 Introduction

Let Ω be a convex bounded domain in $\mathbb{R}^d$, $d = 2, 3$, with boundary $\partial\Omega$. Consider an elliptic problem

$$\begin{aligned} -\mathrm{div}\,(K \cdot \nabla u) &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma_0, \\ (K\nabla u, \mathbf{n}) &= 0 && \text{on } \Gamma_1, \end{aligned} \tag{4.1}$$

where $K(\mathbf{x})$ is a positive definite, uniformly bounded symmetric tensor, $f(\mathbf{x}) \in L^2(\Omega)$, $\overline{\Gamma_0 \cup \Gamma_1} = \partial\Omega$, $\Gamma_0 \cap \Gamma_1 = \emptyset$. We shall consider the case when $\Gamma_0 \equiv \overline{\Gamma_0} \neq \emptyset$. The pure Neumann problem ($\Gamma_0 = \emptyset$) can be treated in a similar way but for the sake of simplicity is not described here.

Let the bilinear form $a(\cdot,\cdot)$ be defined by

$$a(u,v) = (K \cdot \nabla u, \nabla v), \qquad u, v \in V_0(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_0\},$$

where $(\cdot,\cdot)$ denotes the inner product in $L^2(\Omega)$. Then the usual weak form of (4.1) for the solution $u \in V_0(\Omega)$ is

$$a(u,v) = (f,v), \qquad \forall v \in V_0(\Omega). \tag{4.2}$$

Let $\mathcal{T}_h$ be a regular partitioning of Ω into simplices τ with mesh-size h and let $V_h(\Omega)$ be the P_1–nonconforming finite element space of functions $v \in L^2(\Omega)$ [5] such that $v|_\tau$ are linear for all $\tau \in \mathcal{T}_h$, v are continuous at the barycenters of $\tau \in \mathcal{T}_h$ and vanish at the barycenters of the boundary faces on Γ_0 (defined by (2.49)). Note that the space $V_h(\Omega)$ is not a subspace of $H^1(\Omega)$.

Define the bilinear form on $V_h(\Omega)$ by

$$a_h(u,v) = \sum_{\tau \in \mathcal{T}_h} (K\nabla u, \nabla v)_\tau, \qquad \forall\, u, v \in V_h(\Omega), \tag{4.3}$$

where $(\cdot,\cdot)_\tau$ is the inner product in $L^2(\tau)$, $\tau \in \mathcal{T}_h$. Then the P_1–nonconforming finite element discretization of (4.1) has the form: *find $u_h \in V_h$ such that*

$$a_h(u_h, v) = (f, v), \qquad \forall v \in V_h(\Omega). \tag{4.4}$$

Once a nodal basis $\{\varphi_i(\mathbf{x})\}_{i=1}^N$ for $V_h(\Omega)$ is chosen, (4.4) leads to a system of linear algebraic equations. Write $u(\mathbf{x}) = \sum_{i=1}^N u_i \varphi_i(\mathbf{x})$. Then (4.4) becomes

$$\sum_{i=1}^N u_i a_h(\varphi_i, \varphi_j) = (f, \varphi_j), \qquad j = 1, \ldots, N,$$

or in matrix representation

$$A\mathbf{u} = \mathbf{f}, \tag{4.5}$$

where $A_{ji} = a_h(\varphi_i, \varphi_j)$, $f_j = (f, \varphi_j)$, $i, j = 1, \ldots, N$.

Although the methods of solving (4.5) have been extensively studied in the past few years (see, e.g., [5, 15, 20, 22, 35]), their efficiency depends on the coefficient matrix $K(\mathbf{x})$, and in the case of strong anisotropy in the coefficients the question of constructing effective solution techniques is still open.

In this chapter we will describe and analyze a method of constructing the preconditioner for (4.5) using an idea of algebraic substructuring which can be described as follows [71].

Let us partition the domain Ω into subdomains Ω_s, $s = 1, \ldots, n$, such that each Ω_s is a union of simplices $\tau \in \mathcal{T}_h$,

$$\Omega = \bigcup_{s=1}^{n} \Omega_s, \qquad \Omega_s = \bigcup_{l=1}^{n_s} \{\tau_l \in \mathcal{T}_h : \tau_l \subset \Omega_s\}.$$

Below these subdomains Ω_s are called superelements.

Let us introduce local stiffness matrices A_s on each superelement Ω_s as follows:

$$(A_s\mathbf{u}_s, \mathbf{v}_s) = \sum_{\tau_l \subset \Omega_s} (K(\mathbf{x})\nabla u_h, \nabla v_h)_{\tau_l}, \qquad \forall u_h, v_h \in V_h(\Omega_s).$$

All these matrices are at least positive semidefinite, and the global stiffness matrix is determined by assembling the local stiffness matrices over all the superelements:

$$(A\mathbf{u}, \mathbf{v}) = \sum_{s=1}^{n} (A_s\mathbf{u}_s, \mathbf{v}_s), \qquad \forall \mathbf{u}, \mathbf{v} \in \mathrm{I\!R}^N.$$

We can symbolically write

$$A = \{A_s\}_{s=1}^{n},$$

where $\{\cdot\}_{s=1}^{n}$ denotes assembling with respect to the partitioning $\{\Omega_s\}_{s=1}^{n}$ of Ω.

In the above notation each superelement matrix A_s can be represented in terms of local stiffness matrices over simplices τ_l from Ω_s, i.e. $A_s = \{A_{sl}\}_{\tau_l \subset \Omega_s}$. Note that matrices A_{sl} are also at least positive semidefinite.

Following [71, 73], let us introduce on each simplex $\tau \in \mathcal{T}_T$ another matrix $\hat{A}_{sl}$ which has the same kernel as A_{sl} (i.e. Ker A_{sl} = Ker $\hat{A}_{sl}$). Define the matrix $\hat{A}_s$ on each superelement Ω_s by assembling $\hat{A}_{sl}$:

$$\hat{A}_s = \left\{\hat{A}_{sl}\right\}_{\tau_l \subset \Omega_s}.$$

Then it can easily be shown that Ker A_s = Ker $\hat{A}_s$ and the matrices $\hat{A}_s$ are also at least positive semidefinite.

Now let us define an $N \times N$ matrix $\hat{A}$ by assembling $\hat{A}_s$ over all the superelements

$$\hat{A} = \left\{\hat{A}_s\right\}_{s=1}^{n}.$$

To obtain an estimate of the condition number of $\hat{A}^{-1}A$ we use the so-called superelement analysis which we outline here. Suppose we have two sequences of nonnegative numbers $\{a_i\}_{i=1}^{n}$ and $\{b_i\}_{i=1}^{n}$ such that a_i and b_i, $i = 1, \ldots, n$, are simultaneously either positive numbers or zeroes. And suppose we seek for estimates of the ratio $\sum_{i=1}^n a_i / \sum_{i=1}^n b_i$ from below and from above. The solution of this problem is well-known [64]:

$$\min_{\substack{i \\ b_i \neq 0}} \frac{a_i}{b_i} \leq \frac{\sum\limits_{i=1}^{n} a_i}{\sum\limits_{i=1}^{n} b_i} \leq \max_{\substack{i \\ b_i \neq 0}} \frac{a_i}{b_i} \; .$$

Then we can formulate the following lemma:

Lemma 4.1 *The following relations hold.*

$$\max_{(\hat{A}\mathbf{u},\mathbf{u})\neq 0} \frac{(A\mathbf{u},\mathbf{u})}{(\hat{A}\mathbf{u},\mathbf{u})} = \max_{(\hat{A}\mathbf{u},\mathbf{u})\neq 0} \frac{\sum\limits_{s=1}^{n} (A_s\mathbf{u}_s,\mathbf{u}_s)}{\sum\limits_{s=1}^{n} (\hat{A}_s\mathbf{u}_s,\mathbf{u}_s)} \leq \max_{\substack{s=1,\ldots,n \\ (\hat{A}_s\mathbf{u}_s,\mathbf{u}_s)\neq 0}} \frac{(A_s\mathbf{u}_s,\mathbf{u}_s)}{(\hat{A}_s\mathbf{u}_s,\mathbf{u}_s)}, \qquad (4.6)$$

and

$$\min_{(\hat{A}\mathbf{u},\mathbf{u})\neq 0} \frac{(A\mathbf{u},\mathbf{u})}{(\hat{A}\mathbf{u},\mathbf{u})} = \min_{(\hat{A}\mathbf{u},\mathbf{u})\neq 0} \frac{\sum\limits_{s=1}^{n} (A_s\mathbf{u}_s,\mathbf{u}_s)}{\sum\limits_{s=1}^{n} (\hat{A}_s\mathbf{u}_s,\mathbf{u}_s)} \geq \min_{\substack{s=1,\ldots,n \\ (\hat{A}_s\mathbf{u}_s,\mathbf{u}_s)\neq 0}} \frac{(A_s\mathbf{u}_s,\mathbf{u}_s)}{(\hat{A}_s\mathbf{u}_s,\mathbf{u}_s)}. \qquad (4.7)$$

From Lemma 4.1 it is easy to see that to estimate the extreme eigenvalues of $\hat{A}^{-1}A$ it is sufficient to consider the local problems

$$A_s\mathbf{u}_s = \mu^{(s)}\hat{A}_s\mathbf{u}_s, \qquad \mathbf{u}_s \perp \text{Ker } \hat{A}_s,$$

on all the superelements Ω_s, $s = 1, \ldots, n$. Thus, the superelement analysis is a very useful and rather simple tool for estimating the condition numbers of preconditioned matrices (see, e.g., [7, 70, 53, 73]). It can be shown that to estimate the extreme eigenvalues of $\hat{A}^{-1}A$ it is sufficient to consider the worst cases when the superelements Ω_s have no common faces with Γ_0.

Thus, if the superelement matrices A_s and $\hat{A}_s$ are spectrally equivalent with respect to Ker A_s, i.e. there exist constants $c_{0,s}$ and $c_{1,s}$ such that

$$c_{0,s}(\hat{A}_s\mathbf{u}_s,\mathbf{u}_s) \leq (A_s\mathbf{u}_s,\mathbf{u}_s) \leq c_{1,s}(\hat{A}_s\mathbf{u}_s,\mathbf{u}_s), \qquad \forall \mathbf{u}_s \in \mathbb{R}^{N_s}, \qquad N_s = \dim \Omega_s,$$

where constants $c_{0,s}$, $c_{1,s}$ do not depend on mesh-size parameter h, then matrices $\hat{A}$ and A are also spectrally equivalent, i.e.

$$c_0(\hat{A}\mathbf{u},\mathbf{u}) \leq (A\mathbf{u},\mathbf{u}) \leq c_1(\hat{A}\mathbf{u},\mathbf{u}), \qquad \forall \mathbf{u} \in \mathbb{R}^N,$$

with $c_0 = \min\limits_s c_{0,s}$ and $c_1 = \max\limits_s c_{1,s}$.

Now let us partition all the unknowns in (4.5) into two groups:

$$\mathbf{u} = (\mathbf{u}_1^T, \mathbf{u}_2^T)^T, \qquad \dim \mathbf{u}_1 = N_1, \qquad \dim \mathbf{u}_2 = N - N_1,$$

so that matrix $\hat{A}$ is represented in a block form:

$$\hat{A} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} \qquad (4.8)$$

such that block $\hat{A}_{22}$ is easily invertible. Then introducing the Schur complement $S = \hat{A}_{11} - \hat{A}_{12}\hat{A}_{22}^{-1}\hat{A}_{21}$, we can rewrite matrix $\hat{A}$ as

$$\hat{A} = \begin{bmatrix} S + \hat{A}_{12}\hat{A}_{22}^{-1}\hat{A}_{21} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}. \tag{4.9}$$

Following [17, 70, 69], we construct a matrix $\tilde{S}$ which is spectrally equivalent to S, i.e.

$$d_0(\tilde{S}\mathbf{v},\mathbf{v}) \le (S\mathbf{v},\mathbf{v}) \le d_1(\tilde{S}\mathbf{v},\mathbf{v}), \qquad \forall \mathbf{v} \in \mathrm{I\!R}^{N_1},$$

where constants $0 < d_0 \le d_1$ are independent of mesh-size parameter h. Then the matrix

$$B = \begin{bmatrix} \tilde{S} + \hat{A}_{12}\hat{A}_{22}^{-1}\hat{A}_{21} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} \tag{4.10}$$

is spectrally equivalent to matrix A, i.e.

$$r_0(B\mathbf{u},\mathbf{u}) \le (A\mathbf{u},\mathbf{u}) \le r_1(B\mathbf{u},\mathbf{u}), \qquad \forall \mathbf{u} \in \mathrm{I\!R}^{N},$$

where $r_0 = c_0 \min\{1; d_0\}$, $r_1 = c_1 \max\{1; d_1\}$. To construct such a matrix $\tilde{S}$, again, we can use the idea of the algebraic substructuring described above.

Concluding this overview, we can say that the algebraic substructuring procedure consists of the following main steps:

(A) the reconstruction of the directed graph of matrix A from (4.5) in such a way that the resulting matrix $\hat{A}$ has the same kernel and is still positive definite (or positive semidefinite if matrix A is singular);

(B) the representation of matrix $\hat{A}$ in 2×2 block form (4.8) in such a way that one of the blocks, $\hat{A}_{11}$ or $\hat{A}_{22}$, is easily invertible;

(C) the replacement of the Schur complement S in (4.9) by a spectrally equivalent matrix $\tilde{S}$; we can use steps (A) and (B) to construct such a matrix $\tilde{S}$.

Note that we can first represent matrix A in 2×2 block form (4.9) and then use steps (A)–(C) to construct a preconditioner for the Schur complement $S = A_{11} - A_{12}A_{22}^{-1}A_{21}$. Implementing a finite number of these steps, we can get matrix B which is spectrally equivalent to the given matrix A.

Because of the algebraic nature of such a procedure this approach strongly depends on the structure of the graph of matrix A and consequently on the type of the nonconforming finite element space V_h. In this chapter we consider in a different way two- and three-dimensional problems with both constant and almost constant matrix coefficient $K(\mathbf{x})$. Most of the theory developed in this chapter is based on results published by the author in [77, 78], and in joint works with R. Ewing, R. Lazarov, Yu. Kuznetsov, and Z. Chen in [52, 55, 33, 51, 73].

The outline of the reminder of the chapter is as follows. In Section 4.2 we consider a two-dimensional problem with diagonal matrix coefficient $K(\mathbf{x})$. A detailed description of constructing algebraic substructuring preconditioners for three-dimensional problems is given in Sections 4.3 and 4.4. We consider there a formulation of the model problem with a diagonal constant tensor, develop an algebraic substructuring preconditioner for the resulting linear system, and give an implementation algorithm. In Section 4.3 we define partitioning $\mathcal{T}_h$ of the whole domain, subdividing it into topological parallelepipeds and splitting each parallelepiped in turn into **six** tetrahedra. The case of splitting each topological parallelepiped into **five** tetrahedra when $K(\mathbf{x})$ is a diagonal tensor is considered in Section 4.4. In Section 4.5 we consider the case of full tensor function $K(\mathbf{x})$ and domain Ω being a topological parallelepiped. We develop here a variant of the fictitious domain method for anisotropic problems.

4.2 Two-dimensional problem

Consider a model problem on a unit square:

$$\begin{aligned} -k_x \frac{\partial^2 u}{\partial x^2} - k_y \frac{\partial^2 u}{\partial y^2} + c_0 u = f, \qquad & \text{in } \Omega \equiv [0,1]^2, \\ u = 0, \qquad & \text{on } \partial\Omega, \end{aligned} \tag{4.11}$$

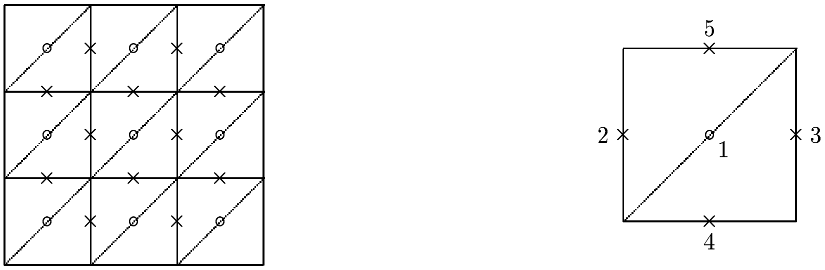
where coefficients $k_x > 0$, $k_y > 0$, and $c_0 \geq 0$, are constants in Ω. It is clear that a method developed for this model problem can be easily generalized for the case of rectangular domain and mixed boundary conditions.

Let $\mathcal{C}_h = \{C^{(i,j)}\}$ be a partition of Ω into uniform squares with the length of the side $h = 1/n$, where (x_i, y_j) is the lower left corner of the square $C^{(i,j)}$. We enumerate the squares in a lexicographical order, first, in the y-direction, then in the x-direction. Next, we divide each square $C^{(i,j)}$ into 2 triangles as shown in Figure 4.1a. The partitioning of Ω into triangles is denoted by $\mathcal{T}_h$.

We introduce the set of centers of all the edges of the triangulation of Ω, and the set Q_h of those centers that are not on the Dirichlet boundary $\Gamma_0 = \partial\Omega$ (see Fig. 4.1a). The Crouzeix-Raviart P_1–nonconforming finite element space V_h is defined by

$$\begin{aligned} V_h = \{v \in L^2(\Omega) : \quad & v|_\tau \in P_1(\tau),\ \forall \tau \in \mathcal{T}_h;\ v \text{ is continuous at the points} \\ & \text{from } Q_h \text{ and vanishes at the middle points of edges on } \Gamma_0\}. \end{aligned} \tag{4.12}$$

Let the dimension of V_h be N. Obviously, $N \approx 3n^2$.



(a) *Triangulation of the domain* Ω.

(b) *Local enumeration of the degrees of freedom.*

Figure 4.1: *2D problem. Triangulation and partition of the degrees of freedom.*

Now we define the bilinear form on V_h by

$$a_h(u,v) = \sum_{\tau \in \mathcal{T}_h} \int_\tau \left(k_x \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + k_y \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} + c_0 uv \right) d\mathbf{x}, \qquad \forall\, u,v \in V_h. \tag{4.13}$$

Thus the nonconforming discretization of problem (4.11) is given by seeking $u_h \in V_h$ such that

$$a_h(u_h, v) = (f, v), \qquad \forall\, v \in V_h. \tag{4.14}$$

For any function $v \in V_h$ we denote by $\mathbf{v} \in \mathrm{I\!R}^N$ its representation with respect to the basis in V_h.

Let $(\mathbf{u}, \mathbf{v})_N$ be a standard bilinear form defined on $\mathrm{I\!R}^N$ by $(\mathbf{u}, \mathbf{v})_N = \sum_{\mathbf{x} \in Q_h} u(\mathbf{x})v(\mathbf{x})$, $\forall u, v \in V_h$. Then define symmetric and positive definite operator $A : \mathrm{I\!R}^N \to \mathrm{I\!R}^N$ by

$$(A\mathbf{u}, \mathbf{v})_N = a_h(u, v), \qquad u, v \in V_h. \tag{4.15}$$

For each square $C = C^{(i,j)} \in \mathcal{C}_h$, we denote by V_h^C the subspace of the restriction of the functions from V_h into C. For each $v \in V_h^C$, we indicate by $\mathbf{v}_c$ the corresponding vector. The dimension of V_h^C is denoted by N_c. Obviously, for a square without faces on Γ_0 we have $N_c = 5$.

The local stiffness matrix A^C on a square $C \in \mathcal{C}_h$ is given by

$$\begin{aligned}(A^C\mathbf{u}_c, \mathbf{v}_c)_{N_c} = \sum_{\tau \subset C} \Big(k_x (\tfrac{\partial u}{\partial x}, \tfrac{\partial v}{\partial x})_\tau + k_y (\tfrac{\partial u}{\partial y}, \tfrac{\partial v}{\partial y})_\tau + c_0 (u, v)_\tau \Big), \\ \forall u, v \in V_h^C.\end{aligned} \tag{4.16}$$

Note that the matrices A^C are positive definite when $\partial C \cap \Gamma_0 \neq 0$ and at least semidefinite otherwise (if $c_0 \neq 0$ then all the matrices A^C are positive definite). The global stiffness matrix is determined by assembling the local stiffness matrices:

$$(A\mathbf{u}, \mathbf{v})_N = \sum_{C \in \mathcal{C}_h} (A^C\mathbf{u}_c, \mathbf{v}_c)_{N_c}, \qquad \forall \mathbf{u}, \mathbf{v} \in \mathrm{I\!R}^N. \tag{4.17}$$

To define the solution procedure we divide all the unknowns in the system into two groups:

1. The first group consists of the unknowns corresponding to the edges of the triangles that are internal for each square (these are the unknowns corresponding to the nodes marked by "∘" in Fig. 4.1). We denote these unknowns by $vc_{i,j}$, $i, j = 1, \ldots, n$.

2. The second group consists of all the unknowns corresponding to the edges of the squares in partition $\mathcal{C}_h$, without the faces on Γ_0 (Fig. 4.1, the nodes marked by "×").

   (a) First, we enumerate the unknowns on the edges perpendicular to the x-axis (nodes 2 and 3 in Fig. 4.1b). We denote these unknowns by $vx_{i,j}$, $i = 1, \ldots, n-1$, $j = 1, \ldots, n$.

   (b) Second, we enumerate the unknowns on the edges perpendicular to the y-axis (nodes 4 and 5 in Fig. 4.1b). We denote these unknowns by $vy_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, n-1$.

Now we consider a square C that has no face on the boundary $\partial\Omega$ and enumerate the edges s_j, $j = 1, \ldots, 5$, of the triangles in this square in correspondence with the partitioning introduced above as is shown in Figure 4.1b. Then the local stiffness matrix for this square has the following form:

$$A^C = \begin{bmatrix} A_{11,c} & A_{12,c} \\ A_{21,c} & A_{22,c} \end{bmatrix}. \tag{4.18}$$

Introducing parameter $c = c_0 h^2/12$ we can write

$$A_{11,c} = 4\left[k_x + k_y + c\right], \qquad A_{12,c} = A_{21,c}^T = \left[-2k_x, -2k_x, -2k_y, -2k_y\right], \tag{4.19}$$

$$A_{22,c} = \begin{bmatrix} 2k_x & & & \\ & 2k_x & & \\ & & 2k_y & \\ & & & 2k_y \end{bmatrix} + 2c \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}.$$

The splitting of the space $\mathrm{I\!R}^N$ induces the presentation of the vectors: $\mathbf{v}^T = (\mathbf{v}_1^T, \mathbf{v}_2^T)$, where $\mathbf{v}_1 \in \mathrm{I\!R}^{N_1}$, $\mathbf{v}_2 \in \mathrm{I\!R}^{N_2}$, and $\mathbf{v}_2$ corresponds to the unknowns of the 2-nd group. Obviously, $N_1 = n^2$ and $N_2 = N - n^2$. Then matrix A can be presented in the following block form:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{4.20}$$

where $A_{22} : \mathrm{I\!R}^{N_2} \to \mathrm{I\!R}^{N_2}$ is a diagonal matrix.

Now denote by $\hat{A}_{11} = A_{11} - A_{12} A_{22}^{-1} A_{21}$ the Schur complement of A obtained by elimination of the vector $\mathbf{v}_2$. Then $A_{11} = \hat{A}_{11} + A_{12} A_{22}^{-1} A_{21}$, so matrix A has the form:

$$A = \begin{bmatrix} \hat{A}_{11} + A_{12} A_{22}^{-1} A_{21} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \tag{4.21}$$

To understand the structure of the Schur complement $\hat{A}_{11}$ let us write explicitly the matrix equation

$$A\mathbf{v} = \mathbf{g}$$

in terms of the unknowns $vc_{i,j}$, $vx_{i,j}$, and $vy_{i,j}$. For any square $C^{(i,j)} \cap \partial\Omega \neq 0$ we have:

$$4(k_x + k_y + c) vc_{i,j} - 2k_x (vx_{i,j} + vx_{i+1,j}) - 2k_y (vy_{i,j} + vy_{i,j+1}) = gc_{i,j}, \qquad i,j = 1, \ldots, n,$$

$$4(k_x + c) vx_{i,j} - 2k_x (vc_{i-1,j} + vc_{i,j}) = gx_{i,j}, \qquad i = 2, \ldots, n, \ j = 1, \ldots, n,$$

$$4(k_y + c) vy_{i,j} - 2k_y (vc_{i,j-1} + vc_{i,j}) = gy_{i,j}, \qquad i = 1, \ldots, n, \ j = 2, \ldots, n.$$

After eliminating the unknowns $vx_{i,j}$ and $vy_{i,j}$ we have a 5-point computational scheme for the unknowns $vc_{i,j}$:

$$(2a_x + 2a_y + b) vc_{i,j} - a_x (vc_{i-1,j} + vc_{i+1,j}) - a_y (vc_{i,j-1} + vc_{i,j+1}) = \tilde{gc}_{i,j}, \tag{4.22}$$

where

$$a_x = \frac{k_x}{1 + c/k_x}, \quad a_y = \frac{k_y}{1 + c/k_y}, \qquad b = 4c\left(1 + \frac{1}{1 + c/k_x} + \frac{1}{1 + c/k_y}\right). \tag{4.23}$$

It is easy to see that matrix $\hat{A}_{11}$ can be represented in a tensor product form (according to the enumeration introduced earlier in this section):

$$\hat{A}_{11} = a_x (A_x \otimes I_y) + a_y (I_x \otimes A_y) + b (I_x \otimes I_y), \tag{4.24}$$

where the matrices $I_x, I_y : \mathbb{R}^n \to \mathbb{R}^n$ are identity ones, and the matrices $A_x, A_y : \mathbb{R}^n \to \mathbb{R}^n$ are tridiagonal:

$$A_x = A_y = \begin{bmatrix} 3 & -1 & & & \mathbf{0} \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ \mathbf{0} & & & -1 & 3 \end{bmatrix}. \tag{4.25}$$

To solve the problem with separable matrix $\hat{A}_{11}$ we can use either the discrete fast Fourier transform [106] or an algebraic multigrid method (AMG) [8, 20, 70, 120]. When an implementation cost of the first method is estimated by $O(h^{-2}\ln(h^{-1}))$, the AMG methods have the optimal order of arithmetic complexity $O(h^{-2})$. Since these methods are well described in the literature we are not going to discuss them in greater detail.

4.3 3D problem. Partition of cube into 6 tetrahedra

In this section we consider multilevel preconditioners for (4.5) based on the partitioning of the regular parallelepipeds into tetrahedral substructures, following the ideas in [52, 55]. Here we treat the case where Ω is a unit cube and $K(\mathbf{x})$ is a diagonal tensor.

4.3.1 Two level preconditioners

Let $\mathcal{C}_h = \{C^{(i,j,k)}\}$ be a partition of Ω into uniform cubes with length $h = 1/n$, where (x_i, y_j, z_k) is the right back upper corner of the cube $C^{(i,j,k)}$. Next, each cube $C^{(i,j,k)}$ is divided into two prisms $P_1 = P_1^{(i,j,k)}$ and $P_2 = P_2^{(i,j,k)}$ as shown in Figure 4.2. The resulting partition of Ω is denoted by $\mathcal{P}_h$. Finally, we divide each prism into three tetrahedra as illustrated in Figure 4.2 and denote this partition of Ω into tetrahedra by $\mathcal{T}_h$.

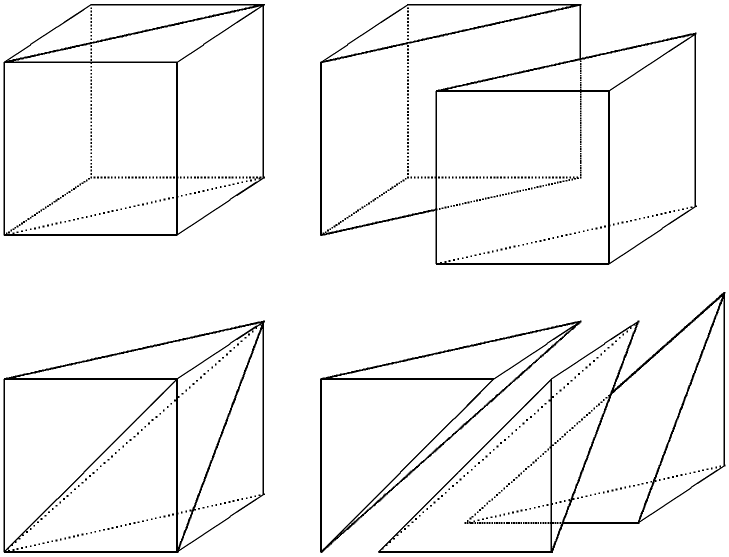


Figure 4.2: *The partition of a cube into 2 prisms and 6 tetrahedra.*

Let $W_{c,h}$ be the space of piecewise constants associated with $\mathcal{C}_h$, and $P_{c,h}$ be the L^2-projection onto $W_{c,h}$. To define our preconditioners, we introduce $C_h = P_{c,h}K^{-1}$ in the hybrid form (2.38) instead of $C_h = P_h K^{-1}$. Obviously, Lemma 2.8 and Proposition 2.1 are still valid for this modification since $\mathcal{T}_h$ is a refinement of $\mathcal{C}_h$. With this modification, C_h^{-1} is a constant on each cube. For notational convenience, we drop the subscript h and simply write $C_h^{-1} = \text{diag}\,\{k_1, k_2, k_3\}$.

Let V_h be the nonconforming finite element space associated with $\mathcal{T}_h$ as defined in (2.49), and let its dimension be N. All the unknowns on the faces of $\partial\Omega$ are excluded. For any function $v_h \in V_h$, we denote by $\mathbf{v} \in \mathbb{R}^N$ the corresponding vector of its degrees of freedom. Introduce the inner product

$$(\mathbf{u}, \mathbf{v})_N = h^3 \sum_{p_i \in \partial\mathcal{T}_h} u_h(p_i) v_h(p_i), \quad u_h, v_h \in V_h, \tag{4.26}$$

where the p_i's are the barycenters of the interior faces. The norm induced by (4.26) is equivalent to the L^2-norm on Ω.

For each prism $P = P^{(i,j,k)} \in \mathcal{P}_h$, denote by V_h^P the subspace of the restriction of the functions from V_h into P. For each $v \in V_h^P$, we indicate by $\mathbf{v}_P$ its corresponding vector. The dimension of V_h^P is denoted by N^P. Obviously, for a prism without faces on $\partial\Omega$ its dimension is 10, i.e. $N^P = 10$.

The local stiffness matrix A^P on the prism $P \in \mathcal{P}_h$ is given by

$$(A^P \mathbf{u}_P, \mathbf{v}_P)_{N^P} = \sum_{\tau \subset P} (C_h^{-1} \nabla u_h, \nabla v_h)_\tau. \tag{4.27}$$

Then the global stiffness matrix is determined by assembling the local stiffness matrices:

$$(A\mathbf{u}, \mathbf{v})_N = \sum_{P \in \mathcal{P}_h} (A^P \mathbf{u}_P, \mathbf{v}_P)_{N^P}. \tag{4.28}$$

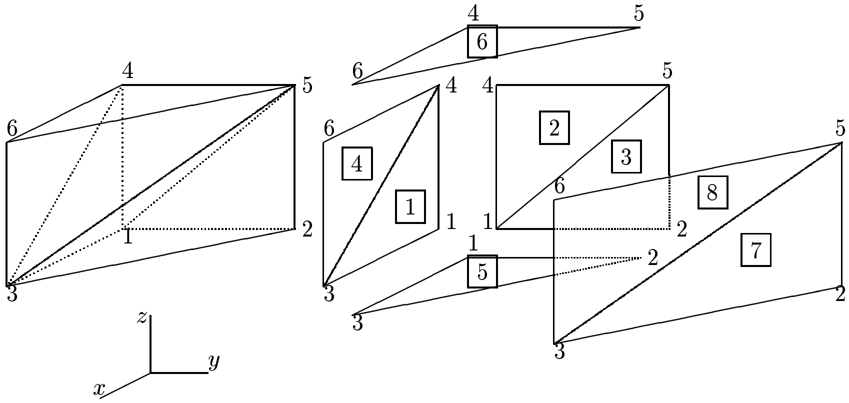
Now we consider a prism P of a cube that has no face on the boundary $\partial\Omega$ and enumerate the faces s_j, $j = 1, \ldots, 10$, of the tetrahedra in this prism as shown in Figure 4.3. Then the local stiffness matrix of this prism has the following form:

$$A^P = \frac{3h}{2} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{4.29}$$

where $A_{11} = \text{diag}\,\{k_2,\ k_1,\ k_1,\ k_2,\ k_3,\ k_3\}$ and

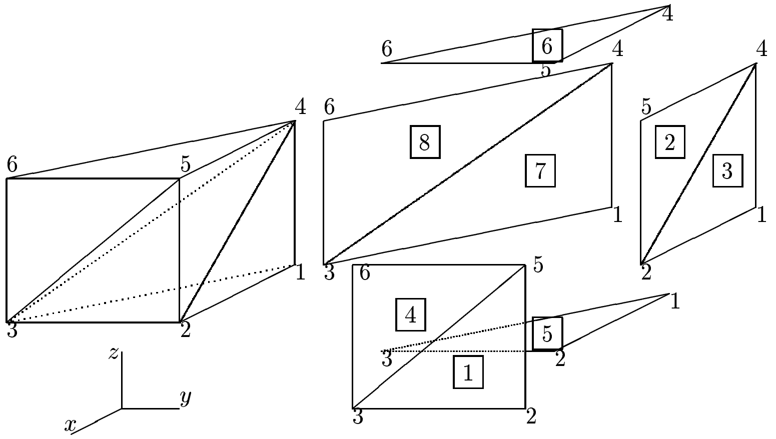
$$A_{21} = A_{12}^T = \begin{bmatrix} 0 & 0 & -k_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -k_2 & 0 & 0 \\ -k_2 & 0 & 0 & 0 & -k_3 & 0 \\ 0 & -k_1 & 0 & 0 & 0 & -k_3 \end{bmatrix},$$

$$A_{22} = \begin{bmatrix} k_1 + k_2 & 0 & -k_2 & 0 \\ 0 & k_1 + k_2 & 0 & -k_1 \\ -k_2 & 0 & 2(k_2 + k_3) & -k_3 \\ 0 & -k_1 & -k_3 & 2(k_1 + k_3) \end{bmatrix}.$$



(a) Prism P_1

$s_1 = (1,4,3)$ $s_3 = (1,2,5)$ $s_5 = (1,2,3)$ $s_7 = (2,5,3)$ $s_9 = (1,5,3)$
$s_2 = (1,4,5)$ $s_4 = (3,4,6)$ $s_6 = (4,5,6)$ $s_8 = (3,5,6)$ $s_{10} = (3,4,5)$



(b) Prism P_2

$s_1 = (2,3,5)$ $s_3 = (1,2,4)$ $s_5 = (1,2,3)$ $s_7 = (1,3,4)$ $s_9 = (2,3,4)$
$s_2 = (2,4,5)$ $s_4 = (3,5,6)$ $s_6 = (4,5,6)$ $s_8 = (3,4,6)$ $s_{10} = (3,4,5)$

Figure 4.3: *Local enumeration of faces in prisms.*

Along with matrix A^P we also introduce a new matrix B^P. The purpose of introducing B^P is to simplify the graph of connectedness in the local stiffness matrix in such a way that the kernel is preserved and the elimination of the internal for the prism unknowns leads to a simplier Schur complement. Matrix B^P is defined on the same space V_h^P by

$$B^P = \frac{3h}{2} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & B_{22} \end{bmatrix}, \tag{4.30}$$

where

$$B_{22} = \begin{bmatrix} k_1+k_2+b & -b & -k_2 & 0 \\ -b & k_1+k_2+b & 0 & -k_1 \\ -k_2 & 0 & 2k_2+k_3 & 0 \\ 0 & -k_1 & 0 & 2k_1+k_3 \end{bmatrix},$$

with a parameter b. This parameter will be chosen in such a way that matrix B^P is spectrally equivalent to A^P (with respect to the kernel) with a possibly smallest relative condition number.

Proposition 4.1 *It holds that* $\text{Ker } A^P = \text{Ker } B^P$.

Proof: It is easy to see from the definitions of A^P and B^P that $\text{Ker } A^P = \text{Ker } B^P = \{\mathbf{v} = (v_1, v_2, \ldots, v_{10})^T \in \mathrm{I\!R}^{10} : v_i = v_1,\ i = 2, \ldots, 10\}$. □

Remark 4.1 If the prism $P \in \mathcal{P}_h$ has a face on $\partial\Omega$, then matrix A^P does not have the rows and columns which correspond to the nodes on that face, and the modification of B_{22} is obvious.

Now we define the $N \times N$ matrix B by the following equality:

$$(B\mathbf{u}, \mathbf{v})_N = \sum_{P \in \mathcal{P}_h} (B^P \mathbf{u}_P, \mathbf{v}_P)_{N^P}, \qquad \forall \mathbf{u}, \mathbf{v} \in \mathrm{I\!R}^N. \tag{4.31}$$

Since matrix B is used for preconditioning the original problem (4.5), it is important to estimate the condition number of $B^{-1}A$. Thus, we consider an eigenvalue problem:

$$A\mathbf{u} = \mu B\mathbf{u}. \tag{4.32}$$

Lemma 4.2 *Let $\mu_P \neq 0$ satisfy the equality*

$$A^P \mathbf{u}_P = \mu_P B^P \mathbf{u}_P, \qquad P \in \mathcal{P}_h,\ \mathbf{u}_P \neq 0. \tag{4.33}$$

Then we have

$$\max_{(B\mathbf{u},\mathbf{u})_N \neq 0} \frac{(A\mathbf{u}, \mathbf{u})_N}{(B\mathbf{u}, \mathbf{u})_N} \leq \max_{P \in \mathcal{P}_h} \mu_P \quad \textit{and} \quad \min_{(B\mathbf{u},\mathbf{u})_N \neq 0} \frac{(A\mathbf{u}, \mathbf{u})_N}{(B\mathbf{u}, \mathbf{u})_N} \geq \min_{P \in \mathcal{P}_h} \mu_P. \tag{4.34}$$

Proof: For each $P \in \mathcal{P}_h$, it follows from (4.33) that

$$(A^P \mathbf{u}_P, \mathbf{u}_P)_{N^P} = \mu_P (B^P \mathbf{u}_P, \mathbf{u}_P)_{N^P}.$$

Then, from the fact that all the local stiffness matrices are nonnegative it follows that

$$\begin{aligned}\sum_{P\in\mathcal{P}_h}(A^P\mathbf{u}_P,\mathbf{u}_P)_{N^P} &= \sum_{P\in\mathcal{P}_h}\mu_P(B^P\mathbf{u}_P,\mathbf{u}_P)_{N^P}\\ &\le \max_{P\in\mathcal{P}_h}\mu_P\sum_{P\in\mathcal{P}_h}(B^P\mathbf{u}_P,\mathbf{u}_P)_{N^P}.\end{aligned}$$

Hence, from the definitions of A and B we see that

$$(A\mathbf{u},\mathbf{u})_N \le \max_{P\in\mathcal{P}_h}\mu_P(B\mathbf{u},\mathbf{u})_N.$$

Consequently, the first inequality in (4.34) is true. The same argument can be used to show the second inequality. □

From Lemma 4.2, we see that, to estimate the condition number of $B^{-1}A$, it suffices to consider local problems (4.33). Using a superelement analysis [69] to estimate $\max\limits_{P\in\mathcal{P}_h}\mu_P$ and $\min\limits_{P\in\mathcal{P}_h}\mu_P$, it suffices to treat the worst case where the prism $P\in\mathcal{P}_h$ has no face on the boundary $\partial\Omega$. From (4.29) and (4.30), direct calculations show that the eigenvalues μ_P are within the interval $[\mu_P^-,\mu_P^+]$, where

$$\mu_P^{\pm}=\frac{1}{2}\left(1+\frac{k_3}{k_1}+\frac{k_3}{k_2}+\frac{k_3}{b}\right)\left(1\pm\sqrt{1-\frac{4k_3/b}{(1+k_3/k_1+k_3/k_2+k_3/b)^2}}\right). \tag{4.35}$$

Obviously, $\mu_P^{\pm}$ depends on the parameter b. We shall choose b to minimize the ratio μ_P^+/μ_P^-, which then gives an upper bound for the condition number Cond $(B^{-1}A)$.

Until the end of the section we shall use the following assumption.

Assumption 4.1 *Assume that the matrix coefficient of equation* (4.1) *is a diagonal tensor* $K(\mathbf{x})=$ *diag* $\{k_1,k_2,k_3\}$, *where* k_i, $i=1,2,3$, *are constants on each prism* $P\in\mathcal{P}_h$, *and there exists a parameter* κ *such that*

$$\max_{P\in\mathcal{P}_h}\left\{\frac{k_3}{k_1},\ \frac{k_3}{k_2}\right\}\le\kappa. \tag{4.36}$$

Remark 4.2 Generally speaking, we need only the assumption that the coefficient k_* in some direction multiplied by some fixed parameter $1/\kappa$ is not greater than the coefficients in the other directions. For the sake of simplicity we assume that this is the “z-direction”.

The optimal choice of b is given in the following theorem.

Theorem 4.1 *The eigenvalues of problem* (4.32) *with the parameter* $b^{-1}=k_1^{-1}+k_2^{-1}+k_3^{-1}$ *belong to the interval*

$$\left[(1+2\kappa)\left(1-\sqrt{\frac{2\kappa}{1+2\kappa}}\right),\ (1+2\kappa)\left(1+\sqrt{\frac{2\kappa}{1+2\kappa}}\right)\right],$$

and the condition number is then estimated as follows:

$$Cond\ (B^{-1}A)\le 3+8\kappa.$$

Proof: With the choice $b^{-1} = k_1^{-1} + k_2^{-1} + k_3^{-1}$, the expression $\mu_P^{\pm}$ can be written as

$$\mu_P^{\pm} = \left(1 + \frac{k_3}{k_1} + \frac{k_3}{k_2}\right)\left(1 \pm \sqrt{1 - \frac{1}{1 + \frac{k_3}{k_1} + \frac{k_3}{k_2}}}\right).$$

Then we consider the functions

$$f_{\pm}(x) = x\left(1 \pm \sqrt{1 - \frac{1}{x}}\right), \quad x \geq 1.$$

Note that f_+ is a nondecreasing function and f_- is a nonincreasing function. Hence, the desired result follows from the definition of κ. □

Remark 4.3 If the parameter b is chosen by the simple relation $b = k_3$, then the eigenvalues of problem (4.32) belong to the interval

$$\left[1 + \kappa - \sqrt{\kappa^2 + 2\kappa},\, 1 + \kappa + \sqrt{\kappa^2 + 2\kappa}\right],$$

and the condition number is thus estimated by

$$\text{Cond}\,(B^{-1}A) \leq 3 + 8\kappa + 4\kappa^2.$$

We stress that the condition number of matrix $B^{-1}A$ is bounded by a constant independent of the step size of mesh h. Since we introduced a two level subdivision, matrix B can be referred to as a two level preconditioner.

Remark 4.4 Because the condition number of matrix $B^{-1}A$ depends on the value of the parameter κ it is very important to choose the "z-direction" in the proper way. Note that we can always rearrange the coordinate axes (make a change of coordinates) to ensure Assumption 4.1.

4.3.2 Three level preconditioners

While preconditioner B has good properties, it is not economical to invert it. In this subsection we propose a modification of matrix B and consider its properties and computational scheme. Toward the end of this section, we divide all unknowns in the system into two groups:

1. The first group consists of all the unknowns corresponding to the faces of the prisms in partition $\mathcal{P}_h$, excluding the faces on $\partial\Omega$ (see Figure 4.3).

2. The second group consists of the unknowns corresponding to the faces of the tetrahedra that are internal for each prism (these are faces s_9 and s_{10} in Figure 4.3).

This splitting of the space $\mathrm{I\!R}^N$ induces the presentation of the vectors: $\mathbf{v} = (\mathbf{v}_1^T, \mathbf{v}_2^T)^T$, where $\mathbf{v}_1 \in \mathrm{I\!R}^{N_1}$ and $\mathbf{v}_2 \in \mathrm{I\!R}^{N_2}$. Obviously, $N_1 = N - 4n^3$. Then matrix B can be presented in the following block form:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \qquad \dim B_{11} = N_1. \tag{4.37}$$

Now we denote by $\hat{B}_{11} = B_{11} - B_{12}B_{22}^{-1}B_{21}$ the Schur complement of B obtained by elimination of vector $\mathbf{v}_2$. Then $B_{11} = \hat{B}_{11} + B_{12}B_{22}^{-1}B_{21}$, and hence matrix B has the form:

$$B = \begin{bmatrix} \hat{B}_{11} + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}. \tag{4.38}$$

Note that for each prism $P \in \mathcal{P}_h$ the unknowns on faces s_9 and s_{10} (see Figure 4.3) are connected only with the unknowns associated with this prism and therefore can be eliminated locally; that is, matrix B_{22} is block diagonal with 2 × 2 blocks and can be inverted locally (prism by prism). Thus, matrix $\hat{B}_{11}$ is easily computable. The proposed modification of matrix B in (4.38) is of the form

$$\tilde{B} = \begin{bmatrix} \tilde{B}_0 + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where $\tilde{B}_0$ is to be defined later.

4.3.2.1 Group partitioning of grid points

For the sake of simplicity of representation of matrices and computational schemes we introduce the partitioning of all nodes in $\partial\mathcal{T}_h$ into the following three groups. Denote by $s_{r,l,m}^{(i,j,k)}$ the face of the cube $C^{(i,j,k)}$ with vertices r, l, m (see Figure 4.4).

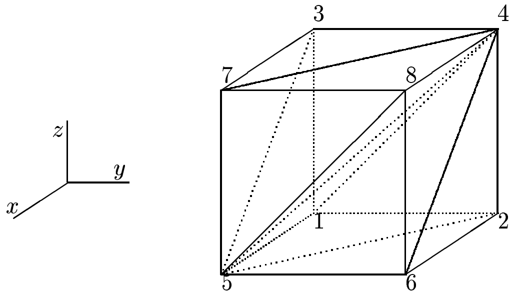


Figure 4.4: *Enumeration of the vertices of a cube* $C^{(i,j,k)}$.

1. First, we group the nodes on the faces

$$s_{2,4,5}^{(i,j,k)} \quad \text{and} \quad s_{4,5,7}^{(i,j,k)}, \qquad i,j,k = \overline{1,n};$$

we denote the unknowns at these nodes by $VI_\ell^{(i,j,k)}$, $\ell = 1,2$, $i,j,k = \overline{1,n}$.

2. Second, we number the nodes on the faces perpendicular to x, y, and z axes:

$$\text{(i)} \quad s_{1,2,4}^{(i,j,k)}, \quad s_{1,3,4}^{(i,j,k)}, \quad i = \overline{2,n}, \quad j,k = \overline{1,n};$$

we denote the unknowns at these nodes by $Vx_\ell^{(i,j,k)}$, $\ell = 1,2$, $i = \overline{2,n}$, $j,k = \overline{1,n}$.

$$\text{(ii)} \quad s_{1,3,5}^{(i,j,k)}, \quad s_{5,3,7}^{(i,j,k)}, \quad j = \overline{2,n}, \quad i,k = \overline{1,n};$$

we denote the unknowns at these nodes by $Vy_\ell^{(i,j,k)}$, $\ell = 1, 2$, $j = \overline{2,n}$, $i, k = \overline{1,n}$.

$$\text{(iii)} \quad s_{1,2,5}^{(i,j,k)}, \quad s_{2,5,6}^{(i,j,k)}, \quad i, j = \overline{1,n}, \quad k = \overline{2,n};$$

we denote the unknowns at these nodes by $Vz_\ell^{(i,j,k)}$, $\ell = 1, 2$, $i, j = \overline{1,n}$, $k = \overline{2,n}$.

3. Finally, we number the remaining nodes on the faces

$$s_{1,4,5}^{(i,j,k)}, \quad s_{3,4,5}^{(i,j,k)}, \quad s_{4,5,6}^{(i,j,k)}, \quad s_{4,5,8}^{(i,j,k)}, \quad i, j, k = \overline{1,n};$$

we denote the unknowns at these nodes by $VA_\ell^{(i,j,k)}$, $\ell = \overline{1,4}$, $i, j, k = \overline{1,n}$.

4.3.2.2 Definition of the preconditioner

We partition each cube $C^{(i,j,k)}$ into left and right prisms $P_p^{(i,j,k)}$, $p = 1, 2$ (see Fig. 4.2). Below we skip the indices '(i, j, k)' and the superscript 'P' when no ambiguity occurs.

In the local numeration (see Fig 4.3) matrices B_1 and B_2, corresponding to the left and right prisms have the form (4.30). We rewrite these matrices in the above group partitioning:

$$B_1 = \frac{3h}{2} \left[\begin{array}{cc|cccccc|cc}
k_1{+}k_2{+}b & -b & -k_1 & 0 & 0 & 0 & 0 & 0 & -k_2 & 0 \\
-b & k_1{+}k_2{+}b & 0 & 0 & 0 & -k_2 & 0 & 0 & 0 & -k_1 \\
\hline
-k_1 & 0 & k_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & k_1 & 0 & 0 & 0 & 0 & 0 & -k_1 \\
0 & 0 & 0 & 0 & k_2 & 0 & 0 & 0 & -k_2 & 0 \\
0 & -k_2 & 0 & 0 & 0 & k_2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & k_3 & 0 & -k_3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & k_3 & 0 & -k_3 \\
\hline
-k_2 & 0 & 0 & 0 & -k_2 & 0 & -k_3 & 0 & 2k_2{+}k_3 & 0 \\
0 & -k_1 & 0 & -k_1 & 0 & 0 & 0 & -k_3 & 0 & 2k_1{+}k_3
\end{array} \right],$$

$$B_2 = \frac{3h}{2} \left[\begin{array}{cc|cccccc|cc}
k_1{+}k_2{+}b & -b & 0 & 0 & -k_2 & 0 & 0 & 0 & -k_1 & 0 \\
-b & k_1{+}k_2{+}b & 0 & -k_1 & 0 & 0 & 0 & 0 & 0 & -k_2 \\
\hline
0 & 0 & k_1 & 0 & 0 & 0 & 0 & 0 & -k_1 & 0 \\
0 & -k_1 & 0 & k_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-k_2 & 0 & 0 & 0 & k_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & k_2 & 0 & 0 & 0 & -k_2 \\
0 & 0 & 0 & 0 & 0 & 0 & k_3 & 0 & -k_3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & k_3 & 0 & -k_3 \\
\hline
-k_1 & 0 & -k_1 & 0 & 0 & 0 & -k_3 & 0 & 2k_1{+}k_3 & 0 \\
0 & -k_2 & 0 & 0 & 0 & -k_2 & 0 & -k_3 & 0 & 2k_2{+}k_3
\end{array} \right].$$

The partitioning of nodes into the above three groups induces the following block forms of matrices B_p, $p = 1, 2$:

$$B_p = \begin{bmatrix} B_{11,p} & B_{12,p} \\ B_{21,p} & B_{22,p} \end{bmatrix}, \qquad p = 1, 2, \tag{4.39}$$

where blocks $B_{22,p}$ correspond to the unknowns of the last group and blocks $B_{11,p}$ correspond to the unknowns of the first and second groups.

We eliminate the unknowns of the last group from each matrix B_p, $p = 1, 2$, which is done locally on each prism. Then we get the matrices

$$\hat{B}_{11,p} = B_{11,p} - B_{12,p} B_{22,p}^{-1} B_{21,p}, \qquad p = 1, 2,$$

where

$$B_{12,1} B_{22,1}^{-1} B_{21,1} = \frac{3h}{2} \left[\begin{array}{cc|cccccc} \frac{k_2^2}{2k_2+k_3} & 0 & 0 & 0 & \frac{k_2^2}{2k_2+k_3} & 0 & \frac{k_2k_3}{2k_2+k_3} & 0 \\ 0 & \frac{k_1^2}{2k_1+k_3} & 0 & \frac{k_1^2}{2k_1+k_3} & 0 & 0 & 0 & \frac{k_1k_3}{2k_1+k_3} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{k_1^2}{2k_1+k_3} & 0 & \frac{k_1^2}{2k_1+k_3} & 0 & 0 & 0 & \frac{k_1k_3}{2k_1+k_3} \\ \frac{k_2^2}{2k_2+k_3} & 0 & 0 & 0 & \frac{k_2^2}{2k_2+k_3} & 0 & \frac{k_2k_3}{2k_2+k_3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ \frac{k_2k_3}{2k_2+k_3} & 0 & 0 & 0 & \frac{k_2k_3}{2k_2+k_3} & 0 & \frac{k_3^2}{2k_2+k_3} & 0 \\ 0 & \frac{k_1k_3}{2k_1+k_3} & 0 & \frac{k_1k_3}{2k_1+k_3} & 0 & 0 & 0 & \frac{k_3^2}{2k_1+k_3} \end{array} \right],$$

and a similar expression holds for $B_{12,2} B_{22,2}^{-1} B_{21,2}$.

Following [52], we introduce on each prism a modification of matrices $\hat{B}_{11,p}$:

$$B_0 = \frac{3h}{2} \left[\begin{array}{cc|cccccc} k_1+k_2+b+s_2 & -b & -k_1 & 0 & -k_2 & 0 & -s_2/2 & -s_2/2 \\ -b & k_1+k_2+b+s_1 & 0 & -k_1 & 0 & -k_2 & -s_1/2 & -s_1/2 \\ \hline -k_1 & 0 & k_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -k_1 & 0 & k_1 & 0 & 0 & 0 & 0 \\ -k_2 & 0 & 0 & 0 & k_2 & 0 & 0 & 0 \\ 0 & -k_2 & 0 & 0 & 0 & k_2 & 0 & 0 \\ -s_2/2 & -s_1/2 & 0 & 0 & 0 & 0 & \frac{s_1+s_2}{2} & 0 \\ -s_2/2 & -s_1/2 & 0 & 0 & 0 & 0 & 0 & \frac{s_1+s_2}{2} \end{array} \right],$$

with some parameters s_1 and s_2.

Proposition 4.2 *Matrices $\hat{B}_{11,p}$, $p = 1, 2$, and B_0 have the same kernel, i.e.*

$$\text{Ker } \hat{B}_{11,p} = \text{Ker } B_0.$$

Proof: It can be easily checked that Ker $\hat{B}_{11,p}$ = Ker $B_0 = \{\mathbf{v} = (v_1, v_2, \ldots, v_8)^T \in \mathbb{R}^8 : v_i = v_1,\ i = 2, \ldots, 8\}, \quad p = 1, 2.$ □

Now we consider the eigenvalue problem

$$\hat{B}_{11,p}\mathbf{u} = \mu B_0 \mathbf{u}, \quad \mathbf{u} \in \mathbb{R}^8 \setminus \text{Ker } B_0, \qquad p = 1, 2, \tag{4.40}$$

with the following choices of s_1 and s_2.

Proposition 4.3 *For the case of $s_i = 2k_ik_3/(2k_i + k_3)$, $i = 1, 2$, the eigenvalues of problem* (4.40) *belong to the interval*

$$\left[\frac{3+2\kappa}{4+2\kappa}(1-\frac{1}{\sqrt{3}}),\ \frac{3+2\kappa}{4+2\kappa}(1+\frac{1}{\sqrt{3}})\right].$$

If we choose $s_i = \max\{k_i, k_3\}$, $i = 1, 2$, the eigenvalues of problem (4.40) *are within the interval*

$$\left[\frac{3+\kappa}{4+2\kappa}(1-\frac{1}{\sqrt{3}}),\ \frac{3+\kappa}{4+2\kappa}(1+\frac{1}{\sqrt{3}})\right].$$

Both cases have the same estimate of the condition number

$$Cond\ (B_0^{-1}\hat{B}_{11,p}) \leq 2 + \sqrt{3},$$

where the condition number is defined as the ratio of the biggest and the smallest nonzero eigenvalues of problem (4.40).

Proof: A direct calculation shows that $\mu \in [\mu^-, \mu^+]$ where

$$\mu^- = \min_{i=1,2}\left\{\frac{k_i}{4k_i+2k_3}\left(1+\frac{k_3}{k_i}+\frac{2k_3}{s_i}\right)\left(1-\sqrt{1-\frac{k_3^2/(k_is_i)+2k_3/s_i}{1+k_3/k_i+2k_3/s_i}}\right)\right\},$$

and

$$\mu^+ = \max_{i=1,2}\left\{\frac{k_i}{4k_i+2k_3}\left(1+\frac{k_3}{k_i}+\frac{2k_3}{s_i}\right)\left(1+\sqrt{1-\frac{k_3^2/(k_is_i)+2k_3/s_i}{1+k_3/k_i+2k_3/s_i}}\right)\right\}.$$

With $s_i = 2k_ik_3/(2k_i + k_3)$, $i = 1, 2$, and the definition of κ, it can be seen as in Theorem 4.1 that

$$\mu_- \geq \frac{3+2\kappa}{4+2\kappa}\left(1-\sqrt{1-\frac{2+3\kappa/2+\kappa^2/2}{3+2\kappa}}\right),$$

and

$$\mu_+ \geq \frac{3+2\kappa}{4+2\kappa}\left(1+\sqrt{1-\frac{2+3\kappa/2+\kappa^2/2}{3+2\kappa}}\right).$$

Note that

$$1-\frac{2+3\kappa/2+\kappa^2/2}{3+2\kappa} \leq \frac{1}{3},$$

so that the first case follows. The same argument applies to the second case. □

Now we define a new matrix on each prism:

$$\tilde{B}_p = \begin{bmatrix} B_0 + B_{12,p}B_{22,p}^{-1}B_{21,p} & B_{12,p} \\ B_{21,p} & B_{22,p} \end{bmatrix}, \qquad p = 1, 2. \tag{4.41}$$

As we noted in Remark 4.1 on page 46, when the cube C has nonempty intersection with $\partial\Omega$, matrices B_0, $B_{12,p}$, and $B_{21,p}$, $p = 1, 2$, do not have the rows and columns corresponding to the nodes on the boundary.

For each prism $P \in \mathcal{P}_h$ we now consider the eigenvalue problem:

$$B^P\mathbf{u} = \mu\tilde{B}^P\mathbf{u}, \tag{4.42}$$

where $B^P = B_p^P$ is defined in (4.39) and $\tilde{B}^P = \tilde{B}_p^P$ in (4.41), $p = 1, 2$. Below we consider only the simplest choice: $s_i = \max\{k_i, k_3\}$, $i = 1, 2$.

Proposition 4.4 *The eigenvalues of problem* (4.42) *belong to the interval*

$$\left[\frac{3+\kappa}{4+2\kappa}(1-\frac{1}{\sqrt{3}}), \frac{3+\kappa}{4+2\kappa}(1+\frac{1}{\sqrt{3}})\right].$$

Moreover, on each prism $P \in \mathcal{P}_h$ the eigenvalues of the problem

$$A^P \mathbf{u} = \mu \tilde{B}^P \mathbf{u}, \tag{4.43}$$

are within the interval $[\mu_-, \mu_+]$, where

$$\mu_\pm = (1+2\kappa)\left(1 \pm \sqrt{\frac{2\kappa}{1+2\kappa}}\right)\frac{3+\kappa}{4+2\kappa}\left(1 \pm \frac{1}{\sqrt{3}}\right).$$

Proof: The first statement follows directly from Proposition 4.3, and the second one then follows from Theorem 4.1. □

Now we define the symmetric positive-definite $N_1 \times N_1$ matrix $\tilde{B}_0$ by

$$(\tilde{B}_0 \mathbf{u}_1, \mathbf{v}_1) = \sum_{P \in \mathcal{P}_h} (B_0 \mathbf{u}_{1,P}, \mathbf{v}_{1,P}),$$

where $\mathbf{v}_1, \mathbf{u}_1 \in \mathrm{I\!R}^{N_1}$, and $\mathbf{u}_{1,P}$ and $\mathbf{v}_{1,P}$ are their respective restrictions on prism P. As in (4.38), we introduce the matrix

$$\tilde{B} = \begin{bmatrix} \tilde{B}_0 + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}. \tag{4.44}$$

Using Proposition 4.4 and the same proof as in Theorem 4.1, we have the following theorem.

Theorem 4.2 *Matrix $\tilde{B}$ defined in* (4.44) *is spectrally equivalent to matrix A, i.e.*

$$\mu_* \tilde{B} \leq A \leq \mu^* \tilde{B}.$$

Moreover,

$$Cond\,(\tilde{B}^{-1}A) \leq \overline{\mu} \equiv \mu^*/\mu_* \leq (3+8k)(2+\sqrt{3}). \tag{4.45}$$

Instead of matrix B in the form from (4.38) we take matrix $\tilde{B}$ from (4.44) as a preconditioner for matrix A. Because we have introduced a two-level subdivision of matrix $\tilde{B}_0$, matrix $\tilde{B}$ can be considered a three-level preconditioner.

As we noted earlier, matrix B_{22} is block-diagonal and can be inverted locally on prisms. So we concentrate on the linear system

$$\tilde{B}_0 \mathbf{u} = \mathbf{G}. \tag{4.46}$$

In terms of the group partitioning in Section 4.3.2.1, matrix $\tilde{B}_0$ has the block form:

$$\tilde{B}_0 = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \tag{4.47}$$

where block C_{22} corresponds to the nodes from the second group, which are on the faces of tetrahedra perpendicular to the coordinate axes. From the definition of B_0, it can be seen that matrix C_{22} is diagonal. In the above partitioning, we present $\mathbf{u}$ and $\mathbf{G}$ in (4.46) in the form:

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}.$$

Then, after elimination of the second group of unknowns:

$$\mathbf{u}_2 = C_{22}^{-1}(\mathbf{G}_2 - C_{21}\mathbf{u}_1),$$

we get the system of linear equations

$$(C_{11} - C_{12}C_{22}^{-1}C_{21})\mathbf{u}_1 = \mathbf{G}_1 - C_{12}C_{22}^{-1}\mathbf{G}_2 \equiv \tilde{\mathbf{G}}_1,$$

where vector $\mathbf{u}_1$ and block C_{11} correspond to the unknowns from the first group, which have only two unknowns per cube. The dimension of vectors $\mathbf{u}_1$ and $\mathbf{G}_1$ is equal to dim $(\mathbf{u}_1) = 2n^3$. The above simplification of (4.46) is carried out in detail in the next subsection.

Remark 4.5 We note that all the estimates in this section depend on parameter κ introduced in Assumption 4.1 (see page 47). Hence, it is very important to arrange the coordinate axes in such a way that parameter κ has the smallest value.

Remark 4.6 Note that the estimate of the condition number of the preconditioned matrix (4.45) is proportional to the value of parameter κ. In some sense we benefit from anisotropy. The smaller the coefficient k_3 of matrix K (the coefficient in the "z-direction") the better the preconditioner B.

4.3.2.3 Computational scheme

We now consider the computational scheme for (4.46). In terms of the unknowns introduced in Section 4.3.2.1:

$$\begin{aligned}
&uI_\ell^{(i,j,k)}, \quad GI_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad i,j,k = \overline{1,n}, \\
&ux_\ell^{(i,j,k)}, \quad Gx_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad i = \overline{2,n}, \qquad j,k = \overline{1,n}, \\
&uy_\ell^{(i,j,k)}, \quad Gy_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad j = \overline{2,n}, \qquad i,k = \overline{1,n}, \\
&uz_\ell^{(i,j,k)}, \quad Gz_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad k = \overline{2,n}, \qquad i,j = \overline{1,n},
\end{aligned}$$

system (4.46) with $K(x) \equiv \mathrm{diag}\,\{1,1,1\}$ can be written as

$$\begin{aligned}
\frac{1}{3}\begin{bmatrix} 20 & -2 \\ -2 & 20 \end{bmatrix} uI^{(i,j,k)} &- \Big((1-\delta_{i1})ux^{(i-1,j,k)} + (1-\delta_{in})ux^{(i,j,k)}\Big) \\
&- \Big((1-\delta_{j1})uy^{(i,j-1,k)} + (1-\delta_{jn})uy^{(i,j,k)}\Big) \\
-\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} &\Big((1-\delta_{k1})uz^{(i,j,k-1)} + (1-\delta_{kn})uz^{(i,j,k)}\Big) \\
&= \frac{2}{3h}GI^{(i,j,k)}, \qquad i,j,k = \overline{1,n},
\end{aligned} \tag{4.48}$$

and

$$\begin{aligned}
&2ux^{(i,j,k)} - uI^{(i+1,j,k)} - uI^{(i,j,k)} = \frac{2}{3h}Gx^{(i,j,k)}, \qquad i = \overline{1, n-1},\ j,k = \overline{1,n},\\
&2uy^{(i,j,k)} - uI^{(i,j+1,k)} - uI^{(i,j,k)} = \frac{2}{3h}Gy^{(i,j,k)}, \qquad j = \overline{1, n-1},\ i,k = \overline{1,n},\\
&2uz^{(i,j,k)} - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} uI^{(i,j,k+1)} - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} uI^{(i,j,k)} = \frac{2}{3h}\, Gz^{(i,j,k)},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k = \overline{1, n-1},\ i,j = \overline{1,n},
\end{aligned} \tag{4.49}$$

where δ_{ij} (the Kronecker symbol) is introduced to take into account the Dirichlet boundary conditions. Eliminating unknowns $ux_\ell^{(i,j,k)}$, $uy_\ell^{(i,j,k)}$, $uz_\ell^{(i,j,k)}$, $\ell = 1, 2$, from equations (4.48), we obtain the block "seven-point" scheme with 2×2-blocks for the unknowns $uI_\ell^{(i,j,k)}$, $\ell = 1, 2$, $i, j, k = \overline{1, n}$. From (4.49) we have

$$\begin{aligned}
ux^{(i,j,k)} &= \frac{1}{3h}Gx^{(i,j,k)} + \frac{1}{2}\left(uI^{(i+1,j,k)} + uI^{(i,j,k)}\right), \qquad i = \overline{1,n-1}, \quad j,k = \overline{1,n},\\
uy^{(i,j,k)} &= \frac{1}{3h}Gy^{(i,j,k)} + \frac{1}{2}\left(uI^{(i,j+1,k)} + uI^{(i,j,k)}\right), \qquad j = \overline{1,n-1}, \quad i,k = \overline{1,n},\\
uz^{(i,j,k)} &= \frac{1}{3h}Gz^{(i,j,k)} + \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left(uI^{(i,j,k+1)} + uI^{(i,j,k)}\right), \quad k = \overline{1,n-1}, \quad i,j = \overline{1,n}.
\end{aligned} \tag{4.50}$$

Substituting (4.50) into (4.48), we see that

$$\begin{aligned}
&\frac{1}{3}\begin{bmatrix} 20 & -2 \\ -2 & 20 \end{bmatrix} uI^{(i,j,k)}\\
&\qquad -\frac{1}{2}\left((1-\delta_{i1})\left(uI^{(i-1,j,k)} + uI^{(i,j,k)}\right) + (1-\delta_{in})\left(uI^{(i+1,j,k)} + uI^{(i,j,k)}\right)\right)\\
&\qquad -\frac{1}{2}\left((1-\delta_{j1})\left(uI^{(i,j-1,k)} + uI^{(i,j,k)}\right) + (1-\delta_{jn})\left(uI^{(i,j+1,k)} + uI^{(i,j,k)}\right)\right)\\
&-\frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left((1-\delta_{k1})\left(uI^{(i,j,k-1)} + uI^{(i,j,k)}\right) + (1-\delta_{kn})\left(uI^{(i,j,k+1)} + uI^{(i,j,k)}\right)\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = g^{(i,j,k)}, \quad i,j,k = \overline{1,n},
\end{aligned} \tag{4.51}$$

where

$$\begin{aligned}
g^{(i,j,k)} = \frac{2}{3h}\Big\{ & GI^{(i,j,k)} + \frac{1}{2}\left((1-\delta_{i1})Gx^{(i-1,j,k)} + (1-\delta_{in})Gx^{(i,j,k)}\right)\\
& + \frac{1}{2}\left((1-\delta_{j1})Gy^{(i,j-1,k)} + (1-\delta_{jn})Gy^{(i,j,k)}\right)\\
& + \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left((1-\delta_{k1})Gz^{(i,j,k-1)} + (1-\delta_{kn})Gz^{(i,j,k)}\right)\Big\}.
\end{aligned} \tag{4.52}$$

To solve system (4.51) we introduce the rotation matrix

$$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$

and new vectors $\mathbf{v}^{(i,j,k)} = (v_1^{(i,j,k)},\ v_2^{(i,j,k)})^T$, $i, j, k = \overline{1, n}$, given by

$$\mathbf{v}^{(i,j,k)} = Q \cdot uI^{(i,j,k)}, \qquad i,j,k = \overline{1,n}. \tag{4.53}$$

Then multiplying both sides of matrix equation (4.51) by matrix Q and using the relation

$$uI^{(i,j,k)} = Q^T \cdot \mathbf{v}^{(i,j,k)}, \qquad i,j,k = \overline{1,n}, \tag{4.54}$$

we obtain the following problem for the unknowns $\mathbf{v}^{(i,j,k)}$:

$$\begin{aligned}
&\begin{bmatrix} 6 & 0 \\ 0 & 22/3 \end{bmatrix} \mathbf{v}^{(i,j,k)} \\
&\qquad -\frac{1}{2}\left((1-\delta_{i1})\left(\mathbf{v}^{(i-1,j,k)} + \mathbf{v}^{(i,j,k)}\right) + (1-\delta_{in})\left(\mathbf{v}^{(i+1,j,k)} + \mathbf{v}^{(i,j,k)}\right)\right) \\
&\qquad -\frac{1}{2}\left((1-\delta_{j1})\left(\mathbf{v}^{(i,j-1,k)} + \mathbf{v}^{(i,j,k)}\right) + (1-\delta_{jn})\left(\mathbf{v}^{(i,j+1,k)} + \mathbf{v}^{(i,j,k)}\right)\right) \\
&-\frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\left((1-\delta_{k1})\left(\mathbf{v}^{(i,j,k-1)} + \mathbf{v}^{(i,j,k)}\right) + (1-\delta_{kn})\left(\mathbf{v}^{(i,j,k+1)} + \mathbf{v}^{(i,j,k)}\right)\right) \\
&\qquad\qquad = Q \cdot g^{(i,j,k)} \equiv \tilde{g}^{(i,j,k)}, \qquad i,j,k = \overline{1,n}.
\end{aligned} \tag{4.55}$$

It is easy to see that problem (4.55) can be decomposed into the following two independent problems:

$$\begin{aligned}
6v_1^{(i,j,k)} \; & -(1-\delta_{i1})\tfrac{1}{2}\left(v_1^{(i-1,j,k)} + v_1^{(i,j,k)}\right) - (1-\delta_{in})\tfrac{1}{2}\left(v_1^{(i+1,j,k)} + v_1^{(i,j,k)}\right) \\
& -(1-\delta_{j1})\tfrac{1}{2}\left(v_1^{(i,j-1,k)} + v_1^{(i,j,k)}\right) - (1-\delta_{jn})\tfrac{1}{2}\left(v_1^{(i,j+1,k)} + v_1^{(i,j,k)}\right) \\
& -(1-\delta_{k1})\tfrac{1}{2}\left(v_1^{(i,j,k-1)} + v_1^{(i,j,k)}\right) - (1-\delta_{kn})\tfrac{1}{2}\left(v_1^{(i,j,k+1)} + v_1^{(i,j,k)}\right) = \tilde{g}_1^{(i,j,k)}, \\
& \qquad\qquad i,j,k = \overline{1,n},
\end{aligned} \tag{4.56}$$

and

$$\begin{aligned}
\frac{22}{3}v_2^{(i,j,k)} \; & -(1-\delta_{i1})\tfrac{1}{2}\left(v_2^{(i-1,j,k)} + v_2^{(i,j,k)}\right) - (1-\delta_{in})\tfrac{1}{2}\left(v_2^{(i+1,j,k)} + v_2^{(i,j,k)}\right) \\
& -(1-\delta_{j1})\tfrac{1}{2}\left(v_2^{(i,j-1,k)} + v_2^{(i,j,k)}\right) - (1-\delta_{jn})\tfrac{1}{2}\left(v_2^{(i,j+1,k)} + v_2^{(i,j,k)}\right) = \tilde{g}_2^{(i,j,k)}, \\
& \qquad\qquad i,j = \overline{1,n}, \quad \forall\, k = \overline{1,n}.
\end{aligned} \tag{4.57}$$

Hence, we reduced linear system (4.55) of dimension $(2n^3)$ to one linear system of equations (4.56) of dimension n^3 and n linear systems of equations (4.57) of dimension n^2.

Again, for all these problems we can use either the method of separation of variables [106] or an algebraic multigrid method [8, 20, 70, 120]. An implementation cost of the first method is estimated by $O(h^{-3}\ln(h^{-1}))$. The AMG methods have the optimal order of arithmetic complexity $O(h^{-3})$. For completeness we describe below the method of separation of variables.

After we find the solution of problems (4.56) and (4.57) we easily retrieve vectors $uI^{(i,j,k)}$ by using relations (4.54).

4.3.2.4 A method of separation of variables

In this section we consider a method of separation of variables for solving problems (4.56) and (4.57). Problem (4.56) can be represented in the form:

$$C^{(3)}v_1 = \tilde{g}_1, \qquad v_1,\, \tilde{g}_1 \in \mathrm{I\!R}^{n^3}, \tag{4.58}$$

where

$$C^{(3)} = C_0 \otimes I_0 \otimes I_0 + I_0 \otimes C_0 \otimes I_0 + I_0 \otimes I_0 \otimes C_0,$$

I_0 is the $(n \times n)$-identity matrix, $\otimes$ denotes the tensor product of matrices, and C_0 has the form:

$$C_0 = \frac{1}{2} \begin{bmatrix} 3 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 3 \end{bmatrix}. \qquad (4.59)$$

If C_0 is factorized by

$$C_0 = Q_0 \Lambda_0 Q_0^T,$$

where Λ_0 is an $(n \times n)$-diagonal matrix and Q_0 is an $(n \times n)$-orthogonal matrix ($Q_0^{-1} = Q_0^T$), then matrix $C^{(3)}$ can be rewritten as follows:

$$C^{(3)} = Q^{(3)} \Lambda^{(3)} Q^{(3)},$$

where

$$Q^{(3)} = Q_0 \otimes Q_0 \otimes Q_0,$$

$$\Lambda^{(3)} = \Lambda_0 \otimes I_0 \otimes I_0 + I_0 \otimes \Lambda_0 \otimes I_0 + I_0 \otimes I_0 \otimes \Lambda_0.$$

Note that $Q^{(3)}$ is an $(n^3 \times n^3)$-orthogonal matrix and $\Lambda^{(3)}$ is an $(n^3 \times n^3)$-diagonal matrix. We can now use the following method to solve system (4.56):

$$\begin{aligned} &(1) \qquad \tilde{f}_1 = \left(Q^{(3)}\right)^T \tilde{g}_1, \\ &(2) \qquad \Lambda^{(3)} w = \tilde{f}_1, \\ &(3) \qquad v_1 = Q^{(3)} w. \end{aligned} \qquad (4.60)$$

The same argument can be exploited to solve (4.57). The problem can be rewritten as

$$C^{(2)} v_2 = \tilde{g}_2, \qquad v_2, \tilde{g}_2 \in \mathbb{R}^{n^2}, \qquad (4.61)$$

where

$$C^{(2)} = K_0 \otimes I_0 + I_0 \otimes K_0,$$

and the $(n \times n)$-matrix K_0 is given by

$$K_0 = \frac{1}{6} \begin{bmatrix} 19 & -3 & & & \\ -3 & 16 & -3 & & \\ & \ddots & \ddots & \ddots & \\ & & -3 & 16 & -3 \\ & & & -3 & 19 \end{bmatrix}.$$

Again, if we write K_0 as

$$K_0 = R_0 D_0 R_0^T,$$

where D_0 is an $(n\times n)$-diagonal matrix and R_0 is an $(n\times n)$-orthogonal matrix, we can rewrite matrix $C^{(2)}$ as follows:

$$C^{(2)} = Q^{(2)}\Lambda^{(2)}Q^{(2)^T},$$

where $Q^{(2)} = R_0 \otimes R_0$ and $\Lambda^{(2)} = D_0 \otimes I_0 + I_0 \otimes D_0$. Then system (4.57) can be solved with the following method:

$$\begin{array}{ll} (1) & \tilde{f}_2 = \left(Q^{(2)}\right)^T \tilde{g}_2, \\ (2) & \Lambda^{(2)}w = \tilde{f}_2, \\ (3) & v_2 = Q^{(2)}w. \end{array} \tag{4.62}$$

4.4 3D problem. Partition of cube into 5 tetrahedra

To explain our approach in this case we again consider the model problem when Ω is a unit cube in $\mathbb{R}^3$, Γ_0 is a union of some faces of Ω, the boundary conditions are homogeneous, and $K(\mathbf{x})$ satisfies the following assumption.

Assumption 4.2 *Assume that the coefficient matrix of equation* (4.1) *is a diagonal tensor* $K(\mathbf{x}) = diag\{k_1, k_2, k_3\}$*, where* k_i*,* $i = 1, 2, 3$*, are constants over the cube* Ω *such that* $\kappa = \min\{k_3/k_1, k_3/k_2\} \geq 1$.

Remark 4.7 In fact, we need only the assumption that coefficient k_* in some direction is not less then the coefficients in the other directions. For the sake of definiteness we assume that this is the "z-direction".

Note that the extension of the method to the case in which Ω is a union of parallelepipeds is straightforward.

Let $\mathcal{C}_h = \{C^{(i,j,k)}\}$ be a partition of Ω into uniform cubes with edge length $h = 1/n$; here (x_i, y_j, z_k) is the right back upper corner of cube $C^{(i,j,k)}$. Next, we divide each cube $C^{(i,j,k)}$ into 5 tetrahedra as shown in Figure 4.5. We denote this partitioning of Ω into tetrahedra by $\mathcal{T}_h$. Note that we have two types of partitioning of cubes $C^{(i,j,k)}$ into tetrahedra, the cube with one type of partitioning having all the adjacent cubes of another type.

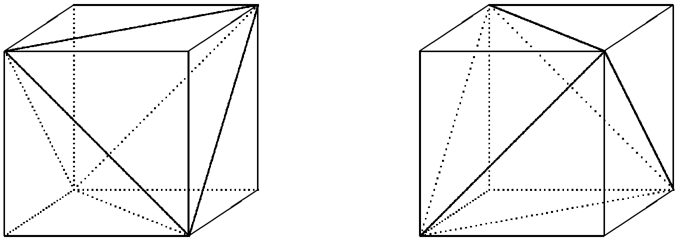


Figure 4.5: *Partition of cubes* $C^{(i,j,k)}$ *into 5 tetrahedra.*

We introduce the set of barycenters of all the faces of the tetrahedral partition of Ω and the set Q_h of those barycenters that do not belong to Γ_0. The Crouzeix-Raviart P_1–nonconforming finite element space V_h is defined by

$$V_h = \Big\{ v \in L^2(\Omega) : \quad v|_T \in P_1(T),\ \forall T \in \mathcal{T}_h;\ v \text{ is continuous at the barycenters from } Q_h \text{ and vanishes at the barycenters of faces on } \Gamma_0 \Big\}. \tag{4.63}$$

Let its dimension be N. Note that $N \approx 10n^3$. Remember that in the case of splitting each cube into **six** tetrahedra the number of degrees of freedom is $N \approx 12n^3$.

Below we use the same notation as in Sections 4.1 and 4.3. For each cube $C = C^{(i,j,k)} \in \mathcal{C}_h$, we denote by V_h^C the subspace of the restriction of the functions in V_h into C. For each $v_h \in V_h^C$, we indicate by $\mathbf{v}_c$ the corresponding vector. The dimension of V_h^C is denoted by N_c. Obviously, for a cube without faces on Γ_0 we have $N_c = 16$.

The local stiffness matrix A^C for a cube $C \in \mathcal{C}_h$ is given by

$$(A^C \mathbf{u}_c, \mathbf{v}_c)_{N_c} = \sum_{\tau \subset C} (K(\mathbf{x}) \nabla u_h, \nabla v_h)_\tau, \qquad \forall u_h, v_h \in V_h^C. \tag{4.64}$$

Note that matrices A^C are positive definite when $C \cap \Gamma_0 \neq 0$ and semidefinite otherwise. The global stiffness matrix is determined by assembling the local stiffness matrices:

$$(A\mathbf{u}, \mathbf{v})_N = \sum_{C \in \mathcal{C}_h} (A^C \mathbf{u}_c, \mathbf{v}_c)_{N_c}, \qquad \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^N. \tag{4.65}$$

4.4.1 Algebraic substructuring preconditioner

In this section we construct the algebraic substructuring preconditioner outlined in Introduction 4.1 of this chapter. Toward the end of the section, we divide all the unknowns in the system into two groups:

1. The first group consists of the unknowns corresponding to the faces of the tetrahedra that are internal for each cube (these are the unknowns on faces 1, 2, 3 and 4 in Figure 4.6). We denote these unknowns by $VI_l^{(i,j,k)}$, $l = 1, 2, 3, 4$, $i, j, k = \overline{1, n}$.

2. The second group consists of all the unknowns corresponding to the faces of the cubes in partition $\mathcal{C}_h$, without the faces on Γ_0 (Figure 4.6, faces $5, 6, \dots, 16$).

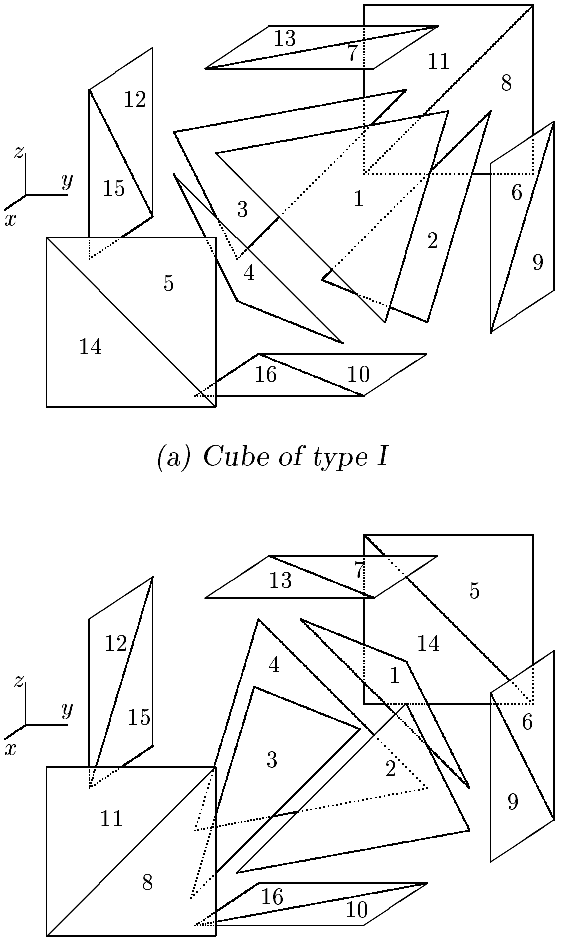
   (a) First, we number the unknowns on the faces perpendicular to the x-axis (faces 5, 8, 11, 14). We denote these unknowns by $Vx_l^{(i,j,k)}$, $l = 1, 2$, $i = \overline{1, n-1}$, $j, k = \overline{1, n}$.

   (b) Second, we number the unknowns on the faces perpendicular to the y-axis (faces 6, 9, 12, 15). We denote these unknowns by $Vy_l^{(i,j,k)}$, $l = 1, 2$, $j = \overline{1, n-1}$, $i, k = \overline{1, n}$.

   (c) Finally, we number the unknowns on the faces perpendicular to the z-axis (faces 7, 10, 13, 16). We denote these unknowns by $Vz_l^{(i,j,k)}$, $l = 1, 2$, $k = \overline{1, n-1}$, $i, j = \overline{1, n}$.

Now we consider a cube C that has no face on the boundary $\partial\Omega$ and number the faces s_j, $j = 1, \dots, 16$, of the tetrahedra in this cube in accordance with the partitioning introduced above as is shown in Figure 4.6. Then the local stiffness matrix of this cube has the following form:

$$A^C = \frac{3h}{2} \begin{bmatrix} A_{11,c} & A_{12,c} \\ A_{21,c} & A_{22,c} \end{bmatrix}, \tag{4.66}$$



(a) Cube of type I

(b) Cube of type II

Figure 4.6: *Local enumeration of faces in cubes.*

where

$$A_{11,c} = (k_1 + k_2 + k_3)\; I_c + \frac{1}{2}\left[k_1 T_1 + k_2 T_2 + k_3 T_3\right], \tag{4.67}$$

$$T_1 = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}, \quad T_3 = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix},$$

$$I_c = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \qquad A_{22,c} = \begin{bmatrix} D & & & \\ & D & & \\ & & D & \\ & & & D \end{bmatrix}, \qquad D = \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix},$$

$$A_{12,c} = \begin{bmatrix} -k_1 & -k_2 & -k_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -k_1 & -k_2 & -k_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -k_1 & -k_2 & -k_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -k_1 & -k_2 & -k_3 \end{bmatrix}.$$

Along with matrix A^C we introduce on each cube $C \in \mathcal{C}_h$ matrix B^C

$$B^C = A^C + \frac{3h}{2} \begin{bmatrix} \tilde{B}_{11,c} & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.68}$$

where $\tilde{B}_{11,c} = (k_1 + k_2)T_3$. Thus, matrix B^C can be represented in the form:

$$B^C = \frac{3h}{2} \begin{bmatrix} B_{11,c} & A_{12,c} \\ A_{21,c} & A_{22,c} \end{bmatrix},$$

where $B_{11,c} = (k_1 + k_2 + k_3)\; I_{11,c} + \frac{1}{2}\left[k_1(T_1 + T_3) + k_2(T_2 + T_3) + k_3 T_3\right]$.

Note that Ker $A^C =$ Ker B^C.

We now define the $N \times N$ matrix B by the following equality:

$$(B\mathbf{u}, \mathbf{v})_N = \sum_{C \in \mathcal{C}_h} (B^C \mathbf{u}_c, \mathbf{v}_c)_{N^C}, \qquad \forall \mathbf{u}, \mathbf{v} \in \mathrm{I\!R}^N. \tag{4.69}$$

From Lemma 4.1 we see that to estimate the condition number of $B^{-1}A$, it is sufficient to consider the local eigenvalue problems for $\mu_c \neq 0$

$$A^C \mathbf{u}_c = \mu_c B^C \mathbf{u}_c, \qquad \mathbf{u}_c \neq \mathbf{0}, \qquad \mathbf{u}_c \in \mathrm{I\!R}^{N_c}.$$

By direct calculations, from (4.66) and (4.68), we find that the eigenvalues μ_c belong to the interval $[1/3, 1]$ provided Assumption 4.2.

Then the inequalities (4.6) and (4.7) yield:

Proposition 4.5 *Suppose that the coefficient matrix of equation* (4.1) *is a diagonal tensor* $K(\mathbf{x}) = diag\{k_1, k_2, k_3\}$, *where* k_i, $i = 1, 2, 3$, *are constants over cube* Ω *such that* $\kappa = \min\{k_3/k_1, k_3/k_2\} \geq 1$.

Then eigenvalues of the problem

$$A\mathbf{u} = \mu B\mathbf{u} \tag{4.70}$$

belong to the interval $[\kappa/(2+\kappa), 1]$ *and thus the condition number is estimated by*

$$Cond\,(B^{-1}A) \leq 1 + 2/\kappa \leq 3.$$

We emphasize that the condition number of matrix $B^{-1}A$ is bounded by a constant independent of mesh-size h and the values of coefficients k_i, $i = 1, 2, 3$, when $k_3 \geq \max\{k_1, k_2\}$.

Splitting the space $\mathrm{I\!R}^N$ into two groups induces a vector presentation: $\mathbf{v}^T = (\mathbf{v}_1^T, \mathbf{v}_2^T)$, where $\mathbf{v}_1 \in \mathrm{I\!R}^{N_1}$ and $\mathbf{v}_2 \in \mathrm{I\!R}^{N_2}$; here $\mathbf{v}_2$ corresponds to the unknowns of the 2-nd group. Obviously, $N_1 = 4n^3$ and $N_2 = N - 4n^3$. Then matrices A and B can be represented in the following block form:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} B_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{4.71}$$

where $B_{11} : \mathrm{I\!R}^{N_1} \to \mathrm{I\!R}^{N_1}$.

Now denote by $\hat{B}_{11} = B_{11} - A_{12}A_{22}^{-1}A_{21}$ the Schur complement of B obtained by elimination of the vector $\mathbf{v}_2$. Then $B_{11} = \hat{B}_{11} + A_{12}A_{22}^{-1}A_{21}$ and hence matrix B have the form:

$$B = \begin{bmatrix} \hat{B}_{11} + A_{12}A_{22}^{-1}A_{21} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \tag{4.72}$$

Note that for each cube $C \in \mathcal{C}_h$ the unknowns of the 2-nd group (unknowns on the faces $5, 6, \dots, 16$, in the local numbering; see Figure 4.6) are only connected with the unknowns of the 1-st group, and therefore matrix A_{22} is diagonal. Thus, matrix $\hat{B}_{11}$ is easily computable. The important fact which can be established by direct computations is that matrix $\hat{B}_{11}$ can be obtained by assembling local matrices $\hat{B}_{11,c} = B_{11,c} - A_{12,c}A_{22,c}^{-1}A_{21,c}$:

$$(\hat{B}_{11}\mathbf{u}_1, \mathbf{v}_1) = \sum_{C \in \mathcal{C}_h} (\hat{B}_{11,c}\mathbf{u}_{1,c}, \mathbf{v}_{1,c}), \qquad \forall \mathbf{u}_1, \mathbf{v}_1 \in \mathrm{I\!R}^{N_1}$$

over all cubes. Here $\mathbf{u}_{1,c}$ is a restriction of $\mathbf{u}_1$ into the nodes of the first group of cube $C \in \mathcal{C}_h$ and $\dim \mathbf{u}_{1,c} = 4$.

Remark 4.8 The dimension of matrix $\hat{B}_{11}$ is approximately 2.5 times smaller than the order of matrix A.

Now we need to develop a preconditioner for matrix $\hat{B}_{11}$. Below we show that using algebraic substructuring we can construct a sparse separable matrix $\tilde{B}_{11}$ spectrally equivalent to $\hat{B}_{11}$ so that the resulting matrix

$$\tilde{B} = \begin{bmatrix} \tilde{B}_{11} + A_{12}A_{22}^{-1}A_{21} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{4.73}$$

is spectrally equivalent to initial matrix A. In this case we shall use the method of separation of variables in order to solve the system of linear equations with matrix $\tilde{B}_{11}$.

First, consider the linear system

$$B\mathbf{v} = \mathbf{g}. \tag{4.74}$$

Let us write explicitly the elements of $B\mathbf{v}$ for the Dirichlet boundary conditions on the whole boundary $\partial\Omega$ in terms of the unknowns introduced earlier in this section, i.e. in terms of

$$\begin{array}{llll} gi_\ell^{(i,j,k)}, \quad VI_\ell^{(i,j,k)}, & \ell = 1,2,3,4 & i,j,k = \overline{1,n}; & \\ gx_\ell^{(i,j,k)}, \quad Vx_\ell^{(i,j,k)}, & \ell = 1,2, & i = \overline{1,n-1}, & j,k = \overline{1,n}; \\ gy_\ell^{(i,j,k)}, \quad Vy_\ell^{(i,j,k)}, & \ell = 1,2, & j = \overline{1,n-1}, & i,k = \overline{1,n}; \\ gz_\ell^{(i,j,k)}, \quad Vz_\ell^{(i,j,k)}, & \ell = 1,2, & k = \overline{1,n-1}, & i,j = \overline{1,n}. \end{array} \tag{4.75}$$

Below we use the function

$$\delta_{ik} = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}$$

to take into account the Dirichlet boundary conditions and a vector

$$\mathbf{vr}^{(i,j,k)} = \begin{bmatrix} vr_1^{(i,j,k)} \\ vr_2^{(i,j,k)} \end{bmatrix} \in \mathrm{I\!R}^2$$

to denote variables (4.75).

Note that the technique of constructing a separable matrix $\tilde{B}_{11}$ in the case of $\Gamma_1 \neq \emptyset$ when Γ_0 is the union of entire faces of cube Ω, is the same.

The above equations are different for different types of cubes. For any cube of type I (see Fig. 4.6) we have

$$\left\{(k_1+k_2+k_3)I_c+\frac{1}{2}\left[k_1(T_1+T_3)+k_2(T_2+T_3)+k_3T_3\right]\right\}\mathbf{VI}^{(i,j,k)}- \tag{4.76}$$

$$-(1-\delta_{i1})k_1\begin{bmatrix}0&0\\1&0\\0&1\\0&0\end{bmatrix}\mathbf{Vx}^{(i-1,j,k)}-(1-\delta_{in})k_1\begin{bmatrix}1&0\\0&0\\0&0\\0&1\end{bmatrix}\mathbf{Vx}^{(i,j,k)}$$

$$-(1-\delta_{j1})k_2\begin{bmatrix}0&0\\0&0\\1&0\\0&1\end{bmatrix}\mathbf{Vy}^{(i,j-1,k)}-(1-\delta_{jn})k_2\begin{bmatrix}1&0\\0&1\\0&0\\0&0\end{bmatrix}\mathbf{Vy}^{(i,j,k)}$$

$$-(1-\delta_{k1})k_3\begin{bmatrix}0&0\\1&0\\0&0\\0&1\end{bmatrix}\mathbf{Vz}^{(i,j,k-1)}-(1-\delta_{kn})k_3\begin{bmatrix}1&0\\0&0\\0&1\\0&0\end{bmatrix}\mathbf{Vz}^{(i,j,k)}=\left(\tfrac{2}{3h}\right)\mathbf{gi}^{(i,j,k)},$$

$$\begin{aligned}
2k_1\mathbf{Vx}^{(i,j,k)}-k_1\begin{bmatrix}1&0&0&0\\0&0&0&1\end{bmatrix}\mathbf{VI}^{(i,j,k)}-k_1\begin{bmatrix}1&0&0&0\\0&0&0&1\end{bmatrix}\mathbf{VI}^{(i+1,j,k)}&=\left(\tfrac{2}{3h}\right)\mathbf{gx}^{(i,j,k)},\\
2k_2\mathbf{Vy}^{(i,j,k)}-k_2\begin{bmatrix}1&0&0&0\\0&1&0&0\end{bmatrix}\mathbf{VI}^{(i,j,k)}-k_2\begin{bmatrix}0&0&1&0\\0&0&0&1\end{bmatrix}\mathbf{VI}^{(i,j+1,k)}&=\left(\tfrac{2}{3h}\right)\mathbf{gy}^{(i,j,k)},\\
2k_3\mathbf{Vz}^{(i,j,k)}-k_3\begin{bmatrix}1&0&0&0\\0&0&1&0\end{bmatrix}\mathbf{VI}^{(i,j,k)}-k_3\begin{bmatrix}0&1&0&0\\0&0&0&1\end{bmatrix}\mathbf{VI}^{(i,j,k+1)}&=\left(\tfrac{2}{3h}\right)\mathbf{gz}^{(i,j,k)}.
\end{aligned} \tag{4.77}$$

For any cube of type II the entries of the unknowns $\mathbf{Vx}^{(i,j,k)}$ are different from the previous ones:

$$\left\{(k_1+k_2+k_3)I_c+\frac{1}{2}\left[k_1(T_1+T_3)+k_2(T_2+T_3)+k_3T_3\right]\right\}\mathbf{VI}^{(i,j,k)}- \tag{4.78}$$

$$-(1-\delta_{i1})k_1\begin{bmatrix}1&0\\0&0\\0&0\\0&1\end{bmatrix}\mathbf{Vx}^{(i-1,j,k)}-(1-\delta_{in})k_1\begin{bmatrix}0&0\\1&0\\0&1\\0&0\end{bmatrix}\mathbf{Vx}^{(i,j,k)}$$

$$-(1-\delta_{j1})k_2\begin{bmatrix}0&0\\0&0\\1&0\\0&1\end{bmatrix}\mathbf{Vy}^{(i,j-1,k)}-(1-\delta_{jn})k_2\begin{bmatrix}1&0\\0&1\\0&0\\0&0\end{bmatrix}\mathbf{Vy}^{(i,j,k)}$$

$$-(1-\delta_{k1})k_3\begin{bmatrix}0&0\\1&0\\0&0\\0&1\end{bmatrix}\mathbf{Vz}^{(i,j,k-1)}-(1-\delta_{kn})k_3\begin{bmatrix}1&0\\0&0\\0&1\\0&0\end{bmatrix}\mathbf{Vz}^{(i,j,k)}=\left(\tfrac{2}{3h}\right)\mathbf{gi}^{(i,j,k)},$$

$$
\begin{aligned}
&2k_1\mathbf{Vx}^{(i,j,k)} - k_1 \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{VI}^{(i,j,k)} - k_1 \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{VI}^{(i+1,j,k)} = \left(\tfrac{2}{3h}\right)\mathbf{gx}^{(i,j,k)}, \\
&2k_2\mathbf{Vy}^{(i,j,k)} - k_2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{VI}^{(i,j,k)} - k_2 \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{VI}^{(i,j+1,k)} = \left(\tfrac{2}{3h}\right)\mathbf{gy}^{(i,j,k)}, \\
&2k_3\mathbf{Vz}^{(i,j,k)} - k_3 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{VI}^{(i,j,k)} - k_3 \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{VI}^{(i,j,k+1)} = \left(\tfrac{2}{3h}\right)\mathbf{gz}^{(i,j,k)}.
\end{aligned}
\tag{4.79}
$$

Note that

$$
\tfrac{1}{2}(T_1+T_3) = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \qquad \tfrac{1}{2}(T_2+T_3) = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}.
$$

After eliminating the unknowns $\mathbf{Vx}^{(i,j,k)}$, $\mathbf{Vy}^{(i,j,k)}$, $\mathbf{Vz}^{(i,j,k)}$ from equations (4.76) and (4.78) we have a block "7-point" computational scheme with 4×4-blocks for the unknowns $\mathbf{VI}^{(i,j,k)}$:

$$
\left(\hat{B}_{11}\mathbf{VI}\right)^{(i,j,k)} \equiv \left\{(k_1+k_2+k_3)I_c + \frac{1}{2}\left[k_1(T_1+T_3) + k_2(T_2+T_3) + k_3T_3\right]\right\}\mathbf{VI}^{(i,j,k)} - \tag{4.80}
$$

$$
-(1-\delta_{i1})\frac{k_1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\left(\mathbf{VI}^{(i,j,k)} + \mathbf{VI}^{(i-1,j,k)}\right)
$$

$$
-(1-\delta_{in})\frac{k_1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\left(\mathbf{VI}^{(i,j,k)} + \mathbf{VI}^{(i+1,j,k)}\right)
$$

$$
-(1-\delta_{j1})\frac{k_2}{2}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\mathbf{VI}^{(i,j,k)} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}\mathbf{VI}^{(i,j-1,k)}\right)
$$

$$
-(1-\delta_{jn})\frac{k_2}{2}\left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\mathbf{VI}^{(i,j,k)} + \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\mathbf{VI}^{(i,j+1,k)}\right)
$$

$$
-(1-\delta_{k1})\frac{k_3}{2}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\mathbf{VI}^{(i,j,k)} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\mathbf{VI}^{(i,j,k-1)}\right)
$$

$$-(1-\delta_{kn})\frac{k_3}{2}\left(\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&0\end{bmatrix}\mathbf{VI}^{(i,j,k)}+\begin{bmatrix}0&1&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{bmatrix}\mathbf{VI}^{(i,j,k+1)}\right),$$

$$i,j,k=1,\dots,n.$$

Along with the Schur matrix $\hat{B}_{11}$ we define matrix $\tilde{B}_{11}$ in the form:

$$\begin{gathered}\left(\tilde{B}_{11}\mathbf{VI}\right)^{(i,j,k)}\equiv\left\{(k_1+k_2+\frac{1}{2}k_3)I_c+\frac{1}{2}\left[k_1(T_1+T_3)+k_2(T_2+T_3)+k_3T_3\right]\right\}\mathbf{VI}^{(i,j,k)}-\\ -(1-\delta_{i1})\frac{k_1}{2}\mathbf{VI}^{(i-1,j,k)}-(1-\delta_{in})\frac{k_1}{2}\mathbf{VI}^{(i+1,j,k)}\\ -(1-\delta_{j1})\frac{k_2}{2}\begin{bmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix}\mathbf{VI}^{(i,j-1,k)}-(1-\delta_{jn})\frac{k_2}{2}\begin{bmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix}\mathbf{VI}^{(i,j+1,k)}\\ -(1-\delta_{k1})\frac{k_3}{2}\begin{bmatrix}0&0&0&0\\1&0&0&0\\0&0&0&0\\0&0&1&0\end{bmatrix}\mathbf{VI}^{(i,j,k-1)}-(1-\delta_{kn})\frac{k_3}{2}\begin{bmatrix}0&1&0&0\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{bmatrix}\mathbf{VI}^{(i,j,k+1)},\end{gathered}\tag{4.81}$$

$$i,j,k=1,\dots,n.$$

Let us consider an eigenvalue problem

$$\hat{B}_{11}\mathbf{u}=\lambda\tilde{B}_{11}\mathbf{u},\qquad \mathbf{u}\in\mathbb{R}^{N_1}.\tag{4.82}$$

Proposition 4.6 *The eigenvalues of problem* (4.82) *belong to the interval* [1/6, 1].

Proof: Note first that matrices $\hat{B}_{11}$ and $\tilde{B}_{11}$ may be represented in the form:

$$\begin{aligned}\hat{B}_{11}&=k_1\hat{B}^{(1)}+k_2\hat{B}^{(2)}+k_3\hat{B}^{(3)},\\ \tilde{B}_{11}&=k_1\tilde{B}^{(1)}+k_2\tilde{B}^{(2)}+k_3\hat{B}^{(3)},\end{aligned}\tag{4.83}$$

where matrices $\hat{B}^{(i)}$, $i=1,2,3$, and $\tilde{B}^{(j)}$, $j=1,2$, do not depend on the coefficients of the problem, k_1,k_2,k_3.

Since all the components on the right-hand sides are nonnegative we can estimate eigenvalues λ of problem (4.82) by inequalities

$$\min_{i=1,2}\left\{\mu^{(i)}_{\min};1\right\}\le\lambda\le\max_{i=1,2}\left\{\mu^{(i)}_{\max};1\right\},\tag{4.84}$$

where $\mu^{(i)}_*$ are the extremal eigenvalues of the auxiliary problems

$$\hat{B}^{(i)}\mathbf{u}=\mu^{(i)}\tilde{B}^{(i)}\mathbf{u},\qquad i=1,2.$$

Direct calculations show that $\mu^{(i)}\in[1/6,1]$. Taking into account inequalities (4.84) we get the above proposition. □

Using Propositions 4.5 and 4.6, and Lemma 4.1 we have the following theorem.

Theorem 4.3 *Suppose that the coefficient matrix of equation* (4.1) *is a diagonal tensor* $K(\mathbf{x}) = diag\{k_1, k_2, k_3\}$, *where* k_i, $i = 1, 2, 3$, *are constants over cube* Ω *such that* $\kappa = \min\{k_3/k_1, k_3/k_2\} \geq 1$.

Then matrix $\tilde{B}$ *defined in* (4.73) *with block* $\tilde{B}_{11}$ *defined in* (4.81) *is spectrally equivalent to matrix A. Moreover,*

$$\mu_* \tilde{B} \leq A \leq \mu^* \tilde{B},$$

where $\mu_* = \kappa/6(2+\kappa)$ *and* $\mu^* = 1$, *hence*

$$Cond\ (\tilde{B}^{-1}A) \leq \overline{\mu} \equiv \mu^*/\mu_* \leq 6(1 + 2/\kappa) \leq 18. \tag{4.85}$$

Instead of matrix B in the form of (4.72) we take matrix $\tilde{B}$ from (4.73) with block $\tilde{B}_{11}$ in the form of (4.81) as a preconditioner for matrix A. As we noted above, matrix A_{22} is block-diagonal and can be inverted locally face-by-face.

Remark 4.9 Again, we note that the condition number depends neither on mesh-size h nor on the value of the coefficients when $k_3 \geq \max\{k_1, k_2\}$. Because the condition number of matrix $\tilde{B}^{-1}A$ depends on the value of parameter κ it is very important to choose the "z-direction" in the proper way. If, for example, we have the problem in which coefficient k_1 is greater than coefficients k_2 and k_3, we rearrange the variables so that the new variable z coincides with the old variable x. It means that we simply rename the axes of the coordinate system.

From representations (4.81), (4.83) it is easy to see that matrix $\tilde{B}_{11}$ is separable. It is also separable for $\Gamma_1 \neq 0$ when Γ_0 is a union of some faces of cube Ω. To solve the system of linear equations with matrix $\tilde{B}_{11}$ from (4.81) we use the method of separation of variables which is described in the next subsection.

4.4.2 Implementation of the method of separation of variables

Since matrix $\tilde{B}_{11}$ is separable, we can use the method of separation of variables to solve the problem

$$\tilde{B}_{11}\mathbf{w} = \mathbf{g}, \qquad \mathbf{w}, \mathbf{g} \in \mathrm{I\!R}^{N_1}. \tag{4.86}$$

Matrix $\tilde{B}_{11}$ can be represented in the form:

$$\tilde{B}_{11} = k_1 B_x + k_2 B_y + k_3 B_z, \tag{4.87}$$

$$B_x = I_z \otimes I_y \otimes (I_x \otimes D_1 + K_x \otimes I_0), \qquad B_y = I_z \otimes (I_y \otimes I_x \otimes D_2 + K_y \otimes I_x \otimes D_0),$$

$$B_z = I_z \otimes I_y \otimes I_x \otimes D_3 + K_{lz} \otimes I_y \otimes I_x \otimes D_{3l} + K_{uz} \otimes I_y \otimes I_x \otimes D_{3u},$$

where I_0, D_* are 4×4-matrices, I_x, I_y, I_z, K_* are $n \times n$-matrices,

$$D_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 2 & 0 & -1 & -1 \\ 0 & 2 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}, \quad D_0 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

$$D_3 = \frac{1}{2}\begin{bmatrix} 2 & -1 & 1 & -1 \\ -1 & 2 & -1 & 1 \\ 1 & -1 & 2 & -1 \\ -1 & 1 & -1 & 2 \end{bmatrix}, \quad D_{3l} = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad D_{3u} = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$K_x = K_y = \frac{1}{2}\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}, \qquad K_{lz} = K_{uz}^T = \begin{bmatrix} 0 & & & \\ -1 & 0 & & \\ & \ddots & \ddots & \\ & & -1 & 0 \end{bmatrix}.$$

We represent matrices $K_x, K_y, D_1, D_2, D_0, D_3$ in the form:

$$\begin{aligned} K_\alpha &= Q_\alpha \Lambda_\alpha Q_\alpha^T, \qquad && \alpha = x, y \\ K_\beta &= Q_0 \Lambda_\beta Q_0^T, \qquad && \beta = 0, 1, 2, 3, \end{aligned} \tag{4.88}$$

where

$$Q_x = Q_y = \{q_{ij}\}_{i,j=1}^n, \qquad q_{ij} = \sqrt{\frac{2}{n+1}}\,\sin\left(\frac{\pi}{n+1}\cdot i\cdot j\right),$$

$$Q_0 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

and Λ_x, Λ_y are $(n \times n)$-diagonal matrices, and $\Lambda_0, \Lambda_1, \Lambda_2, \Lambda_3$ are 4×4-diagonal matrices.

Define a matrix Q as

$$Q = I_z \otimes Q_y \otimes Q_x \otimes Q_0. \tag{4.89}$$

Note that Q is an $(4n^3 \times 4n^3)$-orthogonal matrix.

Then matrix $\tilde{B}_{11}$ can be represented in the form:

$$\tilde{B}_{11} = Q \Lambda Q^T, \tag{4.90}$$

where $\Lambda = Q^T \tilde{B}_{11} Q =$

$$\begin{aligned} &k_1 I_z \otimes I_y \otimes (I_x \otimes \Lambda_1 + \Lambda_x \otimes I_0) + k_2 I_z \otimes (I_y \otimes I_x \otimes \Lambda_2 + \Lambda_y \otimes I_x \otimes \Lambda_0) + \\ &k_3 \Big(I_z \otimes I_y \otimes I_x \otimes \Lambda_3 + K_{lz} \otimes I_y \otimes I_x \otimes (Q_0^T D_{3l} Q_0) + K_{uz} \otimes I_y \otimes I_x \otimes (Q_0^T D_{3u} Q_0)\Big), \end{aligned} \tag{4.91}$$

$$Q_0^T D_{3l} Q_0 = \frac{1}{4}\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix}, \qquad Q_0^T D_{3u} Q_0 = \frac{1}{4}\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

Now we can use the following method to solve system (4.86):

$$\begin{aligned} &(1) \qquad \tilde{\mathbf{g}} = Q^T \mathbf{g}, \\ &(2) \quad \Lambda \tilde{\mathbf{w}} = \tilde{\mathbf{g}}, \\ &(3) \qquad \mathbf{v} = Q \tilde{\mathbf{w}}. \end{aligned} \tag{4.92}$$

We note that due to the form (4.91) of matrix Λ, the solution procedure of stage (2) is equivalent to solving $2n^2$ independent tridiagonal linear systems of the order $2n \times 2n$.

Fast Fourier transform implementation of (4.92) will yield a number of arithmetic operations proportional to $N_1 \ln(N_1)$ or $N \ln(0.4N)$, where the constants of proportionality do not depend on the number of unknowns N and on coefficients k_1, k_2, and k_3.

4.5 Fictitious components method for model problem

Now we consider an elliptic boundary value problem in a domain Ω of general geometric shape (see, e.g., Figure 4.7). Suppose that $\bar{\Omega}$ can be embedded in a larger domain Π ($\bar{\Omega} \subset \Pi$) which has relatively simple form (e.g., a rectangle) so we can effectively solve corresponding grid systems for problems in Π. Π is called a fictitious domain (see Figure 4.7). It is attractive to replace the solution of the original grid systems for Ω by suitable problems in Π. The introduction of such a fictitious domain Π for the approximate solution of elliptic boundary value problems associated with Ω has been used in the method of fictitious domains and its most effective matrix modification, the fictitious components method. As an iterative process for solving systems of mesh equations the latter method was proposed and studied, for example, in [6, 82, 85, 84, 86].

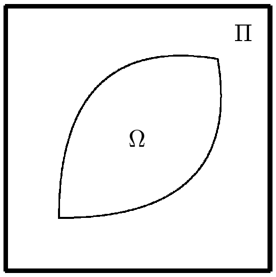


Figure 4.7: *Real domain* Ω *embedded in fictitious domain* Π.

In the fictitious components method, instead of problem (4.5) with $N \times N$ symmetric matrix A, an extended problem is considered:

$$\tilde{A}\tilde{\mathbf{u}} = \tilde{\mathbf{f}}, \tag{4.93}$$

where a square matrix $\tilde{A}$ is of order $M > N$. We assume that there exists a permutation matrix $\tilde{P}$ such that

$$\tilde{P}\tilde{A}\tilde{P}^T = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \tilde{P}\tilde{\mathbf{f}} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}. \tag{4.94}$$

It is obvious that for any solution $\tilde{\mathbf{u}}$ of problem (4.93) solution $\mathbf{u}$ of problem (4.5) can be found by the formula $\mathbf{u} = Q\tilde{P}\tilde{\mathbf{u}}$, where $Q = [I_N\mathbf{0}]$ is an $N \times M$ projection matrix.

The main ingredient of the method is the construction of a preconditioning $M \times M$ matrix B for extended system (4.93).

In this section we propose a variant of the fictitious components method for nonconforming approximations of anisotropic elliptic problems. The method is described in Section 4.5.1. Although this method can be formulated for problems in general domains, here we discuss only a model problem in a unit square. Some generalizations are suggested in the remarks at the end of this section. The proof of optimality of the method considered is based on the theory of the extension of mesh functions from the original domain Ω into the fictitious domain Π [85, 86, 96]. A variant of the extension theorem for nonconforming finite element spaces is given in Section 4.5.2.

4.5.1 Formulation of the method

Consider the model problem in $\Omega = [0,1]^2$:

$$\begin{aligned} -\mathrm{div}\,(K\nabla u) + c\,u &= f, && \text{in } \Omega, \\ (K\nabla u, \mathbf{n}) &= 0, && \text{on } \partial\Omega, \end{aligned} \tag{4.95}$$

where $c > 0$ is a constant and K is a full symmetric matrix in Ω. Let $(k_1, \mathbf{u}_1)$ and $(k_2, \mathbf{u}_2)$ be the eigenpairs of K with $\mathbf{u}_1 = (\alpha, \beta)$, $\mathbf{u}_2 = (-\beta, \alpha)$, $\alpha^2 + \beta^2 = 1$. Then consider a transformation of the coordinates $(\xi, \nu) = F(x,y)$: $\xi = \alpha \cdot x + \beta \cdot y$, $\nu = -\beta \cdot x + \alpha \cdot y$. In coordinates (ξ, ν) problem (4.95) has the diagonal matrix coefficient $\tilde{K} = \mathrm{diag}\,\{k_1, k_2\}$ and is represented in the form:

$$\begin{aligned} -k_1 u_{\xi\xi} - k_2 u_{\nu\nu} + c\,u &= f && \text{in } \tilde{\Omega} \equiv F(\Omega), \\ \frac{\partial u}{\partial n} &= 0 && \text{on } \Gamma \equiv \partial\tilde{\Omega}. \end{aligned} \tag{4.96}$$

Now construct a closed rectangle Π in the (ξ, ν)-plane which contains $\tilde{\Omega}$ in such a way that diam $(\Pi) \approx$ diam $(\tilde{\Omega})$. First, we define a uniform triangular mesh in Π, and then, locally modify it to fit the boundaries of $\tilde{\Omega}$. Denote this mesh by $\mathcal{T}_{h,\Pi}$ and its trace in the domain $\tilde{\Omega}$ by $\mathcal{T}_{h,\tilde{\Omega}}$. The triangulation $\mathcal{T}_h$ of Ω is defined by the inverse transformation $(x,y) = F^{-1}(\xi, \nu)$ of mesh $\mathcal{T}_{h,\tilde{\Omega}}$.

Since problems (4.95) in Ω and (4.96) in $\tilde{\Omega}$ are equivalent, below we consider only problem (4.96).

We use the nonconforming finite element space $V_h(\tilde{\Omega})$ introduced in Section 4.1. Define the bilinear form on $V_h(\tilde{\Omega})$ by

$$a^h_{\tilde{\Omega}}(u,v) = \sum_{\tau \in \mathcal{T}_{h,\tilde{\Omega}}} \int_\tau (k_1 u_\xi v_\xi + k_2 u_\nu v_\nu + c\,uv)\; d\xi\; d\nu, \qquad \forall\; u,v \in V_h(\tilde{\Omega}). \tag{4.97}$$

Then the P_1–nonconforming finite element discretization of (4.96) has the form: *find* $u_h \in V_h(\tilde{\Omega})$ *such that*

$$a^h_{\tilde{\Omega}}(u_h, v) = (f, v), \qquad \forall v \in V_h(\tilde{\Omega}). \tag{4.98}$$

Once a nodal basis $\{\varphi_i(\mathbf{x})\}_{i=1}^N$ for $V_h(\tilde{\Omega})$ has been chosen, equation (4.98) yields a system of linear algebraic equations (see Section 4.1):

$$A\mathbf{u} = \mathbf{f}, \tag{4.99}$$

with $N \times N$ symmetric positive definite matrix A.

Along with problem (4.96) we consider the same problem in rectangle Π with homogeneous Neumann boundary conditions on $\partial\Pi$. First, we define the bilinear form on $V_h(\Pi)$ by

$$a^h_{\Pi}(u,v) = \sum_{\tau \in \mathcal{T}_{h,\Pi}} \int_\tau (k_1 u_\xi v_\xi + k_2 u_\nu v_\nu + cuv)\; d\xi\; d\nu, \qquad \forall\; u,v \in V_h(\Pi). \tag{4.100}$$

Then the symmetric positive definite $M \times M$ matrix B is defined as follows:

$$(B\mathbf{u}, \mathbf{v}) = a^h_\Pi(u,v), \qquad u,v \in V_h(\Pi), \tag{4.101}$$

where M is the dimension of $V_h(\Pi)$ and $\mathbf{u},\ \mathbf{v} \in \mathrm{I\!R}^M$ are vector representations of functions u, v corresponding to the nodal basis $\{\varphi_i(\mathbf{x})\}_{i=1}^M$ of $V_h(\Pi)$.

We partition all degrees of freedom in Π into three groups:

1. The first group consists of the unknowns corresponding to the degrees of freedom in $\tilde{\Omega} \setminus \Gamma$.

2. The second group consists of the unknowns on the boundary Γ of the domain $\tilde{\Omega}$.

3. Finally, we enumerate the unknowns corresponding to the degrees of freedom in $\Pi \setminus \tilde{\Omega}$.

Then matrices A, $\tilde{A}$, and B can be represented in block form:

$$A = \begin{bmatrix} A_1 & A_{1\Gamma} \\ A_{\Gamma 1} & A_\Gamma \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} A_1 & A_{1\Gamma} & \mathbf{0} \\ A_{\Gamma 1} & A_\Gamma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad B = \begin{bmatrix} A_1 & A_{1\Gamma} & \mathbf{0} \\ A_{\Gamma 1} & B_\Gamma & B_{\Gamma 2} \\ \mathbf{0} & B_{2\Gamma} & B_2 \end{bmatrix}, \tag{4.102}$$

where blocks A_1, A_Γ, and B_2 correspond to the unknowns of the first, second, and third groups, respectively.

We note that matrix B_Γ can be represented as a sum $B_\Gamma = B_\Gamma^{(1)} + B_\Gamma^{(2)}$, where $B_\Gamma^{(1)} = A_\Gamma$, and the matrix

$$\begin{bmatrix} B_\Gamma^{(2)} & B_{\Gamma 2} \\ B_{2\Gamma} & B_2 \end{bmatrix} \tag{4.103}$$

corresponds to the nonconformal discretization of equation (4.96) in the domain $\Pi \setminus \tilde{\Omega}$ with the homogeneous Neumann boundary conditions.

Since $(\tilde{A}\mathbf{u}, \mathbf{u}) \leq (B\mathbf{u}, \mathbf{u})$ for any $\mathbf{u} \in \mathrm{I\!R}^M$, an eigenvalue problem

$$\tilde{A}\mathbf{u} = \lambda B\mathbf{u}, \qquad \mathbf{u} \in \operatorname{Im} B, \tag{4.104}$$

has $\lambda_{\max} \leq 1$. To estimate the minimal eigenvalue $\lambda_{\min}$ of problem (4.104) we need the following assumption.

Assumption 4.3 *For any function $u \in V_h(\tilde{\Omega})$ there exists a function $\tilde{u} \in V_h(\Pi)$ such that $\tilde{u}(\mathbf{x}) \equiv u(\mathbf{x})$ for any $\mathbf{x} \in \tilde{\Omega}$ and*

$$a_\Pi^h(\tilde{u}, \tilde{u}) \leq C_0 \ \cdot a_{\tilde{\Omega}}^h(u, u), \tag{4.105}$$

where a positive constant $C_0 > 1$ is not dependent on mesh-size parameter h.

This assumption is very important and is connected with the theory of the extension of mesh functions. Proof of the proposition given below is completely dependent on the statement of the assumption. For the case of conforming finite element spaces the questions of the extension of mesh functions is considered in [87, 86, 96, 93, 119]. For the case of nonconforming finite element spaces we provide the foundation of this assumption and respective theory in the next subsection 4.5.2.

Using this assumption we have the following result.

Proposition 4.7 *The minimal eigenvalue of problem* (4.104) *satisfies the inequality $\lambda_{\min} \geq 1/C_0$ and, hence,*

$$\nu = \lambda_{\max}/\lambda_{\min} \leq C_0. \tag{4.106}$$

Proof: Consider an equality

$$\begin{aligned}\lambda_{\min} &= \min_{\mathbf{v}\in \mathrm{Im}\, B} \frac{(\tilde{A}\mathbf{v},\mathbf{v})}{(B\mathbf{v},\mathbf{v})} = \frac{(\tilde{A}\mathbf{u},\mathbf{u})}{(B\mathbf{u},\mathbf{u})} \\ &= \frac{a^h_{\tilde{\Omega}}(u,u)}{a^h_{\Pi}(u,u)} = \frac{\sum\limits_{\tau\in\mathcal{T}_{h,\tilde{\Omega}}} \int\limits_{\tau} \left(k_1 u_\xi^2 + k_2 u_\nu^2 + cu^2\right) d\mathbf{x}}{\sum\limits_{\tau\in\mathcal{T}_{h,\Pi}} \int\limits_{\tau} \left(k_1 u_\xi^2 + k_2 u_\nu^2 + cu^2\right) d\mathbf{x}},\end{aligned} \tag{4.107}$$

where $\mathbf{u} = (\mathbf{u}_I^T, \mathbf{u}_\Gamma^T, \mathbf{u}_2^T)^T \in \mathrm{Im}\, B$ is the eigenvector of problem (4.104) corresponding to $\lambda_{\min}$ such that $B_{2\Gamma}\mathbf{u}_\Gamma + B_2\mathbf{u}_2 = \mathbf{0}$, and $u \in V_h(\Pi)$ is its corresponding finite element function.

Note that for any finite element function $v \in V_h(\Pi)$ such that $v(\mathbf{x}) = u(\mathbf{x})$, $\mathbf{x} \in \tilde{\Omega}$, we have an equality

$$a^h_{\Pi\setminus\tilde{\Omega}}(v,v) = a^h_{\Pi\setminus\tilde{\Omega}}(v-u, v-u) + a^h_{\Pi\setminus\tilde{\Omega}}(u,u), \tag{4.108}$$

where $a^h_{\Pi\setminus\tilde{\Omega}}(u,v) = a^h_{\Pi}(u,v) - a^h_{\tilde{\Omega}}(u,v)$ for any $u, v \in V_h(\Pi)$. Choosing as a function v the extension $\tilde{u}$ of the function u (which exists according to Assumption 4.3) we get

$$a^h_{\Pi\setminus\tilde{\Omega}}(u,u) \le a^h_{\Pi\setminus\tilde{\Omega}}(\tilde{u},\tilde{u}). \tag{4.109}$$

Using (4.109) and (4.105) we get

$$\begin{aligned} a^h_{\Pi}(u,u) &= \sum_{\tau\in\mathcal{T}_{h,\Pi}} \int\limits_{\tau} \left(k_1 u_\xi^2 + k_2 u_\nu^2 + cu^2\right) d\mathbf{x} = a^h_{\tilde{\Omega}}(u,u) + a^h_{\Pi\setminus\tilde{\Omega}}(u,u) \\ &\le a^h_{\tilde{\Omega}}(u,u) + a^h_{\Pi\setminus\tilde{\Omega}}(\tilde{u},\tilde{u}) = a^h_{\Pi}(\tilde{u},\tilde{u}) \le C_0\, a^h_{\tilde{\Omega}}(u,u). \end{aligned} \tag{4.110}$$

From (4.110) and (4.107) it follows that $\lambda_{\min} \ge 1/C_0$, which completes the proof. □

From Proposition 4.7 and estimate (3.30) it follows that the rate of convergence of the conjugate gradient method in subspace Im B does not depend on mesh-size parameter h. At each step of the conjugate gradient method we need to solve a linear problem with matrix B. For a two-dimensional model problem we can use the method described in Section 4.2. The AMG implementation of the described method has an optimal order of arithmetic complexity $O(h^{-2})$.

Remark 4.10 The fictitious components method can be developed also when the homogeneous Dirichlet boundary condition is posed on some part of the boundary $\partial\Omega$. However, this case requires more careful consideration of the extension theorem for nonconforming approximations and is not considered here.

Remark 4.11 The described fictitious domain method can also be used for three-dimensional model problems provided that Assumption 4.3 holds. To solve the problem with matrix B we can use modifications of the methods described in Sections 4.3 and 4.4. Using the AMG as an internal solver (see Section 4.3), the method has an optimal order of arithmetic complexity $O(h^{-3})$.

Remark 4.12 The analysis provided in this section can be applied also to the case of $c = 0$ in (4.95). To do this we can use, for example, the technique of [95] and analog of Assumption 4.3 for seminorms (see Remark 4.13).

4.5.2 Extension theorem for anisotropic elliptic problems

To justify Assumption 4.3 we consider a model problem in $\mathbb{R}^2$. Let $\Pi = [0,1]^2$ be the unit square and $\Omega \subset \Pi$ be a convex Lipshitz domain in $\mathbb{R}^2$ such that diam $(\Pi) \approx$ diam (Ω) (see Fig. 4.8a).

Consider a problem

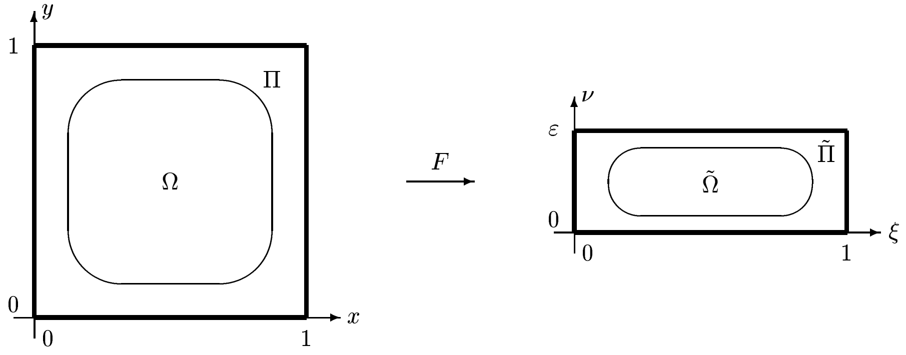
$$\begin{array}{rcll} L\,u \equiv -u_{xx} - ku_{yy} + u &=& f, & \text{in } \Omega, \\ \dfrac{\partial u}{\partial n} &=& 0, & \text{on } \partial\Omega, \end{array} \tag{4.111}$$

where $k \geq 1$. Obviously, after appropriate scaling problem (4.96) can be described by (4.111).

First, we make a transformation of the coordinates in (4.111) by $(\xi, \nu) = F(x,y) \equiv (x, \varepsilon y)$, where $\varepsilon = 1/\sqrt{k}$. Then (4.111) becomes

$$\begin{array}{rcll} -u_{\xi\xi} - u_{\nu\nu} + u &=& \tilde{f}, & \text{in } \tilde{\Omega} \equiv F(\Omega), \\ \dfrac{\partial u}{\partial n} &=& 0, & \text{on } \Gamma \equiv \partial\tilde{\Omega}, \end{array} \tag{4.112}$$

and $\tilde{\Pi} \equiv F(\Pi) = [0,1] \times [0,\varepsilon]$ (see Fig. 4.8b).



(a) *Real domain* Ω.

(b) *Transformed domain* $\tilde{\Omega}$.

Figure 4.8: *Transformation of the real and fictitious domains.*

Next, we define the triangulations $\mathcal{T}_{h,\tilde{\Pi}}$ of $\tilde{\Pi}$ and $\mathcal{T}_{h,\tilde{\Omega}}$ of $\tilde{\Omega}$ as is described in the previous section. On these triangulations we define the nonconforming finite element spaces $V_h(\tilde{\Pi})$ and $V_h(\tilde{\Omega})$ (see Section 4.1) and their norms:

$$\begin{array}{rcll} \|u^h\|^2_{V_h(\tilde{\Omega})} &=& \sum\limits_{\tau \in \mathcal{T}_{h,\tilde{\Omega}}} \|u^h\|^2_{V_h(\tau)}, & \forall u^h \in V_h(\tilde{\Omega}), \\ \|u^h\|^2_{V_h(\tilde{\Pi})} &=& \sum\limits_{\tau \in \mathcal{T}_{h,\tilde{\Pi}}} \|u^h\|^2_{V_h(\tau)}, & \forall u^h \in V_h(\tilde{\Pi}). \end{array} \tag{4.113}$$

Here $\|\cdot\|_{V_h(\tau)}$ means the usual norm in $H^1(\tau)$:

$$\|u^h\|^2_{V_h(\tau)} = \int\limits_{\tau} \left((\nabla u)^2 + u^2\right) d\mathbf{x}.$$

The main result of this section is the following proposition:

Proposition 4.8 (Extension Theorem) *Let $\tilde{\Omega} \subset \tilde{\Pi}$ be a convex Lipshitz domain. Then for any function $u^h \in V_h(\tilde{\Omega})$ there exists a function $\tilde{u}^h \in V_h(\tilde{\Pi})$ such that $\tilde{u}^h(\mathbf{x}) \equiv u^h(\mathbf{x})$ for any $\mathbf{x} \in \tilde{\Omega}$ and*

$$\|\tilde{u}^h\|_{V_h(\tilde{\Pi})} \le C_0 \, \|u^h\|_{V_h(\tilde{\Omega})}, \tag{4.114}$$

where positive constant $C_0 > 1$ does not depend on mesh-size parameter h and the value of ε.

The results analogous to those of Proposition 4.8 for conforming finite element subspaces of Sobolev spaces are well known (see, e.g., [87, 85, 86, 96]). Since we can not directly apply these results for nonconforming spaces, we need a special construction which is based on an isomorphism between the nonconforming and conforming finite element spaces [37, 38, 107].

Namely, we use the following scheme to prove the proposition:

(A) First, we define an equivalence map I_h between the nonconforming space $V_h(\Omega)$ and some Galerkin space of continuous piecewise linear functions $H^1_{h/2}(\Omega)$.

(B) Then, for a given function $u \in V_h(\Omega)$ we apply the theory of extensions of mesh functions in $H^1_{h/2}(\Omega)$ for the function $I_h u$ to get the extension $v^h \in H^1_{h/2}(\Pi)$.

(C) Finally, using the same equivalence between finite element spaces we define a nonconforming function $\tilde{u} \in V_h(\Pi) = P_h v^h$ with an operator P_h, which is conjugate to the operator I_h. The function $\tilde{u}$ defined by this algorithm satisfies the statement of Proposition 4.8.

The rest of the section is divided onto three parts. First, we define an isomorphism between conforming and nonconforming finite element spaces introduced and used in [37, 38, 107]. Then, we provide some necessary facts from the theory of extensions of functions from finite element subspaces of Sobolev spaces (see, e.g., [87, 85, 86, 96]). Finally, we combine these facts to prove Proposition 4.8.

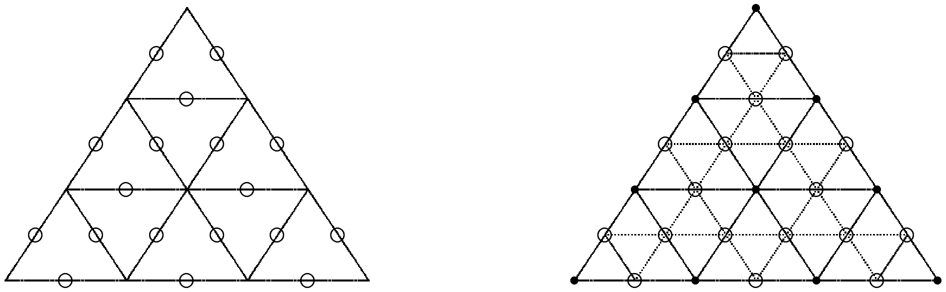
4.5.2.1 Isomorphism between conforming and nonconforming spaces

Let $H^1_{h/2}(\Omega)$ be the conforming space of piecewise linear functions on the triangulation $\mathcal{T}_{h/2,\Omega}$, where the $h/2$-mesh is obtained by joining midpoints of the edges of elements of $\mathcal{T}_{h,\Omega}$.

A vertex of $\mathcal{T}_{h/2,\Omega}$ is called *primary* if it is a nodal point corresponding to a degree of freedom of nonconforming space $V_h(\Omega)$; otherwise the vertex is called *secondary.* We say that two vertices of triangulation $\mathcal{T}_{h/2,\Omega}$ are *adjacent* if there exists an edge of $\mathcal{T}_{h/2,\Omega}$ connecting the vertices. An example of the triangulations $\mathcal{T}_{h,\Omega}$ and $\mathcal{T}_{h/2,\Omega}$ with corresponding degrees of freedom in the two-dimensional case is shown in Figure 4.9.

We define the equivalence map $I_h : V_h(\Omega) \to H^1_{h/2}(\Omega)$ for any function $u \in V_h(\Omega)$ as follows:

$$(I_h u)(\mathbf{x}) = \begin{cases} u(\mathbf{x}), \quad \text{if } \mathbf{x} \text{ is a primary vertex in } \Omega; \\[1ex] \text{The average of all adjacent primary vertices on the} \\ \text{boundary of } \Omega, \text{ if } \mathbf{x} \text{ is a secondary vertex on } \partial\Omega; \\[1ex] \text{The average of all adjacent primary vertices, if } \mathbf{x} \text{ is a} \\ \text{secondary vertex in } \Omega; \\[1ex] \text{The continuous piecewise linear interpolant of the} \\ \text{above vertex values if } \mathbf{x} \text{ is not a vertex of } \mathcal{T}_{h/2,\Omega}. \end{cases} \tag{4.115}$$



(a) Fragment of triangulation $\mathcal{T}_{h,\Omega}$ and degrees of freedom of $V_h(\Omega)$. "∘" denote primary vertices.

(b) Fragment of triangulation $\mathcal{T}_{h/2,\Omega}$ and degrees of freedom of $H^1_{h/2}(\Omega)$. "•" denote secondary vertices.

Figure 4.9: *Fragments of the meshes $\mathcal{T}_{h,\Omega}$ and $\mathcal{T}_{h/2,\Omega}$.*

It can be shown [37, 107] that there exist constants $\check{c}$ and $\hat{c}$ independent of h such that for any function $u \in V_h(\Omega)$ the following inequalities hold true:

$$\check{c}\ \|u\|_{V_h(\Omega)} \leq \|I_h u\|_{H^1_{h/2}(\Omega)} \leq \hat{c}\ \|u\|_{V_h(\Omega)}. \tag{4.116}$$

Also we introduce a projection operator $P_h : H^1_{h/2}(\Omega) \to V_h(\Omega)$ for any function $v^h \in H^1_{h/2}(\Omega)$ as follows:

$$(I_h u, v^h)_{H^1_{h/2}(\Omega)} = (u, P_h v^h)_{V_h(\Omega)}, \qquad \forall u \in V_h(\Omega). \tag{4.117}$$

From (4.116) it is easy to see that the norm of operator P_h is bounded by $\|P_h\| \leq \hat{c}$.

4.5.2.2 Some results for extensions of mesh functions in conforming finite element spaces

First, we state a lemma which makes it possible to extend the function $u \in H^1(\Omega)$ to the space $H^1(\mathbb{R}^d)$, $d = 2, 3$.

Lemma 4.3 *Let Ω be a Lipshitz domain. Then there exists a positive constant C_1 such that for any function $u \in H^1(\Omega)$ there exists a function $\tilde{u} \in H^1(\mathbb{R}^d)$ such that $\tilde{u}(\mathbf{x}) = u(\mathbf{x})$ a.e. in Ω and*

$$\|\tilde{u}\|_{H^1(\mathrm{R}^d)} \leq C_1\ \|u\|_{H^1(\Omega)}. \tag{4.118}$$

The proof of this lemma can be found in [13, 3].

For the functions $\varphi \in H^{1/2}(\partial\Omega)$ instead of (2.4) we introduce the norm

$$\|\varphi\|^2_{H^{1/2}(\partial\Omega)} = |\varphi|^2_{H^{1/2}(\partial\Omega)} + \varepsilon \cdot \|\varphi\|^2_{L^2(\partial\Omega)}. \tag{4.119}$$

Then, the following lemma is valid due to [13, 96]:

Lemma 4.4 *Let Ω be a Lipshitz domain. Then there exists a positive constant C_2' independent of ε such that*

$$\|\varphi\|_{H^{1/2}(\partial\Omega)} \leq C_2'\ \|u\|_{H^1(\Omega)} \tag{4.120}$$

for any function $u \in H^1(\Omega)$, where $\varphi \in H^{1/2}(\partial\Omega)$ is the trace of u on the boundary $\partial\Omega$.

Conversely, there exists a positive constant C_3' independent of ε such that for any function $\varphi \in H^{1/2}(\partial\Omega)$ there exists a function $u \in H^1(\Omega)$ such that $u(\mathbf{x}) = \varphi(\mathbf{x})$ a.e. on $\partial\Omega$, and

$$\|u\|_{H^1(\Omega)} \leq C_3' \; \|\varphi\|_{H^{1/2}(\partial\Omega)}. \tag{4.121}$$

Let Ω^h be a regular triangulation [36] of Ω and the nodes of the triangulation be denoted by z_i, $i = 1, \ldots, N$. Denote by $H_h^1(\Omega^h)$ the space of real-valued continuous functions u^h which are linear on the triangles τ_i of the triangulation Ω^h and by $H_h^{1/2}(\Gamma^h)$ the space of traces of functions from $H_h^1(\Omega^h)$ at the boundary $\Gamma^h \equiv \partial\Omega^h$:

$$H_h^{1/2}(\Gamma^h) = \left\{\varphi^h : \varphi^h = u^h|_{\Gamma^h}, u^h \in H_h^1(\Omega^h)\right\}.$$

To each node $z_i \in \Gamma^h$ let us put into correspondence the number $h_i = |z_i - z_i'|$, where $z_i' \in \Gamma^h$ is a node neighboring z_i, and set

$$\begin{aligned}
|\varphi^h|^2_{H_h^{1/2}(\Gamma^h)} &= \sum_{z_i, z_j \in \Gamma^h,\, z_i \neq z_j} \frac{\left(\varphi^h(z_i) - \varphi^h(z_j)\right)^2}{|z_i - z_j|^d} \; h_i^{d-1} h_j^{d-1}, \\
\|\varphi^h\|^2_{L_h^2(\Gamma^h)} &= \varepsilon \sum_{z_i \in \Gamma^h} \left(\varphi^h(z_i)\right)^2 \; h_i^{d-1}, \\
\|\varphi^h\|^2_{H_h^{1/2}(\Gamma^h)} &= \|\varphi^h\|^2_{L_h^2(\Gamma^h)} + |\varphi^h|^2_{H_h^{1/2}(\Gamma^h)}.
\end{aligned} \tag{4.122}$$

As follows from [96, 93, 41] the norm $\|\cdot\|_{H_h^{1/2}(\Gamma^h)}$ is equivalent to norm (4.119) in the subspace $H_h^{1/2}(\Gamma^h)$.

The next lemma is the mesh counterpart of Lemma 4.4.

Lemma 4.5 *Let Ω be a Lipshitz domain and Ω^h be its regular triangulation. Then there exists a positive constant C_2 independent of Ω^h and ε such that*

$$\|\varphi^h\|_{H_h^{1/2}(\Gamma^h)} \leq C_2 \; \|u^h\|_{H^1(\Omega^h)} \tag{4.123}$$

for any function $u^h \in H_h^1(\Omega^h)$, where $\varphi^h \in H_h^{1/2}(\Gamma^h)$ is the trace of u^h at the boundary Γ^h.

Conversely, there exists a positive constant C_3 independent of Ω^h and ε such that for any function $\varphi^h \in H_h^{1/2}(\Gamma^h)$ there exists a function $u^h \in H_h^1(\Omega^h)$ such that $u^h(\mathbf{x}) = \varphi^h(\mathbf{x})$ for any $\mathbf{x} \in \Gamma^h$, and

$$\|u^h\|_{H^1(\Omega^h)} \leq C_3 \; \|\varphi^h\|_{H_h^{1/2}(\Gamma^h)}. \tag{4.124}$$

4.5.2.3 Proof of the proposition

Now we prove Proposition 4.8.

Proof: Given $u \in V_h(\Omega)$ consider its map $I_h u \in H^1_{h/2}(\Omega)$. According to Lemma 4.5 for the trace φ^h of $I_h u$ on the boundary $\partial\Omega$ we have

$$\|\varphi^h\|_{H^{1/2}_{h/2}(\partial\Omega)} \leq C_2 \; \|I_h u\|_{H^1_{h/2}(\Omega)}. \tag{4.125}$$

Now we use the second part of Lemma 4.5 and define $\bar{u}^h \in H^1_{h/2}(\Omega)$ such that $\bar{u}^h(\mathbf{x}) = \varphi^h(\mathbf{x})$ for any $\mathbf{x} \in \partial\Omega$, and

$$\|\bar{u}^h\|_{H^1_{h/2}(\Omega)} \leq C_3 \; \|\varphi^h\|_{H^{1/2}_{h/2}(\partial\Omega)}. \tag{4.126}$$

Obviously, $\bar{u}^h \in H^1(\Omega)$. By Lemma 4.3 we can construct function $v \in H^1(R^2)$ as an extension of the function $\bar{u}^h$ to R^2 such that $v(\mathbf{x}) = \bar{u}^h(\mathbf{x})$ a.e. in Ω and

$$\|v\|_{H^1(\mathrm{R}^2)} \leq C_1 \; \|\bar{u}^h\|_{H^1(\Omega)}. \tag{4.127}$$

Then, define a continuous piecewise linear function $v^h \in H^1_{h/2}(\Pi)$ through its values in the nodes of triangulation $\mathcal{T}_{h/2,\Omega}$ as follows:

$$\begin{array}{rcll} v^h(\mathbf{x}) & = & (I_h u)(\mathbf{x}), & \forall \mathbf{x} \in \Omega, \\ v^h(\mathbf{x}) & = & (S^h v)(\mathbf{x}), & \forall \mathbf{x} \in \Pi \setminus \Omega, \end{array} \tag{4.128}$$

where S^h is an operator of the Steklov averaging of the function $v \in H^1(\mathbb{R}^2)$. Using the technique described in [96] we get the following inequalities:

$$\|v^h\|_{H^1_{h/2}(\Pi\setminus\Omega)} \leq C_4 \; \|\varphi^h\|_{H^{1/2}_{h/2}(\partial\Omega)} \leq C_5 \; \|I_h u\|_{H^1_{h/2}(\Omega)}. \tag{4.129}$$

Now we define a nonconforming function $\tilde{u} = P_h v^h$. From (4.116) and (4.117) it follows that

$$\|\tilde{u}\|^2_{V_h(\Pi\setminus\Omega)} = (P_h v^h, P_h v^h)_{V_h(\Pi\setminus\Omega)} = (I_h P_h v^h, v^h)_{H^1_{h/2}(\Pi\setminus\Omega)} \leq (\hat{c})^2 \; \|v^h\|^2_{H^1_{h/2}(\Pi\setminus\Omega)}. \tag{4.130}$$

From (4.129) and (4.130) we get

$$\|\tilde{u}\|^2_{V_h(\Pi\setminus\Omega)} \leq C_6 \; \|I_h u\|^2_{H^1_{h/2}(\Omega)} \leq C_7 \; \|u\|^2_{V_h(\Omega)}. \tag{4.131}$$

The result of the proposition follows from this inequality. □

Remark 4.13 Under the conditions of Proposition 4.8 we can state a similar extension result for seminorms in $V_h(\tilde{\Omega})$ and $V_h(\Pi)$. The proof of the corresponding inequalities for seminorms is analogous to the proof of Proposition 4.8 and is based on the analog of Lemma 4.5 for the seminorms in $H^1_h(\Omega)$.

Remark 4.14 The proof of Proposition 4.8 does not depend on the dimension of the space $\mathbb{R}^d$, and therefore holds true for both two- and three-dimensional problems.

Remark 4.15 We stress the fact that Proposition 4.8 is valid only for convex domain Ω. This follows from the remarks made by Nepomnyaschikh in [97] that Lemma 4.5 has various restrictions.

CHAPTER V

DOMAIN DECOMPOSITION PRECONDITIONERS FOR NONCONFORMING APPROXIMATIONS

5.1 Introduction

In the last two decades a lot of interest has been devoted to numerical methods for solving second-order boundary value problems in domains of complex geometric shape which involve a solution of analogous problems in domains of relatively simple form. The known methods of this type are the Schwarz alternating subdomain methods [42, 88, 95, 76], the fictitious components method [6, 82, 85, 86], and methods based on matrix bordering [48, 40, 89, 94, 93, 100]. Methods which are based on the partitioning of the initial domain into subdomains are called domain decomposition methods (DD).

It is believed that the first DD method was proposed by Hermann Schwarz [108]. It was originally used to show the existence of the solution of an elliptic boundary value problem on domains that consist of the union of simple overlapping subdomains.

Recently, DD algorithms have become increasingly popular because they take full advantage of modern parallel computing technology. DD methods make it possible to solve the subdomain problems independently on different processors while exchanging information between them only time to time. DD methods have an advantage of "natural parallelization" in comparison with any other effective method of solving an elliptic boundary value problem. Exhaustive results of the development of DD algorithms in the last decade can be found in the Proceedings of International Conferences on Domain Decomposition methods, and also in numerous papers (see, e.g., [14, 29, 49, 82, 86, 93, 111, 120]).

In general, DD algorithms are based on variational methods for decomposing and solving elliptic problems. Most of the applications use discretization grids which are defined globally over the whole domain and then split into subdomains. In mechanics, this results in an overall conforming approximation of the primary variable field. However, it might be more convenient and efficient to use approximations which are defined independently on each subdomain and which do not match at the interfaces. This allows the user to make local and adaptive changes to the models, the approximation strategies, or the grids in one subdomain without modifying the other ones. This of course is possible if there is an adequate way of imposing the continuity (possibly in a weak sense) of both the fluxes and primary variables across such nonconforming interfaces.

In this chapter we present a construction of the domain decomposition method for solving systems of grid equations approximating boundary value problems for second-order elliptic problems with anisotropic coefficients. We consider problems for which the computational domain Ω can be represented as a union of nonoverlapping subdomains $\Omega = \overset{m}{\underset{i=1}{\cup}} \Omega_i$ inside which the equation coefficients vary insignificantly. We develop two different methods for the nonconforming approximations of the anisotropic problems:

(A) In Section 5.2 we consider a variant of the block bordering method [89, 94] for the anisotropic problem. This algorithm uses the preconditioner developed in Chapter IV for problems in subdomains. For the problem at the interfaces we construct a preconditioner in the form of the inner Chebyshev iterative procedure. More precisely, this is a preconditioner for the Schur complement of the original symmetric positive definite matrix, which results after eliminating the block corresponding to the unknowns in the subdomains.

This approach combines the ideas of domain decomposition methods [14, 18, 29, 111, 120] and the algorithms of multilevel and algebraic multigrid methods [8, 20, 60, 70] with the bordering method for solving systems of mesh equations.

(B) In Section 5.3 we propose iterative methods for solving systems of linear equations which arise under the nonconforming finite element approximation of elliptic PDE's on nonmatching grids. More precisely, we use the technique of mortar finite elements which has been proposed recently (see, e.g., [1, 2, 12, 72, 109, 110]). The mortar element method is an optimal nonconforming domain decomposition method for the discretization of partial differential equations which provides for a maximum of mesh, refinement, and resolution flexibility while simultaneously preserving locality and elemental structure.

Using the results of Section 4.5, in each subdomain we construct its own coordinate system and a grid (triangular one for two-dimensional equations and tetrahedral one for three-dimensional equations) in accordance with the main directions of anisotropy, so that the coefficient matrix is diagonal in the local coordinates. The original elliptic problem is posed as a problem with Lagrange multipliers at the interfaces between subdomains and with the continuity conditions of the solution (in a weak form) at the same interfaces. A mortar finite element subspace is constructed in the space of Lagrange multipliers. The resulting algebraic systems have the form of a saddle-point problem.

The main part of this chapter is based on the results published in [78, 79].

5.2 Block bordering method for anisotropic problem

The outline of this section is as follows. In Subsection 5.2.1 we formulate the problem, present its nonconforming finite element discretization, and outline the construction of a block diagonal preconditioner for the algebraic system. It is shown that for the subdomain problems we can use the method described in Section 4.2. Subsection 5.2.2 is subdivided into three parts. In the first part we construct a preconditioner for the problem at the interface in the form of an inner iterative procedure considering the union of two rectangular subdomains. The second part describes an algorithm for implementing the interface preconditioner. In the third part we construct the interface preconditioner for domains composed of rectangles.

The arithmetic cost of solving the system with the proposed preconditioner is proportional to the number of the unknowns of the original algebraic system, i.e. the preconditioner constructed is of the optimal order of the arithmetical complexity.

5.2.1 Problem formulation

Let Ω be a bounded domain on a plane $\mathbb{R}^2$, which is composed of open rectangles Ω_i whose sides are parallel to the coordinate axes $\Omega = \bigcup_{i=1}^{m} \Omega_i$. Consider an elliptic problem

$$\begin{aligned} -\mathrm{div}\,(\tilde{K}\nabla u) + \tilde{c}_0 \cdot u = f \qquad & \text{in } \Omega, \\ u = 0 \qquad & \text{on } \Gamma_0, \\ (\tilde{K}\nabla u, \mathbf{n}) = 0 \qquad & \text{on } \Gamma_1, \end{aligned} \tag{5.1}$$

where $\tilde{K}(\mathbf{x})$ is a positive definite symmetric coefficient matrix, $\tilde{c}_0(\mathbf{x})$ is a nonnegative bounded function, $f(\mathbf{x}) \in L^2(\Omega)$ is a given function, $\overline{\Gamma_0 \cup \Gamma_1} = \partial\Omega$, $\Gamma_0 \cap \Gamma_1 = \emptyset$. We consider the case of $\Gamma_0 \equiv \overline{\Gamma_0} \neq \emptyset$. The pure Neumann problem ($\Gamma_0 = \emptyset$) can be treated in a similar way but for the sake of simplicity is not described here.

Assume that the interior of each side of the rectangles Ω_i either entirely belongs to Γ_0 or Γ_1, or lies inside Ω. Also assume that $\overline{\Omega}_i$, $i = 1, \ldots, m$, can have either a common side or only a common vertex, or they do not overlap. It is obvious that any domain composed of rectangles can be partitioned by additional lines into subdomains Ω_i satisfying this assumption.

Let the bilinear form $a(\cdot,\cdot)$ be defined by

$$\tilde{a}(u,v) = (\tilde{K}\nabla u, \nabla v) + (\tilde{c}_0 \cdot u, v), \qquad u, v \in V_0(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_0\},$$

where $(\cdot,\cdot)$ denotes the inner product in $L^2(\Omega)$.

Assumption 5.1 *There exist a diagonal coefficient matrix $K(\mathbf{x}) = diag\,\{k_x(\mathbf{x}), k_y(\mathbf{x})\}$ and a piecewise constant function $c_0(\mathbf{x})$ such that*

$$k_x(\mathbf{x}) = k_{x,i}, \quad k_y(\mathbf{x}) = k_{y,i}, \quad c_0(\mathbf{x}) = c_{0,i}, \qquad \mathbf{x} \in \Omega_i, \qquad i == 1, \ldots, m,$$

with constants $k_{x,i} > 0$, $k_{y,i} > 0$, $c_{0,i} \geq 0$, satisfying inequalities

$$\alpha_0\ \left((K\nabla u, \nabla u) + (c_0 \cdot u, u)\right) \leq a(u,u) \leq \alpha_1\ \left((K\nabla u, \nabla u) + (c_0 \cdot u, u)\right), \quad \forall u \in V_0(\Omega), \tag{5.2}$$

with some positive constants α_0, α_1.

The standard weak form of (5.1) is: *find $u \in V_0(\Omega)$ such that*

$$\tilde{a}(u,v) = (f,v), \qquad \forall v \in V_0(\Omega). \tag{5.3}$$

Let $\mathcal{C}_h$ be a rectangular mesh in Ω. Assume that in each rectangle Ω_i the mesh steps $h_{x,i}$, $h_{y,i}$, $i = 1, \ldots, m$, are constant in each direction, and the boundaries $\partial\Omega_i$ of the rectangles belong to the mesh lines. Also assume that there exist constants c_0 and c_1 independent of h such that

$$c_0 h \leq \min_{i=1,\ldots,m} \{h_{x,i}, h_{y,i}\} \leq \max_{i=1,\ldots,m} \{h_{x,i}, h_{y,i}\} \leq c_1 h.$$

Here $h = 1/\sqrt{M}$, where M is the number of mesh nodes belonging to $\overline{\Omega} \setminus \Gamma_0$.

Let $\mathcal{T}_h$ be a regular partitioning of $\mathcal{C}_h$ into triangles τ [36] and let $V_h(\Omega)$ be the P_1–nonconforming finite element space of functions $v \in L^2(\Omega)$ [5]: that is $v|_\tau$ are linear for all $\tau \in \mathcal{T}_h$, v are continuous at the middle points of the sides of $\tau \in \mathcal{T}_h$, and vanish at the middle points of the sides of triangles on Γ_0 (see (4.12)). Note that the space $V_h(\Omega)$ is not a subspace of $H^1(\Omega)$.

Define the bilinear forms on $V_h(\Omega)$ by

$$\begin{aligned}\tilde{a}_\Omega^h(u,v) &= \sum_{\tau\in\mathcal{T}_h} (\tilde{K}\nabla u, \nabla v)_\tau + (\tilde{c}_0 \cdot u, v)_\tau, \qquad \forall\ u,v \in V_h(\Omega),\\ \alpha_\Omega^h(u,v) &= \sum_{\tau\in\mathcal{T}_h} (K\nabla u, \nabla v)_\tau + (c_0 \cdot u, v)_\tau, \qquad \forall\ u,v \in V_h(\Omega),\end{aligned} \tag{5.4}$$

where $(\cdot,\cdot)_\tau$ is the inner product in $L^2(T)$, $\tau \in \mathcal{T}_h$. Then the P_1–nonconforming finite element discretization of (5.1) is: *find* $u_h \in V_h$ *such that*

$$\tilde{a}_\Omega^h(u_h, v) = (f, v), \qquad \forall v \in V_h(\Omega). \tag{5.5}$$

Once a nodal basis $\{\varphi_i(\mathbf{x})\}_{i=1}^N$ for $V_h(\Omega)$ is chosen, where $N = \dim\ V_h(\Omega)$, then (5.5) yields the system of linear algebraic equations (see (4.5)):

$$A\mathbf{u} = \mathbf{f}, \tag{5.6}$$

where $A_{ji} = \tilde{a}_\Omega^h(\varphi_i, \varphi_j)$, $f_j = (f, \varphi_j)$, $i,j = 1,\ldots,N$.

In the same way we define matrix $\hat{A}$ by $\hat{A}_{ji} = a_\Omega^h(\varphi_i, \varphi_j)$, $i,j = 1,\ldots,N$. Then from (5.2) it follows that

$$\alpha_0(\hat{A}\mathbf{u}, \mathbf{u}) \le (A\mathbf{u}, \mathbf{u}) \le \alpha_1(\hat{A}\mathbf{u}, \mathbf{u}), \qquad \forall \mathbf{u} \in \mathrm{I\!R}^N, \tag{5.7}$$

i.e. matrices A and $\hat{A}$ are spectrally equivalent.

The underlying method to solve (5.6) is a preconditioned iterative method. Inequalities (5.7) suggest considering matrix $\hat{A}$ as a preconditioner to A. Therefore, we need to find an efficient method for solving the problem

$$\hat{A}\mathbf{v} = \mathbf{g}. \tag{5.8}$$

Let $\mathbf{u}^{(i)}$ and $\mathbf{v}^{(i)}$ denote the vectors corresponding to the finite element functions u and v from $V_h(\Omega_i)$. Let $\hat{A}^{(i)}$ denote the local stiffness matrix arising from $\alpha_{\Omega_i}^h(\cdot,\cdot)$:

$$(\hat{A}^{(i)}\mathbf{u}^{(i)}, \mathbf{v}^{(i)}) = \alpha_{\Omega_i}^h(u,v), \qquad \forall u,v \in V_h(\Omega_i). \tag{5.9}$$

For each subdomain Ω_i, $i = 1,\ldots,m$, we can partition the degrees of freedom $\mathbf{u}^{(i)}$ into two sets. The first set includes the degrees of freedom at the nodes in the interior of subdomain Ω_i, denoted $\mathbf{u}_I^{(i)}$, and the second set corresponds to the degrees of freedom at the nodes on the boundary $\partial\Omega_i \setminus \Gamma_0$, denoted $\mathbf{u}_\Gamma^{(i)}$. Such a partitioning induces the partitioning of $\hat{A}^{(i)}$ given by

$$(\hat{A}^{(i)}\mathbf{u}^{(i)}, \mathbf{v}^{(i)}) = \left(\begin{bmatrix} \hat{A}_{II}^{(i)} & \hat{A}_{I\Gamma}^{(i)} \\ \hat{A}_{\Gamma I}^{(i)} & \hat{A}_{\Gamma\Gamma}^{(i)} \end{bmatrix} \begin{bmatrix} \mathbf{u}_I^{(i)} \\ \mathbf{u}_\Gamma^{(i)} \end{bmatrix}, \begin{bmatrix} \mathbf{v}_I^{(i)} \\ \mathbf{v}_\Gamma^{(i)} \end{bmatrix} \right). \tag{5.10}$$

Finite element system (5.8) has the obvious algebraic representation:

$$\begin{bmatrix} \hat{A}_{II}^{(1)} & & \mathbf{0} & \hat{A}_{I\Gamma}^{(1)} \\ & \ddots & & \vdots \\ \mathbf{0} & & \hat{A}_{II}^{(m)} & \hat{A}_{I\Gamma}^{(m)} \\ \hat{A}_{\Gamma I}^{(1)} & \cdots & \hat{A}_{\Gamma I}^{(m)} & \hat{A}_{\Gamma\Gamma} \end{bmatrix} \begin{bmatrix} \mathbf{v}_I^{(1)} \\ \vdots \\ \mathbf{v}_I^{(m)} \\ \mathbf{v}_\Gamma \end{bmatrix} = \begin{bmatrix} \mathbf{g}_I^{(1)} \\ \vdots \\ \mathbf{g}_I^{(m)} \\ \mathbf{g}_\Gamma \end{bmatrix}, \tag{5.11}$$

with block $\hat{A}_{\Gamma\Gamma}$ defined by

$$(\hat{A}_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{v}_\Gamma) = \sum_{i=1}^{m}(\hat{A}_{\Gamma\Gamma}^{(i)}\mathbf{u}_\Gamma^{(i)}, \mathbf{v}_\Gamma^{(i)}). \tag{5.12}$$

Note that blocks $\hat{A}_{II}^{(i)}$, $i = 1, \dots, m$, correspond to the boundary value problems in rectangles Ω_i

$$\alpha_{\Omega_i}^{h}(u_h, v) = G(v), \qquad \forall v \in V_h(\Omega_i), \quad i = 1, \dots, m, \tag{5.13}$$

with homogeneous Dirichlet boundary conditions imposed on the boundaries $\partial\Omega_i$. Denote the number of degrees of freedom in Ω_i, $\partial\Omega_i \setminus \Gamma_0$, $\overset{m}{\underset{i=1}{\cup}} \Omega_i$ and $\overset{m}{\underset{i=1}{\cup}} \partial\Omega_i \setminus \Gamma_0$ by $N_I^{(i)}$, $N_\Gamma^{(i)}$, N_I and N_Γ, $i = 1, \dots, m$, respectively.

Eliminating the unknowns $\mathbf{v}_I^{(i)}$, $i = 1, \dots, m$, in (5.11), we obtain the following Schur system:

$$\Lambda_\Gamma \mathbf{v}_\Gamma = \mathbf{G}_\Gamma, \tag{5.14}$$

where

$$\Lambda_\Gamma = \hat{A}_{\Gamma\Gamma} - \sum_{i=1}^{m} \hat{A}_{\Gamma I}^{(i)} \left[\hat{A}_{II}^{(i)}\right]^{-1} \hat{A}_{I\Gamma}^{(i)}, \qquad \mathbf{G}_\Gamma = \mathbf{g}_\Gamma - \sum_{i=1}^{m} \hat{A}_{\Gamma I}^{(i)} \left[\hat{A}_{II}^{(i)}\right]^{-1} \mathbf{g}_I^{(i)}. \tag{5.15}$$

Thus, the solution to system (5.11) can be reduced to the construction of an efficient algorithm for solving systems (5.13) in subdomains and the Schur complement system (5.14).

The algorithm for solving subdomain problems with matrices $\hat{A}_{II}^{(i)}$ is considered in Section 4.2. It is shown that these problems can be solved very efficiently.

The main goal of this section is to construct an easily invertible matrix B which is spectrally equivalent to matrix Λ_Γ:

$$c_0(B_\Gamma \mathbf{v}_\Gamma, \mathbf{v}_\Gamma) \le (\Lambda_\Gamma \mathbf{v}_\Gamma, \mathbf{v}_\Gamma) \le c_1(B_\Gamma \mathbf{v}_\Gamma, \mathbf{v}_\Gamma), \qquad \forall \mathbf{v}_\Gamma \in \mathbb{R}^{N_\Gamma},$$

where constants c_0 and c_1 are independent of mesh size parameter h, the subdomain diameters, and value of the coefficients. This issue is discussed in detail in the next subsection.

5.2.2 Preconditioner for interface problems

In this subsection we construct a preconditioner for the problem at the interface in the form of an inner iterative procedure. More precisely, we construct a preconditioner for the Schur complement of the original matrix.

5.2.2.1 Model problem

For the sake of simplicity, let us consider the model problem

$$\begin{aligned} -k_x\frac{\partial^2 u}{\partial x^2} - k_y\frac{\partial^2 u}{\partial y^2} + c_0 u = f, \qquad & \text{in } \Omega, \\ u = 0, \qquad & \text{on } \partial\Omega, \end{aligned}$$

where Ω is rectangle composed of two squares $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2$, $\Gamma = \bar{\Omega}_1 \cap \bar{\Omega}_2$ (see Fig. 5.1):

$$\begin{aligned} &\Omega = \{(x, y) : 0 < x < 2, 0 < y < 1\} \\ &\Omega_i = \{(x, y) : i - 1 < x < i, 0 < y < 1\}, \qquad i = 1, 2. \end{aligned} \tag{5.16}$$

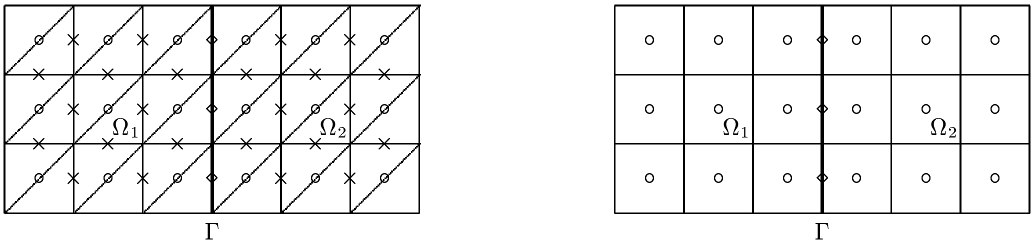
Assume that coefficients k_x, k_y, c_0 are constants in Ω_i, $i = 1, 2$, i.e.

$$k_x \equiv k_{x,i} = const > 0, \qquad k_y \equiv k_{y,i} = const > 0, \qquad c_0 \equiv c_{0,i} = const \geq 0.$$

Let $\mathcal{T}_h$ be a regular triangulation of domain Ω with mesh-size h (as described in Section 4.2). In this case (5.11) has the form:

$$\begin{bmatrix} A_{II}^{(1)} & \mathbf{0} & A_{I\Gamma}^{(1)} \\ \mathbf{0} & A_{II}^{(2)} & A_{I\Gamma}^{(2)} \\ A_{\Gamma I}^{(1)} & A_{\Gamma I}^{(2)} & A_{\Gamma\Gamma} \end{bmatrix} \begin{bmatrix} \mathbf{v}_I^{(1)} \\ \mathbf{v}_I^{(2)} \\ \mathbf{v}_\Gamma \end{bmatrix} = \begin{bmatrix} \mathbf{g}_I^{(1)} \\ \mathbf{g}_I^{(2)} \\ \mathbf{g}_\Gamma \end{bmatrix}. \tag{5.17}$$

Note that blocks $A_{II}^{(i)}$, $i = 1, 2$, correspond to the boundary value problems in subdomains Ω_i and block $A_{\Gamma\Gamma}$ is a diagonal matrix.



(a) Triangulation of the domain Ω. *(b) The degrees of freedom of the reduced problem.*

Figure 5.1: *Degrees of freedom of real and reduced problems.*

Eliminating the unknowns $vx_{i,j}$ and $vy_{i,j}$ in each subdomain as is discussed in Section 4.2 we get the problem

$$\begin{bmatrix} \hat{A}_{II}^{(1)} & \mathbf{0} & \hat{A}_{I\Gamma}^{(1)} \\ \mathbf{0} & \hat{A}_{II}^{(2)} & \hat{A}_{I\Gamma}^{(2)} \\ \hat{A}_{\Gamma I}^{(1)} & \hat{A}_{\Gamma I}^{(2)} & A_{\Gamma\Gamma} \end{bmatrix} \begin{bmatrix} \mathbf{v}_c^{(1)} \\ \mathbf{v}_c^{(2)} \\ \mathbf{v}_\Gamma \end{bmatrix} = \begin{bmatrix} \mathbf{g}_c^{(1)} \\ \mathbf{g}_c^{(2)} \\ \mathbf{g}_\Gamma \end{bmatrix}, \tag{5.18}$$

where blocks $\hat{A}_{II}^{(i)}$, $i = 1, 2$, are separable matrices, and vectors $\mathbf{v}_c^{(i)}$, $i = 1, 2$, consist of the unknowns $vc_{i,j}$ in each subdomain. In Figure 5.1b, the nodes corresponding to these unknowns are marked by "$\circ$".

Matrix $A_{\Gamma\Gamma}$ is defined by equality (5.12)

$$(A_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{v}_\Gamma) = \sum_{i=1}^{2} (A_{\Gamma\Gamma}^{(i)}\mathbf{u}_\Gamma^{(i)}, \mathbf{v}_\Gamma^{(i)}).$$

Introducing the subdomain Schur complements

$$\Lambda_\Gamma^{(i)} = A_{\Gamma\Gamma}^{(i)} - \hat{A}_{\Gamma I}^{(i)} \left[\hat{A}_{II}^{(i)}\right]^{-1} \hat{A}_{I\Gamma}^{(i)}, \qquad i = 1, 2, \tag{5.19}$$

Schur complement (5.15) can be rewritten in the form:

$$\Lambda_\Gamma = A_{\Gamma\Gamma} - \sum_{i=1}^{2} \hat{A}_{\Gamma I}^{(i)} \left[\hat{A}_{II}^{(i)}\right]^{-1} \hat{A}_{I\Gamma}^{(i)} = \Lambda_\Gamma^{(1)} + \Lambda_\Gamma^{(2)}. \tag{5.20}$$

Preconditioner B_Γ for Λ_Γ is constructed by defining preconditioners $B_\Gamma^{(1)}$ and $B_\Gamma^{(2)}$ for $\Lambda_\Gamma^{(1)}$ and $\Lambda_\Gamma^{(2)}$, respectively, and setting

$$B_\Gamma = B_\Gamma^{(1)} + B_\Gamma^{(2)}. \tag{5.21}$$

Let us consider subdomain Ω_1, matrices $\Lambda_\Gamma^{(1)}$ and $A_{\Gamma\Gamma}^{(1)}$, and omit the index "(1)" for simplicity. Boundary nodes belonging to Γ (marked by "$\diamond$") and internal nodes (marked by "$\circ$") are schematically shown in Figure 5.2.

The following lemma is valid.

Lemma 5.1 *There exists an h-independent constant α such that*

$$\alpha \cdot h\ (A_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \le (\Lambda_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \le (A_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma), \qquad \forall \mathbf{u}_\Gamma \in \mathrm{I\!R}^{N_\Gamma}. \tag{5.22}$$

Proof: Consider an eigenvalue problem

$$\Lambda_\Gamma \mathbf{u}_\Gamma = \mu A_{\Gamma\Gamma}\mathbf{u}_\Gamma, \qquad \mathbf{u}_\Gamma \in \mathrm{I\!R}^{N_\Gamma}. \tag{5.23}$$

Since the symmetric matrices Λ_Γ and $A_{\Gamma\Gamma}$ are positive definite problem (5.23) has N_Γ positive eigenvalues.

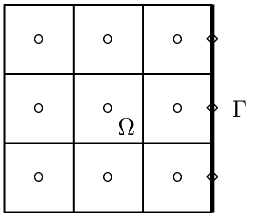


Figure 5.2: *Degrees of freedom of the model subdomain problem.*

It is obvious that eigenvalues μ_i, $i = 1, \dots, N_\Gamma$, of problem (5.23) can be found from the system of equations

$$\begin{aligned} \hat{A}_{II}\mathbf{u}_I + \hat{A}_{I\Gamma}\mathbf{u}_\Gamma &= 0, \\ \hat{A}_{\Gamma I}\mathbf{u}_I + A_{\Gamma\Gamma}\mathbf{u}_\Gamma &= \mu A_{\Gamma\Gamma}\mathbf{u}_\Gamma. \end{aligned} \tag{5.24}$$

Here matrix $\hat{A}_{II}$ is defined by (4.24). The $N_I \times N_\Gamma$ matrix $\hat{A}_{I\Gamma}$ has the form:

$$\hat{A}_{I\Gamma} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ -2k_x I_y \end{bmatrix} \equiv \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -2k_x \end{bmatrix} \otimes I_y,$$

and the diagonal $N_\Gamma \times N_\Gamma$ matrix $A_{\Gamma\Gamma}$ is defined by $A_{\Gamma\Gamma} = 2(k_x + c)I_y$.

Define by $\boldsymbol{\xi}_l \in \mathrm{I\!R}^n$ an eigenvector of $n \times n$ matrix A_y (4.25) corresponding to the eigenvalue λ_l:

$$A_y \boldsymbol{\xi}_l = \lambda_l \boldsymbol{\xi}_l, \qquad l = 1, \dots, n. \tag{5.25}$$

Note that we explicitly know the eigenvalues and eigenvectors of matrix A_y:

$$\lambda_l = 4\sin^2\frac{\pi l}{2n}, \qquad \boldsymbol{\xi}_l = \left\{\sin\frac{\pi l(2i-1)}{2n}\right\}_{i=1}^{n}, \qquad l = 1,\dots,n. \tag{5.26}$$

Fix some $l \in [1,n]$, and define a vector

$$\begin{bmatrix} \mathbf{u}_I \\ \mathbf{u}_\Gamma \end{bmatrix} = \begin{bmatrix} \mathbf{v}\otimes\boldsymbol{\xi}_l \\ \boldsymbol{\xi}_l \end{bmatrix},$$

with some vector $\mathbf{v} \in \mathrm{I\!R}^n$ and substitute it in system (5.24). From the second equation of (5.24) it follows that eigenvalues of problem (5.23) are defined by the expressions

$$\mu_l = 1 - \frac{k_x}{k_x + c}\, v_n^{(l)}, \qquad l = 1,\dots,n, \tag{5.27}$$

where a parameter $v_n^{(l)}$ is the n-th component of vector $\mathbf{v}^{(l)}$ from the system

$$\hat{A}_{II}(\mathbf{v}^{(l)}\otimes\boldsymbol{\xi}_l) = -\hat{A}_{I\Gamma}\boldsymbol{\xi}_l$$

or, using expressions (4.23),

$$(a_x A_x + (\lambda_l a_y + b)I_x)\,\mathbf{v}^{(l)} = 2k_x \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix}^T. \tag{5.28}$$

Introducing notations:

$$d_l = \lambda_l\frac{a_y}{a_x} + \frac{b}{a_x}, \qquad e = \frac{2k_x}{a_x} \equiv 2\left(1 + \frac{c}{k_x}\right), \tag{5.29}$$

system (5.28) can be rewritten in the form:

$$\begin{bmatrix} 3+d_l & -1 & & & \\ -1 & 2+d_l & -1 & & \mathbf{0} \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2+d_l & -1 \\ \mathbf{0} & & & -1 & 3+d_l \end{bmatrix} \cdot \begin{bmatrix} v_1^{(l)} \\ v_2^{(l)} \\ \vdots \\ v_{n-1}^{(l)} \\ v_n^{(l)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ e \end{bmatrix}. \tag{5.30}$$

Set $x_l = 1 + \frac{1}{2}d_l$ and consider the following recurrent sequence:

$$\begin{array}{lcll} \alpha_0 & = & 1 & \\ \alpha_1 & = & 2x_l + 1 & \\ \alpha_{i+1} & = & 2x_l\alpha_i - \alpha_{i-1}, & \quad i = 1,\dots,n-1. \end{array} \tag{5.31}$$

It is easy to see that

$$\alpha_i = U_i(x_l) + U_{i-1}(x_l), \qquad i = 1,\dots,n,$$

where $U_m(x)$ is the Chebyshev polynomial of the 2-nd kind of degree m:

$$U_m(x) = \frac{1}{2\sqrt{x^2-1}}\left(\left(x + \sqrt{x^2-1}\right)^{m+1} - \left(x + \sqrt{x^2-1}\right)^{-(m+1)}\right).$$

By induction it can be shown that

$$v_n^{(l)} = e \cdot \frac{\alpha_{n-1}}{\alpha_{n-1} + \alpha_n} = e \cdot \frac{U_{n-1}(x_l) + U_{n-2}(x_l)}{U_n(x_l) + 2U_{n-1}(x_l) + U_{n-2}(x_l)}. \tag{5.32}$$

Since $x_l \geq 1$ then $v_n^{(l)} \geq 1$ for any $l = 1, \ldots, n$. From (5.27) it follows that the maximal eigenvalue of problem (5.23) is bounded from above by $\mu_{\max} \leq 1$.

Now we have to estimate the minimal eigenvalue $\mu_{\min}$ of problem (5.23) from below. Taking into account that $U_n(x_l) = 2x_l U_{n-1}(x_l) - U_{n-2}(x_l)$ we get

$$v_n^{(l)} = e \cdot \frac{(2x_l + 1)U_{n-1}(x_l) - U_n(x_l)}{(2x_l + 2)U_{n-1}(x_l)} = \frac{e}{2} \cdot \left(\frac{2x_l + 1}{x_l + 1} - \frac{U_n(x_l)}{(x_l + 1)U_{n-1}(x_l)} \right). \tag{5.33}$$

Since

$$\begin{aligned} \frac{U_n(x)}{U_{n-1}(x)} &= \frac{(x + \sqrt{x^2 - 1})^{n+1} - (x + \sqrt{x^2 - 1})^{-(n+1)}}{(x + \sqrt{x^2 - 1})^n - (x + \sqrt{x^2 - 1})^{-n}} = \\ &= x + \sqrt{x^2 - 1} \left(1 + \frac{2}{(x + \sqrt{x^2 - 1})^{2n} - 1} \right), \end{aligned}$$

from (5.27), (5.29), and (5.33) it follows that

$$\begin{aligned} \mu_l &= \sqrt{\frac{x_l - 1}{x_l + 1}} \left(1 + \frac{2}{\left(x_l + \sqrt{x_l^2 - 1} \right)^{2n} - 1} \right) = \\ &= \sqrt{\frac{d_l}{4 + d_l}} \left(1 + \frac{2}{\left(1 + \frac{1}{2}d_l + \sqrt{d_l + \frac{1}{4}d_l^2} \right)^{2n} - 1} \right). \end{aligned} \tag{5.34}$$

To estimate expression (5.34) from below we consider two cases:

1. $y \equiv \left(\frac{1}{2}d_l + \sqrt{d_l + \frac{1}{4}d_l^2} \right) \geq 1/2n$. It means that $d_l + \sqrt{d_l(4 + d_l)} \geq 1/n$, or $d_l \geq \frac{1}{2n(2n+1)}$. Then we have
$$\mu_l \geq \sqrt{\frac{d_l}{4 + d_l}} \geq \frac{1}{\sqrt{1 + 8n(2n + 1)}} = \frac{1}{4n + 1} \geq \frac{h}{5}.$$

2. $y < 1/2n$. In this case $d_l < \frac{1}{2n(2n+1)}$. So, (5.34) is estimated from below as follows

$$\begin{aligned} \mu_l &= \sqrt{\frac{d_l}{4 + d_l}} \left(1 + \frac{2}{(1 + y)^{2n} - 1} \right) \geq \sqrt{\frac{d_l}{4 + d_l}} \left(1 + \frac{2}{e^{2ny} - 1} \right) \\ &\geq \sqrt{\frac{d_l}{4 + d_l}} \left(1 + \frac{2}{(e - 1)2ny} \right) \geq \sqrt{\frac{d_l}{4 + d_l}} \cdot \frac{1 + ny}{2ny} \\ &= \sqrt{\frac{d_l}{4 + d_l}} \left(\frac{\frac{1}{n} + \frac{1}{2}\left(d_l + \sqrt{d_l(4 + d_l)} \right)}{d_l + \sqrt{d_l(4 + d_l)}} \right) = \frac{\frac{1}{n} + \frac{1}{2}\left(d_l + \sqrt{d_l(4 + d_l)} \right)}{4 + \left(d_l + \sqrt{d_l(4 + d_l)} \right)} \\ &\geq \min\left\{ \frac{1}{2}; \frac{1}{4n} \right\} \geq \frac{h}{4}. \end{aligned}$$

From these estimates it follows that $\mu_l \geq h/5$ for any $l = 1, \ldots, n$. Thus, the minimal eigenvalue of problem (5.23) is bounded from below by $\mu_{\min} \geq h/5$ and we have

$$\frac{h}{5}\,(A^{(1)}_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \leq (\Lambda^{(1)}_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \leq (A^{(1)}_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma), \qquad \forall \mathbf{u}_\Gamma \in \mathbb{R}^{N_\Gamma}.$$

Analogously, we have the same estimates for matrices $A^{(2)}_{\Gamma\Gamma}$ and $\Lambda^{(2)}_\Gamma$, which completes the proof. □

Remark 5.1 Lemma 5.1 holds true if on some of the other edges of the rectangular subdomain Ω_1 a homogeneous Neumann boundary condition is imposed.

Remark 5.2 If a homogeneous Neumann boundary condition is imposed on the whole remaining part of the boundary of subdomain $\partial\Omega_1 \setminus \Gamma$ then inequalities (5.22) of Lemma 5.1 should be replaced by:

$$\begin{aligned} \alpha \cdot h\,(A_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \leq (\Lambda_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma) &\leq (A_{\Gamma\Gamma}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma), \\ &\forall \mathbf{u}_\Gamma \in \mathbb{R}^{N_\Gamma} \setminus \mathrm{Ker}\,(\Lambda_\Gamma), \end{aligned} \tag{5.35}$$

where constant α does not depend on mesh size parameter h and coefficients of the subdomain problems.

Next, we proceed with the construction of preconditioner B_Γ. We define it in the form of an inner Chebyshev iterative procedure [8, 18, 63, 70]. From Lemma 5.1 we know that the eigenvalues of matrix $A^{-1}_{\Gamma\Gamma}\Lambda_\Gamma$ belong to segment $[h/5, 1]$. Let $P_L(y)$ be the polynomial of least deviation from zero on this segment and that satisfies the condition $P_L(0) = 1$. Denote by β_l, $l = 1, \ldots, L$, the inverses of the roots of the polynomial $P_L(y)$. The formulae for $P_L(y)$ and its roots $1/\beta_l$, $l = 1, \ldots, L$, are given in Section 3.4. Then preconditioner B_Γ for matrix Λ_Γ is determined by:

$$B^{-1}_\Gamma = \left\{ I_\Gamma - \prod_{l=1}^{L} \left(I_\Gamma - \beta_l A^{-1}_{\Gamma\Gamma}\Lambda_\Gamma \right) \right\} \Lambda^{-1}_\Gamma. \tag{5.36}$$

The procedure for calculating the vector $\mathbf{w}_\Gamma = B^{-1}_\Gamma \mathbf{g}_\Gamma$ for given $\mathbf{g}_\Gamma \in \mathbb{R}^{N_\Gamma}$ has the form:

$$\begin{aligned} \mathbf{w}^{(0)}_\Gamma &= \mathbf{0}, \\ \mathbf{w}^{(l)}_\Gamma &= \mathbf{w}^{(l-1)}_\Gamma - \beta_l A^{-1}_{\Gamma\Gamma}\left(\Lambda_\Gamma \mathbf{w}^{(l-1)}_\Gamma - \mathbf{g}_\Gamma\right), \qquad l = 1, \ldots, L, \\ \mathbf{w}_\Gamma &= \mathbf{w}^{(L)}_\Gamma. \end{aligned} \tag{5.37}$$

For computational stability, instead of (5.37), we can use the three-term recurrence relation [114]. We return to the realization of the iterative procedure (5.37) in the next subsection.

Lemma 5.1 and the theory of Chebyshev iterative methods imply the following basic result.

Theorem 5.1 *Let $L \geq (5/h)^{1/2}$. Then matrix B_Γ in* (5.36) *is spectrally equivalent to matrix Λ_Γ with constants of equivalence independent of mesh size parameter h and the value of coefficients $k_{x,i}$, $k_{y,i}$, $c_{0,i}$, $i = 1, 2$, in the subdomains.*

Remark 5.3 Clearly, in the theory L is chosen to be of the order $(5/h)^{1/2}$. In practice it is calculated explicitly after the boundaries of the spectrum of matrix $A^{-1}_{\Gamma\Gamma}\Lambda_\Gamma$ have been calculated by an appropriate iterative procedure [62].

Remark 5.4 If we replace Λ_Γ in (5.36) by a spectrally equivalent matrix we can introduce another preconditioner $\tilde{B}_\Gamma$ for Λ_Γ by

$$\tilde{B}_\Gamma^{-1} = \left\{ I_\Gamma - \prod_{l=1}^{L} \left(I_\Gamma - \beta_l A_{\Gamma\Gamma}^{-1} S_\Gamma \right) \right\} S_\Gamma^{-1}. \tag{5.38}$$

This matrix also will be spectrally equivalent to Λ_Γ with constants of equivalence independent of mesh size parameter h and the value of coefficients $k_{x,i}$, $k_{y,i}$, $c_{0,i}$, $i = 1, 2$, in the subdomains.

5.2.2.2 Arithmetical complexity of the interface preconditioner

In order to estimate the arithmetic complexity of multiplying a vector by matrix B_Γ^{-1} it is sufficient to estimate the arithmetic complexity of multiplying a vector by matrix Λ_Γ because we know that matrix $A_{\Gamma\Gamma}$ is diagonal and the number of iterations L in the inner Chebyshev iterative procedure (5.37) is of the order $(5n)^{1/2}$.

The procedure of finding the product $\Lambda_\Gamma \mathbf{u}_G$ is based on a partial solution technique [9, 10, 68, 103], which we outline here. We consider the terms $\Lambda_\Gamma^{(1)} \mathbf{u}_\Gamma$ and $\Lambda_\Gamma^{(2)} \mathbf{u}_\Gamma$ in the expression $\mathbf{v}_\Gamma = \Lambda_\Gamma^{(1)} \mathbf{u}_\Gamma + \Lambda_\Gamma^{(2)} \mathbf{u}_\Gamma$ separately. Below we define the multiplication procedure only for the first term $\Lambda_\Gamma^{(1)} \mathbf{u}_\Gamma$. The second term is treated in the same manner. Again, we skip the index "(1)" for simplicity.

First, vector $\mathbf{v}_\Gamma = \Lambda_\Gamma \mathbf{u}_\Gamma = A_{\Gamma\Gamma} \mathbf{u}_\Gamma - \hat{A}_{\Gamma I} \hat{A}_{II}^{-1} \hat{A}_{I\Gamma} \mathbf{u}_\Gamma$ is rewritten as

$$\mathbf{v}_\Gamma = A_{\Gamma\Gamma} \mathbf{u}_\Gamma + \hat{A}_{\Gamma I} \mathbf{u}_I, \tag{5.39}$$

where vector $\mathbf{u}_I$ is defined from the system of equations

$$\hat{A}_{II} \mathbf{u}_I = -\hat{A}_{I\Gamma} \mathbf{u}_\Gamma \equiv \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ 2k_x \mathbf{u}_\Gamma \end{bmatrix}, \tag{5.40}$$

with matrix $\hat{A}_{II}$ defined by (4.24).

Next, we denote by W_y an orthogonal $n \times n$ matrix of eigenvectors of problem (5.25) and by L_y a diagonal $n \times n$ matrix of the eigenvalues of matrix A_y (5.26):

$$W_y = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}, \qquad L_y = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}.$$

Then we define an $N_I \times N_I$ orthogonal matrix $Q = I_x \otimes W_y$. Introducing a vector $\mathbf{v}_I = Q^T \mathbf{u}_\Gamma$ and multiplying both parts of equation (5.40) by matrix Q^T we get the following matrix equation:

$$Q^T \hat{A}_{II} Q \mathbf{v}_I \equiv (a_x(A_x \otimes I_y) + a_y(I_x \otimes L_y) + b(I_x \otimes I_y))\, \mathbf{v}_I = 2k_x \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ W_y^T \mathbf{u}_\Gamma \end{bmatrix}. \tag{5.41}$$

Note that this system can be decoupled into n independent linear systems:

$$(a_x A_x + (b + \lambda_l a_y) I_x) \begin{bmatrix} v_1^{(l)} \\ \vdots \\ v_{n-1}^{(l)} \\ v_n^{(l)} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 2k_x w_l \end{bmatrix}, \qquad l = 1, \dots, n, \tag{5.42}$$

where the component w_l is the l-th component of the backward Fourier transform $\mathbf{w} = W_y^T \mathbf{u}_\Gamma$ of vector $\mathbf{u}_\Gamma$.

As soon as we find vector $\mathbf{v}_I$ from systems (5.42) we compute the product

$$\hat{A}_{\Gamma I} \mathbf{u}_I = \hat{A}_{\Gamma I} Q \mathbf{v}_I = -2k_x W_y \begin{bmatrix} v_n^{(1)} \\ \vdots \\ v_n^{(n)} \end{bmatrix}. \tag{5.43}$$

From (5.43) it follows that to define product $\Lambda_\Gamma \mathbf{u}_\Gamma$ we need to know only the last components $v_n^{(l)}$, $l = 1, \dots, n$, of the solution vectors $\mathbf{v}^{(l)}$ from systems (5.42).

Now we define the partial solution algorithm:

(0) On the initial step we solve n systems with three-diagonal $n \times n$ matrices:

$$(a_x A_x + (b + \lambda_l a_y) I_x) \begin{bmatrix} x_1^{(l)} \\ \vdots \\ x_{n-1}^{(l)} \\ x_n^{(l)} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

and store the last components of vectors $\mathbf{x}^{(l)}$, i.e. parameters $x_n^{(l)}$, $l = 1, \dots, n$. It requires $O(n^2)$ operations and $O(n)$ elements of memory.

(1) Given vector $\mathbf{u}_\Gamma$ we use the discrete fast Fourier transform algorithm to compute a vector $\mathbf{w} = 2k_x W_y^T \mathbf{u}_\Gamma$ for only $O(n \ln n)$ operations.

(2) Then, we compute a vector $\mathbf{v} = \left[v_n^{(1)}, \dots, v_n^{(n)} \right]$ from (5.43) by the formulae $v_n^{(l)} = w_l \cdot x_n^{(l)}$, $l = 1, \dots, n$.

(3) Again, we use the discrete fast Fourier transform algorithm to compute a vector $\mathbf{p} = -2k_x W_y \mathbf{v}$.

(4) Finally, we compute the vector $\mathbf{v}_\Gamma = A_{\Gamma\Gamma} \mathbf{u}_\Gamma + \mathbf{p}$.

As a result of this algorithm the procedure of multiplying a vector by matrix B_Γ^{-1} can be implemented for $O(n^2 + n^{3/2} \ln n)$ operations. Note that this estimate does not depend on the coefficients of the problem.

Remark 5.5 We provided computations of the complexity for so-called "parallel" partial solutions, that is, when the grid line where we need to find a solution is parallel to the grid line where the right-hand side is nonzero. For the case of so-called "perpendicular" partial solutions, when the grid line where we need the solution is perpendicular to the grid line with the nonzero right-hand side we can use an algorithm of the approximate partial solution [9, 68, 103]. Instead of computing the exact value of $\mathbf{v}_\Gamma = \Lambda_\Gamma \mathbf{u}_\Gamma$, we calculate the approximation $\mathbf{v}_\Gamma^{(\varepsilon)} = \Lambda_\Gamma^{(\varepsilon)} \mathbf{u}_\Gamma$. It can be shown [9, 103] that taking the accuracy parameter $\varepsilon \sim h^p$, $p \geq 2$, the matrix $\Lambda_\Gamma^{(\varepsilon)}$ is spectrally equivalent to Λ_Γ.

As a consequence one can develop the partial solution algorithm for the general case when we need to find the partial solution on the entire boundary of the rectangular subdomain. Using the results of [9, 68] the partial solution on the boundary $\partial\Omega_i$ of the rectangular subdomain Ω_i can be found for $O(n^2 + pn \ln^2(n))$, where the first term of this estimate corresponds to the initial step and can be obtained before starting the iterative process. If we use the algorithm of the approximate partial solution developed in [103] we have an estimate of computational cost $O(n^2 + pn^{3/2} \ln(n))$.

Therefore, the algorithm of multiplying a vector by matrix B_Γ^{-1} can be implemented for $O(n^2 + pn^{3/2} \ln^2(n))$ operations. This estimate does not depend on the coefficients of the problem.

5.2.2.3 Interface preconditioner for general problem

Let us now consider problem (5.1) in the domain $\Omega = \bigcup_{i=1}^{m} \Omega_i$, being a union of m rectangles Ω_i, $i = 1, \dots, m$, as is described in Section 5.2.1.

Using the same arguments as in Subsection 5.2.2.1 it is easy to show that the statement of Lemma 5.1 holds true even for a general domain Ω composed of rectangles. Since preconditioner B_Γ is constructed by defining subdomain preconditioners $B_\Gamma^{(i)}$ it is sufficient to consider only the model problem in the rectangular subdomain Ω_i with homogeneous Neumann boundary condition on the whole boundary $\partial\Omega \equiv \Gamma = \bigcup_{i=1}^{4} \Gamma_i$. The boundary nodes belonging to Γ ("$\diamond$") and the internal nodes ("$\circ$") are schematically shown in Figure 5.3.

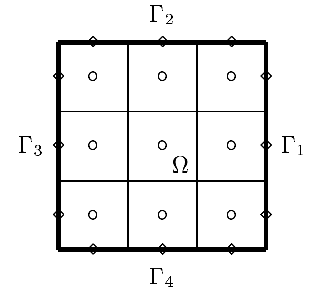


Figure 5.3: *Degrees of freedom of the Neumann subdomain problem.*

Denote by $\mathbf{u}_\Gamma$ and by $\mathbf{u}_{\Gamma_i}$, $i = 1, \dots, 4$, the vectors of the degrees of freedom on the boundaries Γ and Γ_i, $i = 1, \dots, 4$, respectively. Following [89], one can show that for any vector $\mathbf{v}_\Gamma \in \mathrm{I\!R}^{N_\Gamma}$ such that $\mathbf{v}_\Gamma \perp \mathrm{Ker}\,(\Lambda_\Gamma)$ the following is valid:

$$(\Lambda_\Gamma \mathbf{v}_\Gamma, \mathbf{v}_\Gamma) = \inf_{\substack{v^h \in V_h(\Omega) \\ v^h|_\Gamma = v^h_\Gamma}} b^h_\Omega(v^h, v^h) \geq \inf_{\substack{w^h_i \in V_h(\Omega) \\ w^h_i|_{\Gamma_i} = v^h_{\Gamma_i}}} b^h_\Omega(w^h_i, w^h_i) = (\Lambda_{\Gamma_i} \mathbf{v}_{\Gamma_i}, \mathbf{v}_{\Gamma_i}), \qquad i = 1, \dots, 4, \tag{5.44}$$

where u^h_Γ, $u^h_{\Gamma_i}$, $i = 1,\ldots,4$, are piecewise constant functions defined on Γ, Γ_i, $i = 1,\ldots,4$, and generated by vectors $\mathbf{u}_\Gamma$, $\mathbf{u}_{\Gamma_i}$, $i = 1,\ldots,4$, respectively.

From (5.21), (5.35), and (5.44) it follows that for any vector $\mathbf{v}_\Gamma \in \mathrm{I\!R}^{N_\Gamma}$ such that $\mathbf{v}_\Gamma \perp$ Ker (Λ_Γ) we have

$$\begin{gathered}(A_{\Gamma\Gamma}\mathbf{u}_\Gamma,\mathbf{u}_\Gamma) \geq (\Lambda_\Gamma\mathbf{u}_\Gamma,\mathbf{u}_\Gamma) \geq \tfrac{1}{4}\sum_{i=1}^{4}(\Lambda_{\Gamma_i}\mathbf{u}_{\Gamma_i},\mathbf{u}_{\Gamma_i}) \geq \\ \geq \alpha h \cdot \sum_{i=1}^{4}(A_{\Gamma_i\Gamma_i}\mathbf{u}_{\Gamma_i},\mathbf{u}_{\Gamma_i}) \geq \alpha h \cdot (A_{\Gamma\Gamma}\mathbf{u}_\Gamma,\mathbf{u}_\Gamma).\end{gathered} \tag{5.45}$$

Thus, we can define preconditioner B_Γ for the Schur complement (5.15) in the form (5.36). By Theorem 5.1 matrix B_Γ is spectrally equivalent to matrix Λ_Γ provided that the degree L of the matrix polynomial (5.36) is chosen to be $O(h^{-1/2})$.

Using the partial solution technique described in Subsection 5.2.2.2 and Remark 5.5 one can show that the procedure of multiplying a vector by matrix B_Γ^{-1} can be implemented for $O(h^{-2} + h^{-3/2}\ln^2(h^{-1}))$.

Now assume that we use the AMG methods to solve the subdomain problems. Then problem (5.8) with matrix $\hat{A}$ can be solved for $O(h^{-2} + h^{-3/2}\ln^2(h^{-1}))$ operations. As was mentioned in Section 5.2.1 we use matrix $\hat{A}$ as the preconditioner in a preconditioned iterative method to solve problem (5.6).

Summarizing the results of Sections 5.2.1 and 5.2.2 we can formulate the following proposition.

Proposition 5.1 *Under the above assumptions the arithmetic complexity of the proposed algorithm for solving problem* (5.6) *is estimated from above by*

$$C \cdot (h^{-2} + h^{-3/2}\ln^2(h^{-1})),$$

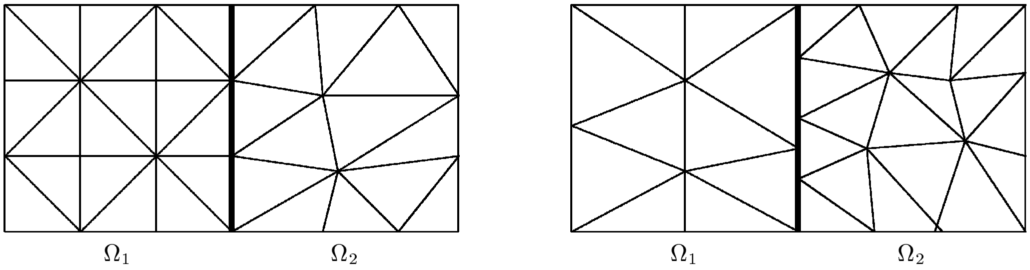
where constant C is independent of mesh size parameter h and coefficients of the problem, $\tilde{K}(\mathbf{x})$ *and* $\tilde{a}_0(\mathbf{x})$.

5.3 Domain decomposition method on nonmatching grids

In this section we describe an algorithm for solving systems of linear algebraic equations arising from nonconforming finite element approximations of the anisotropic diffusion equations on nonmatching grids. We stress that the corresponding matrix is symmetric but indefinite. The iterative method to be considered involves a block diagonal preconditioner with the inner Chebyshev iterative procedure and the preconditioned Lanczos method as an outer iterative procedure.

First, the original differential problem is represented in the hybrid-mixed form using the Arnold-Brezzi formulation (see Section 2.4) via nonoverlapping domain decomposition using additional Lagrange multipliers to enforce the necessary continuity of the solution on the interfaces between subdomains [28, 58, 74]. Next, using the equivalence between hybrid-mixed and nonconforming finite element methods we replace the original three-field formulation in each subdomain with the simple nonconforming one. The original elliptic problem is thus imposed as a nonconforming discrete problem with Lagrange multipliers at the interfaces between the subdomains, into which the original domain is decomposed. At these interfaces certain continuity conditions on the solution are imposed. This construction is done to inherit the properties of the Lagrange multiplier space defined on the interfaces between the subdomains. The Dirichlet boundary conditions, if any, are also given by the Lagrange multipliers.

In each subdomain we introduce its own grid, namely, a triangular one in two dimensions and a tetrahedral one in three dimensions (see an example in Figure 5.4), and corresponding P_1-nonconforming finite element space. A mortar finite element space is constructed in the space of the Lagrange multipliers. The error analysis of this finite element approximation is not discussed in the dissertation.



(a) Matching grids

(b) Nonmatching grids

Figure 5.4: *Example of matching and nonmatching grids.*

The section is logically divided into two subsections. In the first subsection we formulate the problem, present its discretization using nonoverlapping domain decomposition, and give the algebraic formulation of the finite element problem in a saddle-point form.

In the second subsection we construct the block-diagonal preconditioner, analyze its properties, and discuss the implementation costs. The motivation for such a choice for the preconditioner is given in Section 3.2. It is shown that for subdomains we can choose the preconditioners constructed in Chapter IV. For the problem on the interfaces the preconditioner is introduced in the form of the inner Chebyshev procedure for the matrix which is spectrally equivalent to the Schur complement. The construction and the analysis of this preconditioner is based on the new approach recently developed in [72] for solving finite element problems on nonmatching grids with Lagrange multipliers on the interfaces between the subdomains.

It is shown that the proposed block-diagonal preconditioner is spectrally equivalent to the original saddle-point matrix (in the sense of the definition given in Section 3.2) with constants independent of mesh-size parameter and coefficients of the problem.

5.3.1 Mortar finite element method with Lagrange multipliers

Let Ω be a bounded domain in the space $\mathbb{R}^d$, $d = 2, 3$, with boundary $\partial\Omega$. We consider the problem:

$$\begin{aligned} -\mathrm{div}\,(K\nabla u) + c \cdot u &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma_0, \\ (K\nabla u, \mathbf{n}) &= 0 && \text{on } \Gamma_1, \end{aligned} \tag{5.46}$$

where $K(\mathbf{x})$ is a positive definite symmetric coefficient matrix, c is a nonnegative bounded function, and $f \in L^2(\Omega)$ is a given function. Here Γ_0 is a closed subset of $\partial\Omega$ and Γ_1 is another subset of $\partial\Omega$ such that $\Gamma_0 \cup \Gamma_1 = \emptyset$, and $\Gamma_0 \cup \Gamma_1 = \partial\Omega$.

Let the bilinear form $a(\cdot, \cdot)$ be defined by

$$a(u, v) = (K\nabla u, \nabla v) + (c \cdot u, v), \qquad u, v \in V_0(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_0\},$$

where $(\cdot, \cdot)$ denotes the inner product in $L^2(\Omega)$.

Assume that domain Ω is partitioned into the nonoverlapping simply connected subdomains: $\Omega = \bigcup\limits_{i=1}^{m} \Omega_i$. The sets $\Gamma_{kl} = \bar{\Omega}_k \cap \bar{\Omega}_l$, $|\Gamma_{kl}| \neq 0$, are called the interfaces between the subdomains Ω_k and Ω_l. For convenience, we consider here only a model problem with additional assumptions.

Assumption 5.2 *The partitioning of Ω into polygonal subdomains Ω_k is quasiuniform, with a diameter of the subdomains being $r \sim (1/m)^{1/d}$. Each of the interfaces Γ_{kl} is a simply connected set. The interior of each side of the subdomains Ω_i either entirely belongs to Γ_0 or Γ_1, or lies inside Ω.*

Assumption 5.3 *Assume that there exist a piecewise constant function $\tilde{c}(\mathbf{x}) = c^{(l)}$, $\mathbf{x} \in \Omega_l$, $l = 1, \ldots, m$ and a symmetric coefficient matrix*

$$\tilde{K}(\mathbf{x}) = \left\{\tilde{k}_{ij}(\mathbf{x})\right\}_{ij=1}^{d}$$

with piecewise constant functions

$$\tilde{k}_{ij}(\mathbf{x}) = k_{ij}^{(l)}, \qquad \mathbf{x} \in \Omega_l, \qquad l = 1, \ldots, m,$$

satisfying the inequalities

$$\alpha_0 \, a(u,u) \leq (\tilde{K}\nabla u, \nabla u) + (\tilde{c} \cdot u, u) \leq \alpha_1 \, a(u,u), \qquad \forall u \in V_0(\Omega), \tag{5.47}$$

with some positive constants α_0, α_1.

These assumptions are made in order to obtain complete theoretical results for the proposed subdomain and interface preconditioners.

Thus, below without loss of generality we assume that function $c(\mathbf{x})$ is piecewise constant and entries $k_{ij}(\mathbf{x})$, $i, j = 1, \ldots, d$, of the coefficient matrix $K(\mathbf{x}) = \{k_{ij}(\mathbf{x})\}_{ij=1}^{d}$ are constants in each subdomain.

Let $\mathcal{T}_{kh}$ be a triangular ($d = 2$) or tetrahedral ($d = 3$) partitioning of Ω_k [36], i.e. $\mathcal{T}_{kh} = \{\tau_i\}_{i=1}^{M_k}$, where τ_i is a triangle ($d = 2$) or a tetrahedron ($d = 3$) that is called a grid cell. We denote by $\mathcal{F}_{kh}$ the set of faces e of the grid cells $\tau \in \mathcal{T}_{kh}$, $k = 1, \ldots, m$. Note that the traces of grids $\mathcal{T}_{kh}$ and $\mathcal{T}_{lh}$ at the interface Γ_{kl}, generally speaking, do not coincide. These grids are called nonmatching grids.

In each subdomain Ω_k, $k = 1, \ldots, m$, we introduce the following Raviart-Thomas spaces (see Section 2.3.3):

$$\begin{aligned} \mathbf{V}_{kh} \equiv RT_{-1}^{0}(\mathcal{T}_{kh}) &= \left\{\mathbf{p} : \mathbf{p} \in (L^2(\Omega_k))^3, \mathbf{p}|_\tau \in RT_{-1}^{0}(\tau) \ \forall \tau \in \mathcal{T}_{kh}\right\}, \\ W_{kh} \equiv M_{-1}^{0}(\mathcal{T}_{kh}) &= \left\{v : v \in L^2(\Omega_k), v|_\tau = c_\tau \ \forall \tau \in \mathcal{T}_{kh}\right\}, \end{aligned} \tag{5.48}$$

and the spaces of Lagrange multipliers (see Section 2.4.2):

$$\mathcal{L}_{kh} \equiv M_{-1}^{0}(\mathcal{F}_{kh}) = \left\{\mu \in L^2(\mathcal{F}_{kh}) : \mu|_e = c_e \text{ for each } e \in \mathcal{F}_{kh}\right\}, \tag{5.49}$$

that is space $\mathcal{L}_h$ consists of piecewise smooth functions which are constant on each face e. Note that $\mathbf{v} \cdot \mathbf{n}_e$, $\mathbf{v} \in \mathbf{V}_{kh}$, is constant on each face of a grid cell τ.

Note that the well known hybrid-mixed formulation of (5.46) in the case of matching grids is defined by (see (2.38)): *find* $(\mathbf{q}_h^*, u_h^*, \lambda_h) \in \mathbf{V}_h \times W_h \times \mathcal{L}_h$ *such that*

$$\begin{array}{lcl} a(\mathbf{q}_h^*, \mathbf{p}_h) - b(u_h^*, \mathbf{p}_h) + l(\lambda_h, \mathbf{p}_h) & = & 0, \\ -b(v_h, \mathbf{q}_h^*) - (c \cdot u_h^*, v_h) & = & -f(v_h), \\ l(\mu_h, \mathbf{q}_h^*) & = & 0, \end{array} \tag{5.50}$$

for any $(\mathbf{p}_h, v_h, \mu_h) \in \mathbf{V}_h \times W_h \times \mathcal{L}_h$. Here

$$\mathbf{V}_h = \prod_{k=1}^m \mathbf{V}_{kh}, \qquad W_h = \prod_{k=1}^m W_{kh}, \qquad \mathcal{L}_h = \prod_{k=1}^m \mathcal{L}_{kh}, \tag{5.51}$$

and

$$\begin{array}{ll} a(\mathbf{q}, \mathbf{p}) = \sum\limits_{k=1}^m a_k(\mathbf{q}_k, \mathbf{p}_k), & a_k(\mathbf{q}_k, \mathbf{p}_k) = \sum\limits_{\tau \in \mathcal{T}_{kh}} \int\limits_\tau K^{-1} \mathbf{q}_k \cdot \mathbf{p}_k \, dx, \\ b(u, \mathbf{p}) = \sum\limits_{k=1}^m b_k(u_k, \mathbf{p}_k), & b_k(u_k, \mathbf{p}_k) = \sum\limits_{\tau \in \mathcal{T}_{kh}} \int\limits_\tau u_k \mathrm{div}\, \mathbf{p}_k \, dx, \\ l(\lambda, \mathbf{p}) = \sum\limits_{k=1}^m l_k(\lambda_k, \mathbf{p}_k), & l_k(\lambda_k, \mathbf{p}_k) = \sum\limits_{\tau \in \mathcal{T}_{kh}} \int\limits_{\partial\tau} \lambda_k \, (\mathbf{p}_k \cdot \mathbf{n}_\tau) \, ds, \\ f(v) = \sum\limits_{k=1}^m f_k(v_k), & f_k(v_k) = \int\limits_{\Omega_k} f \; v_k \, dx. \end{array} \tag{5.52}$$

In the case of nonmatching grids we have to impose additionally the continuity of the unknowns on the interfaces between subdomains. Using [28, 74], we introduce a four field formulation of problem (5.46). First, we define the spaces:

$$\Lambda = \prod_{k=1}^m \Lambda_k, \qquad \Lambda_k = \prod_{\substack{l=1 \\ |\Gamma_{kl}| \neq 0}}^m \Lambda_{kl}, \qquad \Phi = \prod_{1 \le l < k \le m} \Phi_{kl}, \tag{5.53}$$

where Λ_{kl} are the subspaces of $\mathcal{L}_{kh}$, $k = 1, \ldots, m$, defined by

$$\Lambda_{kl} = \{\mu \in \mathcal{L}_{kh} : \mu_e = 0, e \not\subseteq \Gamma_{kl}\}\,, \tag{5.54}$$

and Φ_{kl} is the dual space of Λ_{kl}. Following [74] we associate Φ_{kl} with one of the subdomains Ω_k or Ω_l. Note that in the case of matching grids, i.e. when $\mathcal{T}_{kh}|_{\Gamma_{kl}} = \mathcal{T}_{lh}|_{\Gamma_{kl}}$, we have $\Lambda_{kl} = \Lambda_{lk}$. However, here we include also the case $\Lambda_{kl} \neq \Lambda_{lk}$.

We also specify the sets $\Gamma_{k0} = \Omega_{kh} \cap \Gamma_0$, $|\Gamma_{k0}| \neq 0$, $k = 1, \ldots, m$. These grid sets are called the interfaces between the subdomains and the set Γ_0, where the Dirichlet boundary conditions are imposed. Following [72] we associate spaces Φ_{k0} with subdomains Ω_k.

Now we define new bilinear forms:

$$\begin{array}{ll} d(\varphi, \mu) = \sum\limits_{k=1}^m d_k(\varphi, \mu_k), & d_k(\varphi, \mu_k) = \sum\limits_{\substack{l=1 \\ |\Gamma_{kl}| \neq 0}}^m d_{kl}(\varphi_{kl}, \mu_k), \\ d_{kl}(\varphi_{kl}, \mu_k) = -\int\limits_{\Gamma_{kl}} \varphi_{kl} \; \mu_k \; ds. & \end{array} \tag{5.55}$$

Here $\mu_k \in \mathcal{L}_{kh}$, $\varphi_{kl} \in \Phi_{kl}$, and $\varphi_{kl} = -\varphi_{lk}$ for $l > k$, $k, l = 1, \ldots, m$.

The new hybrid-mixed formulation of (5.46) on nonmatching grids is defined as follows: *find* $(\mathbf{q}, u, \lambda, \varphi) \in \mathbf{V}_h \times W_h \times \mathcal{L}_h \times \Phi$ *such that*

$$\begin{array}{rlrcl} a(\mathbf{q},\mathbf{p}) & - \, b(u,\mathbf{p}) + \;\; l(\lambda,\mathbf{p}) & & = & 0, \\ -b(v,\mathbf{q}) & - \, (c \cdot u, v) & & = & -f(v), \\ l(\mu,\mathbf{q}) & & + \, d(\varphi,\mu) & = & 0, \\ & \qquad\qquad d(\psi,\lambda) & & = & 0, \end{array} \tag{5.56}$$

for any $(\mathbf{p}, v, \mu, \psi) \in \mathbf{V}_h \times W_h \times \mathcal{L}_h \times \Phi$.

It is easy to see that

$$l(\mu,\mathbf{q}) + d(\varphi,\mu) = \sum_{k=1}^{m} \left(l_k(\mu_k,\mathbf{q}_k) + \sum_{\substack{l=1 \\ |\Gamma_{kl}| \neq 0}}^{m} d_{kl}(\varphi_{kl},\mu_k) \right) \tag{5.57}$$

$$= \sum_{k=1}^{m} \left(\sum_{\tau \in \mathcal{T}_{kh}} \int\limits_{\partial\tau \setminus \Gamma_k} \mu_k \, (\mathbf{q}_k \cdot \mathbf{n}_\tau) \, ds + \sum_{\substack{l=1 \\ |\Gamma_{kl}| \neq 0}}^{m} \int\limits_{\Gamma_{kl}} \mu_k \, ((\mathbf{q}_k \cdot \mathbf{n}_k) - \varphi_{kl}) \, ds \right).$$

Note that the unknowns $(\mathbf{q}_k, \mathbf{u}_k, \lambda_k)$ in each subdomain Ω_k are connected with the unknowns in the other subdomains only through the interface elements φ_{kl}. Thus, the system of equations for $(\mathbf{q}_k, \mathbf{u}_k, \lambda_k)$ in each subdomain corresponds to a homogeneous Neumann problem. Using the results of Section 2.4 (see Lemma 2.8, Proposition 2.1, and Corollary 2.1) we can replace the hybrid-mixed formulation for the triple $(\mathbf{q}_k, \mathbf{u}_k, \lambda_k)$ in each subdomain with the nonconforming formulation for the unknown $u_k \in V_{kh}$, where $V_{kh} \equiv V_h(\Omega_k)$ is the P_1 nonconforming space defined on the domain Ω_k.

After defining the nonconforming space $V_h = \prod\limits_{k=1}^{m} V_{kh}$ problem (5.56) is replaced by the following problem: *find* $(u, \varphi) \in V_h \times \Phi$ *such that*

$$\begin{array}{rcl} a_h(u,v) + \tilde{d}(\varphi,v) & = & g(v), \\ \tilde{d}(\psi,u) & = & 0, \end{array} \tag{5.58}$$

for any $(v, \psi) \in V_h \times \Phi$.

Here the bilinear form $a_h(\cdot,\cdot)$ is given (similarly to (2.50)) by:

$$a_h(u,v) = \sum_{k=1}^{m} a_{kh}(u_k,v_k), \qquad a_{kh}(u_k,v_k) = \sum_{\tau \in \mathcal{T}_{kh}} (K\nabla u_k, \nabla v_k)_\tau + (c \cdot u_k, v_k)_\tau, \tag{5.59}$$

and the bilinear form $\tilde{d}(\cdot,\cdot)$ is given on $V_h \times \Phi$ by:

$$\tilde{d}(\psi,v) = d(\psi, P_h v), \qquad \forall \psi \in \Phi, \quad \forall v \in V_h, \tag{5.60}$$

where $d(\cdot,\cdot)$ is defined by (5.55). Here the projection operator $P_h : V_h \to \mathcal{L}_h$ is defined by (2.52).

In the operator form the finite element problem (5.58) can be represented as follows:

$$\begin{aligned} A_h u_h + D_h^T \varphi_h &= g_h, \\ D_h u_h &= 0, \end{aligned} \tag{5.61}$$

or, equivalently,

$$\begin{aligned} A_{kh} u_{kh} + D_{kh}^T \varphi_{kh} &= g_{kh}, \qquad k = 1, \ldots, m \\ \sum_{k=1}^{m} D_{kh} u_{kh} &= 0, \end{aligned} \tag{5.62}$$

where

$$D_{kh}^T \varphi_{kh} = \sum_{\substack{l=1, l \neq k \\ |\Gamma_{kl}| \neq 0}}^{m} D_{klh}^T \varphi_{klh} \tag{5.63}$$

and $\varphi_{klh} = -\varphi_{lkh}$, $k < l$.

Remark 5.6 When the grids are conforming on the interfaces, i.e. $\Lambda_{klh} = \Lambda_{lkh}$ the finite element problem (5.62) can be reduced to the form

$$\tilde{A}_h \tilde{u}_h = \tilde{g}_h, \tag{5.64}$$

where $\tilde{A}_h$ is obtained by assembling operators A_{kh} and finite element functions $\tilde{u}_h$ and $\tilde{g}_h$ are obtained by assembling finite element functions u_{kh} and g_{kh} under a continuity condition on the interfaces. Thus, finite element problem (5.62) can also be obtained from standard P_1-nonconforming finite element approximations of problem (2.50) when the grids match on the interfaces. It follows that the methods to be considered can also be automatically applied to solving finite element systems arising from discretization of (2.50).

We divide the degrees of freedom in each subdomain Ω_k into two groups so that the first group contains the degrees of freedom from the inner part of Ω_k and the second group contains the degrees of freedom from $\partial\Omega_k$. If the degrees of freedom of the first group carry a subscript I and those of the second group carry a subscript Γ, system (5.62) can be represented in a more detailed form:

$$\begin{aligned} A_{I,kh} \mathbf{u}_{I,kh} + A_{I\Gamma,kh} \mathbf{u}_{\Gamma,kh} &= \mathbf{g}_{I,kh}, \\ A_{\Gamma I,kh} \mathbf{u}_{I,kh} + A_{\Gamma,kh} \mathbf{u}_{\Gamma,kh} + D_{\Gamma,kh}^T \varphi_{kh} &= \mathbf{g}_{\Gamma,kh}, \qquad k = 1, \ldots, m \\ \sum_{k=1}^{m} D_{\Gamma,kh} \mathbf{u}_{\Gamma,kh} &= 0. \end{aligned} \tag{5.65}$$

Here

$$D_{\Gamma,kh}^T \varphi_{kh} = \sum_{\substack{l=1, l \neq k \\ |\Gamma_{kl}| \neq 0}}^{m} D_{\Gamma,klh}^T \, \varphi_{klh}$$

and again $\varphi_{klh} = -\varphi_{lkh}$, $k < l$.

The finite element problem (5.61) results in an algebraic system

$$\mathcal{A}\mathbf{x} = \begin{bmatrix} A & D^T \\ D & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varphi} \end{bmatrix} = \begin{bmatrix} A_1 & & \mathbf{0} & D_1^T \\ & \ddots & & \vdots \\ \mathbf{0} & & A_m & D_m^T \\ D_1 & \ldots & D_m & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_m \\ \boldsymbol{\varphi} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_m \\ \mathbf{0} \end{bmatrix}, \tag{5.66}$$

where $\mathbf{x} = \left[\mathbf{u}^T \boldsymbol{\varphi}^T\right]^T$, $\mathbf{u} = \left[\mathbf{u}_1^T \dots \mathbf{u}_m^T\right]^T$, and

$$A_k = \begin{bmatrix} A_{I,k} & A_{I\Gamma,k} \\ A_{\Gamma I,k} & A_{\Gamma,k} \end{bmatrix}, \quad D_k^T = \begin{bmatrix} \mathbf{0} \\ D_{\Gamma,k}^T \end{bmatrix}, \quad \mathbf{u}_k = \begin{bmatrix} \mathbf{u}_{I,k} \\ \mathbf{u}_{\Gamma,k} \end{bmatrix}, \quad \mathbf{g}_k = \begin{bmatrix} \mathbf{g}_{I,k} \\ \mathbf{g}_{\Gamma,k} \end{bmatrix}, \tag{5.67}$$

$$k = 1, \dots, m.$$

Note that all matrices A_k are at least positive semidefinite.

Proposition 5.2 *Assume that $\Gamma_0 \neq \emptyset$. Then, matrix $\mathcal{A}$ is nonsingular.*

Proof: We consider the equation

$$\begin{bmatrix} A & D^T \\ D & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varphi} \end{bmatrix} = \mathbf{0}. \tag{5.68}$$

If we multiply it scalarly by the vector $\left[\mathbf{u}^T \boldsymbol{\varphi}^T\right]^T$ and take into account the condition $D\mathbf{u} = \mathbf{0}$, we get

$$(A\mathbf{u}, \mathbf{u}) \equiv \sum_{k=1}^{m} (A_k \mathbf{u}_k, \mathbf{u}_k) = 0 \tag{5.69}$$

and, hence $(A_k \mathbf{u}_k, \mathbf{u}_k) = 0$, $k = 1, \dots, m$.

Assume first, that all matrices A_k are positive definite. Then $\mathbf{u} = \mathbf{0}$, and $\boldsymbol{\varphi} = 0$ according to equation $D^T \boldsymbol{\varphi} = \mathbf{0}$. Thus, matrix $\mathcal{A}$ is nonsingular.

Now we consider the case in which matrix A is singular. That is, there is at least one block A_k which is positive semidefinite. Equality (5.69) in this case holds for the nonzero vector $\mathbf{u}$ if and only if at least one of its subvectors $\mathbf{u}_k \in \operatorname{Ker} A_k$, $k \in [1, m]$. This implies that each of the components of $\mathbf{u}_k$ is the same nonzero constant, that is $\mathbf{u}_k = \alpha \cdot \mathbf{e}_k$, where $\mathbf{e}_k^T = (1, \dots, 1)$. On the other hand, equation $D\mathbf{u} = \mathbf{0}$ in the finite element representation implies that

$$\int_{\Gamma_{kl}} \varphi_{klh} \, (P_h u_{kh} - P_h u_{lh}) \, ds = 0, \qquad \forall \varphi_{klh} \in \Phi_{klh}, \quad l < k, \quad k = 1, \dots, m. \tag{5.70}$$

Hence, the vector $\mathbf{u}$ can be chosen in such a way that $\mathbf{u}_k = \mathbf{e}_k$ for all $k = 1, \dots, m$, i.e. each of the components of $\mathbf{u}_k$ is unity. Since we assumed that $\Gamma_0 \neq \emptyset$ then there is at least one subdomain Ω_l for which block A_l is positive definite. It means that $(A_l \mathbf{u}_l, \mathbf{u}_l) > 0$ which contradicts (5.69). Thus, $\mathbf{u} = \mathbf{0}$. Then $\boldsymbol{\varphi} = \mathbf{0}$ according to equation $D^T \boldsymbol{\varphi} = \mathbf{0}$, and hence (5.68) may hold only for $\mathbf{u}$ and $\boldsymbol{\varphi}$ equal to zero. This implies that det $\mathcal{A} \neq 0$. □

5.3.2 Design and analysis of the preconditioner

Based on the results of Section 3.2 it is sufficient to construct a spectrally equivalent preconditioner $\mathcal{B}$ for matrix $\mathcal{A}$ in the form of the block diagonal matrix:

$$\mathcal{B} = \begin{bmatrix} B_A & \\ & B_\varphi \end{bmatrix}, \tag{5.71}$$

where

$$B_A = \begin{bmatrix} B_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & B_m \end{bmatrix}, \tag{5.72}$$

blocks B_k, $k = 1, \dots, m$, are preconditioners for matrices A_k, $k = 1, \dots, m$, respectively, and block B_φ is the preconditioner for the Schur complement $S_\varphi = DA^+D^T$. Here matrix A^+ denotes the pseudo-inverse of matrix A [59].

To develop such a preconditioner in each subdomain Ω_k, $k = 1, \dots, m$, we construct a coordinate system having as axes the eigenvectors of coefficient matrix K_k (see Section 4.5.1 for details). In these coordinates matrix A_k is a P_1-nonconforming approximation of the problem:

$$\begin{aligned} -\sum_{i=1}^{d} k_i^{(k)} \frac{\partial^2 u}{\partial x_i^2} + c^{(k)} u &= f, \qquad \text{in } \Omega_k, \\ \frac{\partial u}{\partial n} &= 0, \qquad \text{on } \partial\Omega_k. \end{aligned} \tag{5.73}$$

We also construct rectangles Π_k, $k = 1, \dots, m$, in the local coordinate systems connected with subdomains Ω_k in such a way that diam $(\Pi_k) \approx$ diam (Ω_k), Π_k contains $\bar{\Omega}_k$, and the boundaries of Π_k are parallel to the corresponding coordinate axis, $k = 1, \dots, m$. Define in each Π_k the uniform mesh $\mathcal{T}_{\Pi,kh}$, $k = 1, \dots, m$.

Assumption 5.4 *Assume that mesh $\mathcal{T}_{kh}$ is a trace on Ω_k of a regular mesh $\mathcal{T}_{\Pi,kh}$ constructed in the embedding rectangle Π_k, $k = 1, \dots, m$.*

Also we assume that we can construct such an extended grid domain Π_k for each subdomain Ω_k, $k = 1, \dots, m$, and that the number of degrees of freedom in Π_k is proportional to N_k, which denotes the number of degrees of freedom in Ω_k. Since the subdomains were assumed to be quasiregular, the number of degrees of freedom on the boundary of subdomain Ω_k denoted by $N_{\Gamma,k}$ is of order $O(N_k^{1/2})$, $k = 1, \dots, m$. We also assume that mesh-size parameter h is of order $h \sim N^{1/d}$, where N is the order of matrix A.

5.3.2.1 Preconditioner for matrix A_k

This is the simplest task. Using previous assumptions we can apply the fictitious components method described in Section 4.5.

To solve the problem in the extended subdomains Π_k we use substructuring methods which are described in Chapter IV.

Thus, the arithmetic cost of product $B_k^+ \mathbf{u}_k$ is $O(N_k)$ in the two-dimensional case and either $O(N_k)$ (if we use the method described in Section 4.3) or $O(N_k \ln(N_k))$ (if we use the method described in Section 4.4) in the three-dimensional case.

5.3.2.2 Preconditioner for matrix S_φ

According to block partitionings (5.65) we have

$$S_\varphi = DA^+D^T = \sum_{k=1}^{m} D_k A_k^+ D_k^T = \sum_{k=1}^{m} D_{\Gamma,k} S_{\Gamma,k}^+ D_{\Gamma,k}^T, \tag{5.74}$$

where $N_{\Gamma,k} \times N_{\Gamma,k}$ matrices $S_{\Gamma,k}$ are Schur complements corresponding to the unknowns on boundaries Γ_k:

$$S_{\Gamma,k} = A_{\Gamma,k} - A_{\Gamma I,k} A_{I,k}^{-1} A_{I\Gamma,k}, \qquad k = 1, \ldots, m. \tag{5.75}$$

Taking into account Remark 5.4 and the results of [72] we can develop a preconditioner for S_φ by constructing for each subdomain Ω_k, $k = 1, \ldots, m$, two matrices:

(1) A diagonal $N_{\Gamma,k} \times N_{\Gamma,k}$ matrix $\tilde{B}_{\Gamma,k}$ such that for each nonzero vector $\mathbf{u}_{\Gamma,k} \in \mathbb{R}^{N_{\Gamma,k}} \setminus$ Ker $(S_{\Gamma,k})$ we have inequalities

$$\alpha_0 \cdot (h/r)\ (\tilde{B}_{\Gamma,k}\mathbf{u}_{\Gamma,k}, \mathbf{u}_{\Gamma,k}) \le (S_{\Gamma,k}\mathbf{u}_{\Gamma,k}, \mathbf{u}_{\Gamma,k}) \le \alpha_1\,(\tilde{B}_{\Gamma,k}\mathbf{u}_{\Gamma,k}, \mathbf{u}_{\Gamma,k}), \tag{5.76}$$

where constants α_0 and α_1 do not depend on mesh-size parameter h and the coefficients of the problem.

(2) A matrix $\tilde{S}_{\Gamma,k}$ which is spectrally equivalent to matrix $S_{\Gamma,k}$. The main requirement is that the matrix-vector multiplication $\tilde{S}_{\Gamma,k}\ \mathbf{u}_{\Gamma,k}$ is easier to compute than the expression $S_{\Gamma,k}\ \mathbf{u}_{\Gamma,k}$.

Having constructed these two matrices for each subdomain Ω_k we define the matrices

$$\tilde{S}_\varphi = \sum_{k=1}^{m} D_{\Gamma,k} \tilde{S}_{\Gamma,k}^{+} D_{\Gamma,k}^{T}, \qquad \tilde{B}_\varphi = \sum_{k=1}^{m} D_{\Gamma,k} \tilde{B}_{\Gamma,k}^{-1} D_{\Gamma,k}^{T}. \tag{5.77}$$

Finally, the preconditioner for matrix S_φ will be defined in the form of a matrix polynomial

$$B_\varphi^{+} = \left\{ I_\varphi - \prod_{l=1}^{L} \left(I_\varphi - \beta_l \tilde{B}_\varphi^{-1} \tilde{S}_\varphi \right) \right\} \tilde{S}_\varphi^{+}. \tag{5.78}$$

5.3.2.2.1 Construction of a diagonal matrix $\tilde{B}_{\Gamma,k}$. For simplicity we consider below one domain Ω_k and skip the index "k" when no ambiguity occurs.

First, we consider a model problem

$$\begin{aligned} -k_x \frac{\partial^2 u}{\partial x^2} - k_y \frac{\partial^2 u}{\partial y^2} + c_0 u = f, &\qquad \text{in } \Omega, \\ u = 0, &\qquad \text{on } \Gamma_0, \\ \frac{\partial u}{\partial n} = 0, &\qquad \text{on } \Gamma_1 = \partial\Omega \setminus \Gamma_0, \end{aligned} \tag{5.79}$$

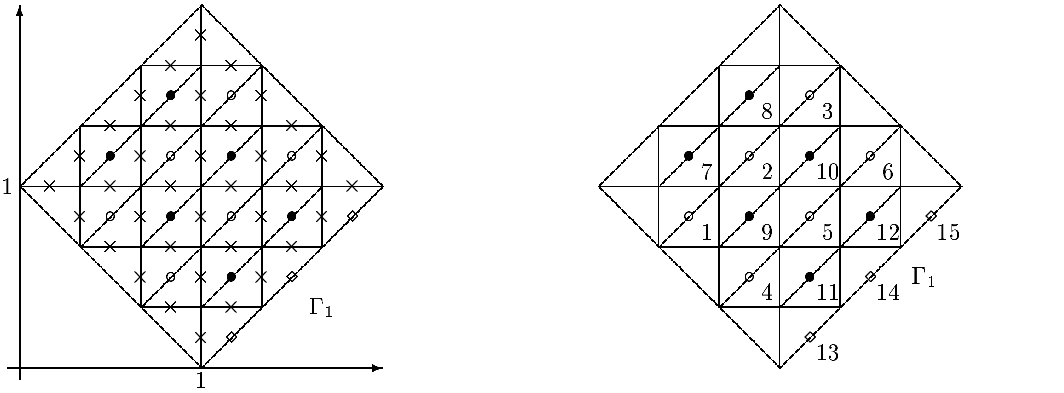
where Ω is the square:

$$\Omega = \{(x, y) :\ 1 \le x + y \le 3,\ -1 \le y - x \le 1\},$$

and the Neumann boundary is $\Gamma_1 = \{(x, y) :\ y - x = 1\}$ (see Figure 5.5a). Assume that coefficients k_x, k_y, c_0 are constants in Ω.

Let $\mathcal{T}_h$ be a regular triangulation of Ω with mesh-size h (see, e.g., Fig. 5.5a). Following the construction in Section 4.2 define a P_1-nonconforming finite element space and introduce a nodal basis in this space. Thus we obtain the following matrix representation of the problem (see Section 4.2 for details):

$$A\mathbf{u} = \mathbf{g}, \tag{5.80}$$



(a) Triangulation of the domain Ω.

(b) The degrees of freedom of the reduced problem.

Figure 5.5: *Degrees of freedom of a model problem.*

where A is an $N \times N$ symmetric at least positive semidefinite matrix and $\mathbf{u}, \mathbf{g} \in \mathbb{R}^N$. Here $N \sim h^{-2}$. Denote also the number of degrees of freedom on Γ_1 by N_Γ. Obviously, $N_\Gamma \sim N^{1/2}$.

Eliminating unknowns $vx_{i,j}$ and $vy_{i,j}$ (marked "×" in Fig. 5.5a) we get the problem

$$\hat{A}\hat{\mathbf{u}} = \begin{bmatrix} \hat{A}_I & \hat{A}_{I\Gamma} \\ \hat{A}_{\Gamma I} & \hat{A}_\Gamma \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_I \\ \mathbf{u}_\Gamma \end{bmatrix} = \begin{bmatrix} \mathbf{g}_I \\ \mathbf{g}_\Gamma \end{bmatrix}. \tag{5.81}$$

Here subvector $\mathbf{u}_\Gamma$ of vector $\hat{\mathbf{u}}$ corresponds to the unknowns on boundary Γ_1 (nodes marked "◇" in Fig. 5.5b) and subvector $\mathbf{u}_I$ corresponds to the unknowns in the domain (nodes marked "○" and "•" in Fig. 5.5b). For more details see Section 4.2.

Now we partition internal nodes into two groups: the first group consists of the nodes marked by "○" in Figure 5.5b and the second group consists of the nodes marked by "•". We enumerate each group of nodes in lexicographic order, first in direction $\mathbf{n}_1 = (1, 1)$ and then in direction $\mathbf{n}_2 = (1, -1)$ (see example in Figure 5.5b). Then problem (5.81) can be represented in another block form:

$$\hat{A}\hat{\mathbf{u}} = \begin{bmatrix} \hat{A}_1 & \hat{A}_{12} & \mathbf{0} \\ \hat{A}_{21} & \hat{A}_2 & \hat{A}_{2\Gamma} \\ \mathbf{0} & \hat{A}_{\Gamma 2} & \hat{A}_\Gamma \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_\Gamma \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_\Gamma \end{bmatrix}. \tag{5.82}$$

Here vectors $\mathbf{u}_1$ and $\mathbf{g}_1$ correspond to nodes of the first group, and vectors $\mathbf{u}_2$ and $\mathbf{g}_2$ correspond to nodes of the second group, respectively.

For the model problem defined on the mesh domain shown in Figure 5.5 these matrices, for example, are:

$$A_1 = (2a_x + 2a_y + b)\, I_1, \qquad A_2 = (2a_x + 2a_y + b)\, I_2, \qquad A_\Gamma = (a_x + a_y + b/2)\, I_\Gamma, \tag{5.83}$$

$$A_{12} = A_{21}^T = (-1) \begin{bmatrix} a_y & 0 & a_x & 0 & 0 & 0 \\ a_x & a_y & a_y & a_x & 0 & 0 \\ 0 & a_x & 0 & a_y & 0 & 0 \\ 0 & 0 & a_y & 0 & a_x & 0 \\ 0 & 0 & a_x & a_y & a_y & a_x \\ 0 & 0 & 0 & a_x & 0 & a_y \end{bmatrix}, \quad A_{2\Gamma} = A_{\Gamma 2}^T = (-1) \begin{bmatrix} a_y & a_x & 0 \\ 0 & a_y & a_x \end{bmatrix},$$

where $I_1, I_2 \in \mathbb{R}^{6\times 6}$, and $I_\Gamma \in \mathbb{R}^{3\times 3}$ are identity matrices, and coefficients a_x, a_y, b are defined by (4.23).

After eliminating the unknowns of the second group system (5.82) becomes:

$$\tilde{A}\tilde{\mathbf{u}} = \begin{bmatrix} \hat{A}_1 - \hat{A}_{12}\hat{A}_2^{-1}\hat{A}_{21} & -\hat{A}_{12}\hat{A}_2^{-1}\hat{A}_{2\Gamma} \\ -\hat{A}_{\Gamma 2}\hat{A}_2^{-1}\hat{A}_{21} & \hat{A}_\Gamma - \hat{A}_{\Gamma 2}\hat{A}_2^{-1}\hat{A}_{2\Gamma} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_\Gamma \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{g}}_1 \\ \tilde{\mathbf{g}}_\Gamma \end{bmatrix}. \tag{5.84}$$

To understand the structure of matrix $\tilde{A}$ we consider a little square domain C containing only one node of the second group (see Figure 5.6). The domain is bounded by the lines connecting the nodes 1, 2, 3, and 4.

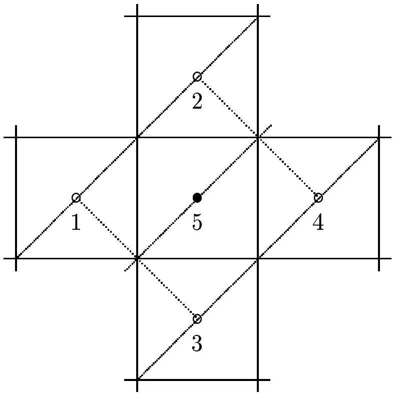


Figure 5.6: *Grid subdomain C of the domain Ω.*

Remark 5.7 The grid domain Ω can be viewed as a union of such subdomains $\Omega = \cup_{i=1}^{M} C_i$.

The submatrix corresponding to a homogeneous Neumann problem on subdomain C which lies inside Ω has the form:

$$\hat{A}_C = \left[\begin{array}{cccc|c} a_x + b/4 & & & & -a_x \\ & a_y + b/4 & & & -a_y \\ & & a_y + b/4 & & -a_y \\ & & & a_x + b/4 & -a_x \\ \hline -a_x & -a_y & -a_y & -a_x & 2a_x + 2a_y + b \end{array}\right]. \tag{5.85}$$

Here the fifth line corresponds to the unknown in node 5. After eliminating unknown 5 we obtain the matrix

$$\tilde{A}_C = \frac{b}{4}\, I + \begin{bmatrix} a_x & & & \\ & a_y & & \\ & & a_y & \\ & & & a_x \end{bmatrix} - \frac{1}{2a_x + 2a_y + b} \begin{bmatrix} a_x^2 & a_x a_y & a_x a_y & a_x^2 \\ a_x a_y & a_y^2 & a_y^2 & a_x a_y \\ a_x a_y & a_y^2 & a_y^2 & a_x a_y \\ a_x^2 & a_x a_y & a_x a_y & a_x^2 \end{bmatrix} = \tag{5.86}$$

$$= \frac{1}{2a_x + 2a_y + b} \left\{ \frac{b}{4} \begin{bmatrix} 6a_x + 2a_y + b & & & \\ & 2a_x + 6a_y + b & & \\ & & 2a_x + 6a_y + b & \\ & & & 6a_x + 2a_y + b \end{bmatrix} \right.$$

$$+a_x a_y \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix} + a_x^2 \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} + a_y^2 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Bigg\}.$$

It is easy to see that after assembling all these local matrices over all squares C_i, $i = 1, \ldots, M$, we obtain matrix $\tilde{A}$ of the following form:

$$\tilde{A} = \frac{1}{2a_x + 2a_y + b} \left\{ b \cdot D + a_x a_y \cdot A_{xy} + a_x^2 \cdot A_{xx} + a_y^2 \cdot A_{yy} \right\}, \tag{5.87}$$

where matrix D is a diagonal one whose diagonal entries are linear combinations of a_x, a_y, and b; the entries of matrices A_{xy}, A_{xx}, and A_{yy} do not depend on the coefficients of the problem. Matrix A_{xy} is a separable matrix in the numbering introduced in this section (see Fig. 5.5). Matrices A_{xx} and A_{yy} have rather complicated structure in this numbering but they can be simplified if we introduce another numbering of the nodes. First, let us number the nodes marked by "$\circ$" in Figure 5.5 in the x-direction, and then in the y-direction. Then matrix A_{xx} is block diagonal, where each block corresponds to the unknowns on one grid line in the x-direction and is an approximation of the one-dimensional Laplace operator in x: $L\,u \equiv -u_{xx}$. If we number nodes, first, in the y-direction then in the x-direction, matrix A_{yy} will be block diagonal where each block corresponds to the unknowns on one grid line in the y-direction and is an approximation of the one-dimensional Laplace operator in y: $L\,u \equiv -u_{yy}$.

Now, using the same arguments as in the proof of Lemma 5.1 we can show that for any vector $\mathbf{u}_\Gamma \notin \text{Ker}\,(S_\Gamma)$, $\mathbf{u}_\Gamma \neq 0$, we have

$$2\frac{(S_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma)}{(A_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma)} \geq \frac{\min\limits_{\hat{\mathbf{u}}|_\Gamma = \mathbf{u}_\Gamma} (\hat{A}\mathbf{u}, \mathbf{u})}{(\hat{A}_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma)} \geq \frac{\min\limits_{\tilde{\mathbf{u}}|_\Gamma = \mathbf{u}_\Gamma} (\tilde{A}\tilde{\mathbf{u}}, \tilde{\mathbf{u}})}{(\hat{A}_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma)} = \tag{5.88}$$

$$= \frac{\min\limits_{\tilde{\mathbf{u}}|_\Gamma = \mathbf{u}_\Gamma} \dfrac{1}{2a_x + 2a_y + b} \Big((b \cdot D + a_x a_y \cdot A_{xy} + a_x^2 \cdot A_{xx} + a_y^2 \cdot A_{yy}) \tilde{\mathbf{u}}, \tilde{\mathbf{u}} \Big)}{(2a_x + 2a_y + b)\ (\mathbf{u}_\Gamma, \mathbf{u}_\Gamma)}$$

$$= \frac{\min\limits_{\tilde{\mathbf{u}}|_\Gamma = \mathbf{u}_\Gamma} \Big((b \cdot D + a_x a_y \cdot A_{xy} + a_x^2 \cdot A_{xx} + a_y^2 \cdot A_{yy}) \tilde{\mathbf{u}}, \tilde{\mathbf{u}} \Big)}{(4a_x^2 + 4a_y^2 + 8a_x a_y + b(4a_x + 4a_y + b))\ (\mathbf{u}_\Gamma, \mathbf{u}_\Gamma)}$$

$$\geq \frac{1}{4} \min_{\tilde{\mathbf{u}}|_\Gamma = \mathbf{u}_\Gamma} \left\{ \frac{(D\tilde{\mathbf{u}}, \tilde{\mathbf{u}})}{(a_x + a_y + b/4)(\mathbf{u}_\Gamma, \mathbf{u}_\Gamma)};\ \frac{(A_{xy}\tilde{\mathbf{u}}, \tilde{\mathbf{u}})}{2(\mathbf{u}_\Gamma, \mathbf{u}_\Gamma)};\ \frac{(A_{xx}\tilde{\mathbf{u}}, \tilde{\mathbf{u}})}{(\mathbf{u}_\Gamma, \mathbf{u}_\Gamma)};\ \frac{(A_{yy}\tilde{\mathbf{u}}, \tilde{\mathbf{u}})}{(\mathbf{u}_\Gamma, \mathbf{u}_\Gamma)}; \right\}.$$

Direct calculations show that for any vector $\mathbf{u}_\Gamma \notin \text{Ker}\,(S_\Gamma)$, $\mathbf{u}_\Gamma \neq 0$, we have an inequality

$$\frac{(S_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma)}{(A_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma)} \geq C \cdot N^{-1/2}, \tag{5.89}$$

where constant C does not depend on the coefficients of the problem.

Taking into account that

$$(S_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \leq (A_\Gamma \mathbf{u}_\Gamma, \mathbf{u}_\Gamma)$$

for any $\mathbf{u}_\Gamma \in \mathbb{R}^{N_\Gamma}$, we show that Lemma 5.1 is valid for the model problem (5.79).

Now we consider a general problem in the polygonal domain Ω with homogeneous Neumann boundary conditions on the entire boundary $\Gamma = \partial\Omega$:

$$\begin{aligned} -k_x \frac{\partial^2 u}{\partial x^2} - k_y \frac{\partial^2 u}{\partial y^2} + c_0 u = f, &\qquad \text{in } \Omega, \\ \frac{\partial u}{\partial n} = 0, &\qquad \text{on } \partial\Omega, \end{aligned} \tag{5.90}$$

To give the P_1-nonconforming approximation to this problem we introduce a uniform triangular mesh $\mathcal{T}_{h,\Omega}$ in the domain Ω and define nonconforming finite element space $V_h(\Omega)$ described in Section 4.2. Next, we define the bilinear form corresponding to the operator of (5.90) on $V_h(\Omega)$ by

$$a_\Omega^h(u,v) = \sum_{\tau \in \mathcal{T}_{h,\Omega}} \int_\tau (k_1 u_\xi v_\xi + k_2 u_\nu v_\nu + c_0\ uv)\ d\xi\ d\nu, \qquad \forall\ u, v \in V_h(\Omega). \tag{5.91}$$

Once a nodal basis $\{\varphi_i(\mathbf{x})\}_{i=1}^N$ for $V_h(\Omega)$ has been chosen, then the bilinear form (5.91) defines the symmetric positive semidefinite (positive definite if $c_0 > 0$) $N \times N$ matrix A_Ω by

$$(A_\Omega \mathbf{u}, \mathbf{v}) = a_\Omega^h(u,v), \qquad u, v \in V_h(\Omega). \tag{5.92}$$

Here $\mathbf{u}$, $\mathbf{v} \in \mathrm{I\!R}^N$ are the vector representations of functions u, v.

Let the number of the unknowns on boundary Γ be N_Γ. Note that by previous assumptions $N_\Gamma \sim N^{1/2}$. Denote by $\mathbf{u}_I$ and $\mathbf{u}_\Gamma$ the vectors of the unknowns in the interior of Ω and on boundary Γ, respectively. Then matrix A_Ω is represented in the form:

$$A_\Omega = \begin{bmatrix} A_I & A_{I\Gamma} \\ A_{\Gamma I} & A_\Gamma^{(\Omega)} \end{bmatrix}. \tag{5.93}$$

The Schur complement corresponding to the unknowns on boundary $\Gamma = \partial\Omega$ we denote by

$$S_\Gamma^{(\Omega)} = A_\Gamma^{(\Omega)} - A_{\Gamma I} A_I^{-1} A_{I\Gamma}. \tag{5.94}$$

Using the results of Subsection 5.2.2.3 and the considerations outlined above we have the following lemma.

Lemma 5.2 *For any vector* $\mathbf{v}_\Gamma \in \mathrm{I\!R}^{N_\Gamma}$ *such that* $\mathbf{v}_\Gamma \perp \mathrm{Ker}\ (S_\Gamma^{(\Omega)})$ *the following inequalities hold true:*

$$\alpha \cdot h \cdot (A_\Gamma^{(\Omega)} \mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \le (S_\Gamma^{(\Omega)} \mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \le (A_\Gamma^{(\Omega)} \mathbf{u}_\Gamma, \mathbf{u}_\Gamma), \tag{5.95}$$

where constant α *does not depend on mesh-size parameter* $h \sim N^{-1/2}$ *or the coefficients of the problem.*

Thus, we can choose as a diagonal matrix $\tilde{B}_{\Gamma,k}$ in (5.76) diagonal matrix $A_{\Gamma,k}$, $k = 1, \ldots, m$.

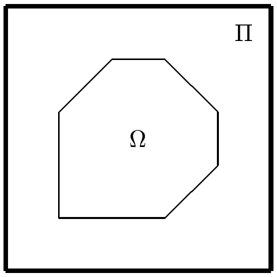


Figure 5.7: *Polygonal domain* Ω *embedded in rectangle* Π.

5.3.2.2.2 Construction of matrix $\tilde{S}_{\Gamma,k}$. The construction of matrix $\tilde{S}_\Gamma$ which is spectrally equivalent to matrix $S_\Gamma^{(\Omega)}$ is based on Extension Theorem 4.8 from Section 4.5.2. Consider domain Ω embedded in rectangle Π (see Figure 5.7). We denote a uniform triangular mesh in Π by $\mathcal{T}_{h,\Pi}$ and its trace in domain Ω by $\mathcal{T}_{h,\Omega}$.

Along with the bilinear form (5.91) we define a bilinear form in rectangle Π by

$$a_\Pi^h(u,v) = \sum_{\tau\in\mathcal{T}_{h,\Pi}} \int_\tau (k_1 u_\xi v_\xi + k_2 u_\nu v_\nu + c_0\ uv)\ d\xi\ d\nu, \qquad \forall\ u,v \in V_h(\Pi). \tag{5.96}$$

Let M be the dimension of $V_h(\Pi)$. Then the symmetric positive semidefinite (positive definite if $c_0 > 0$) $M \times M$ matrix A_Π is defined as follows:

$$(A_\Pi \mathbf{u}, \mathbf{v}) = a_\Pi^h(u,v), \qquad u,v \in V_h(\Pi). \tag{5.97}$$

Here $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^M$ are vector representations of functions u, v corresponding to the nodal basis $\{\varphi_i(\mathbf{x})\}_{i=1}^M$ of the space $V_h(\Pi)$.

Again, as in Section 4.5.1, we partition all the degrees of freedom in Π into three groups:

1. The first group consists of the unknowns corresponding to the degrees of freedom in $\Omega \setminus \Gamma$.

2. The second group consists of the unknowns on boundary Γ of domain Ω.

3. Finally, we enumerate the unknowns corresponding to the degrees of freedom in $\Pi \setminus \bar{\Omega}$.

This partition induces the following block representations of matrices A_Ω and A_Π:

$$A_\Omega = \begin{bmatrix} A_I & A_{I\Gamma} \\ A_{\Gamma I} & A_\Gamma^{(\Omega)} \end{bmatrix}, \qquad A_\Pi = \begin{bmatrix} A_I & A_{I\Gamma} & \mathbf{0} \\ A_{\Gamma I} & A_\Gamma^{(\Pi)} & A_{\Gamma O} \\ \mathbf{0} & A_{O\Gamma} & A_O \end{bmatrix}, \tag{5.98}$$

where blocks A_I, $A_\Gamma^{(*)}$, and A_O correspond to the unknowns of the first, second, and third groups, respectively.

We note that matrix $A_\Gamma^{(\Pi)}$ can be represented as the sum $A_\Gamma^{(\Pi)} = A_\Gamma^{(\Omega)} + A_\Gamma^{(\Pi\setminus\Omega)}$, and the matrix

$$\begin{bmatrix} A_\Gamma^{(\Pi\setminus\Omega)} & A_{\Gamma O} \\ A_{O\Gamma} & A_O \end{bmatrix} \tag{5.99}$$

corresponds to the nonconforming discretization of equation (5.90) in domain $\Pi \setminus \Omega$ with homogeneous Neumann boundary conditions.

For matrix A_Π we introduce the Schur complement corresponding to the unknowns on Γ:

$$S_\Gamma^{(\Pi)} = A_\Gamma^{(\Pi)} - A_{\Gamma I}A_I^{-1}A_{I\Gamma} - A_{\Gamma O}A_O^{-1}A_{O\Gamma} = S_\Gamma^{(\Omega)} + A_\Gamma^{(\Pi\setminus\Omega)} - A_{\Gamma O}A_O^{-1}A_{O\Gamma}. \tag{5.100}$$

The following lemma gives the main result of this section:

Lemma 5.3 *Matrices $S_\Gamma^{(\Omega)}$ and $S_\Gamma^{(\Pi)}$ are spectrally equivalent, that is there exists a constant α independent of mesh-size parameter h and coefficients of the problem, k_x, k_y, and c_0, such that for any vector $\mathbf{v}_\Gamma \in \mathrm{I\!R}^{N_\Gamma}$ the following inequalities hold true:*

$$\alpha \cdot (S_\Gamma^{(\Pi)}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \le (S_\Gamma^{(\Omega)}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma) \le (S_\Gamma^{(\Pi)}\mathbf{u}_\Gamma, \mathbf{u}_\Gamma). \tag{5.101}$$

Proof: Since matrix (5.99) is at least positive semidefinite (positive definite if $c_0 > 0$) then the matrix

$$S_\Gamma^{(\Pi\setminus\Omega)} = A_\Gamma^{(\Pi\setminus\Omega)} - A_{\Gamma O}A_O^{-1}A_{O\Gamma}$$

is also at least positive semidefinite. Thus, the right inequality in (5.101) follows from representation (5.100).

By Proposition 4.8 and Remark 4.13 for seminorms, for any function $v_\Omega^h \in V_h(\Omega)$ there exists a function $v_\Pi^h \in V_h(\Pi)$ such that it coincides with v_Ω^h in Ω and the following inequality holds true:

$$a_\Pi^h(v_\Pi^h, v_\Pi^h) \le C_0 \cdot a_\Omega^h(v_\Omega^h, v_\Omega^h), \tag{5.102}$$

where constant $C_0 > 1$ does not depend on mesh-size parameter h and the coefficients of the problem.

Note that since $v_\Omega^h = v_\Pi^h$ in Ω, then $v_\Omega^h = v_\Pi^h$ on $\Gamma = \partial\Omega$.

Using the basis representations of the functions in $V_h(\Omega)$ and $V_h(\Pi)$ we can also formulate a matrix analog of (5.102). Namely, for any vector $\mathbf{v}_\Omega \in \mathrm{I\!R}^N$ there exists a vector $\mathbf{v}_\Pi \in \mathrm{I\!R}^M$ such that it coincides with $\mathbf{v}_\Omega$ in the nodes of Ω and the following inequality holds true:

$$(A_\Pi\mathbf{v}_\Pi, \mathbf{v}_\Pi) \le C_0 \cdot (A_\Omega\mathbf{v}_\Omega, \mathbf{v}_\Omega). \tag{5.103}$$

Following [94] one can show that for any vector $\mathbf{v}_\Gamma \in \mathrm{I\!R}^{N_\Gamma}$ the following equalities for the corresponding Schur complements on Γ are valid:

$$\begin{aligned} (S_\Gamma^{(\Omega)}\mathbf{v}_\Gamma, \mathbf{v}_\Gamma) &= \min_{\mathbf{v}_\Omega \in \mathrm{I\!R}^N,\, \mathbf{v}_\Omega|_\Gamma = \mathbf{v}_\Gamma} (A_\Omega\mathbf{v}_\Omega, \mathbf{v}_\Omega), \\ (S_\Gamma^{(\Pi)}\mathbf{v}_\Gamma, \mathbf{v}_\Gamma) &= \min_{\mathbf{v}_\Pi \in \mathrm{I\!R}^M,\, \mathbf{v}_\Pi|_\Gamma = \mathbf{v}_\Gamma} (A_\Pi\mathbf{v}_\Pi, \mathbf{v}_\Pi). \end{aligned} \tag{5.104}$$

Now fix any vector $\mathbf{v}_\Gamma$ of unknowns on boundary Γ. Let $\mathbf{u}_\Omega \in \mathrm{I\!R}^N$ be such a vector that satisfies the equality

$$(S_\Gamma^{(\Omega)}\mathbf{v}_\Gamma, \mathbf{v}_\Gamma) = (A_\Omega\mathbf{u}_\Omega, \mathbf{u}_\Omega). \tag{5.105}$$

Denote by $\mathbf{u}_\Pi \in \mathrm{I\!R}^M$ the corresponding extension vector. Note that $\mathbf{u}_\Pi|_\Gamma = \mathbf{v}_\Gamma$. From (5.103) it follows that

$$(A_\Pi\mathbf{u}_\Pi, \mathbf{u}_\Pi) \le C_0 \cdot (A_\Omega\mathbf{u}_\Omega, \mathbf{u}_\Omega). \tag{5.106}$$

Taking into account (5.104), (5.106), and (5.105) we get

$$(S_\Gamma^{(\Pi)}\mathbf{v}_\Gamma, \mathbf{v}_\Gamma) \le (A_\Pi\mathbf{u}_\Pi, \mathbf{u}_\Pi) \le C_0 \cdot (A_\Omega\mathbf{u}_\Omega, \mathbf{u}_\Omega) = C_0 \cdot (S_\Gamma^{(\Omega)}\mathbf{v}_\Gamma, \mathbf{v}_\Gamma). \tag{5.107}$$

Thus, the left inequality in (5.101) with the constant $\alpha = 1/C_0$ follows from (5.107). □

5.3.2.2.3 Construction of preconditioner B_φ. We proceed with the construction of preconditioner B_φ. We define it in the form of the inner Chebyshev iterative procedure [8, 18, 63, 70]. For each subdomain Ω_k, $k = 1, \dots, m$, we define the matrices $S^{(\Pi)}_{\Gamma,k}$. Using Lemmas 5.2 and 5.3 it is easy to show that the nonzero eigenvalues of matrices $A^{-1}_{\Gamma,k} S^{(\Pi)}_{\Gamma,k}$ belong to segment $[\alpha_0 \cdot h/r, \alpha_1]$, where constants α_0 and α_1 do not depend on mesh-size parameter h, the size of subdomain r, and the coefficients of problem (5.46) in the subdomains. Now we define matrices $\tilde{S}_\varphi$ and $\tilde{B}_\varphi$ outlined in (5.77):

$$\tilde{S}_\varphi = \sum_{k=1}^{m} D_{\Gamma,k} [S^{(\Pi)}_{\Gamma,k}]^{+} D^T_{\Gamma,k}, \qquad \tilde{B}_\varphi = \sum_{k=1}^{m} D_{\Gamma,k} A^{-1}_{\Gamma,k} D^T_{\Gamma,k}. \tag{5.108}$$

From the previous considerations it follows that the nonzero eigenvalues of matrix $\tilde{B}^{-1}_\varphi \tilde{S}_\varphi$ belong to segment $[c_0, c_1 \cdot r/h]$, where constants c_0 and c_1 do not depend on mesh-size parameter h and the coefficients of the problem.

Let $P_L(y)$ be a polynomial of least deviation from zero on this segment that satisfies the condition $P_L(0) = 1$. Denote by β_l, $l = 1, \dots, L$, the inverses of the roots of the polynomial $P_L(y)$. The formulae for $P_L(y)$ and its roots $1/\beta_l$, $l = 1, \dots, L$, are given in Section 3.4. Then preconditioner B_φ for matrix S_φ is determined by the formula:

$$B^{+}_\varphi = \left\{ I_\varphi - \prod_{l=1}^{L} \left(I_\varphi - \beta_l \tilde{B}^{-1}_\varphi \tilde{S}_\varphi \right) \right\} \tilde{S}^{+}_\varphi. \tag{5.109}$$

The calculation of vector $\mathbf{w}_\varphi = B^{+}_\varphi \mathbf{g}_\varphi$ given $\mathbf{g}_\varphi \in \mathbb{R}^{N_\varphi}$ can be reduced to:

$$\begin{aligned} \mathbf{w}^{(0)}_\varphi &= \mathbf{0}, \\ \mathbf{w}^{(l)}_\varphi &= \mathbf{w}^{(l-1)}_\varphi - \beta_l \tilde{B}^{-1}_\varphi \left(\tilde{S}_\varphi \mathbf{w}^{(l-1)}_\varphi - \mathbf{g}_\varphi \right), \qquad l = 1, \dots, L, \\ \mathbf{w}_\varphi &= \mathbf{w}^{(L)}_\varphi. \end{aligned} \tag{5.110}$$

For computational stability, instead of (5.110), we can use the three-term formula [114].

To implement this preconditioner we have to develop for each subdomain an algorithm that multiplies vector $\mathbf{u}_{\Gamma,k} \in \mathbb{R}^{N_{\Gamma,k}}$ by matrix $[S^{(\Pi)}_{\Gamma,k}]^{+}$:

$$\mathbf{v}_{\Gamma,k} := [S^{(\Pi)}_{\Gamma,k}]^{+} \, \mathbf{u}_{\Gamma,k}. \tag{5.111}$$

We can define the resulting vector $\mathbf{v}_{\Gamma,k}$ as a solution of the problem

$$S^{(\Pi)}_{\Gamma,k} \, \mathbf{v}_{\Gamma,k} = \mathbf{u}_{\Gamma,k}, \tag{5.112}$$

or, equivalently,

$$\begin{bmatrix} A_{I,k} & A_{I\Gamma,k} & \mathbf{0} \\ A_{\Gamma I,k} & A^{(\Pi)}_{\Gamma,k} & A_{\Gamma O,k} \\ \mathbf{0} & A_{O\Gamma,k} & A_{O,k} \end{bmatrix} \cdot \begin{bmatrix} * \\ \mathbf{v}_{\Gamma,k} \\ * \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{u}_{\Gamma,k} \\ \mathbf{0} \end{bmatrix}. \tag{5.113}$$

Here we use the block representation of matrix $A^{(\Pi)}_k$ corresponding to partition (5.98).

From (5.113) it is easy to see that the procedure of multiplying vector $\mathbf{u}_{\Gamma,k}$ by matrix $[S^{(\Pi)}_{\Gamma,k}]^{+}$ can be considered as a partial solution problem when we have the nonzero right-hand side $\mathbf{u}_{\Gamma,k}$ defined only in the nodes of Γ_k and we need to find the solution only in these nodes. To develop an optimal preconditioner for the initial problem we need the partial solution algorithm for the kind of problem with an implementation cost $O(N_{\Gamma,k}\ln(N_{\Gamma,k}))$. Unfortunately, there is no such algorithm for the general partial solution problem, although for some special cases ("parallel" and "perpendicular" cases) this algorithm is outlined in Section 5.2.2.2 (more information on this issue can be found in [9, 10, 68, 103]). As shown in [68, 103], instead of developing an exact solver we can use an approximate one. That is we can use the approximate partial solution algorithm when we solve problem (5.113) with an accuracy ε. In [103] it is shown that the approximate partial solution of problem (5.113) can be found with accuracy $\varepsilon = ch^p$ for

$$O(p^2 N_{\Gamma,k}^{3/2}(\ln N_{\Gamma,k})^2) \tag{5.114}$$

operations. Here $c > 0$ is a positive constant independent of $N_{\Gamma,k}$ and $p \geq 2$.

Thus, instead of matrix $\tilde{S}_\varphi$ defined by (5.108) we define matrix $\tilde{S}_\varphi^{(\varepsilon)}$ by

$$\tilde{S}_\varphi^{(\varepsilon)} = \sum_{k=1}^{m} D_{\Gamma,k}[S_{\Gamma,k}^{(\Pi)}]_{(\varepsilon)}^{+} D_{\Gamma,k}^T, \tag{5.115}$$

where $[S_{\Gamma,k}^{(\Pi)}]_{(\varepsilon)}^{+}$ means the use of the approximate partial solution algorithm. It can be shown [103] that this matrix is still spectrally equivalent to original matrix S_φ, so preconditioner $B_{\varphi,\varepsilon}$ can be defined by

$$B_{\varphi,\varepsilon}^{+} = \left\{ I_\varphi - \prod_{l=1}^{L}\left(I_\varphi - \beta_l \tilde{B}_\varphi^{-1}\tilde{S}_\varphi^{(\varepsilon)}\right)\right\}\tilde{S}_\varphi^{+}. \tag{5.116}$$

Thus, the previous considerations as well as the theory of Chebyshev iterative methods lead to the important result.

Proposition 5.3 *Let $L \geq C \cdot (r/h)^{1/2}$, where constant C does not depend on mesh-size parameter h, size of the subdomains r, and coefficients of the problem $K(\mathbf{x})$ and $c(\mathbf{x})$. Then matrix $B_{\varphi,\varepsilon}$ in (5.116) with matrices $\tilde{B}_\varphi$ and $\tilde{S}_\varphi^{(\varepsilon)}$ defined in (5.108) and (5.115), respectively, is spectrally equivalent to matrix S_φ.*

Remark 5.8 In theory quantity L is chosen to be of order $(C \cdot r/h)^{1/2}$. In practice it is calculated explicitly after the boundaries of the spectrum of matrix $\tilde{B}_\varphi^{-1}\tilde{S}_\varphi^{(\varepsilon)}$ have been calculated by an appropriate iterative procedure [62].

Thus, we have proved the following theorem

Theorem 5.2 *The constructed block diagonal preconditioner $\mathcal{B}$ is spectrally equivalent to matrix $\mathcal{A}$ with constants independent of mesh-size parameter h and the coefficients of the problem.*

Recall that the spectral equivalence of $\mathcal{B}$ and $\mathcal{A}$ means that the eigenvalues of $\mathcal{B}^{-1}\mathcal{A}$ belong to the union of segments $[d_1, d_2]$ and $[d_3, d_4]$ with $d_1 \leq d_2 < 0 < d_3 \leq d_4$, where d_1, d_2, d_3, and d_4 are independent of h, $K(\mathbf{x})$, and $c(\mathbf{x})$.

5.3.2.2.4 Arithmetic complexity of preconditioner $\mathcal{B}$. In the two dimensional case the arithmetic cost of product $B_k^+\mathbf{u}_k$ is proportional to

$$O(N_k) \sim O(N/m) \sim O(r^{-2}h^{-2}). \tag{5.117}$$

Thus the arithmetic complexity of the multiplication of vector $\mathbf{u} \in \mathrm{I\!R}^N$ by matrix B_A^+ is proportional to N.

Now consider the implementation of the multiplication of vector $\mathbf{u}_\varphi \in \mathrm{I\!R}^{N_\varphi}$ by matrix $B_{\varphi,\varepsilon}^+$. In order to estimate the arithmetic complexity of this multiplication it is sufficient to estimate the arithmetic complexity of multiplying vector $\mathbf{u}_\varphi$ by matrix $\tilde{S}_\varphi^{(\varepsilon)}$ and the arithmetic complexity of solving the system:

$$\tilde{B}_\varphi \mathbf{u}_\varphi = \boldsymbol{\xi}_\varphi \tag{5.118}$$

for a given vector $\boldsymbol{\xi}_\varphi \in \mathrm{I\!R}^{N_\varphi}$. After the estimation of these two operations it is easy to compute the complexity of matrix $B_{\varphi,\varepsilon}^+$ because we know the order of the number of iterations L in the inner Chebyshev iterative procedure (5.116): $L \sim (C \cdot r/h)^{1/2}$.

Since matrices $A_{\Gamma,k}$ are diagonal, problem (5.118) can be solved for

$$O\left(\sum_{k=1}^m N_{\Gamma,k}\right) \sim O\left(m \cdot \sqrt{(N/m)}\right) \sim O\left(r^{-1} \cdot h^{-1}\right) \tag{5.119}$$

operations.

Taking into account estimate (5.114) of the computational complexity $[S_{\Gamma,k}^{(\Pi)}]_{(\varepsilon)}^+$, the multiplication of a vector by matrix $\tilde{S}_\varphi^{(\varepsilon)}$ is estimated by

$$O\left(\sum_{k=1}^m p^2 N_{\Gamma,k}^{3/2} (\ln N_{\Gamma,k})^2\right) \sim O\left(p^2 r^{-2} (r/h)^{3/2} (\ln r/h)^2\right) \sim O\left(p^2 r^{-1/2} h^{-3/2} (\ln r/h)^2\right). \tag{5.120}$$

On the basis of the above analysis we can formulate the assertion.

Lemma 5.4 *Under the above assumptions the arithmetic complexity of the proposed algorithm for solving the problem with matrix B_φ is estimated from above by*

$$C \cdot \left(r^{-1/2}h^{-3/2} + p^2 h^{-2} (\ln r/h)^2\right) \sim C \cdot \left(m^{1/4} N^{3/4} + p^2 N (\ln N/m)^2\right), \tag{5.121}$$

where N is the order of matrix A and constant C does not depend on mesh-size parameter h, size of subdomains r, and coefficients of the problem $K(\mathbf{x})$ and $c(\mathbf{x})$.

Summarizing the results of Sections 5.3.1 and 5.3.2 we can formulate the following proposition.

Proposition 5.4 *Under the above assumptions the arithmetical complexity of the proposed algorithm for solving problem (5.46) in two-dimensional domains is estimated from above by*

$$C \cdot (h^{-2} (\ln (r/h))^2) \sim O(N \ln^2 (N/m)),$$

where constant C is independent of mesh size parameter h, size of subdomains r, and coefficients of the problem $K(\mathbf{x})$ and $c(\mathbf{x})$.

Remark 5.9 Along with preconditioner B_φ for matrix S_φ in the form of the inner Chebyshev iterations (5.109) we can also consider the multigrid method [20] for the Schur complement to precondition S_φ. Although there are some theoretical details which should be addressed before using this method, not considered in the dissertation, we provide an estimate for the arithmetical complexity of the multigrid preconditioner:

$$O(N^{3/4}\ln^2(N/m)).$$

Remark 5.10 The approach described above can be developed for three-dimensional problems provided that we have constructed for each subdomain Ω_k matrices $\tilde{B}_{\Gamma,k}$ and $\tilde{S}_{\Gamma,k}$, $k = 1, \ldots, m$, introduced at the beginning of Section 5.3.2.2.

CHAPTER VI

APPLICATIONS

This chapter is devoted to the applications of the theory developed in Chapters IV and V. We present the results from numerical experiments that illustrate this theory and apply it to the problem of modeling fluid flow in porous media.

To show that the preconditioners developed in the earlier chapters are good and robust we apply them to various problems in two and three dimensions. Our objectives in conducting the numerical experiments were to establish experimentally the conclusions from the theoretical analysis of the algorithms considered and to assess their effectiveness in terms of error reduction after a fixed number of iterations.

The right-hand sides in all model tests have been generated randomly. The condition numbers of the preconditioned matrices have been calculated from the relation between conjugate gradients and the Lanczos algorithm described in Section 3.3. Most of the experiments have been run on Sun workstations, although a simulator of fluid flow in porous media has also been developed in a parallel version which has been run on a Paragon supercomputer.

The rest of the chapter is organized as follows. In the first section we present the experiments for the substructuring methods and the fictitious component method developed in Chapter IV. In the next section we consider experiments with the domain decomposition method on matching and nonmatching grids which illustrate the theory of Chapter V. Finally, we consider an application of the Lagrange multiplier approach described in Chapter II to modeling fluid flow in porous media.

6.1 Experiments with substructuring and fictitious components methods

6.1.1 3D problem. Partition of cube onto 6 tetrahedra

We present three numerical examples. The method of preconditioning on the basis of multilevel substructuring as discussed in Section 4.3 was tested first on the model problem

$$\begin{array}{rcll} -\Delta u & = & f, & \text{in } \Omega = [0,1]^3 \subset \mathbb{R}^3, \\ u & = & 0, & \text{on } \partial\Omega \end{array} \tag{6.1}$$

with the nonconforming finite element method of approximation.

The domain was divided into n^3 cubes (n in each direction) and each cube was partitioned into 6 tetrahedra. The total dimension of the original algebraic system was $N = 12n^3 - 6n^2$.

The original algebraic problem has been solved by the conjugate gradient (CG) method in the form of (3.28) with the preconditioner in the form of (4.44) with accuracy $\varepsilon = 10^{-6}$. For comparison, that problem has been solved by the same method without preconditioning.

These results are summarized in Table 6.1. Here *Iter* is the number of CG iterations, *Cond* is the condition number, and *time* is the computational time needed to achieve the required accuracy.

Table 6.1: *6 tetrahedra per cube. Experiments with Laplace equation.*

		CG without Preconditioning			CG with Preconditioning		
n	N	Iter	Cond	time (sec)	Iter	Cond	time (sec)
4	672	40	66	0.18	22	9.84	0.22
8	5760	73	265	2.18	24	10.7	1.27
16	47616	130	1062	49.2	24	11.94	15.7
32	387072	200<	—	1248	25	12.2	163
40	758400				25	12.26	376
50	1485000				25	12.33	771

In the second example:

$$\begin{aligned} -\sum_{i=1}^{3} \frac{\partial}{\partial x_i}\left(a_i \frac{\partial u}{\partial x_i}\right) &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{6.2}$$

we have considered the dependency of the condition number on the coefficients of the problem. The coefficients a_i, $i = 1, 2, 3$, were constants on each cube. The results are summarized in Table 6.2, where Iter and Cond denote the iteration number and condition number, respectively.

From Table 6.2 we see that the condition number depends on the maximal ratio $\kappa = \max\limits_{C \in \mathcal{C}_h} \left\{ \frac{a_3}{a_1}, \frac{a_3}{a_2} \right\}$. The case of $\kappa < 1$ has a better convergence than the case of the Poisson equation (i.e. $a_1 = a_2 = a_3 = 1$) as is predicted by the theory (see estimate (4.45)).

We note that the condition numbers in all experiments depend on parameter κ introduced in Assumption 4.1 (see page 47). Namely, the estimate of the condition number of the preconditioned matrix (4.45) is proportional to the value of parameter κ. Obviously, it is important to arrange the coordinate axis in such a way that parameter κ has the smallest value. In some sense we can benefit from anisotropy. The smaller coefficient a_3 (the coefficient in the "z-direction") leads to a better preconditioner B.

In the third example we treat the Poisson equation on the domain Ω as shown in Figure 6.1. The domain is subdivided into $90 \times 90 \times 10$ cubes and the number of the unknowns is then $N = 955440$. The algebraic problem is solved with accuracy $\varepsilon = 10^{-6}$. Twenty iterations are needed to achieve the desired accuracy; the computed condition number of matrix $B^{-1}A$ is equal to 10.

6.1.2 3D problem. Partition of cube onto 5 tetrahedra

We present three numerical examples. The preconditioner based on multilevel substructuring as discussed in Section 4.4 was tested first on the model problem

$$\begin{aligned} -\text{div}\,(a(x)\nabla u) &= f, \qquad \text{in } \Omega = [0,1]^3 \\ u &= 0, \qquad \text{on } \partial\Omega \end{aligned} \tag{6.3}$$

with piecewise constant coefficient.

Table 6.2: *6 tetrahedra per cube. Dependency on parameter κ.*

			$N = 47,\,616$		$N = 387,\,072$	
a_1	a_2	a_3	Iter	Cond	Iter	Cond
1	1	1	18	7.5	17	7.6
1	1	0.1	13	3.7	13	3.8
1	1	0.01	10	2.8	11	3.0
10	1	1	16	6	16	6.2
1	10	1				
100	1	1	14	4.7	14	5.2
1	100	1				
1	1	10	34	41	34	42
1	1	100	75	315	80	328
0.1	1	1	32	30	31	29
1	0.1	1				
0.01	1	1	68	198	72	203
1	0.01	1				

Again, the domain is divided into $M = n^3$ cubes (n in each direction). Each cube is partitioned into 5 tetrahedra. The dimension of the original algebraic system is $N = 10n^3 - 6n^2$. The right-hand side is generated randomly, and the accuracy parameter is taken as $\varepsilon = 10^{-6}$. The coefficient $a(x)$ is piecewise constant and is defined to be

$$a(x, y, z) = \begin{cases} a, & (x, y, z) \in [0.5, 1] \times [0.5, 1] \times [0.5, 1] \\ 1, & \text{elsewhere.} \end{cases} \tag{6.4}$$

The computational results are summarized in Table 6.3.

In the second experiment, the method of preconditioning described in Section 4.4 was used to solve problem (6.3) with constant right-hand-side function $f(x)$ by the mixed finite element method with function $a(x)$ in the following form (see also Figure 6.2):

$$a(x, y, z) = \begin{cases} a = 0.01, & (x, y, z) \in \left\{ \begin{array}{l} [0.2, 0.4] \times [0.2, 0.4] \times [0.2, 0.4] \cup \\ [0.6, 0.8] \times [0.2, 0.4] \times [0.2, 0.4] \cup \\ [0.2, 0.4] \times [0.6, 0.8] \times [0.2, 0.4] \cup \\ [0.6, 0.8] \times [0.6, 0.8] \times [0.2, 0.4] \cup \\ [0.2, 0.4] \times [0.2, 0.4] \times [0.6, 0.8] \cup \\ [0.6, 0.8] \times [0.2, 0.4] \times [0.6, 0.8] \cup \\ [0.2, 0.4] \times [0.6, 0.8] \times [0.6, 0.8] \cup \\ [0.6, 0.8] \times [0.6, 0.8] \times [0.6, 0.8] \end{array} \right\}. \\ 1, & \text{elsewhere} \end{cases} \tag{6.5}$$

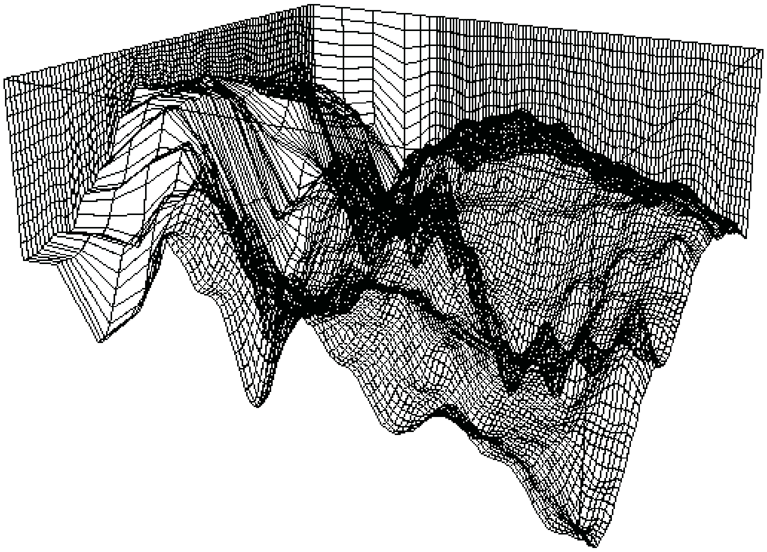


Figure 6.1: *An example of the grid domain* Ω.

Table 6.3: *5 tetrahedra per cube. Dependency on jump parameter* a.

	$20 \times 20 \times 20$		$30 \times 30 \times 30$		$40 \times 40 \times 40$		$50 \times 50 \times 50$	
	$N = 77\ 600$		$N = 264\ 600$		$N = 630\ 400$		$N = 1\ 235\ 000$	
a	Iter	Cond	Iter	Cond	Iter	Cond	Iter	Cond
1	14	5.32	14	5.30	14	5.29	14	5.28
10	17	6.59	17	6.53	16	6.37	16	6.29
100	17	6.94	17	6.90	16	6.89	16	6.88
1000	17	6.98	16	6.96	16	6.95	16	6.93
10^4	16	6.98	16	6.96	16	5.95	16	6.94
0.1	16	5.97	16	5.96	16	5.96	15	5.94
0.01	16	6.02	16	6.02	16	6.00	15	5.97
0.001	16	6.02	16	6.01	16	6.00	15	5.97
10^{-4}	16	6.02	16	6.01	16	6.00	15	5.97

Again, the domain Ω is the unit cube; the domain is divided into $M = 40^3 = 64000$ cubes. The dimension of the original algebraic system for the Lagrange multipliers is $N = 630400$.

The preconditioner (4.73) needs $n_{\text{iter}} = 22$ iterations to solve (6.3).

In both experiments it takes less then 12 minutes to obtain resulting vectors $\mathbf{q}$ and $\mathbf{u}$. The slices of solution $\mathbf{u}$ by planes parallel to the xy–plane are shown in Figure 6.3.

Finally, in this section the method of preconditioning presented in Section 4.4 is tested on

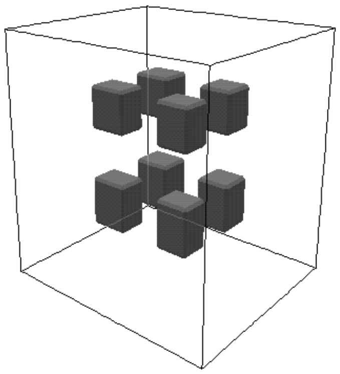
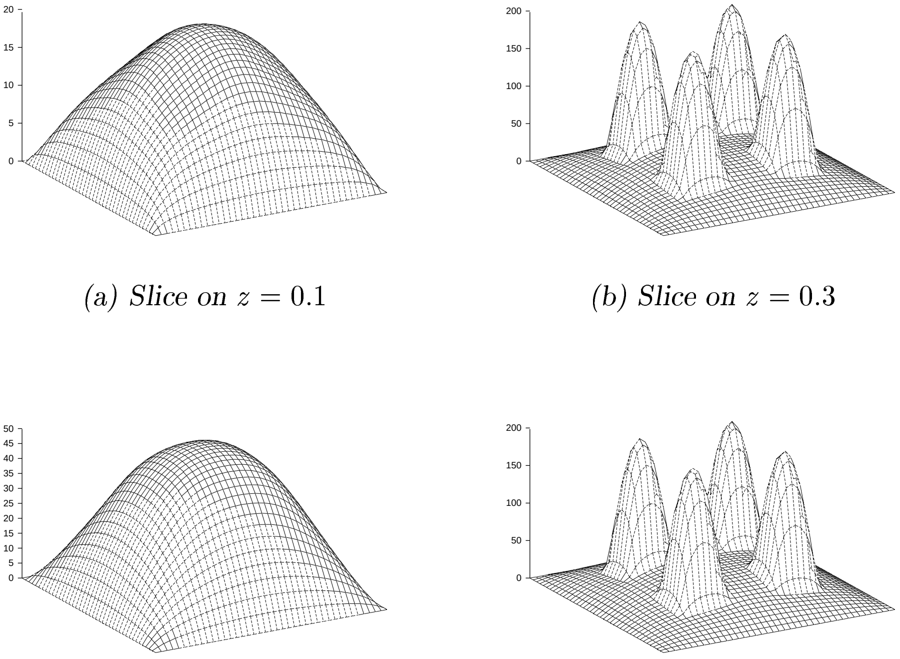


Figure 6.2: *Function $a(x)$ for the model problem (6.3).*



(a) Slice on $z = 0.1$

(b) Slice on $z = 0.3$

(c) Slice on $z = 0.5$

(d) Slice on $z = 0.7$

Figure 6.3: *Slices of the solution parallel to xy-plane.*

the model problem

$$-\sum_{i=1}^{3} \frac{\partial}{\partial x_i}\left(k_i \frac{\partial u}{\partial x_i}\right) = f \quad \text{in } \Omega = [0,1]^3, \qquad u = 0 \quad \text{on } \partial\Omega. \tag{6.6}$$

Again, the domain is divided into n^3 cubes (n in each direction) and each cube is partitioned into 5 tetrahedra. The right-hand side is generated randomly, and the accuracy parameter is taken as $\varepsilon = 10^{-6}$. Coefficients k_i, $i = 1, 2, 3$, are constants on each cube. The results are summarized in Table 6.4.

Table 6.4: *5 tetrahedra per cube. Dependency on parameter* κ.

			$16 \times 16 \times 16$		$20 \times 20 \times 20$		$30 \times 30 \times 30$	
			$N = 39424$		$N = 77600$		$N = 264600$	
k_1	k_2	k_3	Iter	Cond	Iter	Cond	Iter	Cond
1	1	1	14	4.87	14	4.93	14	5.03
1	1	10	12	3.72	12	3.94	12	4.28
1	1	100	9	2.28	10	2.55	10	3.00
1	1	1000	8	1.55	8	1.58	8	1.73
1	1	10000	8	1.48	8	1.49	8	1.51
1	1	0.1	31	19.4	31	19.6	31	19.8
1	1	0.01	62	133.	71	149.	82	168.
10	1	1	24	12.0	25	12.1	25	12.1
1	10	1	24	12.1	24	12.1	24	12.0
100	1	1	58	99.3	63	100.	62	100.
1	100	1	62	100.	60	100.	60	99.5
1	10	10	14	4.72	14	4.81	14	4.94
1	10	100	12	3.62	12	3.85	12	4.25
1	10	1000	9	2.14	10	2.42	10	2.92
1	100	10000	9	2.20	10	2.42	10	2.92

From Table 6.4 we see that the condition number depends on the maximal ratio $\kappa = \max\limits_{C\in\mathcal{C}_h} \left\{ \frac{k_1}{k_3},\ \frac{k_2}{k_3} \right\}$. The numerical results are in full agreement with the theoretical estimates. One can see that the proposed preconditioner is optimal if $\kappa \leq 1$. In the case of $\kappa < 1$ the method has a better convergence than in the case of the Poisson equation (i.e. $k_1 = k_2 = k_3 = 1$). If $\kappa > 1$, the preconditioner looses its optimal order and the corresponding relative condition numbers increased strongly with κ. It is a rather predictable result since we defined local preconditioning matrices B^C in (4.68) on each cube, taking some "additional positiveness" from the direction with the dominated anisotropy (z-direction) to other directions. Experiments show that this procedure is "well behaved" if the coefficient in the z-direction (k_3) is greater than coefficients k_1 and k_2. And the method loses its effectiveness if we choose the wrong direction, i.e. coefficient k_3 is small compared with coefficients k_1 and k_2.

Remember that in the method described in Section 4.4 we need only an assumption that coefficient k_* in some direction is not less then the coefficients in the other directions. Thus, if, for example, we have the problem where coefficient k_1 is not less than coefficients k_2 and k_3 we can simply rename variables in such a way that a new z variable corresponds to the old x variable. The results will be the same.

6.1.3 Fictitious components method

In this section we consider the results of the numerical experiments with the fictitious components method for two- and three-dimensional problems.

First, we consider a model problem

$$\begin{aligned} -\text{div } (K\nabla u) &= f \qquad && \text{in } \Omega = [0,1]^2, \\ (K\nabla u, \mathbf{n}) &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{6.7}$$

where

$$K(x) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is a constant symmetric tensor and $f(x) \in L^2(\Omega)$.

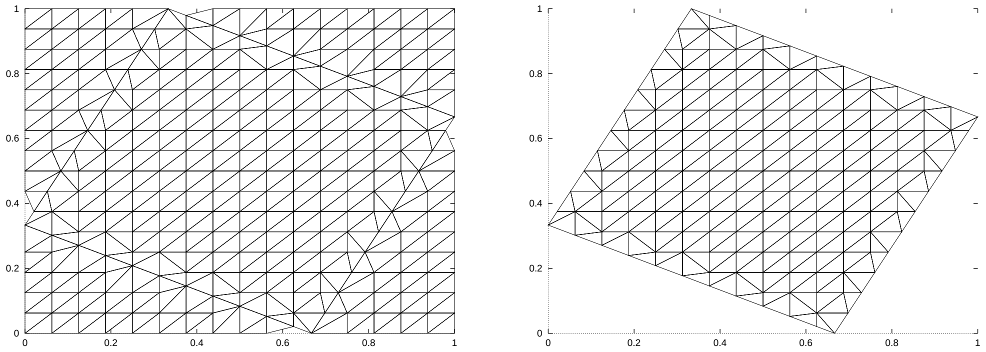
Let K have eigenpairs $(k_1, \mathbf{u}_1)$ and $(k_2, \mathbf{u}_2)$, where $\mathbf{u}_1 = (\alpha, \beta)$, $\mathbf{u}_2 = (-\beta, \alpha)$, $\alpha^2 + \beta^2 = 1$. Then consider a transformation of the coordinates $(\xi, \nu) = F(x, y)$:

$$\xi = \alpha \cdot x + \beta \cdot y, \qquad \nu = -\beta \cdot x + \alpha \cdot y.$$

In coordinates (ξ, ν) problem (6.7) has the form:

$$\begin{aligned} -k_1 \cdot u_{\xi\xi} - k_2 \cdot u_{\nu\nu} &= \tilde{f} \qquad && \text{in } \tilde{\Omega}, \\ \frac{\partial u}{\partial n} &= 0 && \text{on } \partial\tilde{\Omega}. \end{aligned} \tag{6.8}$$

Now construct a rectangle Π in the (ξ, ν) plane which contains $\tilde{\Omega}$ and a uniform triangular mesh in Π. Then locally modify the mesh in Π to fit the boundaries of $\tilde{\Omega}$ (see Figures 6.4, 6.5).



(a) *Fictitious domain* Π (b) *Real domain* Ω

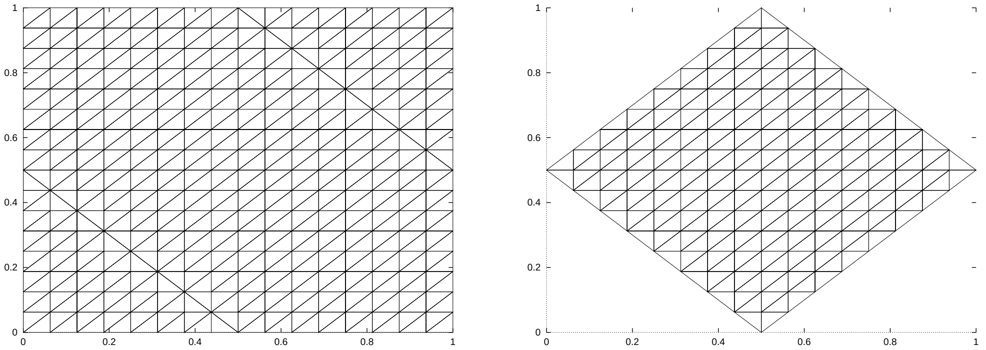
Figure 6.4: *Example 1. Triangulations of the fictitious and real domains.*

Then consider the P_1-nonconforming approximation of (6.7) as is described in Section 4.2:

$$A\mathbf{u} = \mathbf{f}. \tag{6.9}$$

To precondition A we use the fictitious components method described in Section 4.5. In this method we have to solve the problem:

$$\begin{aligned} -k_1 \cdot u_{\xi\xi} - k_2 \cdot u_{\nu\nu} &= \tilde{f} \qquad && \text{in } \Pi, \\ \tfrac{\partial u}{\partial n} &= 0 && \text{on } \partial\Pi. \end{aligned} \tag{6.10}$$



(a) Fictitious domain Π *(b) Real domain* Ω

Figure 6.5: *Example 2. Triangulations of the fictitious and real domains.*

To solve problem (6.10) we use the substructuring method of Section 4.2.

In all examples below we have used the preconditioned conjugate gradient method to solve (6.9). The stopping criterion was relative residue $\leq \varepsilon = 10^{-6}$. The fictitious domain was partitioned onto $M \times M$ rectangles; N was a dimension of the real triangulation. The number of PCG iterations *Iter* and computed condition numbers *Cond* are shown in Tables 6.5 and 6.6.

In the first example the coefficient matrix and its corresponding eigenpairs were

$$K = \begin{bmatrix} 401 & 198 \\ 198 & 104 \end{bmatrix}, \qquad \begin{array}{ll} k_1 = 5, & \mathbf{u}_1 = \frac{1}{\sqrt{5}}(1,-2), \\ k_2 = 500, & \mathbf{u}_2 = \frac{1}{\sqrt{5}}(2,1). \end{array} \tag{6.11}$$

The domain is shown in Figure 6.4, and the results of the experiments are given in Table 6.5.

Table 6.5: *Example 1. Results of the experiments.*

M	N	Iter	Cond
16	169	10	7.463
32	605	13	9.827
64	2377	13	10.007
128	9245	15	14.080
256	36809	15	14.841
512	146205	16	14.346

In the second example the coefficient matrix and its corresponding eigenpairs were

$$K = \begin{bmatrix} 1001 & 999 \\ 999 & 1001 \end{bmatrix}, \qquad \begin{array}{ll} k_1 = 2, & \mathbf{u}_1 = \frac{1}{\sqrt{2}}(1,-1), \\ k_2 = 2000, & \mathbf{u}_2 = \frac{1}{\sqrt{2}}(1,1). \end{array} \tag{6.12}$$

The domain is shown in Figure 6.5, and the results of the experiments are given in Table 6.6.

Finally, we consider a three-dimensional problem

$$\begin{array}{rl} -\text{div }(K\nabla u) = f & \text{in } \Omega = [0,1]^3, \\ (K\nabla u, \mathbf{n}) = 0 & \text{on } \partial\Omega, \end{array} \tag{6.13}$$

Table 6.6: *Example 2. Results of the experiments.*

M	N	Iter	Cond
16	145	7	6.091
32	545	9	7.920
64	2113	9	7.295
128	8321	10	10.988
256	33025	11	12.500
512	131585	10	11.206

with a constant symmetric tensor

$$K(x) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

We construct the coordinate system on the eigenvectors of matrix K and a parallelepiped Π embedding domain Ω. Then we define a uniform cubic mesh in Π and locally modify the mesh in Π to fit the boundaries of $\tilde{\Omega}$. We subdivide each cube into 5 tetrahedra and define the P_1-nonconforming finite element space on this tetrahedral partitioning as is described in Section 4.4.

Again, we use the fictitious components method to precondition problem (6.9). To solve the problem in the parallelepiped we use the substructuring method of Section 4.4.

We considered the problem with coefficient matrices

$$K = \begin{bmatrix} 12 & -19 & 10 \\ -19 & 41 & -19 \\ 10 & -19 & 12 \end{bmatrix}, \qquad \begin{array}{ll} k_1 = 3, & \mathbf{u}_1 = \frac{1}{\sqrt{3}}(1,1,1), \\ k_2 = 2, & \mathbf{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1), \\ k_3 = 60, & \mathbf{u}_3 = \frac{1}{\sqrt{6}}(1,-2,1), \end{array} \tag{6.14}$$

and

$$K = \begin{bmatrix} 102 & -199 & 100 \\ -199 & 401 & -199 \\ 100 & -199 & 102 \end{bmatrix}, \qquad \begin{array}{ll} k_1 = 3, & \mathbf{u}_1 = \frac{1}{\sqrt{3}}(1,1,1), \\ k_2 = 2, & \mathbf{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1), \\ k_3 = 600, & \mathbf{u}_3 = \frac{1}{\sqrt{6}}(1,-2,1). \end{array} \tag{6.15}$$

The fictitious domain was partitioned into $M \times M \times M$ cubes; N was a dimension of the real problem. In Table 6.7 the number of PCG iterations *Iter* and computed condition numbers *Cond* are shown for both problems.

Again, from Table 6.7 we see that the numerical results are in full agreement with the theoretical estimates. The bigger the coefficient k_3 the better the rate of convergence.

Table 6.7: *Fictitious components method in 3D.*

		$k_3 = 60$		$k_3 = 600$	
M	N	Iter	Cond	Iter	Cond
20	77 600	14	12.91	10	10.31
30	264 600	16	15.32	11	11.44
40	630 400	17	15.85	11	11.53
50	1 235 000	17	15.98	12	11.80

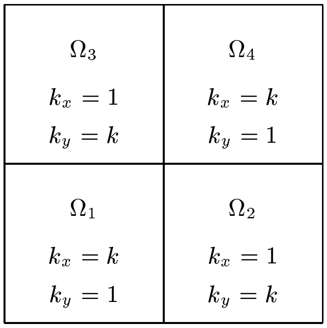
6.2 Experiments with domain decomposition methods

6.2.1 Domain decomposition for problem with diagonal matrix of coefficients

In this section the domain decomposition method presented in Section 5.2 is tested on the model problem in the unit square $\Omega = [0,1]^2$:

$$\begin{aligned} -k_x u_{xx} - k_y u_{yy} &= f, \qquad && \text{in } \Omega \equiv [0,1]^2, \\ u &= 0, && \text{on } \partial\Omega. \end{aligned} \tag{6.16}$$

Domain Ω is composed of 4 subdomains as shown in Figure 6.6. Coefficients k_x and k_y are constants in each subdomain.



Ω_3 $k_x = 1$ $k_y = k$	Ω_4 $k_x = k$ $k_y = 1$
Ω_1 $k_x = k$ $k_y = 1$	Ω_2 $k_x = 1$ $k_y = k$

Figure 6.6: *Coefficients in the subdomains for a model problem.*

The domain is divided into n^2 squares (n in each direction) and each square is partitioned into 2 triangles. The dimension of the original algebraic system is $N = 3n^2 - 2n$ and the dimension of the Schur complement after elimination of the subdomain problems is $N_\Gamma = 4n$. The right-hand side is generated randomly, and the accuracy parameter is taken as $\varepsilon = 10^{-6}$. The degree of matrix polynomial (5.36) equals $L = \left[\sqrt{2.5n}\right] + 1$, where $[\eta]$ is an integer part of η. The condition number of matrix $A_{\Gamma\Gamma}^{-1}\Lambda_\Gamma$ is calculated by the relation between the conjugate gradient and the Lanczos algorithm [59]. The results are summarized in Table 6.8.

Table 6.8: *Results of experiments with bordering method.*

	100×100 $N = 29800$		200×200 $N = 119600$		400×400 $N = 479200$	
k	Iter	Cond	Iter	Cond	Iter	Cond
1	23	10.7	25	10.9	26	10.9
10	23	9.2	24	9.8	26	10.2
100	20	8.3	19	7.9	20	8.1
1000	12	4.2	14	6.2	14	6.4
10000	6	1.5	7	2.0	7	2.1
100000	3	1.1	4	1.1	4	1.1

6.2.2 Domain decomposition on nonmatching grids

In this section the domain decomposition method on nonmatching grids presented in Section 5.3 is tested on the model problem in unit square $\Omega = [0,1]^2$:

$$\begin{aligned} -\sum_{i,j=1}^{2} \frac{\partial}{\partial x_i}(k_{ij}\frac{\partial u}{\partial x_j}) &= f, & &\text{in } \Omega \equiv [0,1]^2, \\ u &= 0, & &\text{on } \partial\Omega. \end{aligned} \tag{6.17}$$

Domain Ω is composed of 4 subdomains. The coefficients k_{ij}, $i,j = 1,2$, are constants in each subdomain. The main directions of the anisotropy in each subdomain and an example of the grids used in an experiment are shown in Figures 6.7 and 6.8, respectively.

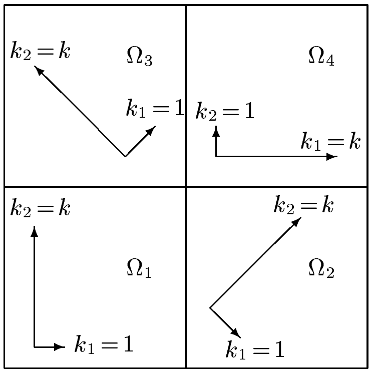


Figure 6.7: *Main directions of the anisotropy in the subdomains.*

Each subdomain Ω_k is embedded in a rectangle Π_k constructed in the local coordinate system as described in Section 5.3. Each rectangle Π_k is divided into n^2 squares (n in each direction) and each square is partitioned into 2 triangles. The dimension of the original algebraic system is N and the dimension of the Schur complement after elimination of the subdomain problems is N_Γ. The right-hand side is generated randomly, and the accuracy parameter is taken as $\varepsilon = 10^{-6}$. The degree of the matrix polynomial (5.116) equals $L = [\sqrt{n}] + 1$, where $[\eta]$ is an integer part of η.

The condition number of matrix $B^{+}_{\varphi,\varepsilon}S_\varphi$ is calculated by the relation between the conjugate gradient and the Lanczos algorithm [59]. The results are summarized in Table 6.9. Here *Iter*

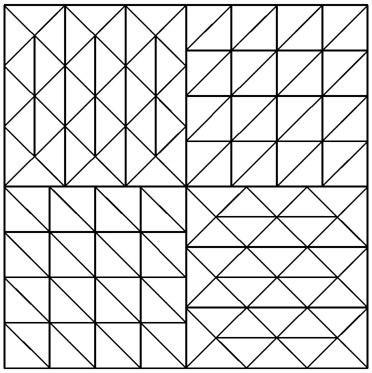


Figure 6.8: *Nonmatching grids.*

denotes the number of iterations of the generalized Lanczos method and *Cond* denotes the condition number of matrix $B^{+}_{\varphi,\varepsilon}S_{\varphi}$.

Table 6.9: *Results of experiments with nonmatching grids.*

	$n = 100$ $N = 90\ 600$		$n = 200$ $N = 361\ 200$		$n = 400$ $N = 1\ 442\ 400$	
k	Iter	Cond	Iter	Cond	Iter	Cond
1	25	12.3	28	13.9	29	14.3
10	23	10.2	25	11.8	27	11.2
100	20	9.4	22	9.9	22	10.1
1000	17	6.2	19	7.2	19	7.4
10000	16	3.5	18	4.0	18	4.1
100000	16	2.1	18	2.1	17	2.1

6.3 Applications to simulation of fluid flow in porous media

In this section we discuss briefly the development of a large scale numerical simulator of multi-phase fuid flow in porous media. The author of this dissertation was one of the code developers in the Partnership in Computational Sciences (PICS) project where he worked on the flow module (see [31]). PICS is an initiative sponsored by the U.S. Department of Energy. The goal of this project is to develop a state of the art computer simulator of groundwater flow and contaminant transport (GCT).

GCT simulates the flow and reactive transport of substance fluids through a heterogeneous porous medium of irregular geometry. This simulator is designed to run on both massively parallel, distributed memory computers and on conventional sequential machines. The flow module is a major part of the PICS code development effort. In numerical simulation of fluid flow in porous media there are two important practical requirements for the approximation method for the corresponding mathematical problem: the method should conserve mass locally on any element and should produce accurate velocities (fluxes) even for highly heterogeneous and anisotropic media with large variations of physical properties. The mixed finite element method considered in Chapter II has all these properties. For this reason in the current version of GCT code the mixed discretization is used for the pressure equations of the two-phase model, whereas the saturation equation is discretized by an upstream weighted Galerkin method. A detailed description of this code can be found in [31].

Triangulation is done first by introducing a logically rectangular grid. Using such a grid essentially simplifies many coding issues and still allows the handling of rather complex geometries. This logically rectangular grid is used to define the upstream-weighted Galerkin method for the saturation equation. To define the mixed method for the pressure equations each grid cell is split into five tetrahedra. When the lowest-order Raviart-Thomas spaces are used, one pressure and four velocity unknowns are attached to every tetrahedron in the grid. As shown in Chapter II the mixed method results in the system of linear equations of a saddle-point form.

In the earliest versions of GCT code the generalized method of minimal residuals was used to solve that systems. In the version GCT 1.3 the Lagrange multiplier technique described in Section 2.4 has been used to replace the saddle-point system by the system with SPD matrix:

$$A\mathbf{u} = \mathbf{f},$$

where matrix A corresponds to the P_1-nonconforming approximation of problem (2.13). Then the iterative methods and preconditioning techniques developed in Chapters III, IV, and V are used.

The results of the experiments with GCT code (written in C) on Sun workstation are presented in Table 6.10. The first column shows the number of cubes in the computational domain. The parameter N in the second column denotes the total dimension of the problem. In columns 3 and 4 of this Table we show the number of iterations and the time (in seconds) required to solve the initial saddle-point system by the minimal residual method (MINRES) with accuracy $\varepsilon = 10^{-6}$. In the next two columns we show the number of conjugate gradient iterations and the time required to solve the same system using the Lagrange multipliers. In the last two columns we show the results for preconditioned conjugate gradient method (PCG) with substructuring preconditioner described in Section 4.4.

Table 6.10: *Results of experiments on Sun workstation.*

		GCT WITH MINRES		GCT WITH CG		PCG METHOD	
$m \times m \times m$	N	Iter	time	Iter	time	Iter	time
$2 \times 2 \times 2$	56	66	4	22	4	12	5
$4 \times 4 \times 4$	544	349	17	52	7	14	7
$8 \times 8 \times 8$	4736	871	252	100	43	15	20
$16 \times 16 \times 16$	39424	1492	2358	189	861	15	43

We have also experimented with the GCT code on distributed memory architectures such as Intel's Paragon. A domain decomposition approach is used in order to use the specific architecture of these machines. The computational domain is decomposed into a set of logically rectangular structures each of which is attached to a single processor. Then a corresponding parallel algorithm for solving the problem is applied. The domain decomposition methods developed in Chapter V of this dissertation make it possible to handle the problems in very heterogeneous and highly anisotropic porous medium.

The computational times required to implement one time step of the GCT code on Paragon supercomputer using 4, 8, and 16 processors are presented in Tables 6.11, 6.12, and 6.13, respectively.

From this experiments one can see that using the Lagrange multiplier reformulation of the mixed system and the nonoverlapping domain decomposition methods are essential for improving the efficiency of the iterative solution and the GCT simulator as a whole.

Table 6.11: *Results of experiments with GCT code on Paragon; 4 processors.*

		GCT with minimal residuals	GCT with conjugate gradient
$m \times m \times m$	N	time (sec)	time (sec)
$4 \times 4 \times 4$	544	10	7
$8 \times 8 \times 8$	4 736	94 (1m34)	19
$16 \times 16 \times 16$	39 424	702 (11m42)	179 (2m59)
$32 \times 32 \times 32$	321 536		2405 (40m05)

Table 6.12: *Results of experiments with GCT code on Paragon; 8 processors.*

		GCT with minimal residuals	GCT with conjugate gradient
$m \times m \times m$	N	time (sec)	time (sec)
$4 \times 4 \times 4$	544	6	9
$8 \times 8 \times 8$	4 736	45	13
$16 \times 16 \times 16$	39 424	359 (5m59)	96 (1m36)
$32 \times 32 \times 32$	321 536	4774 (1h19m34)	1204 (20m04)
$64 \times 64 \times 64$	2 596 864		16579 (4h36m19)

Table 6.13: *Results of experiments with GCT code on Paragon; 16 processors.*

		GCT with minimal residuals	GCT with conjugate gradient
$m \times m \times m$	N	time (sec)	time (sec)
$4 \times 4 \times 4$	544	2	1
$8 \times 8 \times 8$	4 736	22	5
$16 \times 16 \times 16$	39 424	175 (2m55)	51
$32 \times 32 \times 32$	321 536	2401 (40m01)	685 (11m25)
$64 \times 64 \times 64$	2 596 864		9431 (2h37m11)

CHAPTER VII
CONCLUSIONS

The main objectives of this dissertation were:

(1) Development and study of the efficient iterative techniques for nonconforming finite element approximations to boundary value problems of second order self-adjoint linear elliptic PDE's with a special emphasis on problems in three dimensions with possibly large anisotropy in the coefficients of the PDE's.

(2) Construction of an iterative method based on domain decomposition for algebraic systems that occur when using nonmatching grids in subdomains.

(3) Experimental verification of conclusions from the theoretical analysis of the algorithms considered and application of the developed methods to the simulation of fluid flow in porous media.

Based on the research conducted in this dissertation, the following main results are presented for the defense:

(1) New preconditioning techniques for nonconforming approximations of two- and three-dimensional anisotropic problems are developed and studied. It is shown that the preconditioners are spectrally equivalent to the original matrices; the constants of equivalence are independent of mesh size and the coefficients of the problem. In particular, we have proposed preconditioners based on:

   (a) algebraic substructuring method; estimates of computational complexity of the implementation of constructed preconditioners are obtained and optimal arithmetic complexity is shown.

   (b) fictitious components method; the proof of an optimality of the considered method is based on the theory of the extension of mesh functions from the original domain into the fictitious embedded domain; a variant of the extension theorem for nonconforming finite element spaces is given.

   (c) nonoverlapping domain decomposition method based on block bordering; it is shown that the preconditioner constructed has an optimal arithmetic complexity.

   (d) domain decomposition method on nonmatching grids; based on the technique of domain decomposition and the fictitious components methods a construction of block diagonal preconditioners for algebraic systems arising in the mortar finite element method is developed.

(2) Using an equivalence between nonconforming finite element methods and hybrid-mixed methods the constructed iterative methods for algebraic systems with symmetric positive definite matrices are extended to saddle-point problems which arise from mixed finite element approximations.

(3) Extensive testing of the newly developed iterative methods and preconditioning techniques are considered on model and real problems. In particular, these methods are applied in the simulator of fluid flow in porous media.

REFERENCES

[1] Y. Achdou and Y. Kuznetsov, *Substructuring preconditioners for finite element methods on nonmatching grids*, East-West J. Numer. Math., 3 (1995), pp. 1–28.

[2] Y. Achdou, Y. Kuznetsov, and O. Pironneau, *Substructuring preconditioners for q_1 mortar element method*, Numer. Math., 71 (1995), pp. 419–449.

[3] R. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[4] T. Arbogast and Z. Chen, *On the implementation of mixed methods as nonconforming methods for second order elliptic problems*, IMA Preprint 1172, Institute for Mathematics and its Applications, Minneapolis, MN, 1993.

[5] D. Arnold and F. Brezzi, *Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates*, RAIRO Model. Math. Anal. Numer., 19 (1985), pp. 7–32.

[6] G. Astrakhantsev, *Fictitious domain method for second order elliptic equation with natural boundary conditions*, U.S.S.R. J. Comput. Maths. Math. Phys., 18 (1978), pp. 118–125.

[7] O. Axelsson, *On multigrid methods of the two-level type*, in Multigrid Methods, W. Hackbush and U. Trottenberg, eds., Lecture Notes in Mathematics, 960, Springer-Verlag, Berlin, 1981, pp. 352–367.

[8] O. Axelsson and P. Vassilevski, *Algebraic multilevel preconditioning methods, I*, Numer. Math., 56 (1989), pp. 157–177.

[9] N. Bakhvalov and M. Orekhov, *Fast methods of solving Poisson's equation*, U.S.S.R. J. Comput. Maths. Math. Phys., 22 (1982), pp. 107–114.

[10] A. Banegas, *Fast Poisson solvers for problems with sparsity*, Math. Comp., 32 (1978), pp. 441–446.

[11] R. Bank, B. Welfert, and H. Yserentant, *A class of iterative methods for solving saddle-point problems*, Numer. Math., 56 (1990), pp. 645–666.

[12] C. Bernardi, Y. Maday, and A. Patera, *A new nonconforming approach to domain decomposition: the mortar element method*, in Nonlinear Partial Differential Equations and Their Applications, H. Brezis and J. Lions, eds., Pitman Research Notes on Mathematics, 290, Longman Scientific&Technical, Harlow, U.K., 1989, pp. 13–51.

[13] O. Besov, V. Iljin, and S. Nikolskij, *Integral Representation of Functions and Embedding Theorems*, vol. 2, Halsted Press, New York, 1978.

[14] J. Bourgat, R. Glowinski, P. Tallec, and M. Vidrascu, *Variational formulations and algorithms for trace operator in domain decomposition calculations*, in Second Int. Symp. on Domain Decomposition Methods for PDEs, T. Chan, R. Glowinski, J. Periaux, and O. Widlund, eds., SIAM, Philadelphia, PA, 1989, pp. 3–17.

[15] D. Braess and R. Verfürth, *Multigrid methods for nonconforming finite element methods*, SIAM J. Numer. Anal., 27 (1990), pp. 979–986.

[16] J. Bramble and J. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp., 50 (1990), pp. 1–18.

[17] J. Bramble, J. Pasciak, and A. Schatz, *The construction of preconditioners for elliptic problems by substructuring, I*, Math. Comp., 47 (1986), pp. 103–134.

[18] ——, *The construction of preconditioners for elliptic problems by substructuring, III*, Math. Comp., 51 (1988), pp. 415–430.

[19] J. Bramble, J. Pasciak, and A. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle-point problems*, Technical Report ISC-94-09-MATH, Institute for Scientific Computation, Texas A&M University, College Station, TX, 1994.

[20] J. Bramble, J. Pasciak, and J. Xu, *The analysis of multigrid algorithms with non-nested spaces or non-inherited quadratic forms*, Math. Comp., 56 (1991), pp. 1–34.

[21] J. Bramble, R.Ewing, J. Pasciak, and J. Shen, *Analysis of the multigrid algorithms for cell-centered finite difference approximations*, Advances in Comp. Mathematics, (1996), to appear.

[22] S. Brenner, *An optimal-order multigrid method for P_1 nonconforming finite elements*, Math. Comp., 52 (1989), pp. 1–16.

[23] ———, *A multigrid algorithm for the lowest-order Raviart-Thomas mixed triangular finite element method*, SIAM J. Numer. Anal., 29 (1992), pp. 647–678.

[24] F. Brezzi, *On the existence, uniqueness, and approximation of saddle point problems arising from Lagrange multipliers*, RAIRO Math. Model. Numer. Anal., 8 (1974), pp. 129–151.

[25] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York, 1991.

[26] F. Brezzi, J. J. Douglas, R. Duràn, and M. Fortin, *Mixed finite elements for second order elliptic problems in three variables*, Numer. Math., 51 (1987), pp. 237–250.

[27] F. Brezzi, J. J. Douglas, and L. Marini, *Two families of mixed elements for second order elliptic problems*, Numer. Math., 88 (1985), pp. 217–235.

[28] F. Brezzi and L. Marini, *A three-field domain decomposition method*, in Sixth Int. Symp. on Domain Decomposition Methods for PDEs, Y. Kuznetsov, J. Periaux, A. Quarteroni, and O. Widlund, eds., Contemporary Math., Amer. Math. Soc., Providence, R.I., 1994, pp. 27–34.

[29] T. Chan and T. Mathew, *Domain decomposition algorithms*, in Acta Numerica, A. Iserles, ed., Cambridge Univ. Press, Cambridge, MA, 1994, pp. 61–143.

[30] G. Chavent and J. Jaffre, *Mathematical Models and Finite Elements for Reservoir Simulation*, Elsevier Science Publishers B.V., Amsterdam, 1986.

[31] H. Chen, R. Ewing, S. Maliassov, I. Mishev, J. Pasciak, and A. Vassilev, *The TAMU two–phase flow simulator: Programmer's guide*, Technical Report ISC-96-01-MATH, Institute for Scientific Computation, Texas A&M University, College Station, TX, 1996.

[32] Z. Chen, *Analysis of mixed methods using conforming and nonconforming finite element methods*, RAIRO Math. Model. Numer. Anal., 27 (1993), pp. 9–34.

[33] Z. Chen, R. Ewing, Y. Kuznetsov, R. Lazarov, and S. Maliassov, *Multilevel preconditioners for mixed methods for second order elliptic problems*, J. Numer. Lin. Alg., submitted.

[34] Z. Chen, R. Ewing, and R. Lazarov, *Domain decomposition algorithms for mixed methods for second order elliptic problems*, Math. Comp., 65 (1996), to appear.

[35] Z. Chen and D. Kwak, *The analysis of multigrid algorithms for nonconforming and mixed methods for second order elliptic problems*, IMA Preprint 1277, Institute for Mathematics and its Applications, Minneapolis, MN, 1994.

[36] P. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[37] L. Cowsar, *Domain decomposition methods for nonconforming finite elements spaces of Lagrange-type*, Technical Report TR 93-11, Department of Comp. & Appl. Math., Rice University, Houston, TX, 1993.

[38] L. Cowsar, J. Mandel, and M. Wheeler, *Balancing domain decomposition for mixed finite elements*, Technical Report TR 93-08, Department of Comp. & Appl. Math., Rice University, Houston, TX, 1993.

[39] J. Douglas, *Alternating direction iteration for mildly nonlinear elliptic difference equations*, Numer. Math., 3 (1963), pp. 92–98.

[40] M. Dryja, *A finite element capacitance method for the elliptic problem*, SIAM J. Numer. Anal., 20 (1983), pp. 671–680.

[41] ———, *A finite element capacitance method for elliptic problems on regions partitioned into subregions*, Numer. Math., 44 (1984), pp. 153–168.

[42] M. Dryja and O. Widlund, *An additive variant of the Schwarz alternating method for the case of many subregions*, Technical Report 339, Department of Computer Science, Courant Institute, New York, 1987.

[43] R. Durán, *Superconvergence for rectangular mixed finite elements*, Numer. Math., 58 (1990), pp. 287–298.

[44] E. D'yakonov, *On an iterative method for solving finite difference equations*, Soviet Math. Dokl., 2 (1961), pp. 647–650.

[45] ———, *The use of spectrally equivalent operators in solving difference analogs of strongly elliptic systems*, Soviet Math. Dokl., 6 (1965), pp. 1105–1109.

[46] ———, *The construction of iterative methods based on the use of spectrally equivalent operators*, USSR Comp. Math. and Math. Phys., 6 (1966), pp. 14–46.

[47] ———, *On triangulations in the finite element method and efficient iterative methods*, in Topics in Numerical Analysis, III, I. Miller, ed., Academic Press, London, 1977, pp. 103–124.

[48] ———, *On some direct and iterative methods based on matrix bordering*, in Numerical Methods in Mathematical Physics, the USSR Academy of Sciences, Siberian Branch, Novosibirsk, 1979, pp. 45–68. (In Russian).

[49] ———, *Optimization in Solving Elliptic Problems*, CRC Press, Boca Raton, FL, 1996.

[50] H. Elman and G. Golub, *Inexact and preconditioned Uzawa algorithms for saddle-point problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1645–1661.

[51] R. Ewing, Y. Kuznetsov, R. Lazarov, and S. Maliassov, *Preconditioning of nonconforming finite element approximations of second order elliptic problems*, in Third Int. Conf. on Advances in Numerical Methods and Applications, I. Dimov, B. Sendov, and P. Vassilevski, eds., World Scientific, Bulgaria, 1994, pp. 101–110.

[52] ———, *Substructuring preconditioning for finite element approximations of second order elliptic problems. I. Nonconforming linear elements for the Poisson equation in parallelepiped*, IMA Preprint 1280, Institute for Mathematics and its Applications, Minneapolis, MN, 1994.

[53] R. Ewing, R. Lazarov, and P. Vassilevski, *Local refinement techniques for elliptic problems on cell-centered grids, I: Error analysis*, Math. Comput., 56 (1991), pp. 437–462.

[54] R. Ewing, R. Lazarov, and J. Wang, *Superconvergence of the velocity along the gauss lines in mixed finite element methods*, SIAM J. Numer. Anal., 28 (1991), pp. 1015–1029.

[55] R. Ewing, S. Maliassov, Y. Kuznetsov, and R. Lazarov, *Substructure reconditioning for porous flow problems*, in Finite Element Modeling of Environmental Problems, G. Garey, ed., John Wiley & Sons, New York, 1995, pp. 303–332.

[56] R. Ewing and M. Wheeler, *Computational aspects of mixed finite element methods*, in Numerical Methods for Scientific Computing, R. Stepleman, ed., North-Holland, New York, 1983, pp. 163–172.

[57] D. Faddeev and V. Faddeeva, *Computational Methods of Linear Algebra*, Freeman, San Francisco, 1963.

[58] R. Glowinski and M. Wheeler, *Domain decomposition and mixed finite element methods for elliptic problems*, in First Int. Symp. on Domain Decomposition Methods for PDEs, R. Glowinski, G. Golub, G. Meurant, and J. Periaux, eds., SIAM, Philadelphia, PA, 1988, pp. 144–172.

[59] G. Golub and C. V. Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, PA, 1989.

[60] M. Griebel, *Grid- and point- oriented multilevel algorithms*, Technical Report TUM-I9224, Inst. für Informatik, TU München, 1992.

[61] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman Publishing, Boston, MA, 1985.

[62] L. Hageman and D. Young, *Applied Iterative Methods*, Acad. Press, New York, 1983.

[63] Y. Hakopian and Y. Kuznetsov, *Algebraic multigrid/substructuring preconditioners on triangular grids*, Sov. J. Numer. Anal. and Math Modeling, 6 (1991), pp. 453–484.

[64] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*, Cambridge Univ. Press, Cambridge, MA, 1952.

[65] M. Hestenes, *The conjugate gradient method for solving linear systems*, in Proc. Symp. Appl. Math. VI, Amer. Math. Soc., Providence, R.I., 1956, pp. 83–102.

[66] V. Iljin and Y. Kuznetsov, *Iterative methods in numerical solution of differential equations*, in Lecture Notes in Mathematics, 704, R. Glowinski and J.-L. Lions, eds., Springer-Verlag, Berlin, 1979, pp. 23–36.

[67] J. J. Douglas and J. Roberts, *Global estimates for mixed methods for second order elliptic equations*, Math. Comp., 44 (1985), pp. 39–52.

[68] Y. Kuznetsov, *Block-relaxation methods in subspaces, their optimization and application*, in Numerical Methods in Applied Mathematics, (Paris, 1978), G. Marchuk and J.-L. Lions, eds., "Nauka" Sibirsk. Otdel., Novosibirsk, Russia, 1982, pp. 119–143. (In Russian).

[69] ———, *Multigrid domain decomposition methods*, in Third Int. Symp. on Domain Decomposition Methods for PDEs, T. Chan, R. Glowinski, J. Periaux, and O. Widlund, eds., SIAM, Philadelphia, PA, 1989, pp. 290–313.

[70] ———, *Multigrid domain decomposition methods for elliptic problems*, Comput. Meth. Appl. Mech. and Eng., 75 (1989), pp. 185–193.

[71] ———, *Multilevel substructuring preconditioners.* Invited presentation at Seventh Int. Symp. on Domain Decomposition Methods for PDEs, Pennsylvania State University, University Park, PA, October 1993.

[72] ———, *Efficient iterative solvers for elliptic finite element problems on nonmatching grids*, Russian J. Numer. Anal. and Math. Modeling, 10 (1995), pp. 187–211.

[73] Y. Kuznetsov and S. Maliassov, *Substructuring preconditioners for nonconforming finite element approximations of second-order elliptic problems with anisotropy*, Russ. J. Numer. Anal. Math. Modeling, 10 (1995), pp. 511–533.

[74] Y. Kuznetsov and M. Wheeler, *Optimal order substructuring preconditioners for mixed finite element methods on nonmatching grids*, East-West J. Numer. Math., 3 (1995), pp. 127–143.

[75] J. Lions, *Problèmes aux Limites Dans les Equations aux Dérivées Partielles*, Presses de l'Université de Montréal, Montréal, 1962.

[76] P. Lions, *On the Schwarz alternating method. I.*, in First Int. Symp. on Domain Decomposition Methods for PDEs, R. Glowinski, G. Golub, G. Meurant, and J. Périaux, eds., SIAM, Philadelphia, PA, 1988.

[77] S. Maliassov, *Substructuring preconditioning for finite element approximations of second order elliptic problems. II. Mixed method for an elliptic operator with scalar tensor*, Technical Report ISC-94-19-MATH, Institute for Scientific Computation, Texas A&M University, College Station, TX, 1994.

[78] ———, *Substructuring domain decomposition method for nonconforming approximations of elliptic problems with anisotropy*, Technical Report ISC-95-11-MATH, Institute for Scientific Computation, Texas A&M University, College Station, TX, 1995.

[79] ———, *Domain decomposition method for nonconforming finite element approximations of anisotropic elliptic problems on nonmatching grids.* In preparation, 1996.

[80] G. Marchuk and Y. Kuznetsov, *On optimal iteration processes*, Soviet Math. Dokl., 9 (1968), pp. 1041–1045.

[81] ———, *Methodes iteratives et fonctionnelles quadratiques*, in Methodes Mathematiques de L'informatique-4: Sur les Methodes Numeriques en Sciences, Physiques et Economiques, J. Lions and G. Marchouk, eds., Dunod, Paris, 1974, pp. 3–131.

[82] G. Marchuk, Y. Kuznetsov, and A. Matsokin, *Fictitious domain and domain decomposition methods*, Soviet J. Numer. Anal. Math. Modeling, 1 (1986), pp. 3–35.

[83] L. Marini, *An inexpensive method for the evaluation of the solution of the lowest order Raviart-Thomas mixed method*, SIAM J. Numer. Anal., 22 (1985), pp. 493–496.

[84] A. Matsokin, *Method of fictitious components and a modified difference analog of the Schwarz method*, in Computational Methods of Linear Algebra, Computing Center of the USSR Academy of Sciences, Siberian Branch, Novosibirsk, Russia, 1980, pp. 66–77.

[85] ———, *Method of fictitious components and the alternating-subdomains method*, in Vistas in Applied Mathematics: Numerical Analysis, Atmospheric Sciences, Immunology, A. Balakrishnan, A. Dorodnitsyn, and J. Lions, eds., Optimization Software, New York, 1986, pp. 127–144.

[86] ———, *Methods of Fictitious Components and Alternating Domains*, Doctoral dissertation, Computing Center of the USSR Academy of Sciences, Siberian Branch, Novosibirsk, Russia, 1988. (In Russian).

[87] ———, *Norm-preserving prolongation of mesh functions*, Soviet J. Numer. Anal. Math. Modeling, 3 (1988), pp. 137–149.

[88] A. Matsokin and S. Nepomnyaschikh, *The Schwarz alternating method in a subspace*, Soviet Mathematics, 29 (1985), pp. 78–84.

[89] ———, *On using the bordering method for solving systems of mesh equations*, Soviet J. Numer. Anal. and Math. Modeling, 4 (1989), pp. 487–492.

[90] M. Nakata, A. Weiser, and M. Wheeler, *Some superconvergence results for mixed finite element methods for elliptic problems on rectangular domains*, in The Mathematics of Finite Elements and Applications, V, J. Rhiteman, ed., Academic Press, London, 1985, pp. 367–389.

[91] J. Necas, *Les Méthodes Directes en Théorie des Equations Elliptiques*, Masson, Paris, 1967.

[92] J. Nedelec, *Mixed finite elements in* $\mathbf{R}^3$, Numer. Math., 35 (1980), pp. 315–341.

[93] S. Nepomnyaschikh, *Domain Decomposition and Schwarz Method in Subspace for Approximate Solution of Elliptic Boundary Value Problems*, PhD thesis, Computing Center of the USSR Academy of Sciences, Siberian Branch, Novosibirsk, Russia, 1986. (In Russian).

[94] ———, *On the application of the bordering method to the mixed boundary value problem for elliptic equations and on mesh norms in $W_2^{1/2}(S)$*, Soviet J. Numer. Anal. and Math. Modeling, 4 (1989), pp. 493–506.

[95] ———, *Schwartz alternating method for solving the singular Neumann problem*, Soviet J. Numer. Anal. and Math. Modeling, 5 (1990), pp. 69–78.

[96] ———, *Mesh theorems on traces, normalizations of function traces and their inversions*, Sov. J. Numer. Anal. Math. Modeling, 6 (1991), pp. 223–242.

[97] ———, *Preconditioners on unstructured grids.* Invited presentation at Copper Mountain Conference on Multigrid Methods, Copper Mountain, CO, April 1995.

[98] C. Page, *Computational variants of the Lanczos method for eigenproblem*, Numer. Math., 15 (1972), pp. 801–812.

[99] B. Parlett, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs, NJ, 1980.

[100] W. Proskurowski and O. Widlund, *On the numerical solution of Helmholtz's equation by the capacitance matrix method*, Math. Comp., 30 (1976), pp. 433–468.

[101] P. Raviart and J. Thomas, *A mixed finite element method for 2-nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, I. Galligani and E. Magenes, eds., Lecture Notes in Mathematics, 606, Springer-Verlag, New York, 1977, pp. 292–315.

[102] ———, *Primal hybrid finite element methods for second order elliptic equations*, Math. Comp., 31 (1977), pp. 391–413.

[103] T. Rossi, *Fictitious Domain Methods with Separable Preconditioners*, PhD thesis, Department of Mathematics, University of Jyväskyla, Finland, December 1995.

[104] T. Russell and M. Wheeler, *Finite element and finite difference methods for continuous flows in porous media*, in The Mathematics of Reservoir Simulation, R. Ewing, ed., SIAM, Philadelphia, PA, 1983, pp. 35–106.

[105] T. Rusten and R. Winter, *A preconditioned iterative methods for saddle point problems*, SIAM J. Matrix Anal., 13 (1992), pp. 887–904.

[106] A. Samarski and E. Nikolaev, *Méthodes de Résolution des Équations de Mailles*, Mir, Moscow, Russia, 1981.

[107] M. Sarkis, *Schwarz Preconditioners for Elliptic Problems with Discontinuous Coefficients Using Conforming and Non-conforming Elements*, PhD thesis, Courant Institute of Mathematical Sciences, New York University, New York, September 1994.

[108] H. Schwarz, *Gesammelte Mahtematische Abhandlungen*, vol. 2, Springer, Berlin, 1890, pp. 133–143.

[109] P. Tallec and T. Sassi, *Domain decomposition with nonmatching grids: Schur complement approach*, Rapport CEREMADE 9323, Université de Paris Dauphine, Paris, 1993.

[110] P. Tallec, T. Sassi, and M. Vidrascu, *Three-dimensional domain decomposition methods with nonmatching grids and unstructured coarse solvers*, Contemporary Mathematics, 180 (1994), pp. 61–74.

[111] P. L. Tallec, *Domain Decomposition Methods in Computational Mechanics, Computational Mechanics Advances*, vol. 2, North-Holland, Amsterdam, 1994.

[112] R. Temam, *Navier-Stokes equations. Theory and numerical analysis*, Studies in Mathematics and its Applications, vol. 2, North-Holland, Amsterdam, 1977.

[113] J. Thomas, *Sur L'analyse Numérique des Méthodes D'eléments Finis Hybrides et Mixtes*, These de Doctorat d'etat, 'a l'Université Pierre et Marie Curie, Paris, 1977.

[114] R. Varga, *Matrix Iterative Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1961.

[115] P. Vassilevski and R. Lazarov, *Preconditioning saddle-point problems arising from mixed finite element discretization of elliptic equations*, UCLA CAM Report 92-46, UCLA, Los Angeles, CA, 1992.

[116] P. Vassilevski and J. Wang, *Multilevel iterative methods for mixed finite element discretizations of elliptic problems*, Numer. Math., 63 (1992), pp. 503–520.

[117] E. Wachspress, *Extended application of alternating direction implicit iteration. Model problem theory*, SIAM J. Numer. Anal., 11 (1963), pp. 994–1016.

[118] A. Weiser and M. Wheeler, *On convergence of block-centered finite-differences for elliptic problems*, SIAM J. Numer. Anal., 25 (1988), pp. 351–375.

[119] O. Widlund, *An extension theorem for finite element spaces with three applications*, in Numerical Techniques in Continuum mechanics: Notes on Numerical Fluid Mechanics, W. Hackbush and K. Witsch, eds., vol. 16, Friedr. Vieweg und Sohn, Braunschweig/Weisbaden, Germany, 1987, pp. 110–122.

[120] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Rev., 34 (1992), pp. 581–613.

[121] O. Zienkiewicz and K. Morgan, *Finite Elements and Approximations*, Wiley-Interscience Pub., New York, 1983.