NON-CONFORMING MIXED FINITE ELEMENT METHODS FOR DIFFUSION EQUATION

A Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Evgeny Kikinzon

December 2011

NON-CONFORMING MIXED FINITE ELEMENT METHODS FOR DIFFUSION EQUATION

Evgeny Kikinzon

APPROVED:

Dr. Yuri Kuznetsov, Chairman

Dr. Ronald Hoppe,

Dr. Tsorng-Whay Pan,

Dr. Vadim Dyadechko,
ExxonMobil Upstream Research Company

Dr. Mark A. Smith
Dean, College of Natural Sciences
and Mathematics

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitute to my advisor, Dr. Yuri Kuznetsov. He introduced me to the field of numerical methods for partial differential equations and was guiding me from the very first steps I have taken and all the way to the completion of this research. Without his trust, instructions, encouragement, and support, this dissertation would never have been written.

My special thanks go to Dr. Ilya Timofeyev, who convinced me to pursue deeper study of mathematics and invited me to the University of Houston. As my advisor during my first years in the Department of Mathematics, he did a lot to make the transition to a different field and new environment go smoothly for me.

A significant part of this dissertation is based on the problems formulated in collaboration with Dr. Rossen Parashkelov from ExxonMobil Upstream Research Co. I would like to thank him for important suggestions, and I would like to thank ExxonMobil Corporation for a continuous support during the years it took to complete this research.

I am grateful to the members of the committee, Dr. Ronald Hoppe and Dr. Tsorng-Way Pan, who found the time to read the manuscript and made many important remarks. My special thanks go to Dr. Vadim Dyadechko from ExxonMobil Upstream Research Co., who agreed to review this dissertation despite his busy schedule, and provided many interesting comments.

I would like to thank my friends and colleagues Dr. Oleg Boyarkin and Dr. Andrey Prokopenko for their help and advice during the years we have spent working together. I am thankful to Dr. Harbir Antil who was a good friend and an example to follow all the time we know each other.

Finally, my very special thanks go to my parents, who were very enthusiastic in their encouragement and support during all these years.

NON-CONFORMING MIXED FINITE ELEMENT METHODS FOR DIFFUSION EQUATION

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Evgeny Kikinzon

December 2011

ABSTRACT

In this dissertation, we consider new approaches to the construction of meshes, discretization, and preconditioning of the resulting algebraic systems for the diffusion equation with discontinuous coefficients.

In the first part, we discuss mixed finite element approximations of the diffusion equation on general polyhedral meshes. We introduce a non-conforming approximation method for the flux vector functions, and propose a benchmark problem which allows us to analyze its accuracy in the case of 3D diffusion equation with non-homogeneous boundary conditions on domains with oblique parallel layers.

In the second part, we propose a two-stage preconditioning method for the algebraic system resulting from the application of the introduced method to the diffusion equation on the prismatic meshes. We provide the description of the recommended implementation and show the results of numerical experiments used to compare its performance with some well-known preconditioners.

In the third part, we consider application of non-conforming meshes on rectangular domains with oblique parallel or curved concentric layers. We describe possible choices of such meshes for each case, and introduce benchmark problems used to compare the accuracy of finite element methods on conforming and non-conforming meshes. The results of numerical experiments are provided.

Contents

Chapter 1

Introduction

1.1 Review of approximation methods for the diffusion equation

There are many discretization methods which were developed for the second order diffusion equation. Among widely used methods are Finite Differences (FD), Finite Element (FE), Finite Volume (FV), Mimetic Finite Differences, and Mixed Finite Element (MFE) methods.

In the FD method all derivatives are replaced with finite differences. FD are widely used for uniform rectangular grids. The main advantage of this method is in its simplicity, but it also has many disadvantages, for example, its practical applications are restricted to rectangular grids. Also, the implementation of the boundary conditions, especially for domains with curved boundaries, may worsen the existing approximation inside the computational domain. For the complete presentation of the FD method we refer to [41] and references found within.

FE methods are perhaps the most popular and powerful methods in modern numerical applications. One of the first papers on FE is the paper by Courant [13]. The

term "finite element method" was proposed by R.W. Clough in [12].

The main idea of the FE method is based on the approximation of the weak solution of the diffusion equation, i.e. the solution of the variational problem. The weak solution of the diffusion problem belongs to a certain explicitly constructed Hilbert space Q. The existence and uniqueness of the solution is proved by using the properties of Hilbert spaces. We construct some finite dimensional subspace Q_h of Q based on a partitioning of the computational domain Ω. Partitioning means that we split our domain Ω into a set of subdomains, called elements. A solution $p_h \in Q_h$ of the corresponding finite element problem is called a finite element approximation of the solution p of the corresponding differential problem.

The main advantage of the FE method over the FD method is that its applications are not restricted by the geometry, therefore the method can be applied to problems in domains with complex shapes. The application of FE method to domains with curved boundaries is investigated for instance in [4, 5, 30, 48, 49]. Discretization of the second order diffusion equation with the FE method leads to stable and robust algorithms.

The Finite Volume (FV) approximation method allows to obtain locally conservative schemes. The FV method is a Petrov-Galerkin type method for solving boundary value problems, where the solution space is different from the test space. The solution space Q_h is the same as in the mixed finite element method, but the test space Q_h^* is defined on a dual mesh, which is called Voronoi mesh. For further information we refer to [14].

Another extensively used discretization technique for the diffusion equation is the Mimetic Finite Differences method. This discretization methodology is based on the support operator approach, see [17, 32, 34, 35]. This approach requires the constructed discrete operators to preserve main physical properties of the original differential operator, including conservation law, solution symmetries, and so on. In

the case of the linear diffusion problem, the mimetic discretization mimics the Gauss divergence theorem to enforce the local conservation law and preserves the symmetry between the discrete gradient and divergence operators. It also preserves the null spaces of the those operators and guaranties the stability of the discretization.

The term "Mixed Method" is used for problems with two or more physical variables. For the second-order diffusion equation the corresponding formulation can be written as follows:

$$\begin{aligned} K^{-1}\boldsymbol{u} \;+\; \nabla p \;&=\; 0 \text{ in } \Omega, \\ \nabla\cdot\boldsymbol{u} \;+\; cp \;&=\; f \text{ in } \Omega. \end{aligned} \tag{1.1}$$

The unknown vector function $\boldsymbol{u}$ introduced here is called the flux. If mixed formulation is used, both flux $\boldsymbol{u}$ and pressure p solution functions are computed simultaneously.

Applying an MFE method to the diffusion equation of the form (1.1), we replace the first order system of differential equations by a variational problem on two Hilbert spaces, the space $\boldsymbol{V}$ for fluxes and the space Q for pressures. A finite element solution $(\boldsymbol{u}_h, p_h)$ belongs to the space $\boldsymbol{V}_h \times Q_h$, where $\boldsymbol{V}_h$ and Q_h are finite dimensional subspaces of $\boldsymbol{V}$ and Q, respectively.

It is required for the chosen subspaces $\boldsymbol{V}_h$ and Q_h to satisfy so called LBB (Ladyzhenskaya-Babushka-Brezzi) condition,

$$\beta||q||_Q \;\leq\; \sup \frac{(\nabla\cdot\boldsymbol{v},\; q)}{||\boldsymbol{v}||_{H(\mathrm{div},\Omega)}}, \tag{1.2}$$

for all $q \in Q_h$ and a certain constant $\beta > 0$. This condition is required for the stability of the numerical solution.

In classical literature, finite element spaces $\boldsymbol{V}_h$ for fluxes are constructed on "simple" cells, such as triangles and rectangles in 2D, and tetrahedra, triangular prisms, rectangular parallelepipeds in 3D. The examples of such spaces include the Raviart-Thomas spaces RT_m, Brezzi-Douglas-Fortin-Marini spaces $BDFM_m$, and Brezzi-

Douglas-Marini spaces BDM_m, which are introduced and investigated in [7, 8, 11, 36, 38, 46].

Using the change of variables we can generalize the classical FE spaces into ones suitable for general convex quadrilaterals in 2D, and hexahedral or distorted prismatic cells in 3D. Mixed finite element spaces based on the Piola transformation of the vector fields are investigated in [6, 42, 46]. Error estimates are strongly dependent on the properties of the Jacobian of transformation.

In [25, 26] Yu. Kuznetsov and S. Repin introduced a new approach to define a space of fluxes $\boldsymbol{V}_h$ on general polygonal (2D) and polyhedral (3D) meshes. The discretization is based on the partitioning of a particular polygonal or polyhedral macrocell into "simple" cells, the space of fluxes on this macrocell is then defined as a subset of the corresponding RT_0 space on this macrocell. The condition $\nabla \cdot \boldsymbol{u}_h = \text{const}$ imposed on a macrocell allows to eliminate the degrees of freedom on auxiliary interfaces between cells of its local partitioning. Since Q_h is the space of piece-wise constant functions, this condition ensures that the LBB condition is satisfied and the method is stable. In [22], this approach was extended to the mimetic finite difference method.

Yu. Kuznetsov proposed a new discretization method for 2D diffusion equation on polyhedral meshes with mixed cells in [20]. The method was further extended to 3D diffusion problems in [21].

1.2 Review of solution methods

Every discretization scheme (finite elements, finite differences, finite volumes) of the diffusion equation results in an algebraic system with a sparse matrix. In many cases, this matrix is symmetric and positive definite, or positive semi-definite. Often, it is an M-matrix. In each case, producing a solution efficiently on a fine mesh is a challenging task.

The demands of the users in the engineering applications results in systems with tens or hundreds of millions of unknowns. Standard direct methods are usually considered as inappropriately slow for these systems. On the other hand, due to coefficient heterogeneity, coefficient anisotropy, or mesh anisotropy, system matrices have large condition numbers, which results in slow convergence of unpreconditioned iterative solvers.

The use of preconditioners leads to significant improvement in convergence rates. Classical preconditioners, such as Jacobi, Gauss-Seidel, SOR, and SSOR (see *e.g.*, [47]) are effective for a number of simple problems. A combination of these methods with nested iterations was disscussed in [29]. However, these preconditioners are not numerically scalable, i.e. the increase in computational work is not linear with respect to the number of unknowns, therefore they can not meet the efficiency requirement of current applications.

The development of multigrid methods [15, 16, 3, 1, 9, 10] in the 1960s provided a solution to this problem, as such methods, under some restrictions, are numerically scalable. Originally, these methods were tightly connected with the model geometry, specifically, the mesh grid. Geometric multigrids operated on a hierarchy of meshes obtained *a priori* by a coarsening procedure. The increase in complexity of the mesh grids slowed down the development of such methods, which, in turn, resulted in development of algebraic multigrids.

In algebraic multigrid (AMG) methods the coarsening procedure is based on a coefficient matrix instead of the mesh grid. The introductory articles of 1980s [44, 18, 2] have created a new direction in the research of multigrid methods. An important feature of many such methods is that they can be used as a black-box algorithm, i.e. the only input for the coarsening procedures is the coefficient matrix. One such preconditioner was proposed by K.Stüben and his collaborators [44, 33, 40, 45]. One of its versions, `amg1r5` [33], is available to public, and can be used for any symmet-

ric positive semi-definite system. However, the code may stagnate on geometrically anisotropic problems [40], and has a number of other drawbacks. Later versions of the algorithm, `RAMG05` and `SAMG`, solve most of these issues [45].

A different algebraic multigrid method was proposed by Yu. Kuznetsov [18, 19]. The developed preconditioner is spectrally equivalent to the system matrix, and provides the linear increase in the computational work with respect to the number of unknowns. However, this preconditioner requires *a priori* knowledge of the mesh grid.

This multilevel framework was extended in [24] to general systems with symmetric M-matrices with strict diagonal domination. In particular, it can be used with matrices arising from the discretizations of the diffusion equation with heterogeneous coefficients.

A similar approach (referred as algebraic multilevel iteration, AMLI) with an inner Chebyshev iterative procedure was developed by Axelsson and Vassilevski [2] and then extended to anisotropic problems [31]. However, non-uniform meshes are not considered in this approach.

There are very few preconditioners for diffusion equation on meshes with faults. An example of such preconditioner is discussed in [28].

1.3 Dissertation outline

The dissertation is organized as follows. The focus of Chapter 2 is on MFE approximations of the diffusion equation on prismatic meshes. In Section 2.1 we start from the description of the differential diffusion problem and describe the transition to the corresponding macro-hybrid mixed formulation. In Section 2.2 we give a description of the computational domain and the prismatic mesh used. Section 2.3 provides the description of the FE spaces used in Kuznetsov-Repin method. In Section 2.4 the

application of FE method to the macro-hybrid mixed formulation diffusion problem is described. Section 2.5 gives the description of so called piece-wise constant approximation method for the flux vector functions.

In Chapter 3 we describe a benchmark problem for the the 3D diffusion problem on domains with oblique parallel layers. We start with the problem formulation and derive the reference solution. Then, we illustrate its application by comparing the accuracy of the KR and PWC finite element methods on prismatic meshes.

In Chapter 4 we propose a two-stage preconditioning method for the diffusion equation on prismatic meshes. We start from the general framework, propose a particular implementation and compare the performance results in numerical experiments.

Chapter 5 focuses on the application of non-conforming meshes to rectangular domains with parallel oblique layers. We start with the variation of the benchmark problem from Chapter 3, then describe the non-conforming mesh used, and show numerical results illustrating the impact on accuracy its application might have.

In Chapter 6 we extend the application of non-conforming meshes to the rectangular domains with concentric curved layers. We describe a benchmark problem suitable for studying this case, propose a particular implementation of non-conforming mesh, and give the results of numerical experiments used to estimate the resulting accuracy.

Chapter 2

Mixed finite element method on prismatic meshes

2.1 Problem formulation

2.1.1 Differential formulation

We consider the diffusion equation

$$-\nabla \cdot \big(K \nabla p\big) \; + \; cp \; = \; f \quad \text{in} \quad \Omega \; , \tag{2.1}$$

where p is an unknown solution function (pressure), $K = K(x) \in \mathbb{R}^{3\times 3}$ is a diffusion tensor, $c = c(x)$ is a positive function, $f = f(x)$ is a source function, and Ω is a simply connected bounded polyhedral domain in $\mathbb{R}^3$ with boundary $\partial\Omega$. We assume that the functions c and f, as well as the entries of the diffusion tensor K, are piecewise smooth and bounded. We also assume that the matrix (tensor) K is symmetric and positive definite at any point $x \in \Omega$.

Let $\partial\Omega$ be partitioned into two non-overlapping pieces Γ_D and Γ_N, i.e. $\Gamma_D \cup \overline{\Gamma}_N = \partial\Omega$.

Then, equation (2.1) is complemented with the boundary conditions

$$\begin{aligned} p &= g_D \quad \text{on} \quad \Gamma_D, \\ -\big(K\nabla p\big)\cdot \boldsymbol{n} &= g_N \quad \text{on} \quad \Gamma_N, \end{aligned} \tag{2.2}$$

where g_D and g_N are given functions defined on Γ_D (Dirichlet part of $\partial\Omega$) and Γ_N (Neumann part of $\partial\Omega$), respectively, and $\boldsymbol{n}$ is the outward unit normal vector to $\partial\Omega$.

Let us introduce the flux vector-function by

$$\boldsymbol{u} = -K\,\nabla p\ . \tag{2.3}$$

Then, formulation (2.1), (2.2) is equivalent to the boundary value problem for the system of first-order differential equations

$$\begin{aligned} K^{-1}\boldsymbol{u} + \nabla p &= 0 \quad \text{in} \quad \Omega, \\ \nabla\cdot\boldsymbol{u} + cp &= f \quad \text{in} \quad \Omega, \\ p &= g_D \quad \text{on} \quad \Gamma_D, \\ \boldsymbol{u}\cdot\boldsymbol{n} &= g_N \quad \text{on} \quad \Gamma_N\ . \end{aligned} \tag{2.4}$$

In this paper, we shall use only the latter, so called mixed, formulation.

2.1.2 Mixed variational formulation

Let

$$\boldsymbol{V} \;=\; H(\mathrm{div},\ \Omega), \quad Q \;=\; L_2(\Omega), \quad \text{and} \quad \Lambda_N \;=\; L_2(\Gamma_N) \tag{2.5}$$

be the spaces for flux vector-function $\boldsymbol{u}$ and scalar functions p and λ, respectively. Then, the classical mixed formulation of (2.4) is as follows: find $(\boldsymbol{u}, p, \lambda) \in \boldsymbol{V} \times Q \times \Lambda_N$ such that

$$\begin{aligned}
\int_\Omega \left(K^{-1}\boldsymbol{u}\right)\cdot \boldsymbol{v}\,\mathrm{dx} \;-\; \int_\Omega p(\nabla\cdot\boldsymbol{v})\mathrm{dx} \;+\; \int_{\Gamma_N} \lambda(\boldsymbol{v}\cdot\boldsymbol{n})\mathrm{ds} \;&=\; -\int_{\Gamma_D} g_D(\boldsymbol{v}\cdot\boldsymbol{n})\mathrm{ds} \\
\int_\Omega (\nabla\cdot\boldsymbol{u})q\,\mathrm{dx} \;+\; \int_\Omega cpq\,\mathrm{dx} \;&=\; \int_\Omega fq\,\mathrm{dx} \\
\int_{\Gamma_N} (\boldsymbol{u}\cdot\boldsymbol{n})\mu\,\mathrm{ds} \;&=\; \int_{\Gamma_N} g_N\mu\,\mathrm{ds}
\end{aligned} \tag{2.6}$$

for all $\left(\boldsymbol{v}, q, \mu\right) \in \boldsymbol{V} \times Q \times \Lambda_N$.

2.1.3 Macro-hybrid mixed formulation

Let Ω be partitioned into m non-overlapping polyhedral subdomains E_s with boundaries ∂E_s and interfaces between boundaries $\Gamma_{st} = \partial E_s \bigcap \partial E_t,\ s,t = \overline{1,\ m}$. We assume that all nonzero interfaces Γ_{st} are simply connected pieces of piece-wise planar surfaces, $s,t = \overline{1,\ m}$. We denote the union of all nonzero interfaces Γ_{st} by Γ, i.e. $\Gamma = \bigcup\limits_{s,t} \Gamma_{st}$, and denote the intersections of Γ_N with E_s by $\Gamma_{N,s}$, $s = \overline{1,\ m}$.

Let

$$\begin{aligned}
\boldsymbol{V}_s \;&=\; H\big(\mathrm{div},\ E_s\big), \quad Q_s \;=\; L_2\big(E_s\big), \\
\Lambda_{N,s} \;&=\; L_2\big(\Gamma_{N,s}\big), \quad \Lambda_{st} \;=\; L_2\big(\Gamma_{st}\big),
\end{aligned} \tag{2.7}$$

be the spaces of vector-functions $\boldsymbol{u}$ and functions p defined in E_s, functions λ defined on $\Gamma_{N,s}$, and functions λ defined on Γ_{st}, $s,t=\overline{1,\ m}$, respectively.

We define new spaces

$$\begin{aligned}
\boldsymbol{V} &= \boldsymbol{V}_1 \times \boldsymbol{V}_2 \times \cdots \times \boldsymbol{V}_m, \\
Q &= Q_1 \times Q_2 \times \cdots \times Q_m, \\
\Lambda_N &= \Lambda_{N,1} \times \Lambda_{N,2} \times \cdots \times \Lambda_{N,m}, \\
\Lambda_\Gamma &= \prod_{1\le s<t\le m} \Lambda_{st}, \\
\Lambda &= \Lambda_\Gamma \times \Lambda_N.
\end{aligned} \tag{2.8}$$

Then, the macro-hybrid mixed formulation of differential problem (2.4) reads as follows: find $(\boldsymbol{u},p,\lambda)\in \boldsymbol{V}\times Q\times\Lambda$ such that the equations in E_s:

$$\begin{aligned}
\int_{E_s}\Big(K^{-1}\boldsymbol{u}_s\Big)\cdot\boldsymbol{v}_s\,\mathrm{dx} \;-\; \int_{E_s} p_s\big(\nabla\cdot\boldsymbol{v}_s\big)\,\mathrm{dx} \;+\; \int_{\Gamma_s}\lambda\big(\boldsymbol{v}_s\cdot\boldsymbol{n}_s\big)\,\mathrm{ds} \;&=\\
= \;-\int_{\Gamma_{D,s}} g_D\big(\boldsymbol{v}_s\cdot\boldsymbol{n}_s\big)\,\mathrm{ds}, \\
\int_{E_s}(\nabla\cdot\boldsymbol{u}_s)q_s\,\mathrm{dx} \;+\; \int_{E_s} c p_s q_s\,\mathrm{dx} \;&=\; \int_{E_s} f q_s\,\mathrm{dx},
\end{aligned} \tag{2.9}$$

$s=\overline{1,\ m}$, with the variational equations of the continuity of normal fluxes on Γ_{st}:

$$\int_{\Gamma_{st}}\Big[\boldsymbol{u}_s\cdot\boldsymbol{n}_s \;+\; \boldsymbol{u}_t\cdot\boldsymbol{n}_t\Big]\mu_{st}\,\mathrm{ds} \;=\; 0, \tag{2.10}$$

$s,t=\overline{1,\ m}$, and with the variational equations for the Neumann boundary condition:

$$\int_{\Gamma_{N,s}}\big(\boldsymbol{u}_s\cdot\boldsymbol{n}_s\big)\mu_{N,s}\,\mathrm{ds} \;=\; \int_{\Gamma_{N,s}} g_N\mu_{N,s}\,\mathrm{ds}, \tag{2.11}$$

$s = \overline{1,\ m}$, are satisfied for any $(\boldsymbol{v}, q, \mu) \in \boldsymbol{V} \times Q \times \Lambda$. Here, $\boldsymbol{n}_s$ is the unit outward normal to ∂E_s, $\Gamma_s = \partial E_s \setminus \Gamma_D$ and $\Gamma_{D,s} = \partial E_s \bigcap \Gamma_D$ are the non-Dirichlet and the Dirichlet parts of the boundary ∂E_s, respectively, $s = \overline{1,\ m}$.

It is clear that $\boldsymbol{u}_s \in \boldsymbol{V}_s$ and $p_s \in Q_s$ are functional components of $\boldsymbol{u} \in \boldsymbol{V}$ and $p \in Q$ in E_s, respectively, $s = \overline{1,\ m}$.

2.2 Computational domain and mesh

2.2.1 Layered computational domain

Let G be a simply connected polygon in the (x_1, x_2)-plane, and $\widehat{G} = G \times (-\infty;\ +\infty)$ be an unbounded domain in $\mathbb{R}^3$ with vertical planar faces. We introduce a set of continuous piece-wise linear surfaces in $\widehat{G}$ by

$$x_3 \ = \ S_l(x_1,\ x_2), \quad (x_1,\ x_2) \in G, \tag{2.12}$$

where $S_l = S_l(x_1, x_2)$ are single-valued functions, $l = \overline{0,\ L}$.

We assume that

$$S_{l-1}(x_1, x_2) \ \leq \ S_l(x_1, x_2), \quad (x_1, x_2) \in G, \quad l = \overline{1,\ L}\ . \tag{2.13}$$

We define the computational domain Ω as

$$\Omega \ = \ \Big\{ x \,:\, S_0(x_1, x_2) \ < \ x_3 \ < \ S_L(x_1, x_2), \quad (x_1, x_2) \in G \Big\}\ . \tag{2.14}$$

The surfaces $x_3 = S_{l-1}(x_1, x_2)$ and $x_3 = S_l(x_1, x_2)$, $(x_1, x_2) \in G$, $1 \leq l \leq L$, naturally split Ω into subdomains (*e.g.* geological layers) Ω_l, defined by

$$\Omega_l \ = \ \Big\{ x \,:\, S_{l-1}(x_1, x_2) \ < \ x_3 \ < \ S_l(x_1, x_2), \quad (x_1, x_2) \in G \Big\}\ . \tag{2.15}$$

We denote the interface between subdomains Ω_{l-1} and Ω_l by $I_{l-1,l}$, and call the sets

$$P_{l-1,l} \ = \ \Big\{ x \,:\, x_3 = S_{l-1}(x_1, x_2) = S_l(x_1, x_2), \quad (x_1, x_2) \in G \Big\} \tag{2.16}$$

"pinchouts", $l = \overline{1, L}$. By the definition, a pinchout $P_{l-1,l}$ may have nonzero intersection with $P_{l-2,l-1}$, or $P_{l,l+1}$, or both. We also define the sequence of sets $G_{l-1,l}$ in G by

$$G_{l-1,l} \ = \ \Big\{ (x_1, x_2) \,:\, S_{l-1}(x_1, x_2) = S_l(x_1, x_2),\ (x_1, x_2) \in G \Big\}, \quad l = \overline{1,\ L}\ . \tag{2.17}$$

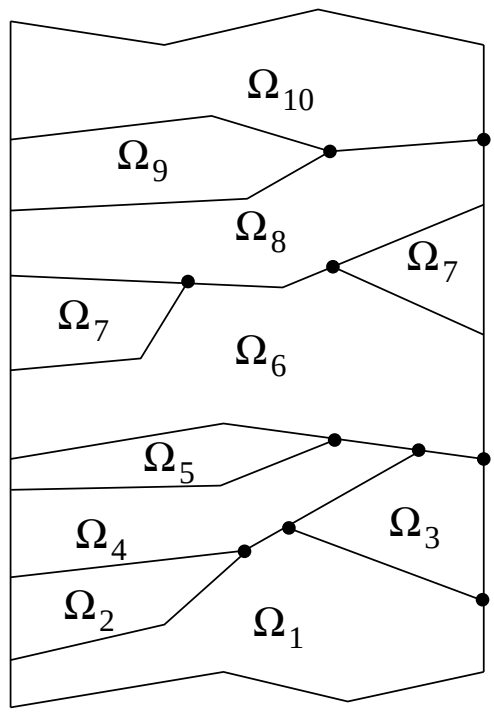

Figure 2.1: An example of partitioning of Ω into ten layered subdomains

A two-dimensional example of Ω partitioned into subdomains Ω_l, $l = \overline{1,\ 10}$, is shown on Figure 2.1.

For the sake of simplicity, we assume that pinchouts $P_{l-1,l}$ are simply connected sets, $l = \overline{1,\ L}$. We denote the boundaries of $P_{l-1,l}$ by $\partial P_{l-1,l}$, $l = \overline{1,\ L}$. In Figure 2.1, they are marked by dots.

2.2.2 Definition of a prismatic mesh

Let G_H be a conforming triangular mesh in G, i.e. any two different triangles in G_H have a common edge, or a common vertex, or do not intersect. We define in Ω a set of continuous piece-wise linear surfaces

$$x_3 \;=\; S_{H,t}(x_1, x_2), \tag{2.18}$$

where $S_{H,t} \equiv S_{H,t}(x_1, x_2)$ are single-valued functions, $t = \overline{0,\ T}$, and T is a positive integer. We always assume that

$$S_{H,0}(x_1, x_2) \;=\; S_0(x_1, x_2), \quad S_{H,T}(x_1, x_2) \;=\; S_L(x_1, x_2) \quad \text{in} \quad G, \tag{2.19}$$

and

$$S_{H,t-1}(x_1, x_2) \;\leq\; S_{H,t}(x_1, x_2) \quad \text{in} \quad G, \quad t = \overline{1,\ T}\ . \tag{2.20}$$

We impose two major restrictions on the set of the surfaces $\{S_{H,t}\}$:

1. For any integer t, $1 \leq t \leq T$, there exists an integer l, $1 \leq l \leq L$, such that

$$S_{l-1}(x_1, x_2) \;\leq\; S_{H,t}(x_1, x_2) \;\leq\; S_l(x_1, x_2) \tag{2.21}$$

for all $(x_1, x_2) \in G$, i.e. the surfaces $\{S_{H,t}\}$ do not cross the surfaces $\{S_l\}$.

2. If the surface $S_{H,t}$, $1 \leq t \leq T$, satisfies inequalities (2.21), then

$$S_{H,t-1}(x_1, x_2) \;<\; S_{H,t}(x_1, x_2) \tag{2.22}$$

for all $(x_1, x_2) \in G \setminus G_{l-1,l}$, i.e. any two neighboring surfaces $S_{H,t-1}$ and $S_{H,t}$, $1 \leq t \leq T$, do not create pinchouts in addition to the pinchouts $P_{l-1,l}$, $1 \leq l \leq L$.

We shall call $S_{H,t}$, $t = \overline{0,\ T}$, the "horizontal" mesh surfaces.

The mesh Ω_H in Ω is defined as an intersection of mesh surfaces $x_3 = S_{H,t}(x_1, x_2)$, $t = \overline{0,\ T}$, with a set of infinite prisms $\Big\{E_G \times (-\infty;\ +\infty)\Big\}$, where E_G is some triangle in G_H. Ω_H is conforming and consists of mesh cells denoted by E. We assume that the surfaces $x_3 = S_l(x_1, x_2)$ and $x_3 = S_{H,t}(x_1, x_2)$ are planar for each cell E_G in G_H. Then, each mesh cell $E \in \Omega_H$ is either a "vertical" prism with two "horizontal" and three vertical nonzero faces, or a degenerated "vertical" prism when there is one or two zero vertical faces. A degenerated mesh cell is either a pyramid (one vertical face is zero), or a tetrahedron (two vertical faces are zero).

Remark. A weaker practical requirement concerning local behavior of surfaces S_l, $0 \le l \le L$, and $S_{H,t}$, $0 \le t \le T$, would be the assumption that they are "almost planar" for each mesh cell $E_G \in G_H$, i.e. they can be approximated with reasonable accuracy by surfaces which are planar for each $E_G \in G_H$.

2.3 Definition of "div-const" FE spaces

To define the FE space for the flux vector-functions we assume that each prismatic mesh cell $e \in \Omega_h$ is partitioned into three tetrahedrons $\triangle_1$, $\triangle_2$, and $\triangle_3$, and each pyramidal mesh cell $e \in \Omega_h$ is partitioned into two tetrahedrons $\triangle_1$ and $\triangle_2$. We denote by $RT_0(e)$ the classical lowest order Raviart-Thomas FE space of vector-functions based on the above partitioning of e into tetrahedrons [6], [39].

Let e be a mesh cell in Ω_h with s planar faces f_i, $i = \overline{1,\ s}$. It is clear that $s = 5$ for "vertical" prisms and pyramids, and $s = 4$ for tetrahedrons. The FE space for the flux vector-functions on e, $\boldsymbol{V}_h(e)$, is defined as follows:

$$\begin{aligned} \boldsymbol{V}_h(e) \ = \ \Big\{ \ & \boldsymbol{v}_h : \boldsymbol{v}_h \in RT_0(e),\ \boldsymbol{v}_h \cdot \boldsymbol{n}_e \equiv \text{const}_i \text{ on } f_i,\ i = \overline{1,s}, \\ & \nabla \cdot \boldsymbol{v}_h \equiv \text{const in } e \ \Big\} . \end{aligned} \tag{2.23}$$

Here, $\boldsymbol{n}_e$ is the outward unit normal to the boundary ∂e of e. The detailed analysis of the space $\boldsymbol{V}_h(e)$ can be found in [25], [27].

We define the FE space $Q_h(e)$ for the solution function p by

$$Q_h(e) \ = \ \Big\{ \ q_h : q_h \equiv \text{const in } e \ \Big\} . \tag{2.24}$$

The global FE spaces for the flux vector-function and the solution function on Ω_h which is partitioned into cells e_s, $s = \overline{1,\ m}$, are defined similar to (2.8) as

$$\boldsymbol{V}_h \ = \ \boldsymbol{V}_{h,1} \times \boldsymbol{V}_{h,2} \times \ldots \times \boldsymbol{V}_{h,m} \tag{2.25}$$

and

$$Q_h \ = \ Q_{h,1} \times Q_{h,2} \times \ldots \times Q_{h,m} \ , \tag{2.26}$$

respectively. Here,

$$\boldsymbol{V}_{h,s} \ = \ \boldsymbol{V}_h(e_s) \quad \text{and} \quad Q_{h,s} \ = \ Q_h(e_s), \quad s = \overline{1,\ m} \ . \tag{2.27}$$

Finally, the FE space $\Lambda_h \equiv \Lambda_h(\Gamma \bigcup \Gamma_N)$ for the Lagrange multipliers is defined as

$$\Lambda_h \;=\; \left\{ \lambda_h \,:\, \lambda_h|_f \equiv \text{const}_f \text{ on any face } f \text{ in } \Omega_h \text{ s.t. } f \subset \Gamma \bigcup \Gamma_N \right\} . \tag{2.28}$$

2.4 Macro-hybrid mixed FE method on prismatic mesh

The macro-hybrid mixed FE discretization of (2.9)-(2.11) reads as follows: find $(\boldsymbol{u}_h, p_h, \lambda_h) \in \boldsymbol{V}_h \times Q_h \times \Lambda_h$ such that the equations in E_s:

$$\begin{aligned}
\int_{E_s} \left(K^{-1}\boldsymbol{u}_{h,s}\right) \cdot \boldsymbol{v}_s \,\mathrm{dx} \;-\; \int_{E_s} p_{h,s}\big(\nabla \cdot \boldsymbol{v}_s\big)\,\mathrm{dx} \;+\; \int_{\Gamma_s} \lambda_h\big(\boldsymbol{v}_s \cdot \boldsymbol{n}_s\big)\,\mathrm{ds} \;&=\\
= \;-\int_{\Gamma_{D,s}} g_D\big(\boldsymbol{v}_s \cdot \boldsymbol{n}_s\big)\,\mathrm{ds}, \qquad &(2.29)\\
\int_{E_s} \big(\nabla \cdot \boldsymbol{u}_{h,s}\big) q_s \,\mathrm{dx} \;+\; \int_{E_s} c p_{h,s} q_s \,\mathrm{dx} \;&=\; \int_{E_s} f q_s \,\mathrm{dx},
\end{aligned}$$

$s = \overline{1,\ m}$, with the variational equations of the continuity of normal fluxes on Γ_{st}:

$$\int_{\Gamma_{st}} \Big[\boldsymbol{u}_{h,s} \cdot \boldsymbol{n}_s \;+\; \boldsymbol{u}_{h,t} \cdot \boldsymbol{n}_t\Big] \mu \,\mathrm{ds} \;=\; 0, \tag{2.30}$$

$s, t = \overline{1,\ m}$, and with the variational equations for the Neumann boundary condition:

$$\int_{\Gamma_{N,s}} \big(\boldsymbol{u}_{h,s} \cdot \boldsymbol{n}_s\big) \mu \,\mathrm{ds} \;=\; \int_{\Gamma_{N,s}} g_N \mu \,\mathrm{ds}, \tag{2.31}$$

$s = \overline{1,\ m}$, are satisfied for any $(\boldsymbol{v}, q, \mu) \in \boldsymbol{V}_h \times Q_h \times \Lambda_h$.

The FE problem (2.29)-(2.31) results in the algebraic equations:

$$\begin{aligned}
M_s \bar{u}_s \;+\; B_s^T \bar{p}_s \;+\; C_s^T \bar{\lambda} \;&=\; \bar{g}_{D,s},\\
B_s \bar{u}_s \;-\; \Sigma_s \bar{p}_s \;&=\; \bar{f}_s,
\end{aligned} \tag{2.32}$$

$s = \overline{1,\ m}$, complemented by the algebraic equations

$$C \begin{pmatrix} \bar{u}_1 \\ \vdots \\ \bar{u}_m \end{pmatrix} \;=\; \bar{g}_N \;. \tag{2.33}$$

The latter equations represent the continuity conditions for the normal fluxes on the interfaces between neighboring cells in Ω_H, and the Neumann boundary conditions on Γ_N.

Here, M_s is a square $n_{u,s} \times n_{u,s}$ symmetric positive definite matrix (the mass matrix in the space of fluxes), B_s is a rectangular $n_{p,s} \times n_{u,s}$ matrix, C_s^T is a rectangular $n_{u,s} \times n_\lambda$ matrix, Σ_s is a diagonal $n_{p,s} \times n_{p,s}$ matrix, where $n_{u,s} = \dim \boldsymbol{V}_{h,s}$, $n_{p,s} = \dim Q_{h,s}$, $s = \overline{1,\ m}$, and $n_\lambda = \dim \Lambda_h$.

In a compact form the system (2.32), (2.33) can be written as

$$\begin{pmatrix} M & B^T & C^T \\ B & -\Sigma & 0 \\ C & 0 & 0 \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{p} \\ \bar{\lambda} \end{pmatrix} = \begin{pmatrix} \bar{g}_D \\ \bar{f} \\ \bar{g}_N \end{pmatrix}, \tag{2.34}$$

where

$$M = M_1 \oplus \ldots \oplus M_m \quad \text{and} \quad B = B_1 \oplus \ldots \oplus B_m \tag{2.35}$$

are $m \times m$ block diagonal matrices,

$$C = \begin{pmatrix} C_1 & \ldots & C_m \end{pmatrix}, \tag{2.36}$$

$$\bar{u} = \begin{pmatrix} \bar{u}_1 \\ \vdots \\ \bar{u}_m \end{pmatrix}, \quad \bar{p} = \begin{pmatrix} \bar{p}_1 \\ \vdots \\ \bar{p}_m \end{pmatrix}, \quad \text{and} \quad \bar{\lambda} \in \mathbb{R}^{n_\lambda}. \tag{2.37}$$

2.5 Piece-wise constant (PWC) flux approximation

In this section, we describe another approach for the approximation of the flux vector function. We construct mass matrices in the space of fluxes using piece-wise constant vector fields.

Let E be a polyhedral cell. Let us assume that there exists a decomposition

$$E = \bigcup_{l=1}^{N_E} e_l \tag{2.38}$$

into polyhedrons (possibly overlapping) such that:

- Each face Γ of e_l is either an inner face with respect to E, or is a face of E;
- For each cell e_l there exists its vertex A such that there are exactly three faces (Γ_1, Γ_2 and Γ_3) of e_l adjacent to it, which are also the faces of E.

Examples of possible partitionings of the cells common for applications in basin modeling can be found in [37].

Let e be one of the the cells e_l from partition (2.38). Let $\bar{v} \in \mathbb{R}^3$ be a vector representing three degrees of freedom. We say that $\boldsymbol{v}_h \in \boldsymbol{V}_e^{(PWC)}$ if and only if the following two conditions hold:

- $\boldsymbol{v}_h \equiv \mathbf{const} \in \mathbb{R}^3$ in e;
- $\frac{1}{|\Gamma_i|} \int_{\Gamma_i} \boldsymbol{v}_h \cdot \boldsymbol{n} \, \mathrm{ds} \; = \; v_i, \quad i = \overline{1,\ 3}.$

Remark: The DOF v_i represents the average normal component $\boldsymbol{v}_h \cdot \boldsymbol{n}$ of $\boldsymbol{v}_h$ on a face Γ_i, $i = \overline{1,\ 3}$.

Explicit formulas. Let

$$\Gamma_i \ = \bigcup_{j=1}^{m_i} \gamma_{ij} \tag{2.39}$$

be a triangular representation of Γ_i. We denote by $\boldsymbol{n}_{ij}$ the unit outward normal vector to ∂e on a triangle γ_{ij} (as Γ_i is also a face of E, this would also be an outward normal vector to ∂E). We define by

$$\boldsymbol{n}_i \ = \sum_{j=1}^{m_i} \frac{|\gamma_{ij}|}{|\Gamma_i|} \boldsymbol{n}_{ij} \tag{2.40}$$

the "effective outward normal vector" to e on Γ_i.

Remark. If Γ_i is planar, then $\boldsymbol{n}_i$ is the outward unit normal vector to e on Γ_i. Otherwise, $\|\boldsymbol{n}_i\| < 1$.

Direct calculations show that these "effective normal vectors" uniquely determine a constant vector field $\boldsymbol{v}_h$ in e. Namely, the constant value of $\boldsymbol{v}_h$ in e is the vector

$$\boldsymbol{v} = N^{-T} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \tag{2.41}$$

where

$$N = \begin{pmatrix} \boldsymbol{n}_1 \ \boldsymbol{n}_2 \ \boldsymbol{n}_3 \end{pmatrix} \tag{2.42}$$

is a three-by-three matrix, and its columns are the corresponding "effective" normals to the sides of e.

In order to construct the mass matrix M, we first introduce a non-overlapping partitioning of E into polyhedral cells

$$E = \bigcup_{k=1}^{N_p} P_k \tag{2.43}$$

such that each subcell e_l from the partitioning (2.38) is a union of several cells P_k.

Let us denote by n_k, $k = 1, ..., N_p$, the number of cells e_l containing P_k. We introduce the functions $\alpha_l(x)$, $l = 1, ..., N_E$ in the following form

$$\alpha_l(x) = \begin{cases} \dfrac{1}{n_k}, & \text{if } x \in P_k \bigcap e_l, \\ \\ 0, & \text{otherwise.} \end{cases} \tag{2.44}$$

Remark. Functions α_l form a unity partition on E, i.e.

$$\sum_{l=1}^{N_E} \alpha_l(x) \equiv 1. \tag{2.45}$$

Let $\bar{u}$ and $\bar{v}$ be vectors of the degrees of freedom corresponding to the cell E. We construct N_E piece-wise constant vector fields $\boldsymbol{u}_h$ and $\boldsymbol{v}_h$ for each e_l according to the above procedure. Then we define the mass matrix M as

$$(Mu, v) = \sum_{l=1}^{N_l} \int_E \alpha_l(x)(K_E^{-1} u_h^l) \cdot v_h^l dx. \tag{2.46}$$

Direct calculations show that

$$M = \sum_{l=1}^{N_E} \mathcal{N}_l \left(\sum_{k: P_k \in e_l} \frac{|P_k|}{n_k} \right) N_l^{-1} K_E^{-1} N_l^{-T} \mathcal{N}_l^T \ , \tag{2.47}$$

where $\mathcal{N}_l$ are assembling matrices.

Chapter 3

Benchmark Problem

3.1 Benchmark problem formulation

In this section, we consider the Neumann boundary value problem for the diffusion equation

$$\begin{aligned} -\nabla \cdot \big(K\ \nabla p\big)\ +\ cp\ &=\ F \quad \text{in} \quad G, \\ -\,(K\ \nabla p)\cdot \boldsymbol{n}\ &=\ G_N \quad \text{on} \quad \partial G, \end{aligned} \tag{3.1}$$

in a parallelepipedal domain G. Here, $K = K(x, y, z)$ is a three-by-three symmetric positive definite matrix (diffusion tensor), $c = c(x, y, z)$ is a non-negative function, ∂G is the boundary of G, $\boldsymbol{n}$ is the unit outward normal to ∂G, $F = F(x, y, z)$ and $G_N = G_N(x, y, z)$ are given functions.

We assume that the domain G can be represented as a union of oblique layers G_l, $l = \overline{1,\ L}$, which are parallel and form an angle α with the x-axis, and an angle β with the y-axis. Our goal is to set the parameters K, c, F, and G_N so that the solution function $p = p(x, y, z)$ of (3.1) can be expressed analytically. For that purpose, we impose a number of restrictions so that the resulting benchmark problem is relevant to basin modeling applications.

First, we assume the diffusion tensor K to be piece-wise constant in G, i.e.

$$K|_{G_s} \equiv K_s \equiv \text{const}_s \in \mathbb{R}^{3\times 3},$$

We also assume that

$$K_s \;=\; W \begin{pmatrix} k_{s,x} & 0 & 0 \\ 0 & k_{s,y} & 0 \\ 0 & 0 & k_{s,z} \end{pmatrix} W^T \quad \text{in} \quad G_s, \quad s = \overline{1,\ L}\ , \tag{3.2}$$

is obtained by an orthogonal transformation of a constant diagonal tensor. The transformation is given by the matrix

$$W = \frac{1}{l_{vp}} \begin{pmatrix} l_{vp}\cos(\alpha) & -\sin(\alpha)\cos(\alpha)\sin(\beta) & \sin(\alpha)\cos(\beta) \\ 0 & \cos(\beta) & \cos(\alpha)\sin(\beta) \\ -l_{vp}\sin(\alpha) & -\cos^2(\alpha)\sin(\beta) & \cos(\alpha)\cos(\beta) \end{pmatrix} \tag{3.3}$$

associated with geological layers, where $l_{vp} = \sqrt{1-\sin^2(\alpha)\sin^2(\beta)}$.

Second, we impose similar restrictions on c, that is

$$c|_{G_s} \equiv c_s \equiv \text{const}_s > 0.$$

The assumptions with regard to the right-hand side F and boundary function G_N are stated in the end of this section and are summarized in (3.12) and (3.13).

Third, we assume that the region G is a part of a larger parallelepipedal domain $\widehat{G}$ such that the subregions G_s, $s = \overline{1,\ L}$, can be extended to horizontal layers $\widehat{G}_s$ in $\widehat{G}$. With $\widehat{G}$, we associate the Cartesian system $(\widehat{x},\ \widehat{y},\ \widehat{z})$, which is obtained by the transformation of system $(x,\ y,\ z)$ given by the matrix W from (3.3).

An example of such transformation is shown on Figures 3.1 and 3.2. Figure 3.1 shows the original domain G in the $(x,\ y,\ z)$ coordinate system, and Figure 3.2 shows the same domain as a part of $\widehat{G}$ in the $(\widehat{x},\ \widehat{y},\ \widehat{z})$ coordinate system.

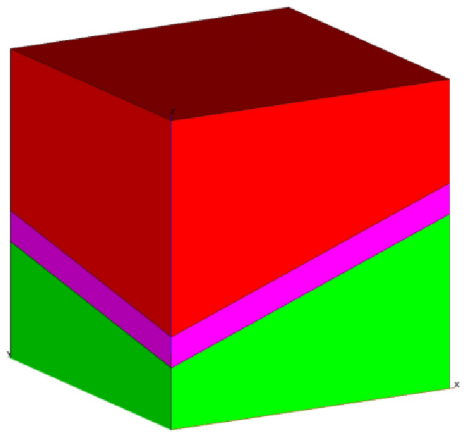

Figure 3.1: Domain G in the original coordinate system

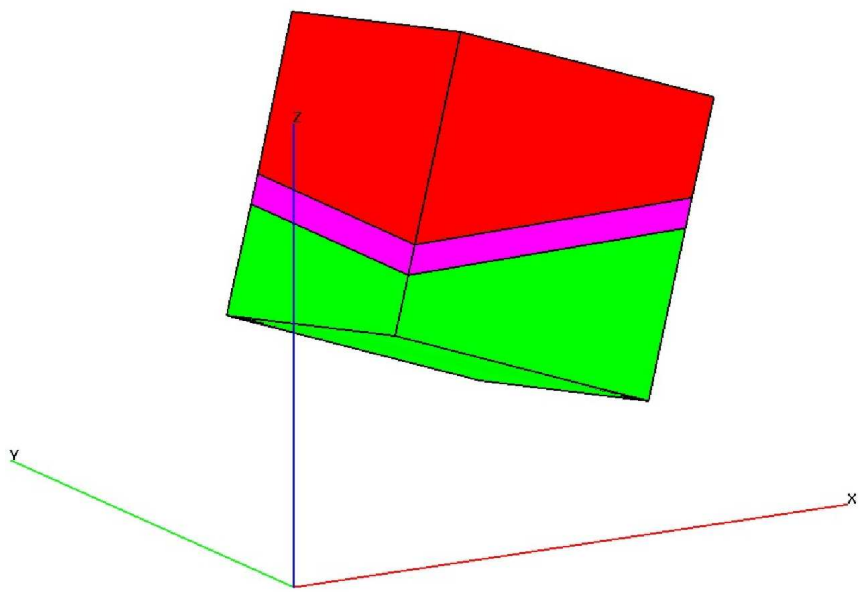

Figure 3.2: Domain G in the reference coordinate system

Now, we consider the diffusion problem with homogeneous Neumann boundary con-

ditions for the region $\widehat{G}$:

$$
\begin{aligned}
-\widehat{\nabla}\cdot\big(\widehat{K}\ \widehat{\nabla}\widehat{p}\big) \ + \ \widehat{c}\widehat{p} \ &= \ \widehat{F} \quad \text{in} \quad \widehat{G}, \\
-\big(\widehat{K}\ \widehat{\nabla}\widehat{p}\big)\cdot\widehat{\boldsymbol{n}} \ &= \ 0 \quad \text{on} \quad \partial\widehat{G},
\end{aligned}
\tag{3.4}
$$

and assume that the values of $\widehat{K}$, $\widehat{c}$, and $\widehat{F}$ in subregions G_s, $s = \overline{1,\ L}$, coincide with the corresponding values of K, c, and F in the original problem (3.1) under the respective change of coordinates, i.e.

$$
\begin{aligned}
\widehat{K}|_{G_s} \ &= \ \begin{pmatrix} k_{x,s} & 0 & 0 \\ 0 & k_{y,s} & 0 \\ 0 & 0 & k_{z,s} \end{pmatrix}, \\
\widehat{c}|_{G_s}(\widehat{x},\ \widehat{y},\ \widehat{z}) \ &= \ c_s, \\
\text{and} \quad \widehat{F}|_{G_s}(\widehat{x},\ \widehat{y},\ \widehat{z}) \ &= \ F|_{G_s}(x,\ y,\ z).
\end{aligned}
\tag{3.5}
$$

The transformation between $(x,\ y,\ z)$-coordinate system associated with the domain G, and the $(\widehat{x},\ \widehat{y},\ \widehat{z})$-coordinates associated with $\widehat{G}$, is given by

$$
\begin{pmatrix} \widehat{x} \\ \widehat{y} \\ \widehat{z} \end{pmatrix} \ = \ \boldsymbol{r}_0 \ + \ W \begin{pmatrix} x \\ y \\ z \end{pmatrix} . \tag{3.6}
$$

Here, $\boldsymbol{r}_0$ is the vector connecting the origins of the corresponding coordinate systems, and W is defined in (3.3).

Then, we extend all the assumptions made for parameters K, c, and F in layers G_s of G onto the parameters $\widehat{K}$, $\widehat{c}$, and $\widehat{F}$ in layers $\widehat{G}_s$ of $\widehat{G}$.

With the restrictions stated above, we can explicitly find the reference solution $\widehat{p}^*$ for problem (3.4) in the domain $\widehat{G}$. Then, we set the boundary conditions for problem

(3.1) as

$$G_N \;=\; \widehat{\boldsymbol{u}}^*(\widehat{x},\ \widehat{y},\ \widehat{z})\cdot\widehat{\boldsymbol{n}} \quad \text{on} \quad \partial G, \tag{3.7}$$

where

$$\widehat{\boldsymbol{u}}^* \;=\; -\widehat{K}\ \widehat{\nabla}\widehat{p}^*$$

is the reference flux for the domain $\widehat{G}$. Consequently, the reference solution of problem (3.1) on the domain G should coincide with the restriction of the reference solution of problem (3.4) to the subdomain G, i.e.

$$\boldsymbol{u}^*(x,\ y,\ z) \;=\; \widehat{\boldsymbol{u}}^*(\widehat{x},\ \widehat{y},\ \widehat{z})|_G \quad \text{and} \quad p^*(x,\ y,\ z) \;=\; \widehat{p}^*(\widehat{x},\ \widehat{y},\ \widehat{z})|_G\ , \tag{3.8}$$

where p^* is the solution of the problem (5.1), and $\boldsymbol{u}^* = -K\nabla p^*$.

In order to obtain the reference solution, we reduce the dimension of the problem by applying the separation of variables. For simplicity in notations, we assume that

$$\widehat{G} \;\equiv\; (0,\ 1)\times(0,\ 1)\times(\widehat{z}_0,\ \widehat{z}_L). \tag{3.9}$$

First, we consider an eigenproblem for the operator $-\dfrac{d^2}{d\xi^2}$, where ξ can stand for either $\widehat{x}$ or $\widehat{y}$:

$$\begin{aligned} -\frac{d^2}{d\xi^2}w \;&=\; \lambda_\xi w, \quad 0<\xi<1, \\ \frac{dw}{d\xi}(0) \;&=\; 0, \quad \frac{dw}{d\xi}(1) \;=\; 0. \end{aligned} \quad . \tag{3.10}$$

The eigenpairs $(\lambda_{\xi,n},\ w_{\xi,n})$ are

$$\begin{aligned} &\lambda_{\xi,0}=0, \quad w_{\xi,0}\equiv 1, \\ &\lambda_{\xi,n}=(n\pi)^2, \quad w_{\xi,n}=\sqrt{2}\cos n\pi\xi, \quad n\geq 1\ . \end{aligned} \tag{3.11}$$

The set of eigenfunctions is an orthonormal basis in $L_2(0;\ 1)$. The right-hand side

function $\widehat{F} = \widehat{F}(\widehat{x}, \widehat{y}, \widehat{z})$ in (5.1) can be expanded in this basis as

$$\begin{aligned}
\widehat{F}(\widehat{x}, \widehat{y}, \widehat{z}) \;&=\; \sum_{i=0}^{\infty} \widehat{f}_i(\widehat{y}, \widehat{z}) w_{x,i}(\widehat{x}) \;= \\
&=\; \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \widehat{f}_{i,j}(\widehat{z}) w_{x,i}(\widehat{x}) w_{y,j}(\widehat{y}) \;, \\
\text{with} \quad \widehat{f}_i(\widehat{y}, \widehat{z}) \;&=\; \int_0^1 \widehat{F}(\widehat{x}, \widehat{y}, \widehat{z}) w_{x,i}(\widehat{x}) d\widehat{x} \\
\text{and} \quad \widehat{f}_{i,j}(\widehat{z}) \;&=\; \int_0^1 \int_0^1 \widehat{F}(\widehat{x}, \widehat{y}, \widehat{z}) w_{x,i}(\widehat{x}) w_{y,j}(\widehat{y}) d\widehat{x} d\widehat{y} \;, \\
F(x, y, z) \;&=\; \widehat{F}(\widehat{x}, \widehat{y}, \widehat{z}) \;.
\end{aligned} \tag{3.12}$$

With that, we can formulate the assumptions imposed on the right-hand side in the benchmark problem formulation. We consider $\widehat{F} = \widehat{F}(\widehat{x}, \widehat{y}, \widehat{z})$ to be admissible if it belongs to the class of functions whose expansion (3.12) satisfies:

$$\begin{aligned}
\widehat{f}_{i,j}(\widehat{z})|_{G_s} \;&\equiv\; \widehat{f}_{s,i,j} \;\equiv\; \text{const}_{s,i,j}, \quad (i,\ j) = (0,\ 0), (1,\ 0), (0,\ 1) \;, \\
\widehat{f}_{i,j}(\widehat{z})|_{G_s} \;&\equiv\; 0, \qquad\qquad \text{otherwise} \;,
\end{aligned} \tag{3.13}$$

i.e. the functions $\widehat{F}_{i,j} = \widehat{f}_{i,j}(\widehat{z})$ are piece-wise constant with respect to layers $\widehat{G}_s$, and $\widehat{F} = \widehat{F}(\widehat{x}, \widehat{y}, \widehat{z})$ has only three nonzero harmonics.

As stated before, we have $\widehat{F}|_{G_s}(\widehat{x},\ \widehat{y},\ \widehat{z}) \;=\; F|_{G_s}(x,\ y,\ z)$, so the right-hand side function $F(x,\ y,\ z)$ in the original problem is assumed to be chosen so that the corresponding function $\widehat{F}(\widehat{x},\ \widehat{y},\ \widehat{z})$ satisfies the conditions in (3.13). With that, we complete the list of assumptions necessary to define the benchmark problem.

From now on, we proceed with solving problem (3.4). The reference solution for original problem (3.1) can be obtained as described in (3.8).

3.2 Solution derivation for the benchmark problem

We expand the solution function $\widehat{p} = \widehat{p}(\widehat{x}, \widehat{y}, \widehat{z})$ of the problem (3.4) with respect to the eigenfunctions in (3.11):

$$\begin{aligned} \widehat{p}(\widehat{x}, \widehat{y}, \widehat{z}) \ &= \sum_{i=0}^{\infty} \widehat{p}_i(\widehat{y}, \widehat{z}) w_{\widehat{x},i}(\widehat{x}) \ = \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \widehat{p}_{i,j}(\widehat{z}) w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \ . \end{aligned} \tag{3.14}$$

Then,

$$\begin{aligned} -\frac{\partial}{\partial \widehat{x}} \left(\widehat{K}_{\widehat{x}} \frac{\partial \widehat{p}}{\partial \widehat{x}} \right) \ &= \ \widehat{K}_{\widehat{x}} \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \lambda_{\widehat{x},i} w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \widehat{p}_{i,j}(\widehat{z}) \ , \\ -\frac{\partial}{\partial \widehat{y}} \left(\widehat{K}_{\widehat{y}} \frac{\partial \widehat{p}}{\partial \widehat{y}} \right) \ &= \ \widehat{K}_{\widehat{y}} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \lambda_{\widehat{y},j} w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \widehat{p}_{i,j}(\widehat{z}) \ , \\ -\frac{\partial}{\partial \widehat{z}} \left(\widehat{K}_{\widehat{z}} \frac{\partial p}{\partial \widehat{z}} \right) \ &= \ -\frac{d}{d\widehat{z}} \left(\widehat{K}_{\widehat{z}} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \frac{d\widehat{p}_{i,j}}{d\widehat{z}}(\widehat{z}) \right) \ . \end{aligned}$$

Therefore, we can write our problem as

$$\begin{aligned} -\frac{d}{d\widehat{z}} \left(\widehat{K}_{\widehat{z}} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \frac{d\widehat{p}_{i,j}}{d\widehat{z}}(\widehat{z}) \right) \ + \\ + \widehat{K}_{\widehat{x}} \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \lambda_{\widehat{x},i} w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \widehat{p}_{i,j}(\widehat{z}) \ + \\ + \widehat{K}_{\widehat{y}} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \lambda_{\widehat{y},j} w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \widehat{p}_{i,j}(\widehat{z}) \ + \\ + \ c \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \widehat{p}_{i,j}(\widehat{z}) w_{\widehat{x},i}(\widehat{x}) w_{\widehat{y},j}(\widehat{y}) \ = \ \widehat{f}_{0,0}(\widehat{z}) w_{\widehat{x},0}(\widehat{x}) w_{\widehat{y},0}(\widehat{y}) \ + \\ + \ \widehat{f}_{1,0}(\widehat{z}) w_{\widehat{x},1}(\widehat{x}) w_{\widehat{y},0}(\widehat{y}) \ + \ \widehat{f}_{0,1}(\widehat{z}) w_{\widehat{x},0}(\widehat{x}) w_{\widehat{y},1}(\widehat{y}) \quad \text{in } \widehat{G} \ , \\ \frac{d\widehat{p}_{i,j}}{d\widehat{z}}(\widehat{z}) \ = \ 0 \quad \text{on } \partial\widehat{G} \quad \text{for } (i, \ j) \geq (0, \ 0) \ . \end{aligned} \tag{3.15}$$

Since the basis $\{w_{\xi,i}\}_{i=0}^{\infty},\ \xi = x, y$, is orthonormal, it can be easily seen seen that

$$\widehat{p}_{i,j}(\widehat{z}) \ \equiv\ 0, \qquad (i,\ j) \neq (0,\ 0), (1,\ 0), (0,\ 1)\ . \tag{3.16}$$

Hence, the resulting system is as follows:

$$\begin{aligned}
&-\frac{d}{d\widehat{z}}\left(\widehat{K}_{\widehat{z}}\frac{d\widehat{p}_{0,0}}{d\widehat{z}}(\widehat{z})\right) \ +\ c\widehat{p}_{0,0}(\widehat{z}) \ =\ \widehat{f}_{0,0}(\widehat{z}) \\
&-\frac{d}{d\widehat{z}}\left(\widehat{K}_{\widehat{z}}\frac{d\widehat{p}_{1,0}}{d\widehat{z}}(\widehat{z})\right) \ +\ \left(c\ +\ \pi^2\widehat{K}_{\widehat{x}}\right)\widehat{p}_{1,0}(\widehat{z}) \ =\ \widehat{f}_{1,0}(\widehat{z}) \\
&-\frac{d}{d\widehat{z}}\left(\widehat{K}_{\widehat{z}}\frac{d\widehat{p}_{0,1}}{d\widehat{z}}(\widehat{z})\right) \ +\ \left(c\ +\ \pi^2\widehat{K}_{\widehat{y}}\right)\widehat{p}_{0,1}(\widehat{z}) \ =\ \widehat{f}_{0,1}(\widehat{z}) \\
&\frac{d\widehat{p}_{i,j}}{d\widehat{z}}(0) \ =\ 0, \quad \frac{d\widehat{p}_{i,j}}{d\widehat{z}}(1) \ =\ 0, \quad (i,j) = (0,0), (1,0), (0,1)\ .
\end{aligned} \tag{3.17}$$

Under the previously stated restrictions on $\widehat{K}$, $\widehat{c}$, and $\widehat{F}$, system (3.17) stands for the reduced formulation of the benchmark problem.

According to the formulation of the problem, the regions $\widehat{G}_s$ are horizontal layers, and therefore we can formally describe them using the notations

$$\widehat{G}_s \ =\ (0,\ 1)\ \times\ (0,\ 1)\ \times\ (\widehat{z}_{s-1},\ \widehat{z}_s) \tag{3.18}$$

with

$$0 \equiv \widehat{z}_0 < \cdots < \widehat{z}_s < \cdots < \widehat{z}_L \equiv 1\ . \tag{3.19}$$

Then, using the assumptions that $\widehat{K}$, $\widehat{c}$, and $\widehat{F}_i$ are piece-wise constant functions, we can write system (3.17) as

$$
\begin{aligned}
-\widehat{K}_{\widehat{z},s}\frac{d^2\widehat{p}_{s,i,j}}{d\widehat{z}^2}(\widehat{z}) + \nu_{s,i,j}{}^2\widehat{p}_{s,i,j}(\widehat{z}) \;&=\; \widehat{f}_{s,i,j}, \quad s=\overline{1,\ L} \\
\frac{d\widehat{p}_{1,i,j}}{d\widehat{z}}(\widehat{z}_0) \;&=\; 0 \\
\frac{d\widehat{p}_{L,i,j}}{d\widehat{z}}(\widehat{z}_L) \;&=\; 0 \\
\widehat{p}_{s,i,j}(\widehat{z}_s-0) \;&=\; \widehat{p}_{s+1,i,j}(\widehat{z}_s+0), \quad s=\overline{1,\ L-1} \\
\widehat{K}_{\widehat{z},s}\frac{d\widehat{p}_{s,i,j}}{d\widehat{z}}(\widehat{z}_s-0) \;&=\; \widehat{K}_{\widehat{z},s+1}\frac{d\widehat{p}_{s+1,i,j}}{d\widehat{z}}(\widehat{z}_s+0), \quad s=\overline{1,\ L-1}
\end{aligned}
\tag{3.20}
$$

for $(i,\ j)=(0,0),(1,0),(0,1)$. Here,

$$
\nu_{s,0,0} \;=\; \sqrt{c_s}, \quad \nu_{s,1,0} \;=\; \sqrt{c_s+\pi^2\widehat{K}_{\widehat{x},s}}, \quad \nu_{s,0,1} \;=\; \sqrt{c_s+\pi^2\widehat{K}_{\widehat{y},s}}\ . \tag{3.21}
$$

The general solution of the second order ODE in (3.20) can be written as

$$
\widehat{p}_{s,i,j}(\widehat{z}) \;=\; B_{s,i,j,1}\cdot e^{\beta_{s,i,j}\widehat{z}} \;+\; B_{s,i,j,2}\cdot e^{-\beta_{s,i,j}\widehat{z}} \;+\; \frac{\widehat{f}_{s,i,j}}{\nu_{s,i,j}{}^2}\ , \tag{3.22}
$$

where

$$
\beta_{s,i,j} \;=\; \frac{\nu_{s,i,j}}{\sqrt{\widehat{K}_{\widehat{z},s}}}, \tag{3.23}
$$

and $B_{s,i,j,1}$, $B_{s,i,j,2}$ are constants dependent on initial conditions. In order to find these constants explicitly, let us first introduce a decomposition of the solution function $\widehat{p}_{s,i,j}(\widehat{z})$:

$$
\begin{aligned}
\widehat{p}_{s,i,j}(\widehat{z}) \;&=\; v_{s-1,i,j}\cdot\varphi_{s,i,j}(\widehat{z}) \;+\; v_{s,i,j}\cdot\psi_{s,i,j}(\widehat{z}) \;+\; \frac{\widehat{f}_{s,i,j}}{\nu_{s,i,j}{}^2}\ , \\
v_{s-1,i,j} \;&=\; \widehat{u}_{\widehat{z},s,i,j}(\widehat{z}_{s-1})\ , \\
v_{s,i,j} \;&=\; \widehat{u}_{\widehat{z},s,i,j}(\widehat{z}_s)\ ,
\end{aligned}
\tag{3.24}
$$

where

$$
\begin{aligned}
\widehat{u}_{\widehat{z},s,i,j}(\widehat{z}) \;&=\; -\widehat{K}_{\widehat{z},s}\frac{d\widehat{p}_{s,i,j}}{d\widehat{z}} \;= \\
&=\; -\nu_{s,i,j}\sqrt{\widehat{K}_{\widehat{z},s}}\left(B_{s,i,j,1}\cdot e^{\beta_{s,i,j}\widehat{z}} \;-\; B_{s,i,j,2}\cdot e^{-\beta_{s,i,j}\widehat{z}}\right), \quad s=\overline{1,L}.
\end{aligned}
\tag{3.25}
$$

We have $v_{s,i,j}$ well-defined for $s = \overline{1,\ L-1}$ due to the condition

$$\widehat{K}_{\widehat{z},s}\frac{d\widehat{p}_{s,i,j}}{d\widehat{z}}(\widehat{z}_s - 0) \ = \ \widehat{K}_{\widehat{z},s+1}\frac{d\widehat{p}_{s+1,i,j}}{d\widehat{z}}(\widehat{z}_s + 0), \quad s = \overline{1,\ L-1}$$

in (3.20). Now, we use expression (3.22) to obtain the system of equations for $B_{s,i,j,1}$ and $B_{s,i,j,2}$:

$$\begin{aligned} -\nu_{s,i,j}\sqrt{\widehat{K}_{\widehat{z},s}}\left(e^{\beta_{s,i,j}\widehat{z}_{s-1}}\cdot B_{s,i,j,1} \ - \ e^{-\beta_{s,i,j}\widehat{z}_{s-1}}\cdot B_{s,i,j,2}\right) &= v_{s-1,i,j} \\ -\nu_{s,i,j}\sqrt{\widehat{K}_{\widehat{z},s}}\left(e^{\beta_{s,i,j}\widehat{z}_{s}}\cdot B_{s,i,j,1} \ - \ e^{-\beta_{s,i,j}\widehat{z}_{s}}\cdot B_{s,i,j,2}\right) &= v_{s,i,j} \end{aligned} \quad . \tag{3.26}$$

The solution of this system in terms of $v_{s-1,i,j}$ and $v_{s,i,j}$ is as follows:

$$\begin{aligned} B_{s,i,j,1} &= \frac{1}{\nu_{s,i,j}\sqrt{\widehat{K}_{\widehat{z},s}}}\cdot\frac{e^{-\beta_{s,i,j}\widehat{z}_{s-1}}\cdot v_s \ - \ e^{-\beta_{s,i,j}\widehat{z}_{s}}\cdot v_{s-1}}{-e^{\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})} \ + \ e^{-\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})}} \\ B_{s,i,j,2} &= \frac{1}{\nu_{s,i,j}\sqrt{\widehat{K}_{\widehat{z},s}}}\cdot\frac{e^{\beta_{s,i,j}\widehat{z}_{s-1}}\cdot v_s \ - \ e^{\beta_{s,i,j}\widehat{z}_{s}}\cdot v_{s-1}}{-e^{\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})} \ + \ e^{-\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})}} \end{aligned} \quad .$$

From this and the definition of $\varphi_{s,i,j}(\widehat{z})$ and $\psi_{s,i,j}(\widehat{z})$ in (3.24), we can obtain explicit expressions for these functions:

$$\begin{aligned} \varphi_{s,i,j}(\widehat{z}) &= \frac{1}{\nu_{s,i,j}\sqrt{\widehat{K}_{\widehat{z},s}}}\cdot\frac{1 \ + \ e^{-2\beta_{s,i,j}(\widehat{z}_s-\widehat{z})}}{1 \ - \ e^{-2\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})}}\cdot e^{-\beta_{s,i,j}(\widehat{z}-\widehat{z}_{s-1})} \ , \\ \psi_{s,i,j}(\widehat{z}) &= -\frac{1}{\nu_{s,i,j}\sqrt{\widehat{K}_{\widehat{z},s}}}\cdot\frac{1 \ + \ e^{-2\beta_{s,i,j}(\widehat{z}-\widehat{z}_{s-1})}}{1 \ - \ e^{-2\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})}}\cdot e^{-\beta_{s,i,j}(\widehat{z}_s-\widehat{z})} \ , \end{aligned} \tag{3.27}$$

and therefore

$$\begin{aligned} \frac{d\varphi_{s,i,j}}{d\widehat{z}}(\widehat{z}) &= -\frac{1}{\widehat{K}_{\widehat{z},s}}\cdot\frac{1 \ - \ e^{-2\beta_{s,i,j}(\widehat{z}_s-\widehat{z})}}{1 \ - \ e^{-2\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})}}\cdot e^{-\beta_{s,i,j}(\widehat{z}-\widehat{z}_{s-1})} \ , \\ \frac{d\psi_{s,i,j}}{d\widehat{z}}(\widehat{z}) &= -\frac{1}{\widehat{K}_{\widehat{z},s}}\cdot\frac{1 \ - \ e^{-2\beta_{s,i,j}(\widehat{z}-\widehat{z}_{s-1})}}{1 \ - \ e^{-2\beta_{s,i,j}(\widehat{z}_s-\widehat{z}_{s-1})}}\cdot e^{-\beta_{s,i,j}(\widehat{z}_s-\widehat{z})} \ . \end{aligned} \tag{3.28}$$

Now, from the condition on continuity of the solution functions $\widehat{p}_{i,j}(\widehat{z})$ given in (3.20), it follows that

$$v_{s,i,j} \cdot \varphi_{s+1,i,j}(\widehat{z}_s) \; + \; v_{s+1,i,j} \cdot \psi_{s+1,i,j}(\widehat{z}_s) \; + \; \frac{\widehat{f}_{s+1,i,j}}{\nu_{s+1,i,j}{}^2} \; =$$

$$= \; v_{s-1,i,j} \cdot \varphi_{s,i,j}(\widehat{z}_s) \; + \; v_{s,i,j} \cdot \psi_{s,i,j}(\widehat{z}_s) \; + \; \frac{\widehat{f}_{s,i,j}}{\nu_{s,i,j}{}^2}, \quad s = \overline{1, \; L-1}.$$

In addition, the Neumann boundary conditions in (3.20) imply that

$$v_{0,i,j} \; = \; \widehat{u}_{1,i,j}(\widehat{z}_0) \; = \; \frac{d\widehat{p}_{1,i,j}}{d\widehat{z}}(\widehat{z}_0) \; = \; 0 \; ,$$

$$v_{L,i,j} \; = \; \widehat{u}_{L,i,j}(\widehat{z}_M) \; = \; \frac{d\widehat{p}_{L,i,j}}{d\widehat{z}}(\widehat{z}_L) \; = \; 0 \; .$$

Therefore, the system for the coefficients v_s is as follows:

$$\begin{aligned}
(\varphi_{2,i,j}(\widehat{z}_1) - \psi_{1,i,j}(\widehat{z}_1)) \cdot v_1 \; + \; \psi_{2,i,j}(\widehat{z}_1) \cdot v_2 \qquad\qquad\qquad\qquad &= \; b_1 \\
\cdots \\
-\varphi_{s,i,j}(\widehat{z}_s) \cdot v_{s-1} \; + \; (\varphi_{s+1,i,j}(\widehat{z}_s) - \psi_{s,i,j}(\widehat{z}_s)) \cdot v_s \; + \; \psi_{s+1,i,j}(\widehat{z}_s) \cdot v_{s+1} \quad &= \; b_s \qquad , \\
\cdots \\
-\varphi_{L-1,i,j}(\widehat{z}_{L-1}) \cdot v_{L-2} \; + \; (\varphi_{L,i,j}(\widehat{z}_{L-1}) - \psi_{L-1,i,j}(\widehat{z}_{L-1})) \cdot v_{L-1} \qquad &= \; b_{L-1}
\end{aligned} \tag{3.29}$$

where

$$b_s \; = \; \frac{\widehat{f}_{s,i,j}}{\nu_{s,i,j}{}^2} \; - \; \frac{\widehat{f}_{s+1,i,j}}{\nu_{s+1,i,j}{}^2}, \quad s = \overline{1, \; L-1}.$$

Resolving this system yields the solution for the initial value problem (3.20), i.e. provides the explicit formulas for $\widehat{p}_{s,i,j}(\widehat{z})$. With that, we can write down the reference solution of our benchmark problem:

$$\begin{aligned}
\widehat{p}_s(\widehat{x}, \widehat{y}, \widehat{z}) \; = \; \widehat{p}_{s,0,0}(\widehat{z}) \; + \sqrt{2}\cos(\pi\widehat{x}) \cdot \widehat{p}_{s,1,0}(\widehat{z}) \; + \\
+ \sqrt{2}\cos(\pi\widehat{y}) \cdot \widehat{p}_{s,0,1}(\widehat{z}), \quad s = \overline{1, L}.
\end{aligned} \tag{3.30}$$

The fluxes U_s with their components $\widehat{U}_{\widehat{x},s}$, $\widehat{U}_{\widehat{y},s}$ and $\widehat{U}_{\widehat{z},s}$ are as follows:

$$\begin{aligned}
\widehat{U}_s(\widehat{x},\widehat{y},\widehat{z}) \;&=\; -\widehat{K}_s \nabla \widehat{p}_s(\widehat{x},\widehat{y},\widehat{z}) \;, \\
\widehat{U}_{\widehat{x},s}(\widehat{x},\widehat{y},\widehat{z}) \;&=\; -\widehat{K}_{\widehat{x},s} \cdot \pi\sqrt{2}\sin(\pi\widehat{x}) \cdot \widehat{p}_{s,1,0}(\widehat{z}) \;, \\
\widehat{U}_{\widehat{y},s}(\widehat{x},\widehat{y},\widehat{z}) \;&=\; -\widehat{K}_{\widehat{y},s} \cdot \pi\sqrt{2}\sin(\pi\widehat{y}) \cdot \widehat{p}_{s,0,1}(\widehat{z}) \;, \\
\widehat{U}_{\widehat{z},s}(\widehat{x},\widehat{y},\widehat{z}) \;&=\; -\widehat{K}_{\widehat{z},s}\big(\frac{d\widehat{p}_{s,0,0}}{d\widehat{z}}(\widehat{z}) + \\
+\sqrt{2}\cos(\pi\widehat{x}) \cdot \frac{d\widehat{p}_{s,1,0}}{d\widehat{z}}(\widehat{z}) \;&+\; \sqrt{2}\cos(\pi\widehat{y}) \cdot \frac{d\widehat{p}_{s,0,1}}{d\widehat{z}}(\widehat{z})\big) \;.
\end{aligned} \tag{3.31}$$

Here,

$$\frac{d\widehat{p}_{s,i,j}}{d\widehat{z}}(\widehat{z}) \;=\; v_{s-1,i,j} \cdot \frac{d\varphi_{s,i,j}}{d\widehat{z}}(\widehat{z}) \;+\; v_{s,i,j} \cdot \frac{d\psi_{s,i,j}}{d\widehat{z}}(\widehat{z}) \tag{3.32}$$

with the expressions for derivatives of $\varphi_{s,i,j}(\widehat{z})$ and $\psi_{s,i,j}(\widehat{z})$ given in (3.28).

3.3 Error analysis of MFE methods on prismatic anisotropic meshes

In this section, we compare the reference solution of the benchmark problem with its discrete approximations.

We consider problem (3.4) in the parallelepipedal domain G with oblique layers having inclination α with respect to x-axis and β with respect to y-axis. We assume that the restrictions stated in the previous sections hold true, and denote the corresponding reference domain by $\widehat{G}$. The reference solution pair is then $(p^*,\ \boldsymbol{u}^*)$. We use a prismatic mesh G_h in the domain G, and denote mesh cells by e_k, $k = \overline{1,\ n}$, where n is the total number of cells. Each cell e_k is a prism divided either into three tetrahedrons when we apply the KR discretization, or into three pyramids and two tetrahedrons if we use the PWC approximation.

For the benchmark problem, the reference solution $\boldsymbol{u}^*$ is known in the entire domain $\widehat{G}$ and, therefore, its entire subdomain G, so for every mesh cell e_k we explicitly know ${\boldsymbol{u}_k}^* \equiv \boldsymbol{u}^*|_{e_k}$, the reference solution for the cell, as well as

$$u^*_{k,i} \;=\; \frac{1}{|\gamma_{k,i}|} \int\limits_{\gamma_{k,i}} {\boldsymbol{u}_k}^*(\boldsymbol{x}) \cdot \boldsymbol{n}_{k,i}\; ds, \tag{3.33}$$

which is the integral average normal component of the reference flux on an interface $\gamma_{k,i}$.

In order to obtain KR or PWC interpolant for every cell e_k, $k = \overline{1,\ n}$, we discretize the equation (3.4) by applying the KR MFE method or the PWC approximation, and obtain the discrete solution pair $(p_{h,k},\ \boldsymbol{w}_{h,k})$. The flux interpolant $\boldsymbol{w}_{h,k}$ can be used to estimate the accuracy of the method.

The absolute error Δ_k over a cell e_k can be computed as

$$\Delta_k \;=\; \left(\int\limits_{e_k} |\boldsymbol{w}_{h,k}(\boldsymbol{x}) \;-\; \boldsymbol{u}^*(\boldsymbol{x})|^2 \; d\boldsymbol{x} \right)^{1/2} , \tag{3.34}$$

and the L_2 norm of the reference flux $\boldsymbol{u}^*$ over the same cell e_k is

$$\|\boldsymbol{u}_k^*\|_2 \;=\; \left(\int\limits_{e_k} |\boldsymbol{u}^*(\boldsymbol{x})|^2 \; d\boldsymbol{x} \right)^{1/2} . \tag{3.35}$$

We define $\omega_{h,s}$ to be a set of cells e_k belonging to the same geological layer G_s, .i.e.

$$\omega_{h,s} \;=\; \{e_k \;:\; e_k \in G_s\} .$$

The relative error in L_2 norm between the interpolant $\boldsymbol{w}_{h,k}$ and the reference solution $\boldsymbol{u}_k^*$ over certain geological layer $\omega_{h,s}$ can be computed as

$$\epsilon_{\omega_{h,s}} \;=\; 100 \cdot \frac{\sum\limits_{e_k \in \omega_{h,s}} \Delta_k}{\sum\limits_{e_k \in \omega_{h,s}} \|\boldsymbol{u}_k^*\|_2} . \tag{3.36}$$

We are particularly interested in the values of errors in thin geological layers. We choose our domain G to be a parallelepiped $(0,\ 0.5) \times (0,\ 0.5) \times (0,\ 0.25)$ with three geological layers. The bottom and top boundary of the thin layer in the middle are two parallel planes passing through the points $(0,\ 0,\ z_1)$ and $(0,\ 0,\ z_2)$, respectively, where $z_1 = 0.05$ and $z_2 = 0.05001$. The inclination of those planes with respect to x and y-axis is given by angles α and β.

The values of the parameters used in the experiment are given in Table 3.1.

The mesh we use is conforming and uniform in x and y coordinates, and is uniform along the z-direction inside each layer. There are $12 \times 12 \times 4$ cells in $\omega_{h,1}$, $12 \times 12 \times 2$ cells in $\omega_{h,2}$, and $12 \times 12 \times 6$ cells in $\omega_{h,3}$.

Table 3.1: Parameters for the chosen test problem

	$\widehat{G}_1$	$\widehat{G}_2$	$\widehat{G}_3$
$k_{x,s}$	5	100	10
$k_{y,s}$	5	100	10
$k_{z,s}$	3	10	5
c_s	1	1	1
$\widehat{f}_{0,0}$	5	1000	1
$\widehat{f}_{1,0}$	0.1	10	0.05
$\widehat{f}_{0,1}$	0.1	10	0.05

The domain and mesh are shown on Figure 3.3 for the case of horizontal layers, and on Figure 3.4 for the case of oblique middle layer.

Relative error values for the case of the horizontal layers are given in Table 3.2, and for the case of the oblique layers in the Table 3.3.

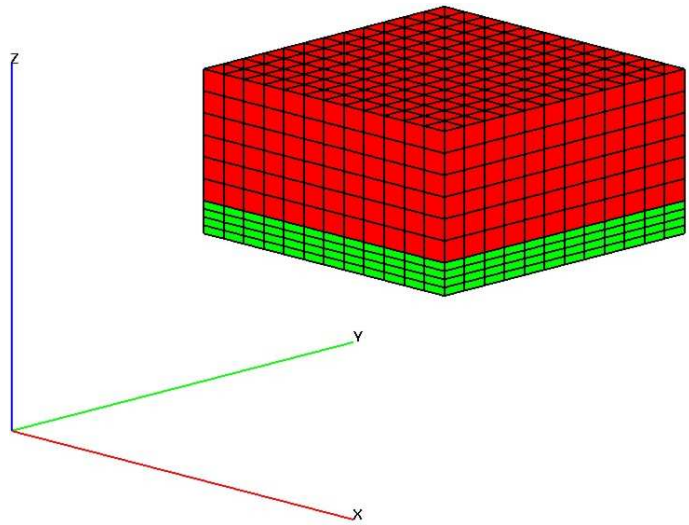

Figure 3.3: Domain and mesh with horizontal layers

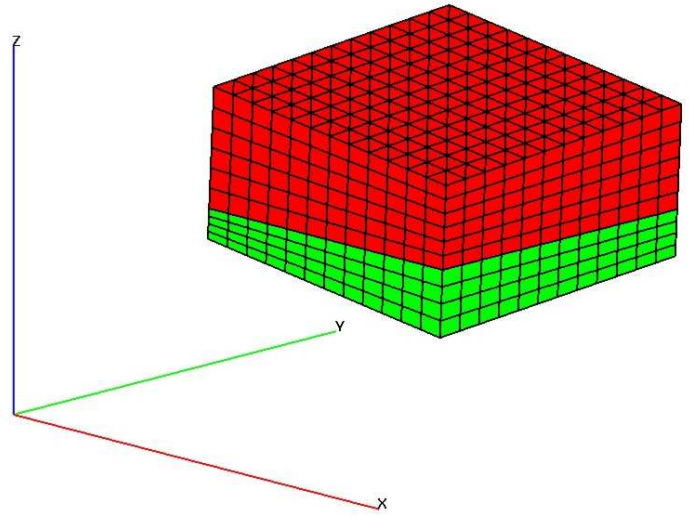

Figure 3.4: Domain and mesh with oblique middle layer

Table 3.2: Relative error in the discrete solutions, %, for angles $\alpha = 0$, $\beta = 0$

	$\omega^s_{h,1}$	$\omega^s_{h,2}$	$\omega^s_{h,3}$
KR	2.60	757.60	1.93
PWC	1.78	1.51	2.82

Table 3.3: Relative error in the discrete solutions, %, for angles $\alpha = 7°$, $\beta = -4°$

	$\omega^s_{h,1}$	$\omega^s_{h,2}$	$\omega^s_{h,3}$
KR	2.91	775.07	1.83
PWC	3.09	28.31	2.63

It is clear that in the case when using the KR-approximation results in unacceptable error values, we can instead use PWC approximation to obtain much better accuracy.

Chapter 4

New preconditioning method for diffusion equations on prismatic meshes

4.1 General description

The Preconditioned Conjugate Gradient (PCG) Method is one of the most efficient algorithms for solving systems with symmetric positive definite matrices. The major problem in application of the PCG method is the design of a symmetric positive definite matrix $\widetilde{S}$, $\widetilde{S} = \widetilde{S}^T > 0$, which is to be used as a reliable and sufficiently cheap preconditioner for the system matrix S as the one given in (4.1).

Consider an algebraic system

$$S\bar{y} \;=\; \bar{g}\;, \tag{4.1}$$

where

$$S \;=\; \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{4.2}$$

is a 2×2 block-diagonal matrix with square diagonal submatrices S_{11} and S_{22}. We assume that the matrix S is symmetric and positive definite.

It can come from a system for the macro-hybrid mixed FE method as shown in Section 2.4, with $n \times n$ submatrix

$$S_{11} \;=\; \Sigma \;+\; BM^{-1}B^T \;, \tag{4.3}$$

$n \times m$ submatrix

$$S_{12} \;=\; S_{21}^T \;=\; BM^{-1}C^T \;, \tag{4.4}$$

and $m \times m$ submatrix

$$S_{22} \;=\; CM^{-1}C^T \;, \tag{4.5}$$

where n is the number of mesh cells, and m is the total number of both the interfaces Γ_{kl} between cells of Ω_h, and the faces of the cells E_k belonging to Γ_N. The size of the matrix S is equal to $N = n + m$. An equivalent definition of S as the assembling of matrices S_k constructed cell-by-cell, is given by

$$S \;=\; \sum_{k=1}^{n} \widetilde{\mathcal{N}}_k \; S_k \; \widetilde{\mathcal{N}}_k^T \;, \tag{4.6}$$

where

$$S_k \;=\; \begin{pmatrix} \Sigma_k & 0 \\ 0 & 0 \end{pmatrix} \;-\; \begin{pmatrix} B_k \\ C_k \end{pmatrix} M_k^{-1} \begin{pmatrix} B_k^T & C_k^T \end{pmatrix}, \quad k = \overline{1,\; n} \;, \tag{4.7}$$

and $\widetilde{\mathcal{N}}_k$ are corresponding assembling matrices.

The convergence rate of the PCG method is estimated using the condition number ν of the matrix $\widetilde{S}^{-1}S$, which is defined by

$$\nu(\widetilde{S}^{-1}S) \;=\; \frac{\lambda_{max}(\widetilde{S}^{-1}S)}{\lambda_{min}(\widetilde{S}^{-1}S)}\;, \tag{4.8}$$

where λ_{max} and λ_{min} are the maximal and minimal eigenvalues of the matrix $\widetilde{S}^{-1}S$.

Then, the energy norm (the S-norm) of the error vector $\bar{z}^k = \bar{y}^k - \bar{y}^*$, where $\bar{y}^* = S^{-1}\bar{g}$, is the solution vector of the system (4.1), and $\bar{y}^k$ is the iterative solution vector on k'th iteration, can be estimated by

$$\|\bar{z}^k\|_S \;\leq\; 2\left(\frac{\sqrt{\nu}\;-\;1}{\sqrt{\nu}\;+\;1}\right)^k \|\bar{z}^0\|_S\;. \tag{4.9}$$

Here, $\|\bar{z}\|_S = (S\bar{z},\;\bar{z})^{1/2}$ is the S-norm of a vector $\bar{z}$.

It follows that the convergence is faster for smaller values of $\nu(\widetilde{S}^{-1}S)$. Thus, we need to design a preconditioner $\widetilde{S}$ which provides a smaller value of $\nu(\widetilde{S}^{-1}S)$.

On each step of the PCG method we have to compute the residual $\bar{\xi}^k = S\bar{y}_k - \bar{g}^k$, and to solve the system

$$\widetilde{S}\bar{\eta}^k \;=\; \bar{\xi}^k \tag{4.10}$$

exactly, or to compute exactly the matrix-vector product $\widetilde{S}^{-1}\bar{\xi}^k$. The implementation of this computation procedure should be sufficiently cheap arithmetically (and logically). This is the second major problem in designing the efficient preconditioner for the matrix S. Another important requirement for an efficient implementation of the preconditioner is that the computation of the product $\widetilde{S}^{-1}\bar{\xi}^k$ should be well-parallelizable.

First, we give a general description of the approach proposed to design an efficient preconditioner for the matrix S resulting from a macro-hybrid mixed FE discretization of the diffusion equation. We design the preconditioner $\widetilde{S}$ for the matrix S in two stages.

Stage 1.

For each matrix M_k in (4.7), we define a diagonal matrix $\widetilde{M}_k$ with positive diagonal entries which is scaled so that

$$\lambda_{k,max}(\widetilde{M}_k^{-1} M_k) \;=\; 1 \;. \tag{4.11}$$

The exact procedure used to obtain the matrix $\widetilde{M}_k$ for the case of PWC discretization is described in the following sections. Then, we introduce matrices

$$\widetilde{S}_k \;=\; \gamma_k \begin{pmatrix} \Sigma_k & 0 \\ 0 & 0 \end{pmatrix} \;-\; \begin{pmatrix} B_k \\ C_k \end{pmatrix} \widetilde{M}_k^{-1} \begin{pmatrix} B_k^T & C_k^T \end{pmatrix} \;, \tag{4.12}$$

where

$$\gamma_k \;=\; \frac{1}{\lambda_{k,min}(\widetilde{M}_k^{-1} M_k)} \;, \tag{4.13}$$

$k = \overline{1,n}$, and the assembled matrix

$$\widetilde{S} \;=\; \sum_{k=1}^{n} \widetilde{\mathcal{N}}_k \; \widetilde{S}_k \; \widetilde{\mathcal{N}}_k^T \;. \tag{4.14}$$

Consider $\widetilde{S}$ as a 2×2 block matrix:

$$\widetilde{S} \;=\; \begin{pmatrix} \widetilde{S}_{11} & \widetilde{S}_{12} \\ \widetilde{S}_{21} & \widetilde{S}_{22} \end{pmatrix} \;, \tag{4.15}$$

where

$$\widetilde{S}_{22} \;=\; \sum_{k=1}^{n} \mathcal{N}_{k,22} \left(C_k \widetilde{M}_k^{-1} C_k^T \right) \mathcal{N}_{k,22}^T \;. \tag{4.16}$$

with the appropriate $m \times \widehat{m}_k$ assembling matrices $\mathcal{N}_{k,22}$, $k = \overline{1,n}$. The matrix $\widetilde{M}_k^{-1}$ is diagonal, so it follows that the matrices $C_k \widetilde{M}_k^{-1} C_k^T$, $k = \overline{1,n}$, and the matrix $\widetilde{S}_{22}$ are also diagonal with positive diagonal entries.

We can also derive the estimates for the minimal and maximal eigenvalues of the matrix $\widetilde{S}^{-1}S$ to estimate the rate of convergence from (4.9). First, we state that

$$\begin{aligned} 1 = \frac{1}{\lambda_{k,max}(\widetilde{M}_k^{-1} M_k)} \leq \lambda_{k,min}(\widetilde{S}_k^{-1} S_k) \leq \\ \lambda_{k,max}(\widetilde{S}_k^{-1} S_k) \leq \frac{1}{\lambda_{k,min}(\widetilde{M}_k^{-1} M_k)} \ . \end{aligned} \tag{4.17}$$

Also,

$$\begin{aligned} \min_k \lambda_{k,min}(\widetilde{S}_k^{-1} S_k) \leq \lambda_{min}(\widetilde{S}^{-1} S) \leq \\ \lambda_{max}(\widetilde{S}^{-1} S) \leq \max_k \lambda_{k,max}(\widetilde{S}_k^{-1} S_k) \ . \end{aligned} \tag{4.18}$$

Therefore,

$$\begin{aligned} \min_k \frac{1}{\lambda_{k,max}(\widetilde{M}_k^{-1} M_k)} \leq \lambda_{min}(\widetilde{S}^{-1} S) \leq \\ \lambda_{max}(\widetilde{S}^{-1} S) \leq \max_k \frac{1}{\lambda_{k,min}(\widetilde{M}_k^{-1} M_k)} \ . \end{aligned} \tag{4.19}$$

From the definition of $\widetilde{M}_k$ and γ_k, it follows that

$$\lambda_{min}(\widetilde{S}^{-1} S) \ \geq \ 1 \ , \quad \lambda_{max}(\widetilde{S}^{-1} S) \ \leq \ \max_{1 \leq k \leq n} \gamma_k \ . \tag{4.20}$$

Consider a system

$$\widetilde{S} \begin{pmatrix} \bar{\eta}_1 \\ \bar{\eta}_2 \end{pmatrix} = \begin{pmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \end{pmatrix} , \tag{4.21}$$

where $\bar{\eta}_1, \bar{\xi}_1 \in \mathbb{R}^n$, and $\bar{\eta}_2, \bar{\xi}_2 \in \mathbb{R}^m$ are corresponding subvectors of $\bar{\eta}$ and $\bar{\xi}$ as seen in (4.10). The block Gauss elimination method for this system can be implemented

in the following way. First, we eliminate by substitution the subvector $\bar{\eta}_2$ from the first block equation:

$$\bar{\eta}_2 \;=\; \widetilde{S}_{22}^{-1}\left(\bar{\xi}_2 \;-\; \widetilde{S}_{21}\bar{\xi}_1\right) \;, \tag{4.22}$$

where the diagonal matrix $\widetilde{S}_{22}$ is easy to invert. Then, we get the system

$$A_{11}\bar{\eta}_1 \;=\; \bar{z}_1 \;, \tag{4.23}$$

where $\bar{z}_1 = \bar{\xi}_1 - \widetilde{S}_{12}\;\widetilde{S}_{22}^{-1}\;\bar{\xi}_2$, and

$$A_{11} \;=\; \widetilde{S}_{11} \;-\; \widetilde{S}_{12}\;\widetilde{S}_{22}^{-1}\;\widetilde{S}_{21} \;. \tag{4.24}$$

It can be shown that A_{11} is a Stieltjes matrix, i.e. all off-diagonal entries of A_{11} are nonpositive, and A_{11} is symmetric positive definite. The Stieltjes matrices are very common in discretization of elliptic equations, for instance, by finite volume method. The major property of irreducible Stieltjes matrices is that all the entries of their inverses are positive. Symmetric and positive definite preconditioners for Stieltjes matrices are much easier to design compared to general positive definite matrices. Solving (4.23) by a direct method is still a very difficult problem, especially in the case of meshes Ω_h relevant to basin modeling.

After solving system (4.23), we can find the remaining solution vector $\bar{\eta}_2$ from (4.22).

Stage 2.

Let $B_{11} \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix which we consider to be a suitable preconditioner for the matrix A_{11} in (4.23). Then, we define a preconditioner $\widehat{S}$ for the matrix S by

$$\widehat{S} \;=\; \begin{pmatrix} B_{11} + \widetilde{S}_{11}\;\widetilde{S}_{11}^{-1}\;\widetilde{S}_{21} & \widetilde{S}_{12} \\ \widetilde{S}_{21} & \widetilde{S}_{22} \end{pmatrix} \;. \tag{4.25}$$

It can be proven that

$$\lambda_{min}(\widehat{S}^{-1}S) \;\geq\; \lambda_{min}\big(B_{11}^{-1}A_{11}\big)\cdot\lambda_{min}\big(\widetilde{S}^{-1}S\big) \tag{4.26}$$

and

$$\lambda_{max}(\widehat{S}^{-1}S) \;\leq\; \max\big\{1,\;\lambda_{max}\big(B_{11}^{-1}A_{11}\big)\big\}\;. \tag{4.27}$$

If we recall that the convergence rate of the PCG method is estimated using the condition number ν of the matrix $\widetilde{S}^{-1}S$ defined by (4.8), it follows that a good choice of matrices $\widetilde{M}_k$ for matrices M_k, $k=\overline{1,\;n}$, and the matrix B_{11} for the matrix A_{11}, provides a good preconditioner $\widehat{S}$ for the matrix S. In other words, if the value of $\lambda_{min}\big(\widetilde{S}^{-1}S\big)$ is not too small, and the ratio

$$\frac{\max\big\{1,\;\lambda_{max}\big(B_{11}^{-1}A_{11}\big)\big\}}{\lambda_{min}\big(B_{11}^{-1}A_{11}\big)}$$

is not too big, then the matrix $\widehat{S}$ is a good preconditioner for the matrix S.

4.2 The choice of the preconditioner B_{11}

The preconditioner B_{11} should be a suitable preconditioner for the matrix A_{11} defined by (4.24). We consider two possible choices of B_{11} which result in two different implementations of the proposed preconditioner. One is the well-known AMG preconditioner, and another is KPMDP, which is based on the preconditioner first introduced in [23].

AMG preconditioner

The description of the AMG preconditioner we use can be found in [45]. It is a well-known preconditioner suitable for matrices in question.

KPMDP preconditioner

The description of the KPMDP preconditioner is given in [24].

4.3 The choice of the diagonal matrix $\widetilde{M}_k$ for the PWC discretization on a prismatic cell.

We introduced the matrix $\widetilde{M}_k$ when we described the first stage of designing the preconditioner in Section 4.1. In this section, we present the procedure used to obtain said matrix in the case of PWC discretization on meshes described in Section 2.2.

For the sake of notation, we assume that the diffusion tensor K is the identity matrix.

4.3.1 Local mass matrix for the PWC discretization on a prismatic cell

Consider a prismatic mesh cell E. We choose a bottom vertex A and the "opposite" top vertex B of the prism. We denote by $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$, $\boldsymbol{\tau}_3$ three unit vectors directed along the edges of the prism and originating from the point A. In a similar way, we denote by $\boldsymbol{\tau}_4$, $\boldsymbol{\tau}_5$, and $\boldsymbol{\tau}_6$ three unit vectors originated from the point B, see Figure 4.1.

The mass matrix $M \in \mathbb{R}^{5\times 5}$ associated with prism E can be written as

$$M \;=\; \text{assembling}\Big\{\text{const}_A M_A,\; \text{const}_B M_B\Big\}\;, \tag{4.28}$$

where matrices $M_A \in \mathbb{R}^{3\times 3}$ and $M_B \in \mathbb{R}^{3\times 3}$ are described below, and the constants satisfy the relation

$$\text{const}_A \;+\; \text{const}_B \;=\; |E|\;. \tag{4.29}$$

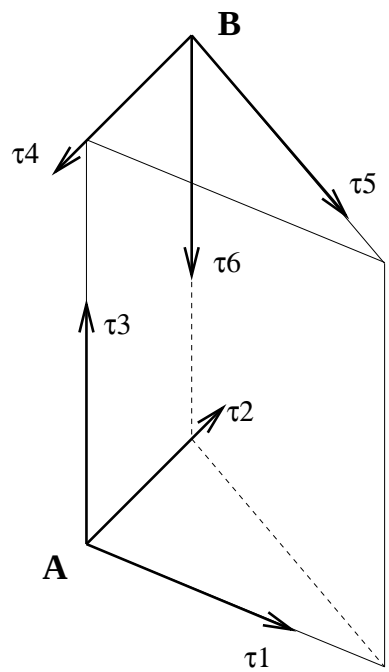

Figure 4.1: Prismatic mesh cell E and two triples of unit vectors directed along the edges of the prism

Exact formulas for matrices M_A and M_B

Let (i, j, k) be a cyclic permutation of the index triple $(1, 2, 3)$, i.e.

$$(i, j, k) \in \Big\{(1, 2, 3),\ (2, 3, 1),\ (3, 1, 2)\Big\}\ . \tag{4.30}$$

We denote by Γ_k the face of the prism containing vectors $\boldsymbol{\tau}_i$ and $\boldsymbol{\tau}_j$, by α_k the angle formed by $\boldsymbol{\tau}_i$ and $\boldsymbol{\tau}_j$, and by $\boldsymbol{n}_k$ the unit outward normal to Γ_k. Then, vectors $\boldsymbol{n}_k$ are given by formulas

$$\boldsymbol{n}_k \ = \ \mp \frac{1}{\|\boldsymbol{\tau}_i \times \boldsymbol{\tau}_j\|}\ \boldsymbol{\tau}_i \times \boldsymbol{\tau}_j\ . \tag{4.31}$$

We should use the sign "$-$" if $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{\tau}_3)$ form a **right** triple of vectors (as shown in Figure 4.1). Otherwise, we should use the sign "$+$".

According to the construction of the PWC vector fields, the inverse of the matrix M_A is given by

$$M_A^{-1} \;=\; N_A^T\, N_A \tag{4.32}$$

where $N_A \in \mathbb{R}^{3\times 3}$ is the matrix whose columns are the normal vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$, and $\boldsymbol{n}_3$:

$$N_A \;=\; \begin{bmatrix} \boldsymbol{n}_1 & \boldsymbol{n}_2 & \boldsymbol{n}_3 \end{bmatrix} \in \mathbb{R}^{3\times 3}\ . \tag{4.33}$$

It is clear that the diagonal entries of M_A^{-1} are equal to one:

$$\left(M_A^{-1}\right)_{11} \;=\; \left(M_A^{-1}\right)_{22} \;=\; \left(M_A^{-1}\right)_{33} \;=\; 1\ . \tag{4.34}$$

Also, it can be easily shown that off-diagonal entries of M_A^{-1} are given by

$$\left(M_A^{-1}\right)_{ij} \;=\; \frac{\cos(\alpha_i)\cos(\alpha_j) - \cos(\alpha_k)}{\sin(\alpha_i)\sin(\alpha_j)}. \tag{4.35}$$

Thus,

$$M_A^{-1} \;=\; \begin{pmatrix} 1 & \frac{c_1c_2 - c_3}{s_1s_2} & \frac{c_1c_3 - c_2}{s_1s_3} \\ \frac{c_2c_1 - c_3}{s_2s_1} & 1 & \frac{c_2c_3 - c_1}{s_2s_3} \\ \frac{c_3c_1 - c_2}{s_3s_1} & \frac{c_3c_2 - c_1}{s_3s_2} & 1 \end{pmatrix}\ , \tag{4.36}$$

where

$$c_k \;=\; \cos(\alpha_k), \quad s_k = \sin(\alpha_k)\ . \tag{4.37}$$

Using the triple of vectors $\boldsymbol{\tau}_4$, $\boldsymbol{\tau}_5$, and $\boldsymbol{\tau}_6$, we find a similar formula for the matrix M_B^{-1}:

$$M_B^{-1} \;=\; \begin{pmatrix} 1 & \frac{c_4c_5 - c_6}{s_4s_5} & \frac{c_4c_6 - c_5}{s_4s_6} \\ \frac{c_5c_4 - c_6}{s_5s_4} & 1 & \frac{c_5c_6 - c_4}{s_5s_6} \\ \frac{c_6c_4 - c_5}{s_6s_4} & \frac{c_6c_5 - c_4}{s_6s_5} & 1 \end{pmatrix}\ . \tag{4.38}$$

4.3.2 Possible choices of the matrix $\widetilde{M}_k$ and related eigenvalue problem

For the matrix M_k described in Section 4.3.1,

$$M_k \;=\; \text{assembling}\Big\{\text{const}_{A_k} M_{A_k},\; \text{const}_{B_k} M_{B_k}\Big\}\;, \tag{4.39}$$

we have to choose a diagonal matrix $\widetilde{M}_k$ represented by

$$\widetilde{M}_k \;=\; \text{assembling}\Big\{\widetilde{\text{const}}_{A_k} \widetilde{M}_{A_k},\; \widetilde{\text{const}}_{B_k} \widetilde{M}_{B_k}\Big\}, \tag{4.40}$$

where the matrix $\widetilde{M}_{A_k} \in \mathbb{R}^{3\times 3}$ is "close" to M_{A_k}, and the matrix $\widetilde{M}_{B_k}$ is "close" to M_{B_k}.

The matrix $\widetilde{M}_k$ is used in the construction of the preconditioner $\widehat{S}$ described in Section 4.1, and has to satisfy the condition $\lambda_{k,max}(\widetilde{M}_k^{-1} M_k) \;=\; 1$. Also, we want $\lambda_{min}(\widetilde{S}^{-1} S)$, $\lambda_{min}(\widetilde{S}^{-1} S) \geq \lambda_{k,min}(\widetilde{M}_k^{-1} M_k)$, to be not too small.

We can either replace the matrices by their diagonals:

$$\widetilde{M}_{A_k} \;=\; \begin{pmatrix} M_{A_k,11} & 0 & 0 \\ 0 & M_{A_k,22} & 0 \\ 0 & 0 & M_{A_k,33} \end{pmatrix}, \quad \widetilde{M}_{B_k} \;=\; \begin{pmatrix} M_{B_k,11} & 0 & 0 \\ 0 & M_{B_k,22} & 0 \\ 0 & 0 & M_{B_k,33} \end{pmatrix}, \tag{4.41}$$

or replace the matrices by diagonal matrices

$$\widetilde{M}_{A_k} \;=\; \begin{pmatrix} d_{A_k,1} & 0 & 0 \\ 0 & d_{A_k,2} & 0 \\ 0 & 0 & d_{A_k,3} \end{pmatrix}, \quad \widetilde{M}_{B_k} \;=\; \begin{pmatrix} d_{B_k,1} & 0 & 0 \\ 0 & d_{B_k,2} & 0 \\ 0 & 0 & d_{B_k,3} \end{pmatrix}, \tag{4.42}$$

where

$$d_{A_k,j} \text{ is "close" to } M_{A_k,jj} \text{ and } d_{B_k,j} \text{ is "close" to } M_{B_k,jj}, \quad j = 1,2,3. \tag{4.43}$$

In order to determine the better choice, we consider the eigenvalue problem

$$\widetilde{M}^{-1/2} M \widetilde{M}^{-1/2}\, \bar{w} \;=\; \lambda \bar{w}\;, \tag{4.44}$$

where $M \in \mathbb{R}^{3\times 3}$ is a symmetric positive definite matrix, and $\widetilde{M}$ is defined as its diagonal part, i.e.

$$\widetilde{M} \;=\; \begin{pmatrix} m_{11} & 0 & 0 \\ 0 & m_{22} & 0 \\ 0 & 0 & m_{33} \end{pmatrix} . \tag{4.45}$$

The matrix $A \;=\; \widetilde{M}^{-1/2} M \widetilde{M}^{-1/2}$ is then an SPD matrix.

4.3.3 The suggested choice of $\widetilde{M}_k$

After performing numerical experiments, we could conclude that the choice of $\widetilde{M}_{A_k}$ different from the diagonal of M_{A_k} doesn't significally improve the condition number of the matrix $\widetilde{M}_{A_k}^{-1} M_{A_k}$, but requires us to solve an additional minimization problem. Therefore, we choose $\widetilde{M}_{A_k}$ and $\widetilde{M}_{B_k}$ to be the diagonal of M_{A_k} and M_{B_k} correspondingly.

We can easily find find the minimal and maximal eigenvalues of the matrices $\widetilde{M}_{A_k}^{-1} M_{A_k}$ and $\widetilde{M}_{B_k}^{-1} M_{B_k}$, so we can write the following spectral inequalities:

$$\lambda_{min}(\widetilde{M}_{A_k}^{-1} M_{A_k}) \cdot \widetilde{M}_{A_k} \leq M_{A_k} \leq \lambda_{max}(\widetilde{M}_{A_k}^{-1} M_{A_k}) \cdot \widetilde{M}_{A_k} \tag{4.46}$$

and

$$\lambda_{min}(\widetilde{M}_{B_k}^{-1} M_{B_k}) \cdot \widetilde{M}_{B_k} \leq M_{B_k} \leq \lambda_{max}(\widetilde{M}_{B_k}^{-1} M_{B_k}) \cdot \widetilde{M}_{B_k} \tag{4.47}$$

As we recall, the matrix M_k is obtained by assembling the matrices M_{A_k} and M_{B_k},

$$M_k \;=\; \text{assembling}\Big\{\text{const}_{A_k} M_{A_k},\; \text{const}_{B_k} M_{B_k}\Big\} \;,$$

Therefore, we obtain the spectral inequality for the matrix M_k:

$$\begin{aligned} &\min\left\{\text{const}_{A_k}\lambda_{min}(\widetilde{M}_{A_k}^{-1} M_{A_k}),\; \text{const}_{B_k}\lambda_{min}(\widetilde{M}_{B_k}^{-1} M_{B_k})\right\} \cdot \widehat{M}_k \leq M_k \\ &\leq \max\left\{\text{const}_{A_k}\lambda_{max}(\widetilde{M}_{A_k}^{-1} M_{A_k}),\; \text{const}_{B_k}\lambda_{max}(\widetilde{M}_{B_k}^{-1} M_{B_k})\right\} \cdot \widehat{M}_k \;, \end{aligned} \tag{4.48}$$

where

$$\widehat{M}_k \;=\; \text{assembling}\Big\{\text{const}_{A_k}\widetilde{M}_{A_k},\;\text{const}_{B_k}\widetilde{M}_{B_k}\Big\}\;.$$

In order to obtain the matrix $\widetilde{M}_k$, we just have to scale the matrix $\widehat{M}_k$:

$$\widetilde{M}_k \;=\; \max\left\{\text{const}_{A_k}\lambda_{max}(\widetilde{M}_{A_k}^{-1}M_{A_k}),\;\text{const}_{B_k}\lambda_{max}(\widetilde{M}_{B_k}^{-1}M_{B_k})\right\}\cdot\widehat{M}_k\;. \tag{4.49}$$

Denoting the minimal eigenvalue bound by α_k,

$$\alpha_k \;=\; \frac{\min\left\{\text{const}_{A_k}\lambda_{min}(\widetilde{M}_{A_k}^{-1}M_{A_k}),\;\text{const}_{B_k}\lambda_{min}(\widetilde{M}_{B_k}^{-1}M_{B_k})\right\}}{\max\left\{\text{const}_{A_k}\lambda_{max}(\widetilde{M}_{A_k}^{-1}M_{A_k}),\;\text{const}_{B_k}\lambda_{max}(\widetilde{M}_{B_k}^{-1}M_{B_k})\right\}}\;, \tag{4.50}$$

we can write

$$\alpha_k\widetilde{M}_k \le M_k \le 1\cdot\widetilde{M}_k\;. \tag{4.51}$$

Hence, the matrix $\widetilde{M}_k$ satisfies the requirement posed in Section 4.1, i.e. $\lambda_{k,max}(\widetilde{M}_k^{-1}M_k) = 1$, and is a good choice among the diagonal matrices with respect to the minimization of $\text{cond}(\widetilde{M}_k^{-1}M_k)$.

4.4 Justification of the proposed choice of the parameter γ_k

In Section 4.1 we introduced the matrix $\widetilde{S}_k$,

$$\widetilde{S}_k \;=\; \gamma_k \begin{pmatrix} \Sigma_k & 0 \\ 0 & 0 \end{pmatrix} \;-\; \begin{pmatrix} B_k \\ C_k \end{pmatrix} \widetilde{M}_k^{-1} \begin{pmatrix} B_k^T & C_k^T \end{pmatrix} \;,$$

the parameter γ_k

$$\gamma_k \;=\; \frac{1}{\lambda_{k,min}(\widetilde{M}_k^{-1} M_k)} \;,$$

and the assembled matrix

$$\widetilde{S} \;=\; \sum_{k=1}^{n} \widetilde{\mathcal{N}}_k \; \widetilde{S}_k \; \widetilde{\mathcal{N}}_k^T \;,$$

which were first defined by (4.12), (4.13), and (4.14). Our reasoning for choosing γ_k in that particular way is as follows. First, from (4.17) and (4.51) we write the spectrum inequalities for the matrix S_k:

$$\min\left\{\gamma_k,\; 1\right\} \cdot \widetilde{S}_k \leq S_k \leq \max\left\{\gamma_k,\; \frac{1}{\alpha_k}\right\} \cdot \widetilde{S}_k \;. \tag{4.52}$$

Then, from (4.18) it follows that

$$\min_k \left(\min\left\{\gamma_k,\; 1\right\}\right) \cdot \widetilde{S} \leq S \leq \max_k \left(\max\left\{\gamma_k,\; \frac{1}{\alpha_k}\right\}\right) \cdot \widetilde{S} \;. \tag{4.53}$$

Therefore, to keep $\mathrm{cond}(\widetilde{S}^{-1}S)$ from increasing, we have to limit our choice of the parameter γ_k to the following interval:

$$1 \;\leq\; \gamma_k \;\leq\; \frac{1}{\alpha_k} \;. \tag{4.54}$$

Let us recall that on every iteration of the PCG method we have to solve the system (4.23) with the matrix A_{11} defined by (4.24) to be

$$A_{11} \;=\; \widetilde{S}_{11} \;-\; \widetilde{S}_{12} \; \widetilde{S}_{22}^{-1} \; \widetilde{S}_{21} \;,$$

where

$$\widetilde{S}_{11} \;=\; \widetilde{\Sigma} \;+\; B\widetilde{M}^{-1}B^T \tag{4.55}$$

with

$$\widetilde{\Sigma} \;=\; \begin{pmatrix} \gamma_1\sigma_1 & & \\ & \ddots & \\ & & \gamma_n\sigma_n \end{pmatrix}, \tag{4.56}$$

and $B\widetilde{M}^{-1}B^T$ is also a diagonal matrix.

Hence, by increasing the values of γ_k, $k = \overline{1,n}$, we increase the diagonal dominance in the matrix A_{11}, thus reducing the number of iterations required to solve system (4.23). Therefore, the best choice for γ_k is the maximum value in the allowed interval (4.54), i.e.

$$\gamma_k \;=\; \frac{1}{\alpha_k} \;=\; \frac{1}{\lambda_{k,min}(\widetilde{M}_k^{-1}M_k)}$$

as it was defined by (4.13) in Section 4.1.

4.5 Numerical experiments

In this section, we provide the description of the domains and corresponding meshes used for the comparison of the preconditioners, and give the performance results obtained for each of the examples.

We assume that the diffusion equation (2.1) comes from the discretization in time variable of the unsteady diffusion equation by the implicit finite difference method with the time step Δt_{imp}. Then, we assume that the coefficient c is a positive constant defined by the formula

$$c = \frac{1}{\Delta t_{imp}} . \tag{4.57}$$

For numerical experiments we choose

$$\Delta t_{imp} = \kappa \sqrt{\Delta t_{exp}} , \tag{4.58}$$

where κ is a positive factor, and Δt_{exp} is chosen to be of the order of the inverse of the minimal mesh step in order to provide the stability of the scheme.

4.5.1 The decription of test domains and meshes

To compare the performance of the proposed preconditioner with existing competitors, we consider three test examples which are relevant to basin modeling. Every geological domain considered contains at least one "thin" layer.

For all of the examples given, the diffusion tensor is diagonal and piece-wise constant, i.e. K_s, the diffusion tensor in the s-th layer, is as follows:

$$K_s = \begin{pmatrix} K_{s,xy} & 0 & 0 \\ 0 & K_{s,xy} & 0 \\ 0 & 0 & K_{s,z} \end{pmatrix} , \tag{4.59}$$

where $K_{s,xy}$ and $K_{s,z}$ are given constants.

Domain with three oblique geological layers

This example features a domain with three geological layers, where the layer in the middle is "thin" and oblique. The mesh is conforming, uniform in the xy-plane, and is uniform along z-direction inside each geological layer.

The domain is a parallelepiped with the dimensions $1.0 \times 1.0 \times 0.25$ in x, y, and z coordinates respectively. The mesh is uniform in x and y coordinates with the step $h_{xy} = 0.3125$, i.e. we have a grid of 32×32 square cells, each of which is then split into two triangles, resulting in 2048 bases for triangular prisms on each horizontal mesh layer.

The geometry of the geological layers is as follows:

Geological layer #1:

At the bottom, it is bordered by the $(x,\ y,\ 0)$ plane. At the top it is bordered by the plane passing through the point $(0,\ 0,\ 0.12)$. This plane formes an angle $\theta_x = 5°$ with the x-axis, and an angle $\theta_y = -5°$ with the y-axis. The mesh is uniform along z-direction with the step $h_{z,1} \in [0.004644,\ 0.029641]$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 7)$ triangular prisms.

Geological layer #2:

At the bottom, it is bordered by the top boundary of the layer #1, at the top by the plane passing through the point $(x,\ y,\ 0.12001)$ and parallel to the one at the bottom. It is a "thin" layer with a thickness of 10^{-5}.

The mesh in z-coordinate is uniform with the step $h_{z,2} = 0.000005$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 2)$ triangular prisms.

Geological layer #3:

At the bottom, it is bordered by the top boundary of the layer #2, at the top by the $(x,\ y,\ 0.25)$ plane.

The mesh is uniform along z-direction with the step $h_{z,3} \in [0.006072,\ 0.031068]$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 7)$ triangular prisms.

Key parameters of the mesh cells in each geological layer are given in Table 4.1.

Table 4.1: Geometrical parameters of the mesh cells

	Layer #1	Layer #2	Layer #3
h_{xy}	0.3125	0.3125	0.3125
h_z	[0.004644, 0.029641]	0.000005	[0.006072, 0.031068]
h_{xy}/h_z	[1.0543, 6.7284]	6250	[1.0058, 5.1469]

The diffusion tensor in each layer is chosen as shown in Table 4.2.

Table 4.2: Diffusion tensor parameters

	Layer #1	Layer #2	Layer #3
$K_{s,xy}$	5	10000	10
$K_{s,z}$	1	1000	1

The domain is pictured on Figure 4.2 with the close-up of the mesh in the "thin" layer given on Figure 4.3.

The total number of mesh cells is $2 \cdot (32 \times 32 \times 16) = 32768$.

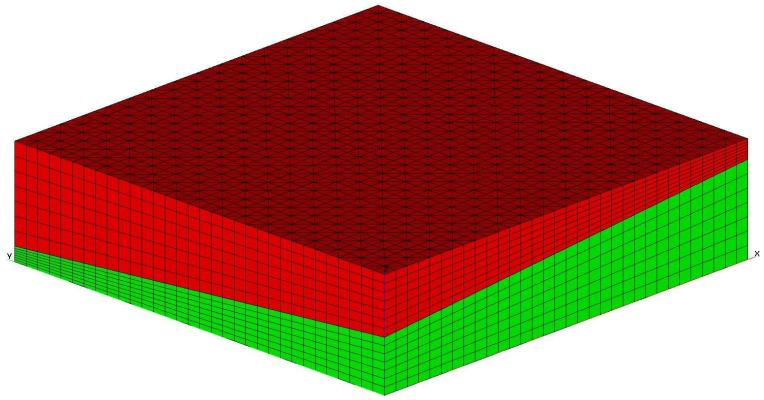

Figure 4.2: Domain with three oblique geological layers

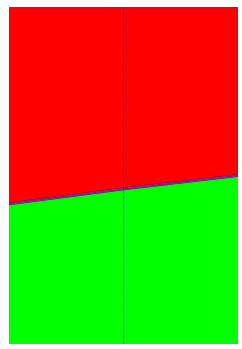

Figure 4.3: Close-up of the oblique "thin" layer

Domain with five oblique geological layers

This example features a domain with five geological layers, where the two of them are "thin", oblique, and parallel to each other. The mesh used is conforming, uniform in the xy-plane, and is uniform along z-direction inside each geological layer.

The domain is a parallelepiped with the dimensions $1.0 \times 1.0 \times 0.25$ in x, y, and z coordinates respectively. The mesh is uniform in x and y coordinates with the step $h_{xy} = 0.3125$, i.e. we have a grid of 32×32 square cells each of which is then split into two triangles, resulting in 2048 bases for triangular prisms on each horizontal

mesh layer.

The geometry of the geological layers is as follows:

Geological layer #1:

At the bottom, it is bordered by the $(x,\ y,\ 0)$ plane. At the top it is bordered by the plane passing through the point $(0,\ 0,\ 0.081426)$. This plane formes an angle $\theta_x = 3^\circ$ with the x-axis, and an angle $\theta_y = -3^\circ$ with the y-axis. The mesh is uniform along z-direction with the step $h_{z,1} \in [0.003627,\ 0.016729]$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 8)$ triangular prisms.

Geological layer #2:

At the bottom, it is bordered by the top boundary of the layer #1, at the top by the plane passing through the point $(x,\ y,\ 0.081436)$ and parallel to the one at the bottom. It is a "thin" layer with a thickness of 10^{-5}.

The mesh in z-coordinate is uniform with the step $h_{z,2} = 0.000005$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 2)$ triangular prisms.

Geological layer #3:

At the bottom, it is bordered by the top boundary of the layer #2, at the top by the plane passing through the point $(x,\ y,\ 0.1601883)$ and parallel to the one at the bottom.

The mesh is uniform along z-direction with the step $h_{z,3} = 0.013125$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 6)$ triangular prisms.

Geological layer #4:

At the bottom, it is bordered by the top boundary of the layer #3, at the top by the plane passing through the point $(x,\ y,\ 0.1602883)$ and parallel to the one at the

bottom. It is a "thin" layer with a thickness of 10^{-4}.

The mesh in z-coordinate is uniform with the step $h_{z,4} = 0.00005$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 2)$ triangular prisms.

Geological layer #5:

At the bottom, it is bordered by the top boundary of the layer #4, at the top by the $(x,\ y,\ 0.25)$ plane.

The mesh is uniform along z-direction with the step $h_{z,5} \in [0.004663,\ 0.017765]$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 8)$ triangular prisms.

Key parameters of the mesh cells in each geological layer are given in Table 4.3.

Table 4.3: Geometrical parameters of the mesh cells

	h_{xy}	h_z	h_{xy}/h_z
Layer #1	0.3125	[0.003627, 0.016729]	[1.868, 8.6154]
Layer #2	0.3125	0.000005	6250
Layer #3	0.3125	0.013125	2.3809
Layer #4	0.3125	0.00005	625
Layer #5	0.3125	[0.004663, 0.017765]	[1.7591, 6.7017]

The diffusion tensor in each layer is chosen as shown in Table 4.4.

The domain is pictured on Figure 4.4.

The total number of mesh cells is $2 \cdot (32 \times 32 \times 26) = 53248$.

Table 4.4: Diffusion tensor parameters

	Layer #1	Layer #2	Layer #3	Layer #4	Layer #5
$K_{s,xy}$	5	10000	10	1000	10
$K_{s,z}$	1	1000	5	500	1

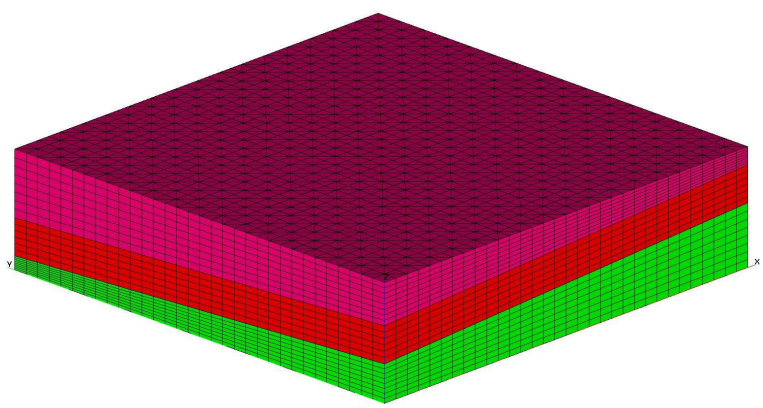

Figure 4.4: Domain with five oblique geological layers

Domain with five oblique and "bended" geological layers

This example features a domain with five geological layers, where the two of them are "thin", oblique, non-parallel, and "bended", i.e. their inclination angle changes at some point. The mesh used is conforming, uniform in the xy-plane, and is uniform along z-direction inside each geological layer.

The domain is a parallelepiped with the dimensions $1.0 \times 1.0 \times 0.25$ in x, y, and z coordinates respectively. The mesh is uniform for x and y coordinates with the step $h_{xy} = 0.3125$, i.e. we have a grid of 32×32 square cells each of which is then split into two triangles, resulting in 2048 bases for the triangular prisms on each horizontal

mesh layer.

The geometry of the geological layers is as follows:

Geological layer #1:

At the bottom, it is bordered by the $(x,\ y,\ 0)$ plane. At the top it is bordered by the "bended" plane passing through the point $(0,\ 0,\ 0.05)$.

This plane formes an angle $\theta_{x,1,1} = 10^\circ$ with the x-axis for $x \in (0,\ 0.25)$, and an angle $\theta_{x,1,2} = -5^\circ$ for $x \in (0.25,\ 1)$.

The angle between the plane and the y-axis is $\theta_{y,1} = 2^\circ$ for $y \in (0,\ 1)$.

The mesh is uniform along z-direction with the step $h_{z,1} \in [0.003558,\ 0.016125]$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 8)$ triangular prisms.

Geological layer #2:

At the bottom, it is bordered by the top boundary of the layer #1, at the top by the "bended" plane passing through the point $(x,\ y,\ 0.05001)$ and parallel to the one at the bottom. It is a "thin" layer with a thickness of 10^{-5}.

The mesh in z-coordinate is uniform with the step $h_{z,2} = 0.000005$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 2)$ triangular prisms.

Geological layer #3:

At the bottom, it is bordered by the top boundary of the layer #2. At the top it is bordered by the "bended" plane passing through the point $(0,\ 0,\ 0.16)$.

This plane formes an angle $\theta_{x,2,1} = 4^\circ$ with the x-axis for $x \in (0,\ 0.625)$, and an angle $\theta_{x,2,2} = -10^\circ$ for $x \in (0.625,\ 1)$.

The angle between the plane and the y-axis is $\theta_{y,2} = -2^\circ$ for $y \in (0,\ 1)$.

The mesh is uniform along z-direction with the step $h_{z,3} \in [0.002258,\ 0.023737]$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 6)$ triangular prisms.

Geological layer #4:

At the bottom, it is bordered by the top boundary of the layer #3, at the top by the "bended" plane passing through the point $(x,\ y,\ 0.1601)$ and parallel to the one at the bottom. It is a "thin" layer with a thickness of 10^{-4}.

The mesh in z-coordinate is uniform with the step $h_{z,4} = 0.00005$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 2)$ triangular prisms.

Geological layer #5:

At the bottom, it is bordered by the top boundary of the layer #4, at the top by the $(x,\ y,\ 0.25)$ plane.

The mesh is uniform along z-direction with the step $h_{z,5} \in [0.005775,\ 0.018405]$.

The mesh for this layer consists of $2 \cdot (32 \times 32 \times 8)$ triangular prisms.

Key parameters of the mesh cells in each geological layer are given in Table 4.5.

Table 4.5: Geometrical parameters of the mesh cells

	h_{xy}	h_z	h_{xy}/h_z
Layer #1	0.3125	[0.003558, 0.016125]	[1.9379, 8.7826]
Layer #2	0.3125	0.000005	6250
Layer #3	0.3125	[0.002258, 0.023737]	[1.3165, 13.839]
Layer #4	0.3125	0.00005	625
Layer #5	0.3125	[0.005775, 0.018405]	[1.6979, 5.4118]

The diffusion tensor in each layer is chosen as shown in Table 4.6.

The domain is pictured on Figure 4.5.

Table 4.6: Diffusion tensor parameters

	Layer #1	Layer #2	Layer #3	Layer #4	Layer #5
$K_{s,xy}$	5	10000	10	1000	10
$K_{s,z}$	1	1000	5	500	1

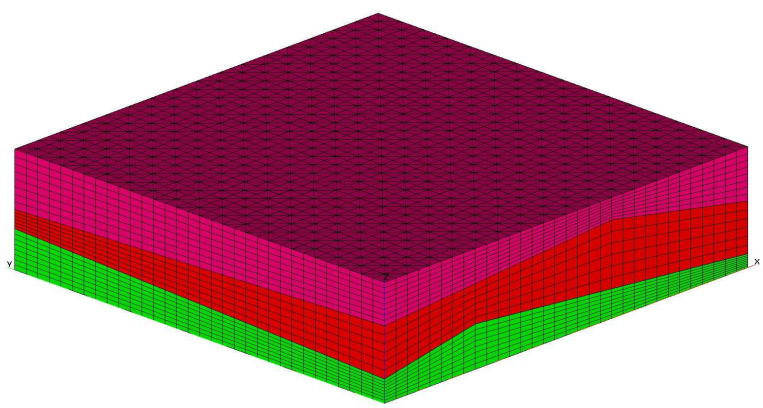

Figure 4.5: Domain with five oblique "bended" geological layers

The total number of mesh cells is $2 \cdot (32 \times 32 \times 26) = 53248$.

4.5.2 Numerical comparison of preconditioners

For each of the examples described in Section 4.5.1, we perform the comparison of the preconditioners used in the PCG method to solve system (4.1) with the right-hand side vector equal to zero, i.e. we solve the system

$$S \begin{pmatrix} \bar{p} \\ \bar{p}_\Gamma \end{pmatrix} = \bar{0} \ , \tag{4.60}$$

or

$$S\bar{z} = \bar{0} \ . \tag{4.61}$$

We compare the performance of the diagonal preconditioner (DIAG), the well-known AMG preconditioner (AMG), and the preconditioner proposed in this report.
In the case of our new preconditioner $H = \widehat{S}^{-1}$, we consider two possible choices for the internal substitution of the matrix A_{11} by its preconditioner B_{11} as shown in (4.25). Using AMG preconditioner for that purpose gives us the first variant ($\widehat{S}_{AMG}$). The alternative is to use KPMDP preconditioner, which gives us the second variant ($\widehat{S}_{KP}$).

We set $\bar{z}^0$ to be a random initial guess such that

$$\|\bar{z}^0\|_S = 1, \tag{4.62}$$

and use the stopping condition $\|\bar{z}^k\|_S < 10^{-6}$.

We perform all the experiments for two choices of the coefficient κ from (4.58), $\kappa = 1$ and $\kappa = 0.1$. This allows us to capture the dynamics of the preconditioners' performance with respect to the choice of the time step.

Table 4.7: Domain with one "thin" layer, $\dim(S) = 115712$

	$\kappa = 1,\ c = 3421.66$		$\kappa = 0.1,\ c = 34216.6$	
	PCG iter.	PCG time, s	PCG iter.	PCG time, s
DIAG	324	2.912	180	1.681
AMG	112	11.29	104	9.687
$\widehat{S}_{AMG}$	16	0.924	14	0.797
$\widehat{S}_{KP}$	29	0.756	27	0.525

Table 4.8: Domain with two parallel "thin" layers, $\dim(S) = 188032$

	$\kappa = 1,\ c = 2216.91$		$\kappa = 0.1,\ c = 22169.1$	
	PCG iter.	PCG time, s	PCG iter.	PCG time, s
DIAG	891	14.72	290	4.939
AMG	93	17.57	87	15.83
$\widehat{S}_{AMG}$	15	1.626	15	1.369
$\widehat{S}_{KP}$	31	1.879	26	0.983

Table 4.9: Domain with two "bended, thin" layers $\dim(S) = 188032$

	$\kappa = 1,\ c = 2237.66$		$\kappa = 0.1,\ c = 22376.6$	
	PCG iter.	PCG time, s	PCG iter.	PCG time, s
AMG	221	37.26	169	27.74
$\widehat{S}_{AMG}$	17	1.719	17	1.509
$\widehat{S}_{KP}$	36	2.196	31	1.125

The obtained results demonstrate the advantage of the new preconditioner. Between two variants of its implementation, the one using KPMDP preconditioner is the better option in the majority of the tests, but the other variant starts to perform better as the time step increases.

Chapter 5

Non-conforming meshes on domains with parallel oblique geological layers

5.1 Benchmark problem formulation

In this section, we consider the Neumann boundary value problem for the diffusion equation

$$\begin{aligned} -\nabla \cdot \big(K\ \nabla p\big)\ +\ cp\ &=\ F \quad \text{in} \quad G\ , \\ -\,(K\ \nabla p)\cdot \boldsymbol{n}\ &=\ G_N \quad \text{on} \quad \partial G\ , \end{aligned} \tag{5.1}$$

in a rectangular domain G. Here, $K = K(x, z)$ is a two-by-two symmetric positive definite matrix, $c = c(x, z)$ is a nonnegative function, ∂G is the boundary of G, $\boldsymbol{n}$ is the unit outward normal to ∂G, $F = F(x, z)$ and $G_N = G_N(x, z)$ are given functions.

We assume that the domain G is a partition of oblique layers G_s, $s = \overline{1,\ L}$, as shown on Figure 5.1. The layers are parallel and form an angle α with the x-axis. In order

to express the solution function $p = p(x, z)$ of (5.1) analytically, we impose a number of restrictions specified below.

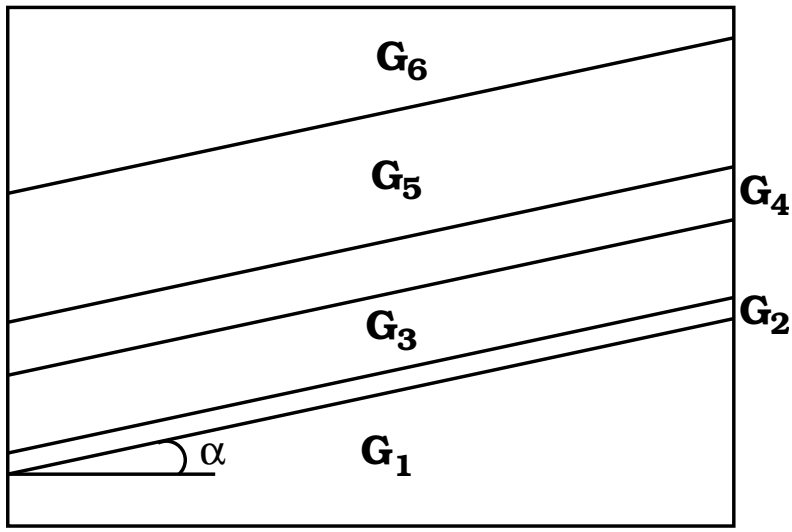

Figure 5.1: Domain with oblique geological layers

First, we assume the diffusion tensor K to be piece-wise constant in G, i.e.

$$K|_{G_s} \equiv K_s \equiv \text{const}_s \in \mathbb{R}^{2\times 2}.$$

We also assume that

$$K_s \;=\; W \begin{pmatrix} k_{s,x} & 0 \\ 0 & k_{s,z} \end{pmatrix} W^T \quad \text{in} \quad G_s, \quad s = \overline{1,\ L}\ , \tag{5.2}$$

is obtained by the rotation of a constant diagonal tensor. Here,

$$W = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} \tag{5.3}$$

is the rotation matrix dependent on the inclination angle of the geological layers.

Second, we impose similar restrictions on c, that is

$$c|_{G_s} \equiv c_s \equiv \text{const}_s > 0.$$

Third, we assume that the region G is a part of a larger rectangular domain $\widehat{G}$ as shown on Figure 5.2. Notice that subregions G_s, $s = \overline{1,\ L}$, can be extended to horizontal layers $\widehat{G}_s$ in $\widehat{G}$. With $\widehat{G}$, we associate the Cartesian system $(\widehat{x},\ \widehat{z})$, which is obtained by a shift and a clockwise rotation of the system $(x,\ z)$, i.e.

$$\begin{pmatrix} \widehat{x} \\ \widehat{z} \end{pmatrix} = \boldsymbol{r}_0 + W \begin{pmatrix} x \\ z \end{pmatrix} , \tag{5.4}$$

where $\boldsymbol{r}_0$ is the vector connecting the origins of the corresponding coordinate systems.

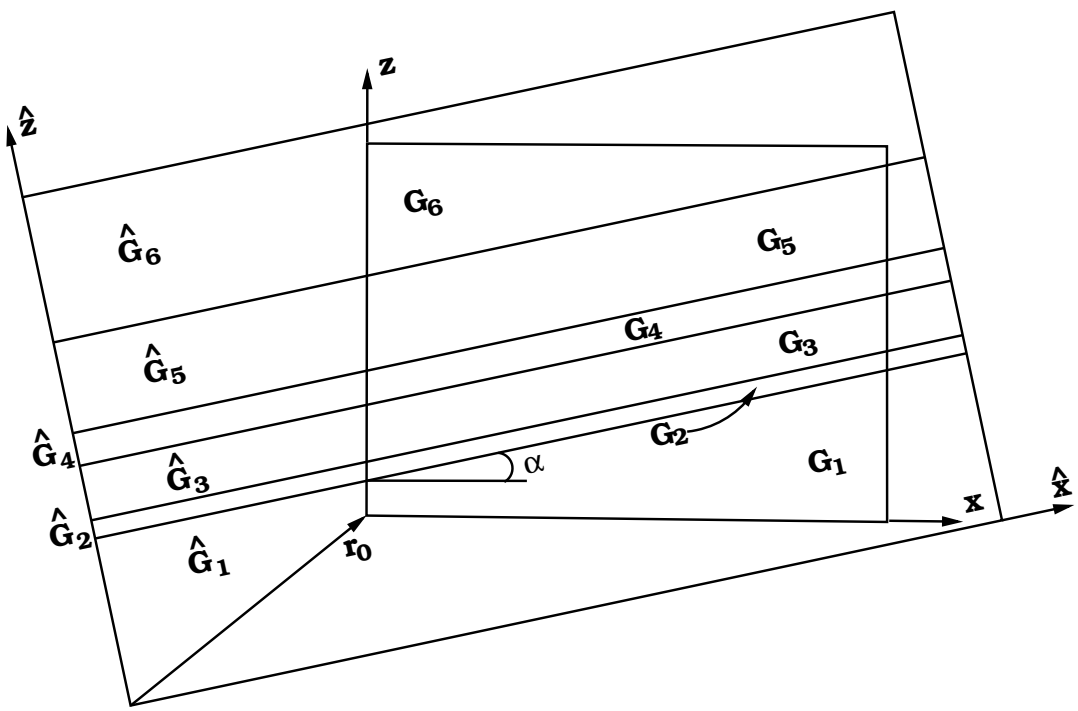

Figure 5.2: Embedding of the domain G into a larger domain $\widehat{G}$

Now, we introduce functions $\widehat{K}$, $\widehat{c}$, and $\widehat{F}$ defined on the domain $\widehat{G}$ so that

$$\widehat{K}|_{G_s}(\widehat{x},\ \widehat{z}) \ = \ \begin{pmatrix} k_{x,s} & 0 \\ 0 & k_{z,s} \end{pmatrix}, \quad \widehat{c}|_{G_s}(\widehat{x},\ \widehat{z}) \ = \ c_s, \tag{5.5}$$
$$\text{and} \quad \widehat{F}|_{G_s}(\widehat{x},\ \widehat{z}) \ = \ F|_{G_s}(x,\ z),$$

i.e. these functions are extensions of the functions K, c, and F from the domain G to the domain $\widehat{G}$, with the values of $\widehat{K}$, $\widehat{c}$, and $\widehat{F}$ coinciding with the corresponding values of K, c, and F in subregions G_s, $s = \overline{1,\ L}$, under the respective change of coordinates.

For simplicity in notations, let us assume that

$$\widehat{G} \ \equiv \ (0,\ 1) \times (\widehat{z}_0,\ \widehat{z}_L) \ , \tag{5.6}$$

and consider the eigenproblem for the operator $-\dfrac{d^2}{d\widehat{x}^2}$:

$$\begin{aligned} -\frac{d^2}{d\widehat{x}^2} w \ &= \ \lambda_{\widehat{x}} w, \quad 0 < \widehat{x} < 1, \\ \frac{dw}{d\widehat{x}}(0) \ &= \ 0, \quad \frac{dw}{d\widehat{x}}(1) \ = \ 0. \end{aligned} \quad . \tag{5.7}$$

The eigenpairs $(\lambda_{\widehat{x},n},\ w_n)$ for this problem are

$$\begin{aligned} &\lambda_{\widehat{x},0} = 0, \quad w_0 \equiv 1, \\ &\lambda_{\widehat{x},n} = (n\pi)^2, \quad w_n = \sqrt{2}\cos n\pi\widehat{x}, \quad n \geq 1 \ . \end{aligned} \tag{5.8}$$

The set of these eigenfunctions is an orthonormal basis in $L_2(0;\ 1)$. The right-hand side function $\widehat{F} = \widehat{F}(\widehat{x}, \widehat{z})$ can then be expanded in this basis as follows:

$$\begin{aligned} \widehat{F}(\widehat{x}, \widehat{z}) \ &= \ \sum_{i=0}^{\infty} \widehat{f}_i(\widehat{z}) w_i(\widehat{x}) \ , \\ \text{with} \quad \widehat{f}_i(\widehat{z}) \ &= \ \int_0^1 \widehat{F}(\widehat{x}, \widehat{z}) w_i(\widehat{x}) d\widehat{x} \ , \\ F(x, z) \ &= \ \widehat{F}(\widehat{x}, \widehat{z}) \ . \end{aligned} \tag{5.9}$$

The assumptions imposed on the right-hand side in the benchmark problem formulation are as follows. We consider $\widehat{F} = F(\widehat{x}, \widehat{z})$ to be admissible if it belongs to the class of functions whose expansion (5.9) satisfies

$$\begin{aligned} \widehat{f}_i(\widehat{z})|_{G_s} \;\equiv\; \widehat{f}_{s,i} \;\equiv\; \mathrm{const}_{s,i}, \quad i = 0, 1, \\ \widehat{f}_i(\widehat{z}) \;\equiv\; 0, \quad i \geq 2, \end{aligned} \tag{5.10}$$

i.e. the functions $\widehat{f}_i = \widehat{f}_i(\widehat{z})$ are piece-wise constant with respect to the layers $\widehat{G}_s$, and $\widehat{F} = \widehat{F}(\widehat{x}, \widehat{z})$ has only two nonzero harmonics.

Recall that the function F is the restriction of the function $\widehat{F}$ to the subregions G_s, $s = \overline{1,\ L}$, under the respective change of coordinates, i.e.

$$\widehat{F}|_{G_s}(\widehat{x},\ \widehat{z}) \;=\; F|_{G_s}(x,\ z)$$

as stated in (5.5); therefore the assumptions imposed on the function $\widehat{F}$ extend to the function F in the original problem (5.11).

With the assumptions above, we consider the diffusion problem with homogeneous Neumann boundary conditions for the region $\widehat{G}$:

$$\begin{aligned} -\widehat{\nabla} \cdot \big(\widehat{K}\ \widehat{\nabla}\widehat{p}\big) \;+\; \widehat{c}\widehat{p} \;&=\; \widehat{F} \quad \text{in} \quad \widehat{G}\ , \\ -\big(\widehat{K}\ \widehat{\nabla}\widehat{p}\big) \cdot \widehat{\boldsymbol{n}} \;&=\; 0 \quad \text{on} \quad \partial\widehat{G}\ . \end{aligned} \tag{5.11}$$

With the imposed restrictions, we can explicitly find the reference solution $\widehat{p}^*$ for the problem (5.11) in the domain $\widehat{G}$. Then, we set the boundary conditions for the problem (5.1) to be

$$G_N \;=\; \widehat{\boldsymbol{u}}^*(\widehat{x},\ \widehat{z}) \cdot \widehat{\boldsymbol{n}} \quad \text{on} \quad \partial G, \tag{5.12}$$

where

$$\widehat{\boldsymbol{u}}^* \;=\; -\widehat{K}\ \widehat{\nabla}\widehat{p}^*$$

is the reference flux for the domain $\widehat{G}$. Consequently, the reference solution of the problem (5.1) on the domain G should coincide with the restriction of the reference solution of the problem (5.11) to the subdomain G, i.e.

$$\boldsymbol{u}^*(x,\ z)\ =\ \widehat{\boldsymbol{u}}^*(\widehat{x},\ \widehat{z})|_G \quad \text{and} \quad p^*(x,\ z)\ =\ \widehat{p}^*(\widehat{x},\ \widehat{z})|_G\ , \tag{5.13}$$

where p^* is the solution of the problem (5.1), and $\boldsymbol{u}^* = -K\nabla p^*$.

With the boundary conditions explicitly defined, the list of assumptions required to describe the benchmark problem is complete.

5.2 Non-conforming mesh description

In this section we describe the construction of non-conforming meshes for domains with thin parallel oblique geological layers.

Let the considered domain G be a rectangle partitioned into a union of oblique subdomains G_s, $s = \overline{1,\ L}$, which correspond to different geological layers. We assume that all the layers are parallel and have an inclination angle α with respect to the x-axis, i.e. the lines separating geological layers inside of G, $z_s(x)$, $s = \overline{1,\ L-1}$, have the slope equal to $\tan(\alpha)$, and we denote the elevation of the geological layer G_s on the left side of the domain G by $z_s(0) = z_{s,0}$.

We start with a conforming mesh G_h which is uniform in variable x, and is such that $G_{h,s} \equiv G_h|_{G_s}$ is uniform in variable z along any vertical line, $s = \overline{1,\ L}$. Mesh cells are quadrilaterals, in particular, mesh cells inside the oblique layers are parallelograms.

We say that the interface $\gamma_{k,i}$ is an interior vertical interface of the cell e_k in the oblique layer G_s, $s = \overline{2,\ L-1}$, provided it doesn't belong to the boundary of the domain, i.e. $\gamma_{k,i} \notin \partial G$, and it's not parallel to the lines separating geological layers inside of G, i.e. $\gamma_{k,i} \nparallel z_s(x)$. We say that the interface $\gamma_{k,i}$ is an interior horizontal interface of the cell e_k provided it doesn't belong to the boundary of the domain, and it's not an interior vertical interface. It follows that $\gamma_{k,i} \parallel z_s(x)$ in this case.

Then, for every mesh cell e_k inside every "thin" geological layer G_s, we replace all the interior vertical interfaces $\gamma_{k,i}$ by interfaces $\widehat{\gamma}_{k,i}$ such that $\widehat{\gamma}_{k,i} \perp z_s(x)$. We do this by shifting the end point of $\gamma_{k,i}$ with a greater z-coordinate along the line parallel to $z_s(x)$ and passing through that point. Then we adjust the length of the interior

horizontal interface at the top of e_k, $\widehat{\gamma}_{k,i}$, so that the cell becomes a rectangle, and denote such interface by $\widetilde{\gamma}_{k,i}$.

Note that if the interior vertical interface $\gamma_{k,i}$ was an interface between mesh cells e_k and e_l, i.e. $\gamma_{k,i} \equiv \gamma_{l,j} \equiv \gamma_{k,l}$, then we have that $\widehat{\gamma}_{k,i} \equiv \widehat{\gamma}_{l,j} \equiv \gamma_{k,l}$. For the interior horizontal interfaces we still write that $\widetilde{\gamma}_{k,i} \equiv \gamma_{l,j} \equiv \gamma_{k,l}$, but implying the equivalence in a logical sense, as geometrically interfaces $\widetilde{\gamma}_{k,i}$ and $\gamma_{l,j}$ no longer coincide.
Taking an example on Figure 5.3, it's the same as saying that even though $\gamma_{k,l} = \gamma_{l,k}$ for conforming cells, and $\gamma_{k,l} \neq \gamma_{l,k}$ for non-conforming cells, we impose the same interface condition

$$u_{k,i}|\gamma_{k,l}| \; + \; u_{l,j}|\gamma_{l,k}| \; = \; 0 \tag{5.14}$$

in both cases.

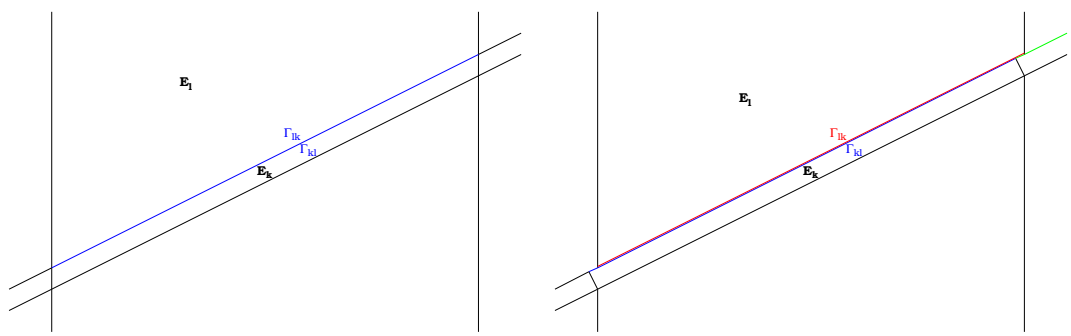

Figure 5.3: An example of conforming and non-conforming cells

Another illustration of the transition to a non-conforming mesh is shown on Figures 5.4 and 5.5.

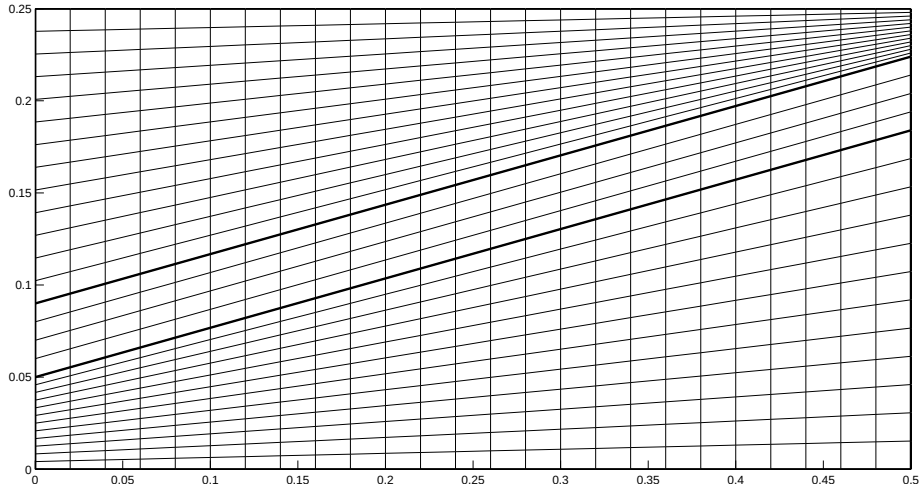

Figure 5.4: An example of the initial conforming mesh

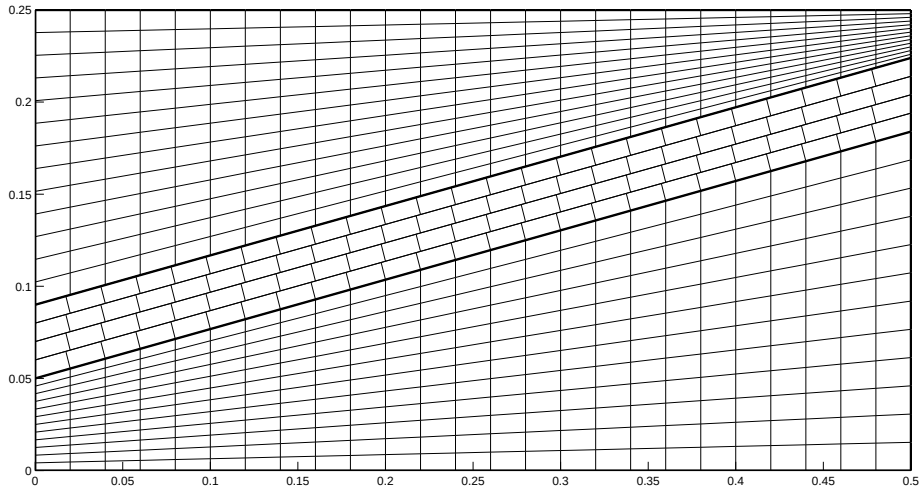

Figure 5.5: An example of the resulting non-conforming mesh

5.3 Error analysis of mixed FE method

In this section, we describe how the benchmark problem for domains with thin parallel oblique layers can be used to compare the accuracy of the mixed finite element methods on conforming and non-conforming meshes.

We consider the benchmark problem (5.11) in the rectangular domain G with oblique layers having inclination angle α. We denote the corresponding reference domain by $\widehat{G}$ and impose all the restrictions listed in the Section 5.1. The reference solution pair is then $(p^*,\ \boldsymbol{u}^*)$. We use a quadrilateral mesh G_h in domain G, and denote mesh cells by e_k, $k = \overline{1,\ n}$, where n is the total number of cells. Each cell e_k is a quadrilateral divided into two triangles: upper triangle $T_k^{(1)}$ and lower triangle $T_k^{(2)}$.

Then, the KR-interpolant of $\boldsymbol{u}^*$ in a cell e_k can be written as

$$\boldsymbol{w}_k^{KR}(\boldsymbol{x}) \;=\; \begin{cases} u_{k,1}\boldsymbol{\phi}_{k,1}^{(1)}(\boldsymbol{x}) \;+\; u_{k,2}\boldsymbol{\phi}_{k,2}^{(1)}(\boldsymbol{x}) \;-\; u_{k,3}\boldsymbol{\phi}_{k,3}^{(1)}(\boldsymbol{x}) & \text{in } T_k^{(1)} \\ u_{k,3}\boldsymbol{\phi}_{k,3}^{(2)}(\boldsymbol{x}) \;+\; u_{k,4}\boldsymbol{\phi}_{k,4}^{(2)}(\boldsymbol{x}) \;+\; u_{k,5}\boldsymbol{\phi}_{k,5}^{(2)}(\boldsymbol{x}) & \text{in } T_k^{(2)} \end{cases}, \tag{5.15}$$

where $\boldsymbol{\phi}_{k,i}^{(j)}(\boldsymbol{x})$ is the RT_0 basis function for the side $\gamma_{k,i}$ in the triangle $T_k^{(j)}$, and $u_{k,i}$ is the average value of normal component of flux on the side $\gamma_{k,i}$.

Similarly, the PWC-interpolant for a cell e_k is as follows:

$$\boldsymbol{w}_k^{PWC}(\boldsymbol{x}) \;=\; \begin{cases} \begin{pmatrix} \boldsymbol{n}_{k,1} & \boldsymbol{n}_{k,2} \end{pmatrix}^{-T} \begin{pmatrix} u_{k,1} \\ u_{k,2} \end{pmatrix} & \text{in } T_k^{(1)} \\ \begin{pmatrix} \boldsymbol{n}_{k,4} & \boldsymbol{n}_{k,5} \end{pmatrix}^{-T} \begin{pmatrix} u_{k,4} \\ u_{k,5} \end{pmatrix} & \text{in } T_k^{(2)} \end{cases}, \tag{5.16}$$

where $\boldsymbol{n}_{k,i}$ is the unit normal vector to the side $\gamma_{k,i}$.

The reference solution $\boldsymbol{u}^*$ is known in the entire domain $\widehat{G}$ and, therefore, its entire subdomain G, so for every mesh cell e_k we explicitly know $\boldsymbol{u}_k{}^* \equiv \boldsymbol{u}^*|_{e_k}$, the reference solution for the cell.

Now, we can discretize the equation (5.11) by applying either KR mixed FE method or PWC approximation, and obtain the solution pair $(p_{h,k},\ \boldsymbol{w}_{h,k})$ using the corresponding interpolant.

An absolute error Δ_{e_k} over a cell e_k can be computed as

$$\Delta_{e_k} \ = \ \left(\int\limits_{e_k} |\boldsymbol{w}_{h,k}(\boldsymbol{x}) \ - \ \boldsymbol{u}^*(\boldsymbol{x})|^2 \ d\boldsymbol{x} \right)^{1/2} \quad , \tag{5.17}$$

and the L_2 norm of the reference solution $\boldsymbol{u}^*$ over the same cell e_k is

$$\|\boldsymbol{u}_k^*\|_2 \ = \ \left(\int\limits_{e_k} |\boldsymbol{u}^*(\boldsymbol{x})|^2 \ d\boldsymbol{x} \right)^{1/2} \quad . \tag{5.18}$$

We can denote the mesh inside the geological layer G_s by $G_{h,s}$. Then the relative error in L_2 norm between the flux interpolant $\boldsymbol{w}_{h,k}$ and the reference solution $\boldsymbol{u}_k^*$ over the geological layer G_s, $s = \overline{1,\ L}$, can be computed as

$$\epsilon_{G_s} \ = \ 100 \cdot \frac{\sum\limits_{e_k \in G_{h,s}} \Delta_{e_k}}{\sum\limits_{e_k \in G_{h,s}} \|\boldsymbol{u}_k^*\|_2} \ . \tag{5.19}$$

5.4 Numerical results

In this section, we consider the benchmark problem for a number of domains with different inclination angles of the oblique geological layer, and estimate the accuracy of PWC and KR approximations as described in Section 5.3. We use both conforming and non-conforming meshes introduced in Section 5.2.

We choose our domain G so that it contains three geological layers with the one in the middle being "thin". The lines separating geological layers inside of G, $z_1(x)$ and $z_2(x)$, have slope equal to $\tan(\alpha)$, where α is the inclination angle of the geological layers with respect to the x-axis, and $z_1(0) = 0.05$, $z_2(0) = 0.05001$. We construct the reference domain $\widehat{G}$ and use the values of the parameters listed in Table 5.1.

Table 5.1: Parameters for the chosen test problem

	$\widehat{G}_1$	$\widehat{G}_2$	$\widehat{G}_3$
$k_{x,s}$	2	100	5
$k_{z,s}$	1	10	4
c_s	1	1	1
$\widehat{f}_0$	5	1000	1
$\widehat{f}_1$	0.1	10	0.05

The mesh G_h is chosen to be uniform in variable x, and is such that $G_{h,s} \equiv G_h|_{G_s}$ is uniform in variable z along any vertical line, $s = \overline{1,\ 3}$. For different values of the inclination angle α and mesh step size, we compute ϵ_{G_s}, $s = \overline{1,\ 3}$, i.e. relative errors in interpolants $\boldsymbol{w}_{h,k}$ over every geological layer G_s.

Domains $\widehat{G}$ and G, as well as the mesh G_h for the intermediate mesh step size, are shown on figures below. The computed values of relative errors are given in Tables 5.2 - 5.9.

Domain 1: G is a $(0,\ 0.4) \times (0,\ 0.5)$ rectangle with an oblique layer having the inclination angle of $45°$. The coarsest mesh has 10×12 cells in $G_{h,1}$, 10×1 cells in $G_{h,2}$, and 10×13 cells in $G_{h,3}$.

Geological layers in the reference domain are $\widehat{G}_1 = (0,\ 1) \times (0,\ 0.320156)$, $\widehat{G}_2 = (0,\ 1) \times (0.320156,\ 0.320163)$, and $\widehat{G}_3 = (0,\ 1) \times (0.320163,\ 0.640312)$.

The finest mesh steps $\boldsymbol{h}_s = (h_{s,x},\ h_{s,z})$ for the layers $G_s,\ s = \overline{1,\ 3}$, are as follows: $\boldsymbol{h}_1 = (0.01, 0.00104167)$, $\boldsymbol{h}_2 = (0.01, 2.5e-006)$, and $\boldsymbol{h}_3 = (0.01, 0.00865365)$

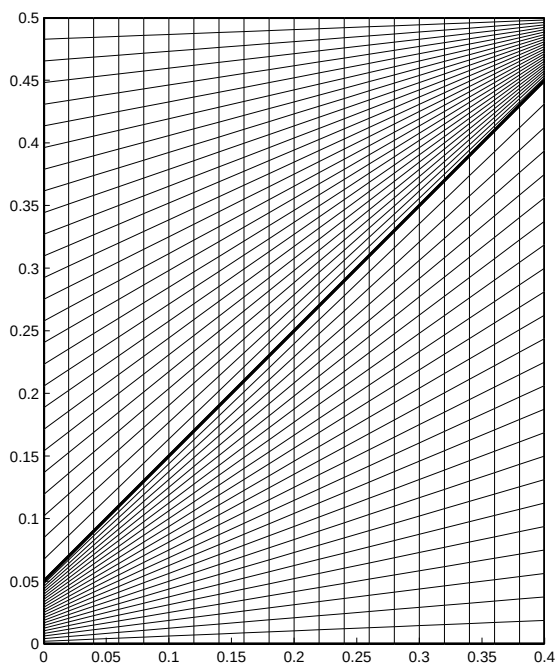

Figure 5.6: Domain G and mesh G_h for angle $\alpha = 45°$ and mesh step size $2\boldsymbol{h}$

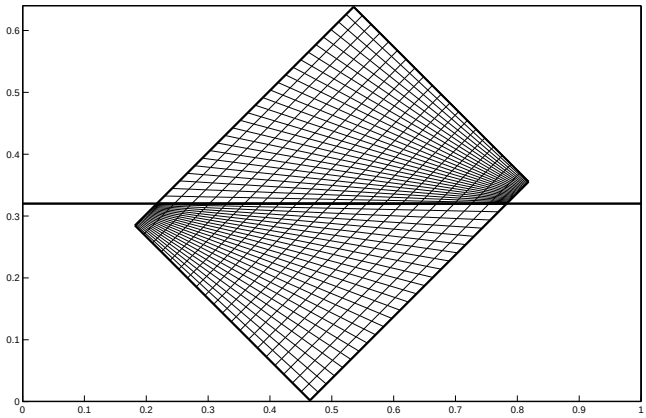

Figure 5.7: Mesh G_h inside the domain $\widehat{G}$ for angle $\alpha = 45°$ and mesh step size $2\boldsymbol{h}$

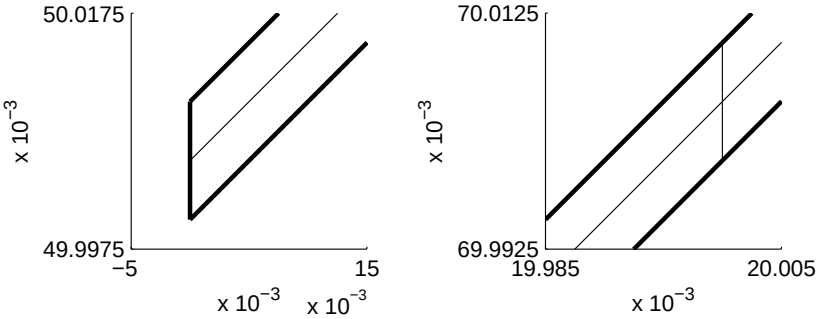

Figure 5.8: Mesh cell inside $G_{h,2}$ for angle $\alpha = 45°$ and mesh step size $2\boldsymbol{h}$, conforming mesh

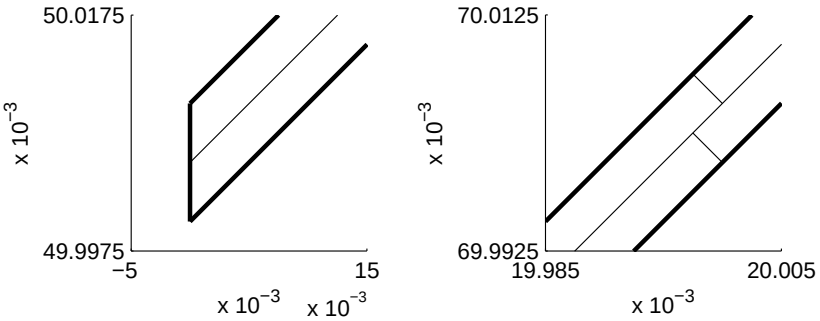

Figure 5.9: Mesh cell inside $G_{h,2}$ for angle $\alpha = 45°$ and mesh step size $2\boldsymbol{h}$, non-conforming mesh

Table 5.2: Relative error in KR interpolant $\boldsymbol{w}_h^{KR}$, %, for angle $\alpha = 45°$, conforming mesh

	G_1	G_2	G_3
$4\boldsymbol{h}$	3.1838	1148.49	3.13859
$2\boldsymbol{h}$	1.5906	574.246	1.56836
$\boldsymbol{h}$	0.795139	287.123	0.784062

Table 5.3: Relative error in KR interpolant $\boldsymbol{w}_h^{KR}$, %, for angle $\alpha = 45°$, non-conforming mesh

	G_1	G_2	G_3
$4\boldsymbol{h}$	3.18381	1148.78	3.13857
$2\boldsymbol{h}$	1.59061	574.398	1.56836
$\boldsymbol{h}$	0.795148	287.201	0.784083

Table 5.4: Relative error in PWC interpolant $\boldsymbol{w}_h^{PWC}$, %, for angle $\alpha = 45°$, conforming mesh

	G_1	G_2	G_3
$4\boldsymbol{h}$	6.47516	5.9941	5.98602
$2\boldsymbol{h}$	3.23569	2.02823	2.98601
$\boldsymbol{h}$	1.61761	0.842736	1.49213

Table 5.5: Relative error in PWC interpolant $\boldsymbol{w}_h^{PWC}$, %, for angle $\alpha = 45°$, non-conforming mesh

	G_1	G_2	G_3
$4\boldsymbol{h}$	6.47512	5.9229	5.98567
$2\boldsymbol{h}$	3.23567	1.96789	2.98584
$\boldsymbol{h}$	1.6176	0.812716	1.49205

Domain 2: G is a $(0,\ 0.025) \times (0,\ 0.5)$ rectangle with an oblique layer having the inclination angle of $85°$. The coarsest mesh has 5×12 cells in $G_{h,1}$, 5×1 cells in $G_{h,2}$, and 5×13 cells in $G_{h,3}$.

Geological layers in the reference domain are $\widehat{G}_1 = (0,\ 1) \times (0,\ 0.232881)$, $\widehat{G}_2 = (0,\ 1) \times (0.232881,\ 0.232882)$, and $\widehat{G}_3 = (0,\ 1) \times (0.232882,\ 0.500625)$.

The finest mesh steps $\boldsymbol{h}_s = (h_{s,x},\ h_{s,z})$ for the layers G_s, $s = \overline{1,\ 3}$, are as follows: $\boldsymbol{h}_1 = (0.00125, 0.00104167)$, $\boldsymbol{h}_2 = (0.00125, 2.5e-006)$, and $\boldsymbol{h}_3 = (0.00125, 0.00865365)$

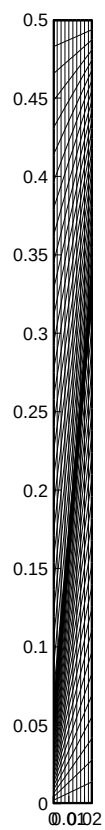

Figure 5.10: Domain G and mesh G_h for angle $\alpha = 85°$ and mesh step size $2\boldsymbol{h}$

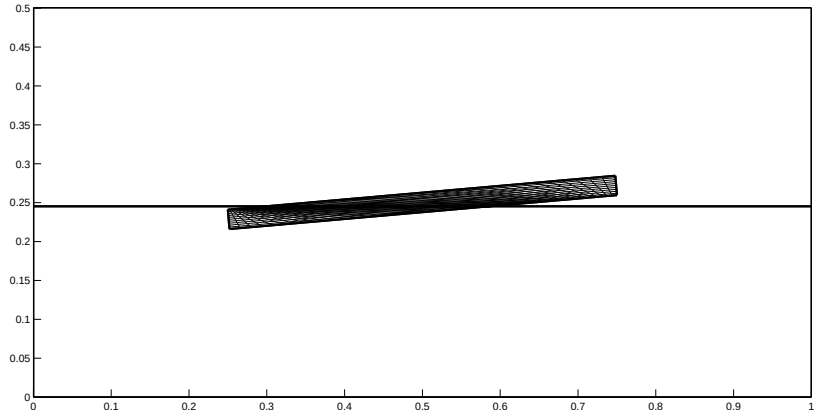

Figure 5.11: Mesh G_h inside the domain $\widehat{G}$ for angle $\alpha = 85°$ and mesh step size $2\boldsymbol{h}$

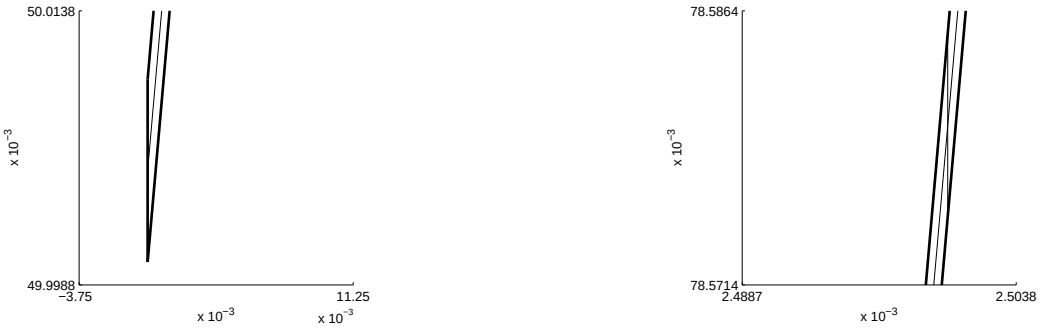

Figure 5.12: Mesh cell inside $G_{h,2}$ for angle $\alpha = 85°$ and mesh step size $2\boldsymbol{h}$, conforming mesh case

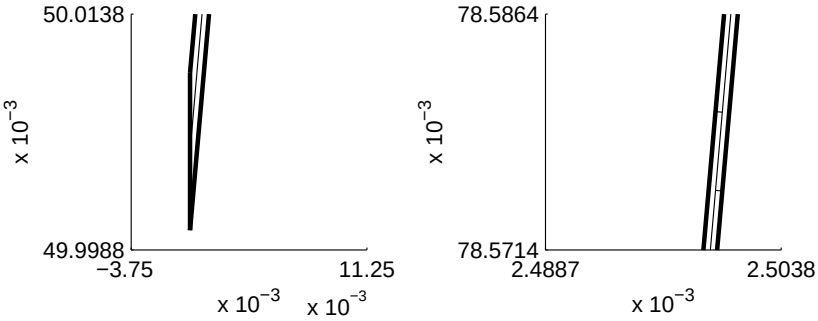

Figure 5.13: Mesh cell inside $G_{h,2}$ for angle $\alpha = 85°$ and mesh step size $2\boldsymbol{h}$, non-conforming mesh case

Table 5.6: Relative error in KR interpolant $\boldsymbol{w}_h^{KR}$, %, for angle $\alpha = 85°$, conforming mesh case

	G_1	G_2	G_3
$4\boldsymbol{h}$	2.29005	1206.12	2.00625
$2\boldsymbol{h}$	1.14379	603.059	1.0056
$\boldsymbol{h}$	0.571822	301.531	0.503148

Table 5.7: Relative error in KR interpolant $\boldsymbol{w}_h^{KR}$, %, for angle $\alpha = 85°$, non-conforming mesh case

	G_1	G_2	G_3
$4\boldsymbol{h}$	2.28981	1207.36	2.00559
$2\boldsymbol{h}$	1.14378	603.747	1.00588
$\boldsymbol{h}$	0.572307	301.889	0.504837

Table 5.8: Relative error in PWC interpolant $\boldsymbol{w}_h^{PWC}$, %, for angle $\alpha = 85°$, conforming mesh case

	G_1	G_2	G_3
$4\boldsymbol{h}$	5.34833	9.79232	4.50218
$2\boldsymbol{h}$	2.66169	3.04768	2.23093
$\boldsymbol{h}$	1.32938	1.07851	1.11288

Table 5.9: Relative error in PWC interpolant $\boldsymbol{w}_h^{PWC}$, %, for angle $\alpha = 85°$, non-conforming mesh case

	G_1	G_2	G_3
$4\boldsymbol{h}$	5.34711	8.81404	4.49751
$2\boldsymbol{h}$	2.661	2.07108	2.22867
$\boldsymbol{h}$	1.32921	0.907528	1.11241

Chapter 6

Non-conforming meshes on domains with concentric curved geological layers

6.1 Benchmark problem formulation

In this section, we consider the Neumann boundary value problem for the diffusion equation

$$\begin{aligned} -\nabla \cdot \big(K\ \nabla p\big)\ +\ cp\ &=\ F \quad \text{in} \quad G \\ (-K\ \nabla p) \cdot \boldsymbol{n}\ &=\ G_N \quad \text{on} \quad \partial G \end{aligned} \tag{6.1}$$

in a rectangular domain G. Here, $K = K(x, z)$ is a two-by-two symmetric positive definite matrix, $c = c(x, z)$ is a nonnegative function, ∂G is the boundary of G, $\boldsymbol{n}$ is the unit outward normal to ∂G, $F = F(x, z)$ and $G_N = G_N(x, z)$ are given functions.

We assume that the domain G is a partition of layers G_s, $s = \overline{1,\ L}$, with the middle layers being curved. The boundaries of all the middle layers are concentric circular

arcs, i.e. parts of the coaxial circles sharing the same center $C(x_C,\ y_C)$. In order to express the solution function $p = p(x, z)$ of (6.1) analytically, we impose a number of restrictions specified below.

First, we associate a polar coordinate system $(r,\ \theta)$ with the center of the circles, i.e. we choose the point C to be its origin. We assume the diffusion tensor K to be piece-wise constant in G with respect to the polar coordinates, i.e.

$$K_{(r,\ \theta)}|_{G_s} \equiv K_{(r,\ \theta),s} \equiv \text{const}_s \in \mathbb{R}^{2\times 2}.$$

We also assume that

$$K_{(r,\ \theta),s} \ = \ \begin{pmatrix} k_{r,s} & 0 \\ 0 & k_{\theta,s} \end{pmatrix} \quad \text{in} \quad G_s, \quad s = \overline{1,\ L}. \tag{6.2}$$

is a constant diagonal tensor.

Second, we impose similar restrictions on c, that is

$$c_{(r,\ \theta)}|_{G_s} \equiv c_{(r,\ \theta),s} \equiv \text{const}_s > 0.$$

Third, we assume that the region G is a part of a larger domain Ω which is bounded by two coaxial circles centered at the point C. With Ω, we associate the polar system $(r,\ \theta)$ defined above, and therefore can describe Ω by writing $\Omega = (r_{min}, r_{max}) \times (0,\ 2\pi)$ in $(r,\ \theta)$ coordinate system. We assume that subregions G_s, $s = \overline{1,\ L}$, can be extended to concentric layers Ω_s in Ω.

Now, we introduce functions $\widehat{K}$, $\widehat{c}$, and $\widehat{F}$ defined on the domain Ω so that

$$\begin{aligned} \widehat{K}|_{G_s} \ &= \ K_{(r,\ \theta),s} \ = \ \begin{pmatrix} k_{r,s} & 0 \\ 0 & k_{\theta,s} \end{pmatrix}, \quad \widehat{c}|_{G_s} \ = \ c_{(r,\ \theta),s}, \\ \text{and} \quad \widehat{F}|_{G_s} \ &= \ F_{(r,\ \theta),s}|_{G_s}, \end{aligned} \tag{6.3}$$

i.e. these functions are extensions of the functions K, c, and F from the domain G to the domain Ω with the values of $\widehat{K}$, $\widehat{c}$, and $\widehat{F}$ coinciding with the corresponding values of K, c, and F in subregions G_s, $s = \overline{1,\ L}$, under the respective change of coordinates.

We can formally describe the partitioning of Ω in the following way:

$$\Omega \ \equiv \ \bigcup_{s=0}^{L-1} (r_s,\ r_{s+1}) \times (0,\ 2\pi) \ , \tag{6.4}$$

where $r_0 = r_{min}$, $r_L = r_{max}$, and r_s is a radius of a circle separating layers Ω_s and Ω_{s+1}, $s = \overline{1,\ L-1}$.

Now we consider the eigenproblem for the operator $-\dfrac{d^2}{d\theta^2}$:

$$\begin{aligned} -\frac{d^2}{d\theta^2} w \ &= \ \lambda_\theta w, \quad 0 < \theta < 2\pi, \\ w(0) \ &= \ w(2\pi), \quad \frac{dw}{d\theta}(0) \ = \ \frac{dw}{d\theta}(2\pi). \end{aligned} \tag{6.5}$$

The eigenpairs $(\lambda_{\theta,n},\ w_n)$ for this problem are

$$\begin{aligned} &\lambda_{\theta,0} = 0, \quad w_0 \equiv 1, \\ &\lambda_{\theta,n} = n^2, \quad w_n = \frac{1}{\sqrt{\pi}} \cos n\theta, \quad n \geq 1 \ . \end{aligned} \tag{6.6}$$

The set of these eigenfunctions is an orthonormal basis in $L_2(0;\ 2\pi)$. The right-hand side function $\widehat{F} = \widehat{F}(r,\theta)$ can then be expanded in this basis as follows:

$$\begin{aligned} \widehat{F}(r,\theta) \ &= \ \sum_{i=0}^{\infty} \widehat{f}_i(r) w_i(\theta) \ , \\ \text{with} \quad \widehat{f}_i(r) \ &= \ \int_0^{2\pi} \widehat{F}(r,\theta) w_i(\theta) d\theta \ , \\ F_{(r,\ \theta)} \ &= \ \widehat{F}(r,\theta) \ . \end{aligned} \tag{6.7}$$

The assumptions imposed on the right-hand side in the benchmark problem formulation are as follows. We consider $\widehat{F} = \widehat{F}(r,\theta)$ to be admissible if it belongs to the class of functions whose expansion (6.7) satisfies:

$$\begin{aligned} \widehat{f}_i(r)|_{\Omega_s} \;&\equiv\; \widehat{f}_{i,s} \;\equiv\; \mathrm{const}_{i,s}, \quad i = 0,1, \\ \widehat{f}_i(r) \;&\equiv\; 0, \quad i \geq 2, \end{aligned} \tag{6.8}$$

i.e. the functions $\widehat{F}_i = \widehat{f}_i(r)$ are piece-wise constant with respect to the layers Ω_s, and $\widehat{F} = \widehat{F}(r,\theta)$ has only two nonzero harmonics.

Recall that the function F is the restriction of the function $\widehat{F}$ to the subregions G_s, $s = \overline{1,\ L}$, under the respective change of coordinates, i.e.

$$\widehat{F}|_{G_s} \;=\; F_{(r,\ \theta)}$$

as stated in (6.3), therefore the assumptions imposed on the function $\widehat{F}$ extend to the function F in the original problem (6.1).

With all the assumptions above, we consider the diffusion problem with homogeneous Neumann boundary conditions for the region Ω:

$$\begin{aligned} -\widehat{\nabla}\cdot\big(\widehat{K}\ \nabla\widehat{p}\big) \;+\; \widehat{c}\widehat{p} \;&=\; \widehat{F} \quad \text{in} \quad \Omega\ , \\ -\big(\widehat{K}\ \widehat{\nabla}\widehat{p}\big)\cdot\widehat{\boldsymbol{n}} \;&=\; 0 \quad \text{on} \quad \partial\Omega\ . \end{aligned} \tag{6.9}$$

With the imposed restrictions, we can explicitly find the reference solution $\widehat{p}^*$ for the problem (6.9) in the domain Ω. Then, we set the boundary conditions for the problem (6.1) to be

$$G_N \;=\; \widehat{\boldsymbol{u}}^*_{(x,\ z)}\cdot\boldsymbol{n} \quad \text{on} \quad \partial G, \tag{6.10}$$

where

$$\widehat{\boldsymbol{u}}^*_{(r,\ \theta)} \;=\; -\widehat{K}\ \widehat{\nabla}\widehat{p}^*$$

is the reference flux for the domain Ω. Consequently, the reference solution of the problem (6.1) on the domain G should coincide with the restriction of the reference solution of the problem (6.9) to the subdomain G, i.e.

$$\boldsymbol{u}^* \;=\; \widehat{\boldsymbol{u}}^*_{(x,\ z)}|_G \quad \text{and} \quad p^* \;=\; \widehat{p}^*_{(x,\ z)}|_G \;, \tag{6.11}$$

where p^* is the solution of the problem (6.1), and $\boldsymbol{u}^* = -K\nabla p^*$.

With the boundary conditions explicitly defined, the list of assumptions required to describe the benchmark problem is complete.

6.2 Obtaining the solution of the benchmark problem

In order to obtain the solution of the benchmark problem described in Section 6.1, we expand the solution function $p = p(r,\ \theta)$ of the problem (6.9) with respect to the eigenfunctions (6.6):

$$p(r,\ \theta)\ =\ \sum_{i=0}^{\infty} p_i(r) w_i(\theta)\ . \tag{6.12}$$

Then, the benchmark problem

$$\begin{gathered} -\frac{1}{r}\frac{\partial}{\partial r}\left(K_r r\ \frac{\partial p}{\partial r}\right)\ -\ \frac{1}{r^2}\frac{\partial}{\partial \theta}\left(K_\theta \frac{\partial p}{\partial \theta}\right)\ +\ \widehat{c}p\ =\ \widehat{F} \quad \text{in} \quad \Omega\ , \\ \frac{\partial p}{\partial r}\ =\ 0 \quad \text{on} \quad \partial\Omega\ , \end{gathered} \tag{6.13}$$

can be written as

$$\begin{gathered} -\frac{d}{dr}\left(K_r \sum_{i=0}^{\infty} \frac{dp_i}{dr}(r) w_i(\theta)\right)\ -\ \frac{1}{r} K_r \sum_{i=0}^{\infty} \frac{dp_i}{dr}(r) w_i(\theta)\ - \\ -\ \frac{1}{r^2}\frac{d}{d\theta}\left(K_\theta \sum_{i=0}^{\infty} p_i(r) \frac{dw_i}{d\theta}(\theta)\right)\ +\ \widehat{c}\sum_{i=0}^{\infty} p_i(r) w_i(\theta)\ =\ \sum_{i=0}^{1} f_i w_i(\theta) \quad \text{in } \Omega\ , \\ \frac{dp_i}{dr}(r)\ =\ 0 \quad \text{on } \partial\Omega \quad \text{for } i \geq 0\ . \end{gathered} \tag{6.14}$$

Recall that $\widehat{K}|_{\Omega_s} \equiv \widehat{K}_s \equiv \text{const}$, where $\Omega\ \equiv\ \bigcup_{s=0}^{L-1}(r_s,\ r_{s+1}) \times (0,\ 2\pi)$ as before. Since the basis $\{w_i\}_{i=0}^{\infty}$ is orthonormal, these two facts imply that

$$p_i(r)\ \equiv\ 0, \qquad i \geq 2\ , \tag{6.15}$$

and the resulting system is as follows:

$$
\begin{aligned}
-k_{r,s}\frac{d^2 p_{i,s}}{dr^2}(r) \;-\; \frac{k_{r,s}}{r}\frac{dp_{i,s}}{dr}(r) \;+\; \sigma_{i,s}(r)p_{i,s}(r) \;&=\; f_{i,s}\ , \\
\frac{dp_{i,1}}{dr}(r_1) \;&=\; 0\ , \\
\frac{dp_{i,L}}{dr}(r_L) \;&=\; 0\ , \qquad (6.16)\\
p_{i,s}(r_s-0) \;=\; p_{i,s+1}(r_s+0), \quad s=\overline{1,\ L-1}\ , \\
K_{r,s}\frac{dp_{i,s}}{dr}(r_s-0) \;=\; K_{r,s+1}\frac{dp_{i,s+1}}{dr}(r_s+0), \quad s=\overline{1,\ L-1}\ ,
\end{aligned}
$$

$s=\overline{1,\ L}$, $i=0,1$. Here,

$$
\sigma_{1,s}(r) \;=\; \widehat{c}_s, \quad \sigma_{2,s}(r) \;=\; \widehat{c}_s + \frac{k_{\theta,s}}{r^2}\ . \qquad (6.17)
$$

The solution to this system, i.e. function $p_{i,s}(r)$, $s=\overline{1,\ L}$, is obtained by applying a finite-difference numerical scheme with sufficiently small step. With that, the reference solution of our benchmark problem can be written as

$$
p_s(r,\ \theta) \;=\; p_{0,s}(r) \;+\; \frac{1}{\sqrt{\pi}}\cos(\theta)\cdot p_{1,s}(r), \quad s=\overline{1,L}. \qquad (6.18)
$$

6.3 Non-conforming mesh description

In this section we describe the construction of non-conforming meshes for domains with concentric thin geological layers.

Let the considered domain G be a rectangle partitioned into a union of subdomains G_s, $s = \overline{1,\ L}$, which correspond to different geological layers. We assume that all the middle layers have coaxial circular boundaries centered at the point $C(x_C,\ y_C)$, and we denote the radiuses of the circular arcs separating the geological layers inside of G by r_s, $s = \overline{1,\ L-1}$. We associate a Cartesian coordinate system $(x,\ y)$ with the lower left corner of the rectangle G, axes directed along its lower and left sides, and introduce a polar coordinate system $(r,\ \theta)$ with the origin at the point $C(x_C,\ y_C)$.

We start with a conforming mesh G_h which is uniform in variable x. Also, the mesh inside of each geological layer, $G_{h,s} \equiv G_h|_{G_s}$, is chosen so that along every vertical line $x = x_j$, it is uniform in variable z with respect to the coordinates of the mesh nodes in the Cartesian system for $s = 1,\ L$, and is uniform in variable r with respect to the coordinates of mesh nodes in the polar system for $s = \overline{2,\ L-1}$. Mesh cells are quadrilaterals, more specifically, they are trapezoids with parallel vertical interfaces.

We say that the interface $\gamma_{k,i}$ is an interior vertical interface of the cell e_k in the middle layer G_s, $s = \overline{2,\ L-1}$, provided it doesn't belong to the boundary of the domain, i.e. $\gamma_{k,i} \notin \partial G$, and the r-coordinates of its endpoints in the polar system are not the same. We say that the interface $\gamma_{k,i}$ is an interior horizontal interface of the cell e_k provided it doesn't belong to the boundary of the domain, and it's not an interior vertical interface. It follows that the r-coordinates of the endpoints of such interface with respect to the polar system are the same.

Then, for every mesh cell e_k inside every "thin" geological layer G_s, we replace all the interior vertical interfaces $\gamma_{k,i}$ by interfaces $\widehat{\gamma}_{k,i}$ such that the θ-coordinates of the endpoints of $\widehat{\gamma}_{k,i}$ are the same with respect to the polar system $(r,\ \theta)$.
To do this, we take the middle point of the interface $\gamma_{k,i}$. Say, the endpoints of $\gamma_{k,i}$ are $\boldsymbol{V}_{k,i}^{(1)}$ and $\boldsymbol{V}_{k,i}^{(2)}$, then the middle point $\boldsymbol{V}_{k,i}^{(M)}$ is just $\frac{1}{2}\left(\boldsymbol{V}_{k,i}^{(1)}+\boldsymbol{V}_{k,i}^{(2)}\right)$. The new interface $\widehat{\gamma}_{k,i}$ is constructed so that it lies on the line connecting the point $\boldsymbol{V}_{k,i}^{(M)}$ and the origin of the polar coordinate system C, i.e. it is orthogonal to the tangent line to the circle centered at C and passing through the point $\boldsymbol{V}_{k,i}^{(M)}$. The endpoints $\widehat{\mathbf{V}}_{k,i}^{(1)}$ and $\widehat{\mathbf{V}}_{k,i}^{(2)}$ have the same r-coordinate as the endpoints $\boldsymbol{V}_{k,i}^{(1)}$ and $\boldsymbol{V}_{k,i}^{(2)}$, respectively, and the same θ-coordinate as the middle point $\boldsymbol{V}_{k,i}^{(M)}$, i.e. $\widehat{v}_{k,i,r}^{(1)}=v_{k,i,r}^{(1)}$, $\widehat{v}_{k,i,r}^{(2)}=v_{k,i,r}^{(2)}$, and $\widehat{v}_{k,i,\theta}^{(1)}=\widehat{v}_{k,i,\theta}^{(2)}=v_{k,i,\theta}^{(M)}$. Then we adjust the interior horizontal interfaces of e_k to account for the shift of the cell's vertices. We denote these new interfaces by $\widetilde{\gamma}_{k,i}$.

Note that if the interior vertical interface $\gamma_{k,i}$ was an interface between mesh cells e_k and e_l, i.e. $\gamma_{k,i}\equiv\gamma_{l,j}\equiv\gamma_{k,l}$, then we have that $\widehat{\gamma}_{k,i}\equiv\widehat{\gamma}_{l,j}\equiv\gamma_{k,l}$. For the interior horizontal interfaces we still write that $\widetilde{\gamma}_{k,i}\equiv\gamma_{l,j}\equiv\gamma_{k,l}$, but implying the equivalence in a logical sense, as geometrically interfaces $\widetilde{\gamma}_{k,i}$ and $\gamma_{l,j}$ no longer coincide.

An example of a conforming mesh is shown on Figure 6.1. A non-conforming mesh constructed in accordance with the described procedure is shown on Figure 6.2.

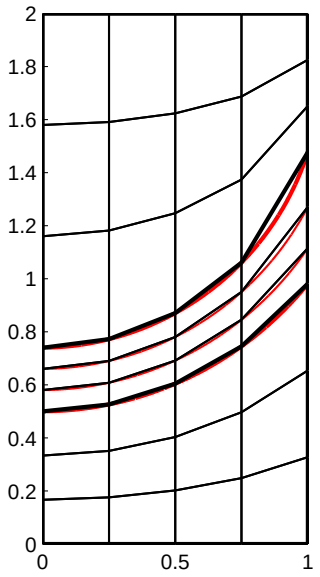

Figure 6.1: An example of a conforming mesh

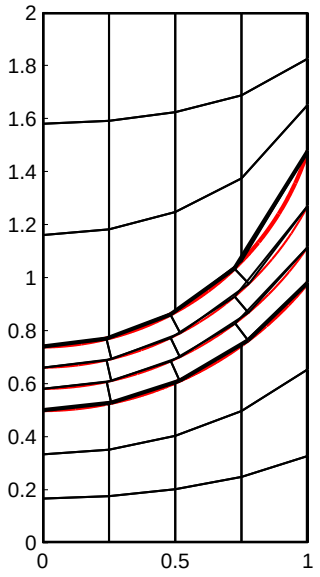

Figure 6.2: An example of the corresponding non-conforming mesh

6.4 Error analysis of mixed FE method

In this section, we describe how the benchmark problem can be used to estimate the accuracy of the discrete solution in the case of domains with concentric curved geological layers.

We consider the benchmark problem (6.9) in the rectangular domain G with curved middle layers. The boundaries of all the middle layers are concentric circular arcs, i.e. parts of the coaxial circles sharing the same center $C(x_C,\ y_C)$. We denote the corresponding reference domain by Ω and impose all the restrictions listed in the Section 6.1. The reference solution pair is then $(p^*,\ \boldsymbol{u}^*)$. We use a quadrilateral mesh G_h in the domain G, either conforming or non-conforming, as described in Section 6.3, and denote mesh cells by e_k, $k = \overline{1,\ n}$, where n is the total number of cells. The interfaces between the cells, as well as the boundary edges, are denoted by γ_j, $j = \overline{1,\ m}$, where m is the total number of interfaces and boundary edges. Γ_s is then a set of all the interfaces and boundary edges in the mesh layer $G_{h,s}$, i.e. such γ_j that $\gamma_j \in e_k$, where $e_k \in G_h$.

The reference solution $\boldsymbol{u}^*$ is known in the entire domain $\widehat{G}$ and, therefore, its entire subdomain G, so we explicitly know $\boldsymbol{u}_j{}^* \equiv \boldsymbol{u}^*|_{\gamma_j}$, the reference solution along every mesh edge. Then, the normal component of the reference solution in the center of the face γ_j is denoted by $u^*_{c,j}$. For all the vertical faces, the boundary edges, and the horizontal interfaces between the mesh cells in the top and bottom mesh layers $G_{h,1}$ and $G_{h,L}$, the center of the interface is taken to be the middle point of the linear segment connecting its two vertices. In the case of the horizontal interfaces where at least one of the cells it belongs to is in one of the middle layers $G_{h,s}$, $s = \overline{2,\ L-1}$, we say that the center of the interface is the center of the arc connecting its two

vertices, i.e. if $\boldsymbol{V}_j^{(1)}$ and $\boldsymbol{V}_j^{(2)}$ are the vertices of the interface γ_j with corresponding polar coordinates $(v_{j,r}^{(i)},\ v_{j,\theta}^{(i)})$, $i = \overline{1,\ 2}$, then $\boldsymbol{V}_j^{(c)}$, the center of the interface γ_j, has coordinates $(v_{j,r}^{(1)},\ \frac{1}{2}(v_{j,\theta}^{(1)} + v_{j,\theta}^{(2)}))$.

Now, we can discretize the equation (6.9) by applying either KR mixed FE method or PWC approximation, and obtain the interpolants p_h and $\boldsymbol{w}_h$ of the solution functions p^* and $\boldsymbol{u}^*$. For every interface γ_j we compute the reference flux and its interpolant in the center of the face as described above, and project it onto the normal vector to the corresponding interface $\widehat{\gamma}_j$ belonging to the non-conforming mesh. We denote the resulting values of normal components by $u_{c,j}^*$ and $w_{h,j}$.

Then, for every geological layers G_s, $s = \overline{1,\ L}$, we compute the following relative error:

$$\epsilon_{G_s} \;=\; 100 \cdot \left(\frac{\displaystyle\sum_{\gamma_j \in \Gamma_s} \left(|\gamma_j| \left(w_{h,j} - u_{c,j}^* \right) \right)^2}{\displaystyle\sum_{\gamma_j \in \Gamma_s} \left(|\gamma_j| u_{c,j}^* \right)^2} \right)^{\frac{1}{2}} . \tag{6.19}$$

6.5 Numerical results

In this section, we consider the benchmark problem for a number of domains with curved geological layers, and estimate the accuracy of KR mixed FE method and PWC approximation as described in Section 6.4. We use both conforming and non-conforming meshes introduced in Section 6.3.

Values of parameters used in experiments are listed in Table 6.1.

Table 6.1: Parameters for the chosen test problem

	$\widehat{G}_1$	$\widehat{G}_2$	$\widehat{G}_3$
$k_{\theta,s}$	1	1000	10
$k_{r,s}$	1	10	5
c_s	1	1	1
$\widehat{f}_0$	5	1000	1
$\widehat{f}_1$	0.1	10	0.05

We choose our domain G so that it contains three geological layers with the one in the middle being "thin". The angle between the tangent line to the arcs separating geological layers inside of G, $z_1(x)$ and $z_2(x)$, and the x-axis is denoted by $\alpha(x)$, and in all the experiments $z_2(0) - z_1(0) = 10^{-5}$.

In the following experiments, G is a $(0,\ 0.2) \times (0,\ 1)$ rectangle with a curved layer, and the finest mesh has 160×80 cells in $G_{h,1}$, 160×4 cells in $G_{h,2}$, and 160×80 cells in $G_{h,3}$. We denote the mesh step corresponding to the finest mesh by $\boldsymbol{h}$.

We consider several choices of the center of curvature for the middle layer, resulting in different inclination angles. Example of such domain is given on Figure 6.3. An example of the embedding of the original domain into the reference domain is shown on Figure 6.4. The results for the different choices of inclination angles are given in the tables below.

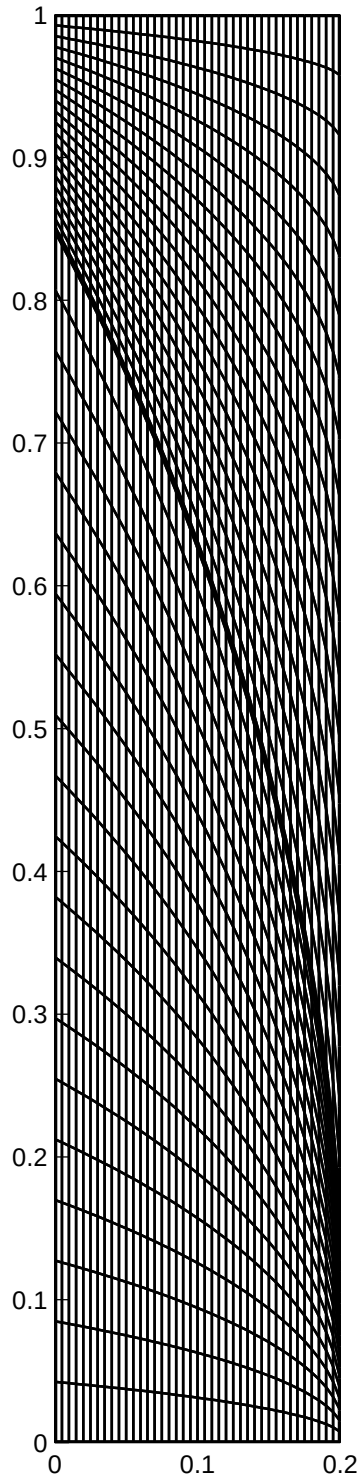

Figure 6.3: Domain G and mesh G_h for mesh step size $4\boldsymbol{h}$, $(x_C,\ y_C) = (-1.475,\ 0.05)$

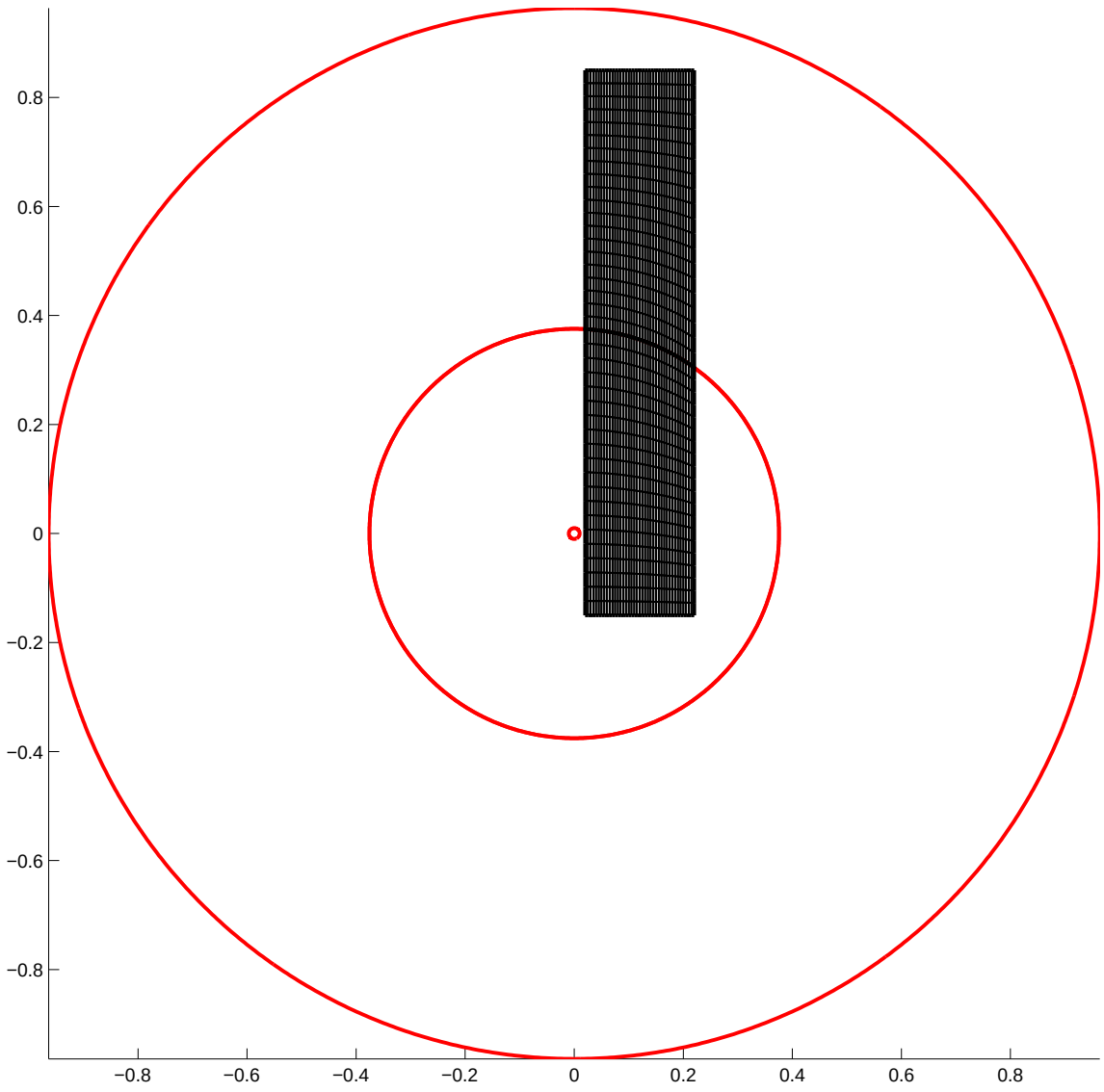

Figure 6.4: Embedding of the original domain G into the reference domain Ω, $(x_C,\ y_C) = (-0.02,\ 0.15)$

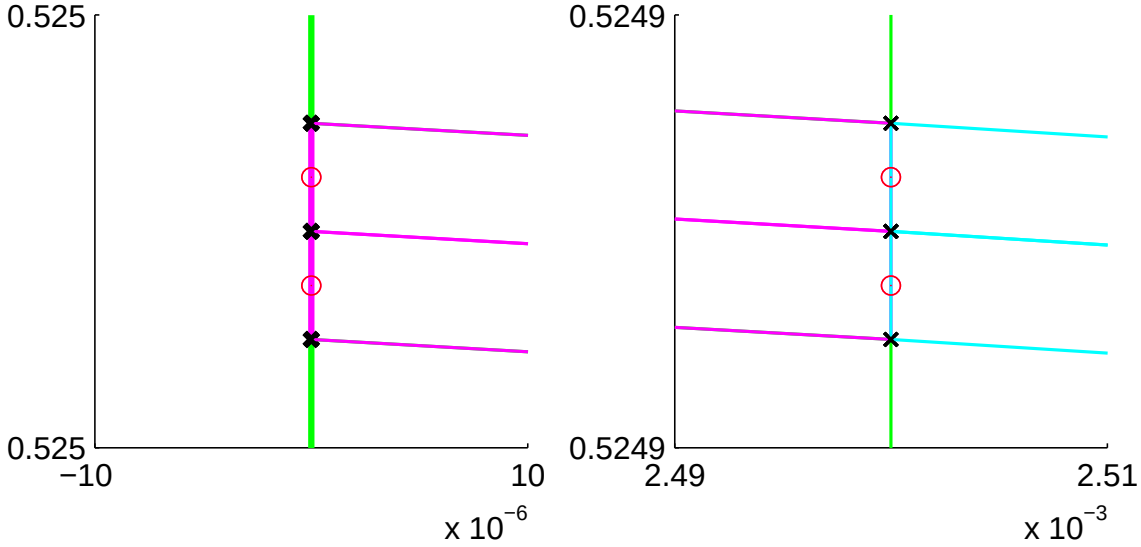

Figure 6.5: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **left** boundary of Ω, $(x_C,\ y_C) = (-0.02,\ 0.15)$, conforming mesh

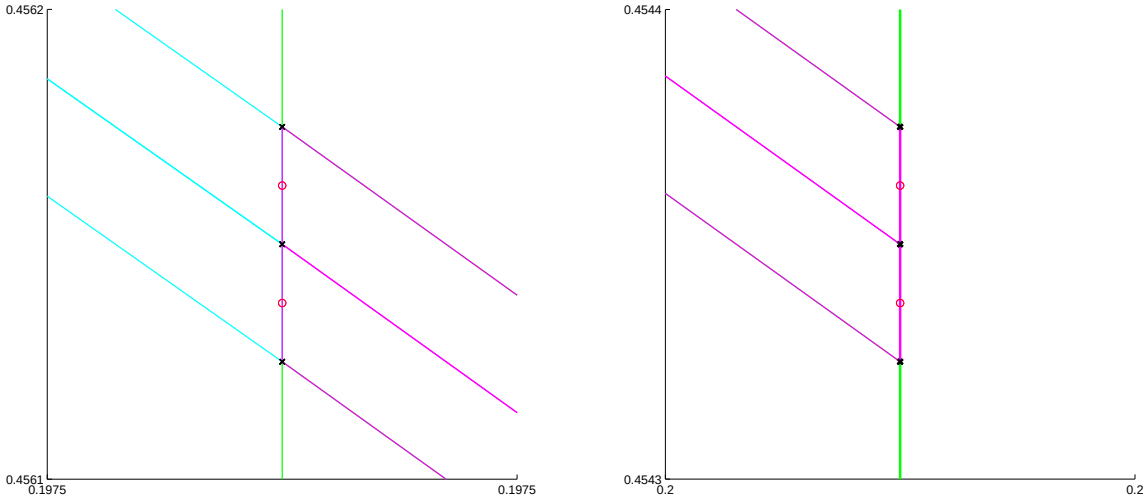

Figure 6.6: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **right** boundary of Ω, $(x_C,\ y_C) = (-0.02,\ 0.15)$, conforming mesh

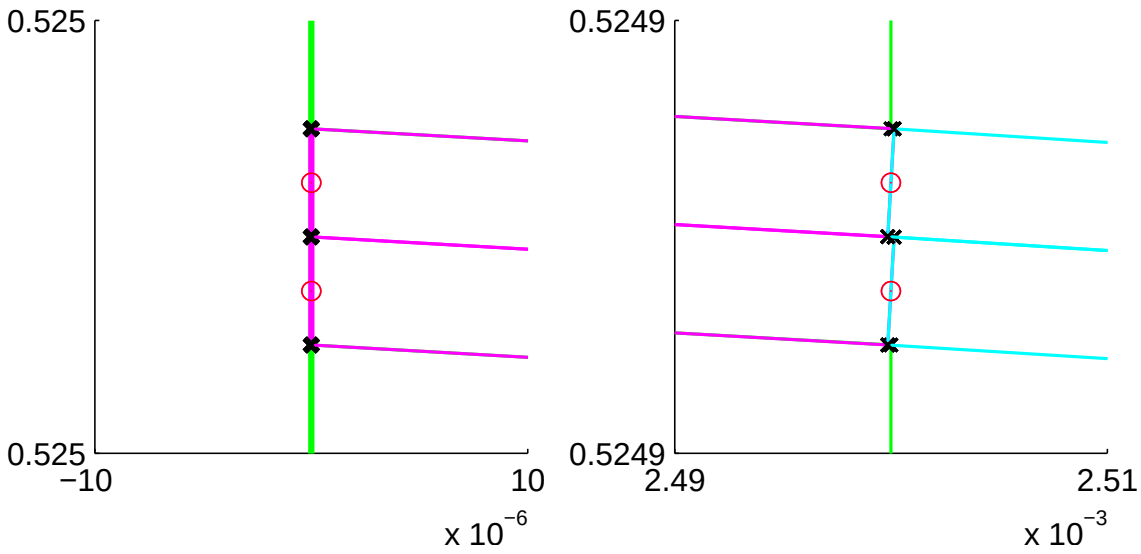

Figure 6.7: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **left** boundary of Ω, $(x_C,\ y_C) = (-0.02,\ 0.15)$, non-conforming mesh

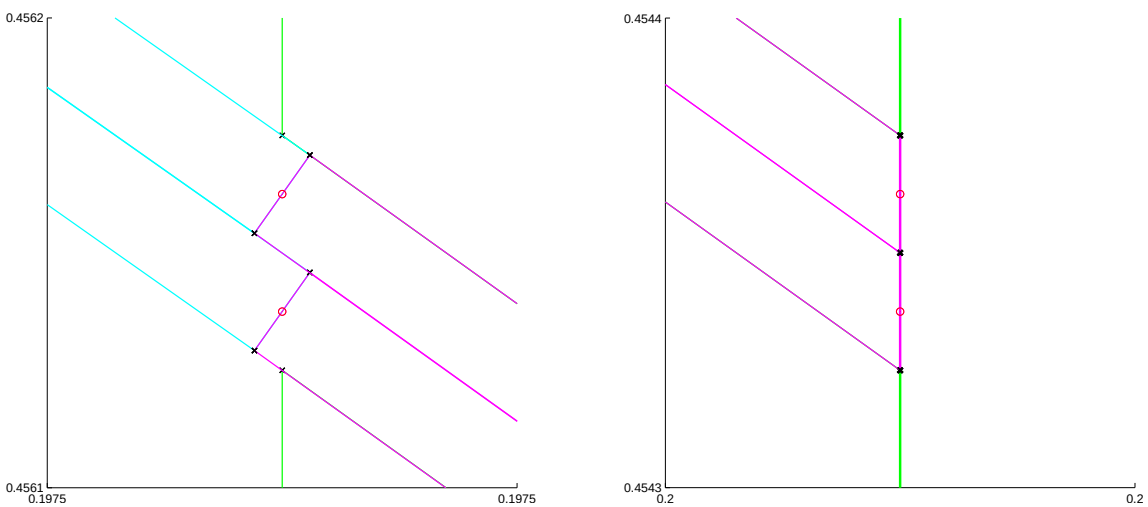

Figure 6.8: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **right** boundary of Ω, $(x_C,\ y_C) = (-0.02,\ 0.15)$, non-conforming mesh

Table 6.2: Relative error in KR interpolant $\overline{w}_h^{KR}$, %, $(x_C,\ y_C) = (-0.02,\ 0.15)$, conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	3.43	35.39	0.037787	0.542793	4.296938
$2\boldsymbol{h}$	3.24	35.63	0.023026	0.318449	2.186214
$\boldsymbol{h}$	3.15	35.74	0.011486	0.175824	1.103039

Table 6.3: Relative error in KR interpolant $\overline{w}_h^{KR}$, %, $(x_C,\ y_C) = (-0.02,\ 0.15)$, non-conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	3.43	35.39	0.037671	0.734999	4.291759
$2\boldsymbol{h}$	3.24	35.63	0.023040	0.371872	2.180725
$\boldsymbol{h}$	3.15	35.74	0.011602	0.187379	1.097456

Table 6.4: Relative error in PWC interpolant $\overline{w}_h^{PWC}$, %, $(x_C,\ y_C) = (-0.02,\ 0.15)$, conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	3.43	35.39	0.026212	0.049375	0.035678
$2\boldsymbol{h}$	3.24	35.63	0.007003	0.014074	0.010013
$\boldsymbol{h}$	3.15	35.74	0.001968	0.003964	0.003150

Table 6.5: Relative error in PWC interpolant $\overline{w}_h^{PWC}$, %, $(x_C,\ y_C) = (-0.02,\ 0.15)$, non-conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	3.43	35.39	0.026857	0.045561	0.029412
$2\boldsymbol{h}$	3.24	35.63	0.007749	0.013327	0.008297
$\boldsymbol{h}$	3.15	35.74	0.002854	0.008940	0.009742

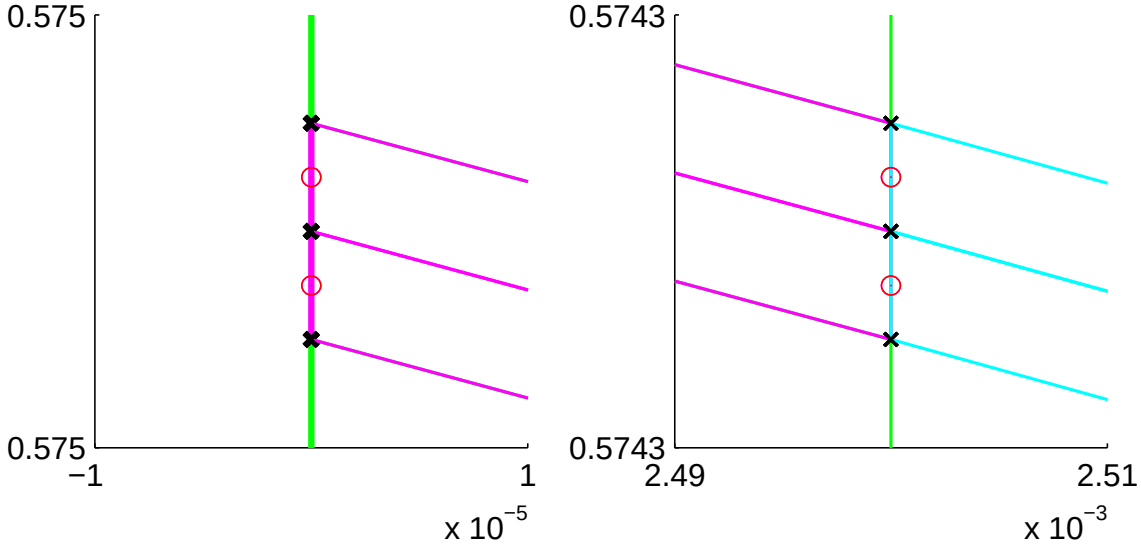

Figure 6.9: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **left** boundary of Ω, $(x_C,\ y_C) = (-0.1,\ 0.2)$, conforming mesh

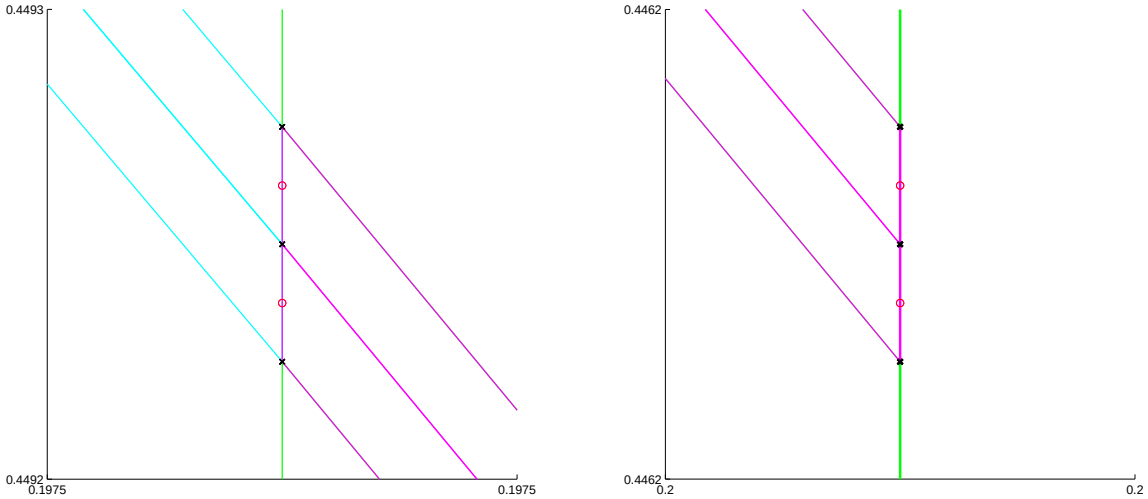

Figure 6.10: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **right** boundary of Ω, $(x_C,\ y_C) = (-0.1,\ 0.2)$, conforming mesh

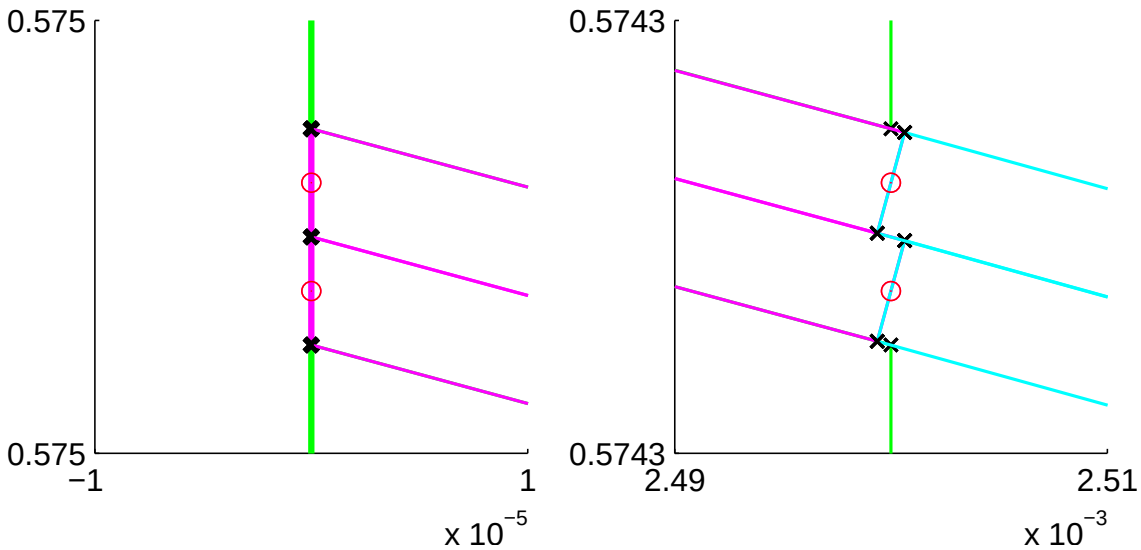

Figure 6.11: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **left** boundary of Ω, $(x_C,\ y_C) = (-0.1,\ 0.2)$, non-conforming mesh

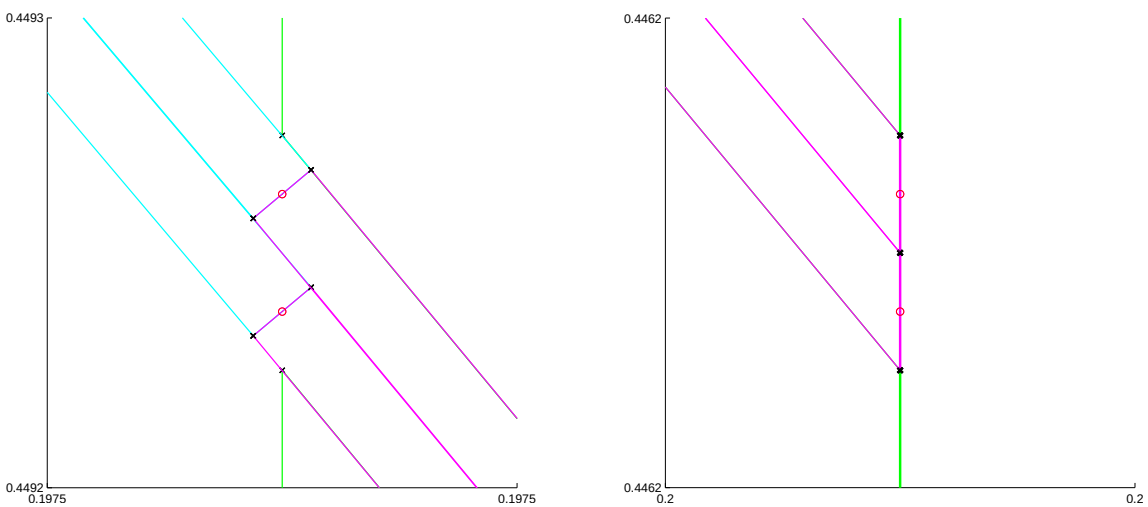

Figure 6.12: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **right** boundary of Ω, $(x_C,\ y_C) = (-0.1,\ 0.2)$, non-conforming mesh

Table 6.6: Relative error in KR interpolant $\overline{w}_h^{KR}$, %, $(x_C,\ y_C) = (-0.1,\ 0.2)$, conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	15.31	50.05	0.124941	0.518597	3.100448
$2\boldsymbol{h}$	15.12	50.33	0.062150	0.305697	1.573861
$\boldsymbol{h}$	15.03	50.48	0.031047	0.169229	0.793169

Table 6.7: Relative error in KR interpolant $\overline{w}_h^{KR}$, %, $(x_C,\ y_C) = (-0.1,\ 0.2)$, non-conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	15.31	50.05	0.124934	0.671095	3.094415
$2\boldsymbol{h}$	15.12	50.33	0.062192	0.333744	1.567483
$\boldsymbol{h}$	15.03	50.48	0.031135	0.165974	0.786714

Table 6.8: Relative error in PWC interpolant $\overline{w}_h^{PWC}$, %, $(x_C,\ y_C) = (-0.1,\ 0.2)$, conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	15.31	50.05	0.033708	0.073180	0.054000
$2\boldsymbol{h}$	15.12	50.33	0.009046	0.020717	0.014866
$\boldsymbol{h}$	15.03	50.48	0.002593	0.005985	0.004385

Table 6.9: Relative error in PWC interpolant $\overline{w}_h^{PWC}$, %, $(x_C,\ y_C) = (-0.1,\ 0.2)$, non-conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	15.31	50.05	0.034445	0.065410	0.046186
$2\boldsymbol{h}$	15.12	50.33	0.009942	0.018719	0.009488
$\boldsymbol{h}$	15.03	50.48	0.003657	0.012302	0.009182

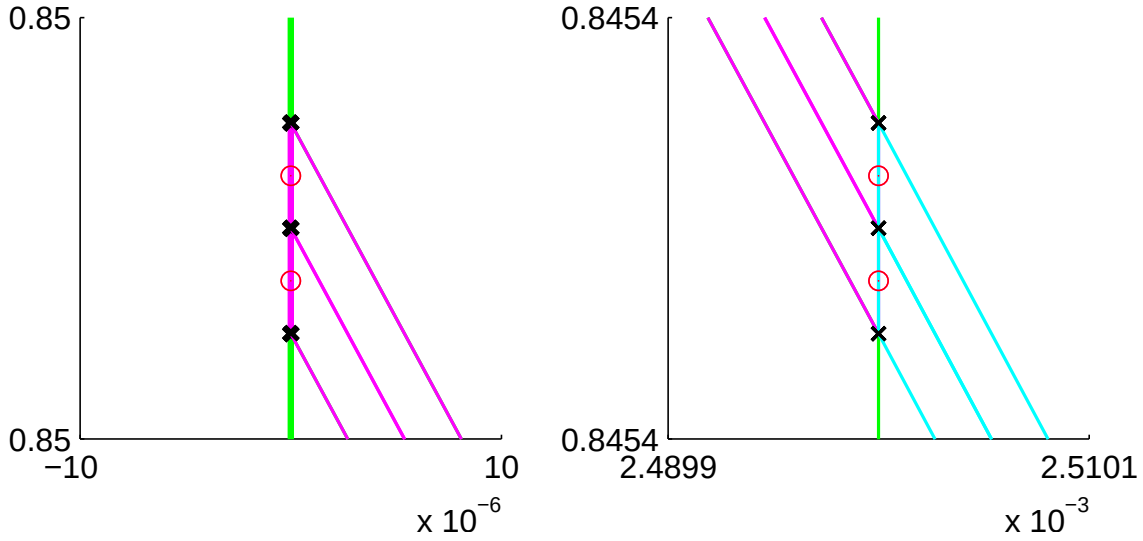

Figure 6.13: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **left** boundary of Ω, $(x_C,\ y_C) = (-1.475,\ 0.05)$, conforming mesh

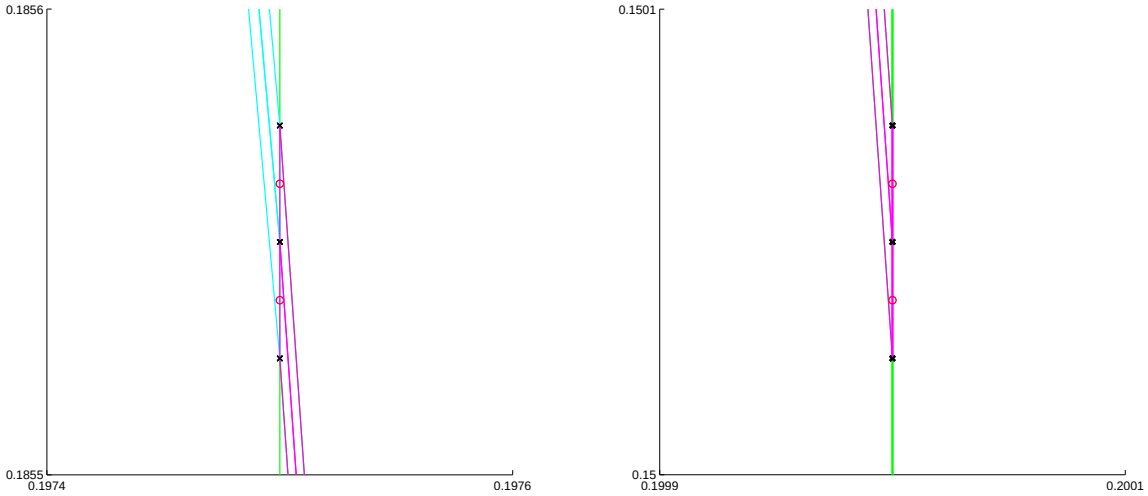

Figure 6.14: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **right** boundary of Ω, $(x_C,\ y_C) = (-1.475,\ 0.05)$, conforming mesh

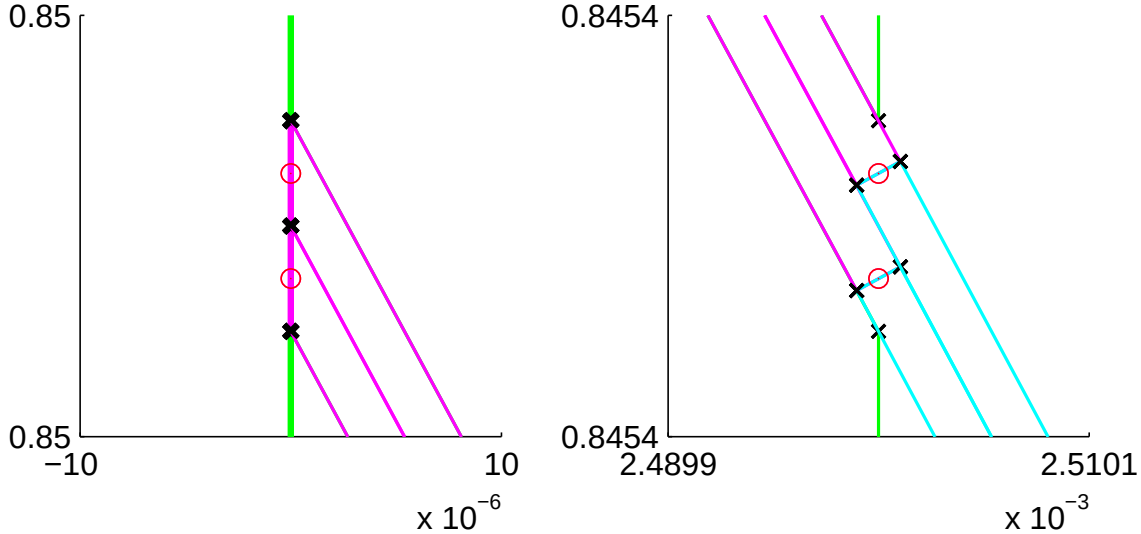

Figure 6.15: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **left** boundary of Ω, $(x_C,\ y_C) = (-1.475,\ 0.05)$, non-conforming mesh

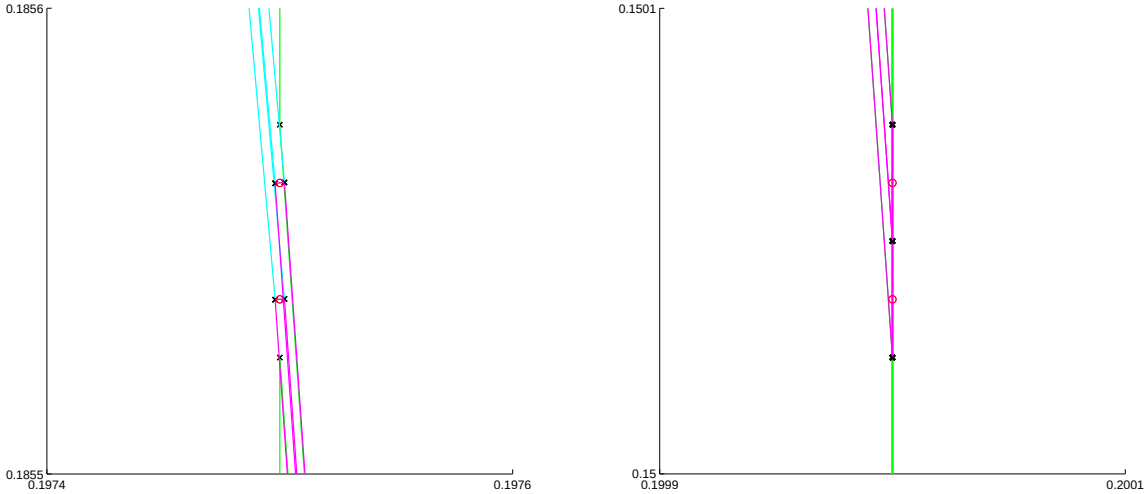

Figure 6.16: Left and right sides of the mesh cells in the thin layer Ω_2 next to the **right** boundary of Ω, $(x_C,\ y_C) = (-1.475,\ 0.05)$, non-conforming mesh

Table 6.10: Relative error in KR interpolant $\overline{w}_h^{KR}$, %, $(x_C,\ y_C) = (-1.475,\ 0.05)$, conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	61.71	85.5	0.874341	0.317107	1.540151
$2\boldsymbol{h}$	61.62	85.97	0.427795	0.176139	0.771348
$\boldsymbol{h}$	61.57	86.26	0.211650	0.095248	0.385448

Table 6.11: Relative error in KR interpolant $\overline{w}_h^{KR}$, %, $(x_C,\ y_C) = (-1.475,\ 0.05)$, non-conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	61.71	85.5	0.874809	0.300527	1.539996
$2\boldsymbol{h}$	61.62	85.97	0.427058	0.128813	0.769416
$\boldsymbol{h}$	61.57	86.26	0.210398	0.058555	0.383108

Table 6.12: Relative error in PWC interpolant $\overline{w}_h^{PWC}$, %, $(x_C,\ y_C) = (-1.475,\ 0.05)$, conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	61.71	85.5	0.133505	0.139052	0.122189
$2\boldsymbol{h}$	61.62	85.97	0.037234	0.043556	0.032024
$\boldsymbol{h}$	61.57	86.26	0.009843	0.013469	0.008046

Table 6.13: Relative error in PWC interpolant $\overline{w}_h^{PWC}$, %, $(x_C,\ y_C) = (-1.475,\ 0.05)$, non-conforming mesh

	$\alpha(0)$	$\alpha(0.2)$	ϵ_{G_1}	ϵ_{G_2}	ϵ_{G_3}
$4\boldsymbol{h}$	61.71	85.5	0.130754	0.141022	0.121639
$2\boldsymbol{h}$	61.62	85.97	0.039405	0.047554	0.029495
$\boldsymbol{h}$	61.57	86.26	0.018608	0.030809	0.010028

Bibliography

[1] G. Astrachancev, *An iterative method for solving elliptic net problems*, USSR Computational Math. and Math. Phys., 11(2) pp. 171–182, 1971.

[2] O. Axelsson, *An algebraic multilevel preconditioning methods. I*, Numer. Math., 56 pp. 157–177, 1989.

[3] N. S. Bachvalov, *On the convergence of relaxation method with natural constrains of the elliptic operator*, USSR Computational Math. and Math. Phys., 6(5) pp. 101–135, 1966.

[4] C. Bernardi, *Optimal finite-element interpolation on curved domains*, SIAM J. Numer. Anal., 5, pp. 1212–1240, 1989.

[5] J.H. Bramble and M. Zlámal, *Triangular elements in the finite element method*, Mathematics of Computation, 112, pp.809–820, 1970.

[6] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods.* Springer Verlag, Berlin, 1991.

[7] F. Brezzi, J. Douglas Jr., M. Fortin, and L.D. Marini, *Two families of mixed elements in two and three space variables*, RAIRO Modeèl. Math. Anal. Numèr., 21, pp. 581–604, 1987.

[8] F. Brezzi, J. Douglas Jr., and L.D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numer. Math., 47, pp. 217–235, 1985.

[9] A. Brandt, *Multilevel adaptive techniques for fast numerical solution to boundary value problems*, In: *Proceedings of the Third International Conference of Numerical Methods in Fluid Dynamics, 1972*, Lecture Notes in Physics, Springer-Verlag, 18 pp. 82–89, 1973.

[10] A. Brandt, *Multilevel adaptive solution to boundary value problems*, Math. Comput., 31 pp. 333–390, 1977.

[11] Z. Chen and J. Douglas Jr., *Prismatic mixed finite methods for second order elliptic problems*, Calcolo, 26, pp. 135–148, 1989.

[12] R.W. Clough, *The finite element method in plane stress analysis, In: Proceedings Second ASCE Conference on Electronic Computation*, Pittsburg, PA, 1960.

[13] R. Courant, *Variational methods for the solution of problems of equilibrium and variations*, Bull. Amer. Math, Comp., 64, pp. 1–23, 1943.

[14] R. Eymard, Th. Gallouët, and R. Herbin, *Finite Volume Methods*, volume VII of *Handbook of Numerical Mathematics*, Elsevier, 2000.

[15] R. P. Fedorenko, *A relaxation method for elliptic difference equations*, USSR Computational Math. and Math. Phys., 1(5) pp. 1092–1096, 1961.

[16] R. P. Fedorenko, *The speed of convergence of one iterative process*, USSR Computational Math. and Math. Phys., 4(3) pp. 227–235, 1964.

[17] J. Hyman, J. Morel, M. Shashkov and S. Steinberg, *Mimetic finite difference methods for diffusion equations.* Comp. Geosciences, 6(3-4), pp. 333–352, 2002.

[18] Yu. Kuznetsov, *Multigrid domain decomposition methods*, the Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, SIAM, pp. 290–313, 1989.

[19] Yu. Kuznetsov, *A new parallel algebraic preconditioner*, Journal of Numerical Linear Algebra with Applications, 1(2) pp. 215–225, 1992.

[20] Yu. A. Kuznetsov, *Mixed finite element method for diffusion equations on polygonal meshes with mixed cells*, J. Numer. Math., 14(4), pp. 305–315, 2006.

[21] Yu. Kuznetsov, Mixed finite element methods on polyhedral meshes for diffusion equations, In: *Partial Differential Equations* (Eds. R.Glowinski and P.Neitaanmäki). Springer, pp. 27–41, 2007.

[22] Y. Kuznetsov, K. Lipnikov, M. Shashkov, *Mimetic finite difference method on polygonal meshes for diffusion-type problems*, Comput. Geosciences, 8, pp. 301–324, 2004.

[23] Yu. Kuznetsov, O. Boyarkin, A. Prokopenko, and N. Yavich. *Preconditioned Iterative Solvers for Diffusion Equation in Heterogeneous Media*. Report on ExxonMobil Project, December 2008.

[24] Yu. Kuznetsov and A. Prokopenko, *A new multilevel algebraic preconditioner for the diffusion equation in heterogeneous media*, Numerical Linear Algebra with Applications, 17(5), pp. 759–769, 2010.

[25] Yu. Kuznetsov and S. Repin, *New mixed finite element method on polygonal and polyhedral meshes*, Russian J. Numer. Anal. Math. Modelling, 18, pp. 261–278, 2003.

[26] Yu. Kuznetsov and S. Repin, *Mixed finite element method on polygonal and polyhedral meshes*, In: Proc. of the 5th ENUMATH conference, Prague, 2003. World Scientific Publ. Co., 2004.

[27] Yu.A. Kuznetsov and S.I. Repin, *Convergence analysis and error estimates for mixed finite element method on distorted meshes*, J. Numer. Math. 13, pp. 33–51, 2005.

[28] Yu. Kuznetsov and M. F. Wheeler, *Optimal order substructuring preconditioners for mixed finite element methods on nonmatching grids*, East West Journal of Numerical Mathematics, 3 pp. 127–144, 1995.

[29] P. J. Lanzkron, D. J. Rose, and D. B. Szyld, *Convergence of nested classical iterative methods for linear systems*, Numerische Mathematik, 58(1) pp. 685–702, 1990.

[30] M. Lenoir, *Optimal isoparametric finite elements and error estimates for domains involving curved boundaries*, SIAM J. Numer. Anal., 23, pp. 562–580, 1986.

[31] S. Margenov and P. Vassilevski, *Algebraic multilevel preconditioning of anisotropic elliptic problems*, SIAM J. Sci. Comput., 15(5) pp. 1026–1037, 1994.

[32] L. Margolin, M. Shashkov, and P. Smolarkiewicz, *A discrete operator calculus for finite difference approximations.* Comput. Meth. Appl. Mech. Engrg., 187, pp. 365–383, 2000.

[33] S. F. McCormick, *Multigrid Methods*, Philadelphia, SIAM, 1987.

[34] J. Morel, M. Hall, and M. Shashkov, *A local support-operators diffusion discretization scheme for hexahedral meshes.* J. Comput. Phys., 170, pp. 338–372, 2001.

[35] J. Morel, M. Hall, and M. Shashkov, *A local support-operators diffusion discretization scheme for quadrilateral $r - z$ meshes.* J. Comput. Phys., 144, pp. 17–51, 1998.

[36] J.C. Nedelec, Mixed Finite Elements in R^3, *Numer. Math.*, 35, pp. 315-341, 1980.

[37] A. Prokopenko, *Multilevel preconditioners and their applications in geoscience*, PhD Thesis, University of Houston, 2011.

[38] P.A. Raviart and J.-M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, In: *Mathematical Aspects of Finite Element Methods* (Eds. I. Galligani and E. Magenes), Springer-Verlag, New York-Berlin, pp. 292–315, 1977.

[39] J.E. Roberts and J.-M. Thomas, Mixed and hybrid methods. In: *Handbook of Numerical Analysis, Vol. 2* (Eds. P.-G. Ciarlet and J.-L. Lions). Elsevier Sci. Publishers, Amsterdam, pp. 523–639, 1991.

[40] A. Cleary, R. Falgout, V. Henson, J. Jones, T. Manteuffel, S. McCormick, G. Miranda, and J. Ruge, *Robustness and scalability of algebraic multigrid*, SIAM J. Sci. Comput, 21 pp. 1886–1908, 1998.

[41] A.A. Samarskii, *The theory of difference schemes*, Marcel Dekker, Inc. New York-Basel, 2001. (Translation from Russian version Teoriya Raznostnych Schem, Moscow, Nauka, 1977).

[42] J. Shen, *Mixed finite element methods on distorted rectangular grids*, Technical Report ISC-94-13-MATH, Institute of Scientific Computation, Texas A&M University, July 1994.

[43] G. Strang, and G.J. Fix, *An Analysis of the Finite Element Method*, Prentice Hall, Englewood Cliffs, NJ, 1973.

[44] K. Stüben, *Algebraic multigrid: experience and comparisons*. Applied Math. and Comp., 13 (3-4) pp. 419–451, 1983.

[45] K. Stüben, *A review of algebraic multigrid*, Journal of Computational and Applied Mathematics, 128(1-2) pp. 281–309, 2001.

[46] J.M. Thomas, *Sur l'analyse numérique des méthodes d'éléments finis hybrides et mixtes*, Thése De Doctorat d'état, Université Pierre et Marie Curie, Paris, France, 1977.

[47] R.S. Varga. *Matrix Iterative Analysis*. Prentice Hall, 1962.

[48] M. Zlámal, *Curved elements in the finite element method, I*, SIAM J. Numer. Anal., 10, pp. 229–240, 1973.

[49] M. Zlámal, *Curved elements in the finite element method, II*, SIAM J. Numer. Anal., 11, pp. 347–362, 1974.