2014-01-01

A Block Precondtioner for a Mixed Finite Element Method for Biot;s Equations

Maranda Lee Bean
University of Texas at El Paso, maranda.bean@gmail.com

A BLOCK PRECONDITIONER FOR A MIXED FINITE ELEMENT METHOD FOR BIOT'S EQUATIONS

MARANDA BEAN

Computational Science

APPROVED:

Son-Young Yi, Chair, Ph.D.

Konstantin Lipnikov, Ph.D.

Natasha Sharma, Ph.D.

Charles Ambler, Ph.D.
Dean of the Graduate School

A BLOCK PRECONDITIONER FOR A MIXED FINITE ELEMENT METHOD FOR BIOT'S EQUATIONS

by

MARANDA BEAN

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Computational Science

THE UNIVERSITY OF TEXAS AT EL PASO

December 2014

Acknowledgements

My deepest thanks go out to my advisor, Dr. Son-Young Yi of the Department of Mathematics at UTEP. Her willingness to answer and re-answer my questions and her enthusiasm for research are a constant source of inspiration. I am very fortunate that she was there to keep me on the right track and help me find and correct my mistakes.

Through the funding of a National Science Foundation grant, I was able to go to Los Alamos National Laboratory in the Summers of 2013 and 1014. I am also eternally gratefully to Dr. Konstantin Lipnikov for mentoring me there. I am in awe of his williness and patience to teach me about topic I knew nothing about.

I would also like to thank Dr. Natasha Sharma for finding the time be on my committee and provide so many helpful comments.

Additionally, I'd like to express my gratitude to all of the faculty at UTEP who have taught me the things I needed to know to complete this project and to the staff at UTEP and Los Alamos National lab, who have helped me with the process of completing this project. I deeply appreciate your dedication.

Finally, I need to thank my family for supporting me, putting up with me and pushing me. Without the constant support of my mother, Barbara, and my husband, Ryan, this task would have been insurmountable.

Abstract

In this thesis, we explore the solution methods for the linear system resulting from a mixed finite element method applied to the Biot's consolidation model. This model describes the coupled interactions between a porous solid and the fluid contained within it. Specifically, we use a method developed by Yi [Numer. Methods for PDEs, 29(5), pp. 1749-1777] that expands Biot's system to include fluid pressure, solid displacement, fluid flux and total stress as primary unknowns.

As the resulting linear system is a large, sparse, saddle point system, we attempt to solve this system via a Schur complement preconditioned iterative method. Using the exact Schur complement preconditioner would require the inversion of the first block of the saddle point system A. Since this can still be computationally expensive, we attempt to use an approximation to the Schur complement based on a spectrally equivalent approximation to A.

To test the preconditioner, we solve problems in homogeneous and heterogeneous layered media. In the homogeneous case, we show that the number of iterations required to solve the system increases only slightly when the element size and time step are decreased at corresponding rates. In the case of heterogeneous material, we require slightly more iterations to solve the problem as the difference in the material parameters of layers is more pronounced. However, the amount of work required to apply the preconditioner within each iteration seems to depend on the difference in the material parameters of the layers and the size of the element.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Consider a solid medium or matrix whose pores or empty spaces are filled with some mobile fluid. For simplicity, the solid is assumed to be isotropic, the same in all directions, and linearly elastic. The theory of poroelasticity describes the mechanical response of this material. Detailed descriptions of poroelasticity can be found, for example, in [5, 58, 56], but we cover only the basic idea below. The two coupled phenomena that are integral to this theory are solid-to-fluid coupling and fluid-to-solid coupling. These phenomena are described by Wang [58] as follows.

- Solid-to-fluid coupling happens when changes in the stress applied to the solid matrix induce changes in the fluid pressure or mass in the system.
- Fluid-to-solid coupling happens when changes in fluid pressure or mass causes changes in the solid matrix.

Consider first the solid-to-fluid coupling. Assume we have an applied stress. This may cause a deformation, which in turn changes the fluid pressure or causes fluid flow. Think for example of squeezing a wet sponge. If we think of a common sponge, the fluid will flow out of the sponge as it is squeezed. However, if the water is not allowed to flow out of the sponge for some reason, then the fluid pressure will be increased as a result of the squeeze.

At the same time, an increase or decrease in fluid mass or fluid pressure may deform the solid matrix. This is the so called fluid-to-solid coupling. A famous example of this is land subsidence, like that of the San Joaquin Valley in California [21]. In this valley, many

years of groundwater pumping and farming have caused consolidation, a reduction of the solid matrix volume as a result of fluid removal, and lowered the land surface.

We assume that the coupling of both of these phenomena has a significant impact on the system. If this is neglected, a simpler problem that accounts only for fluid-to-solid coupling may be solved. It is assumed in this type of model that the changes in fluid pressure or mass affect the solid significantly, but the changes in the solid do not have a large impact on the fluid. When only fluid-to-solid coupling is accounted for, the fluid pressure can first be found, then used to find the stress. These assumptions may work well for certain specific situations, such as when the fluid is highly compressible [58], but in many cases the solid-to-fluid coupling plays a significant role and may not be neglected.

These key phenomena occur simultaneously in a time dependent manner. When the fluid pressure is not uniform throughout the medium, we will have fluid flow that is governed by Darcy's law. This time dependent dissipation of fluid pressure in turn causes a time dependence on the stress in the solid matrix. However, when the internal forces are neglected, the processes can be considered quasi-static. More detailed descriptions of these phenomena can be found, for example, in [58].

Poroelasticity has many applications in the science and engineering fields. Soil consolidation, as described above, is studied for example in [7, 13, 52, 58]. The field of reservoir engineering makes use of poroelasticity to study and predict behavior of coupled geomechanics and flow, see for example [50]. This field is also concerned with behavior of and around boreholes, see e.g. [45]. In environmental engineering, topics such as containment of waste [30] and carbon dioxide sequestration [38] are explored. Additionally, in biomechanical engineering poroelasticity is used to model materials such as the brain [47] and bones [14].

1.2 Biot's model

We consider an incompressible porous medium, which is either nearly saturated by an incompressible fluid or completely saturated by a slightly compressible fluid. The Biot model was originally concerned with soil consolidation. The theoretical basis was formed by Terzaghi [53] and was later generalized into Biot's poroelasticity model [7]. The model consists of an equilibrium equation for momentum and a diffusion equation for Darcy flow.

Consider a bounded, connected, Lipschitz domain Ω in $\mathbb{R}^2$. The governing equations of Biot's model are the following:

$$\frac{\partial}{\partial t}\left(c_o p + \alpha \nabla \cdot \boldsymbol{u}\right) - \nabla \cdot \left(\mathbf{K} \nabla p\right) = h, \tag{1.1a}$$

$$-\left(\lambda + \mu\right) \nabla \left(\nabla \cdot \boldsymbol{u}\right) - \mu \nabla^2 \boldsymbol{u} + \alpha \nabla p = \boldsymbol{f}, \tag{1.1b}$$

where h is the volumetric source or sink term and $\boldsymbol{f}$ is the body force. The unknowns are the fluid pressure, p, and the solid phase displacement, $\boldsymbol{u} = (u_1, u_2)^T$. In general, we will refer to fluid pressure as simply pressure when it is the only pressure concerning us. The parameters λ and μ are the Lam*é* constants. The constant c_0 is the constrained specific storage coefficient which is a ratio between the change in increment of fluid content and the change in pore pressure when the measurements are taken at constant strain [58]. The Biot-Willis constant α is the ratio between volume of fluid content added and the change in bulk volume under the condition of constant pore pressure [58]. The constant, α, is usually near unity. The permeability tensor, $\mathbf{K}$, is assumed to satisfy

$$k_{min} \eta^T \eta \leq \eta^T \mathbf{K}(x) \eta \leq k_{max} \eta^T \eta \quad \forall x \in \Omega \quad \forall \eta \in \mathbb{R}^2, \tag{1.2}$$

where k_{min} and k_{max} are positive constants. Additionally, $\mathbf{K}$ is uniformly positive definite and symmetric. The system (1.1) must have boundary and initial conditions to be solved. Therefore, let the boundary of Ω, $\partial\Omega$, be partitioned into $\{\Gamma_p, \Gamma_f\}$ and $\{\Gamma_d, \Gamma_t\}$ such that $\partial\Omega = \Gamma_p \cup \Gamma_f$ and $\partial\Omega = \Gamma_d \cup \Gamma_t$. Then

$$p = p_0 \text{ on } \Gamma_p, \quad \boldsymbol{q} \cdot \mathbf{n} = q_0 \text{ on } \Gamma_f, \quad \boldsymbol{u} = \boldsymbol{u}_0 \text{ on } \Gamma_d, \quad \widetilde{\boldsymbol{\sigma}}\mathbf{n} = \widetilde{\boldsymbol{\sigma}}_0 \text{ on } \Gamma_t, \tag{1.3}$$

where $\mathbf{n}$ is an outward normal vector. The unknowns $\boldsymbol{q}$ and $\widetilde{\boldsymbol{\sigma}}$ are the volumetric fluid flux and total stress tensor, respectively. The flux is related to the pressure by

$$\boldsymbol{q} = -\mathbf{K}\nabla p. \tag{1.4}$$

Total stress satisfies the constitutive equation

$$\widetilde{\boldsymbol{\sigma}} = 2\mu\epsilon_{ij}(\boldsymbol{u}) + \lambda tr(\epsilon(\boldsymbol{u})) - \alpha p\mathbf{I} \tag{1.5}$$

where $\epsilon(\boldsymbol{u}) = \frac{1}{2}\left[\nabla\boldsymbol{u} + (\nabla\mathbf{u})^T\right]$. The initial conditions are

$$p(0) = p^0 \quad \text{and} \quad \boldsymbol{u}(0) = \boldsymbol{u}^0 \quad \text{in} \quad \Omega. \tag{1.6}$$

More information about the derivation of this model can be found in e.g. [42, 51].

1.3 Known Solution Methods

In practice, analytic solutions of a poroelastic problem are rarely available. Therefore, computational simulations are often used to generate approximate solutions. Many methods have been proposed and studied to solve this complex coupled problem. A few of these methods are described below.

Much research has been conducted concerning the numerical treatment of this poroelasticity model. Traditionally, standard finite element methods have been employed and make use of continuous Galerkin elements for both the displacement and pressure. Lewis and Schrefler [31] provide a general overview and reference of this method. An approach that combines a mixed finite element method for the flow variables, pressure p and flux $\boldsymbol{q}$, and a continuous Galerkin method for displacement, $\boldsymbol{u}$, is also explored by Phillips and Wheeler [42]. However, it is well known that these methods can produce a behavior known as locking. Locking is manifested as nonphysical, oscillatory behavior in the pore pressure. This behavior tends to occur with certain physical assumptions, such as low permeability, and at short time steps. Furthermore, the oscillations tend to dissipate as time increases. More information about locking and its causes can be found for example in [44].

Several methods have been developed to overcome the problem of locking. For example, in [43], a method combining a mixed finite element method for the flow variables, p and $\boldsymbol{q}$, with a discontinuous Galerkin method for displacement, $\boldsymbol{u}$, is explored. This method solves for the same unknowns as, and is a logical extension of, the approach described above that uses a continuous elements for displacement. However, when discontinuous elements are used, Phillips et.al. [43] are able to show theoretical convergence results that are independent of the constrained specific storage coefficient, c_0. This indicates that the method may remedy locking. Additionally, Haga et. al. [25] compared the results of finite element methods based on two-, three- and four-field formulations in cases where locking may be seen. Their two-field formulation solves for just fluid pressure, p, and displacement, $\boldsymbol{u}$. However, they also attempt two different three-field formulations. One formulation uses fluid pressure, fluid velocity and displacement, while the other formulation uses fluid pressure, solid pressure and displacement as its primary unknowns. Their four-field formulation is a combination that uses fluid pressure, solid pressure, fluid velocity and displacement as its unknowns. Each of these is attempted with several different approximation spaces.

Some other possible solution methods include the following. Wan [57] combines a stabilized finite element method and a control-volume finite difference method. He does this on the standard pressure and displacement formulation as well as on a formulation which uses displacement, fluid velocity and fluid pressure as its unknowns. Naumovich [40] explores finite difference, finite volume, and multigrid methods in the case of discontinuous coefficients.

We choose to use a mixed finite element formulation proposed by Yi [60]. This method uses the total stress tensor $\widetilde{\boldsymbol{\sigma}}$, displacement $\boldsymbol{u}$, fluid flux $\boldsymbol{q}$, and pressure p as its primary unknowns. This method has the benefit of coupling two existing, stable mixed finite element. Specifically, we will need one mixed finite element for elasticity that is based on the Hellinger-Reissner principle and a second mixed finite element for flow. It also makes use of the backward Euler time stepping scheme.

There are several benefits to using this method. First,this method is a possible remedy

for the locking problem described above. Additionally, it has the advantage of using stress as a primary variable and the resulting approximation for stress will be more accurate than a result obtained via post processing. Additionally, using stress as a primary variable allows for interface conditions in heterogeneous material, some of which depend on the stress, to be simply enforced. More details about this method will be provided in Chapter 2.

1.4 Overview of Preconditioning

In most cases, we will be using preconditioned iterative methods to solve a linear system. That is, we will seek to solve an equivalent system to $\mathcal{A}x = b$. This is done by either right, left or split preconditioning. We will attempt to form a preconditioner, $\mathcal{P}$, that approximates $\mathcal{A}$ in some manner, with the requirements that a system of the form $\mathcal{P}x = b$ is relatively easy to solve, and the preconditioned system has better properties than the original system [11].

With left preconditioning, the system becomes

$$\mathcal{P}^{-1}\mathcal{A}x = \mathcal{P}^{-1}b. \tag{1.7}$$

In practice, $\mathcal{P}^{-1}$ was not calculated. Therefore, when a preconditioning step was used, a linear system was solved.

In some case, using a right preconditioner is preferable because it leads to the ability to form a flexible variant. More details about a flexible variant will be given in Section 1.5.2. The right preconditioned system takes the form

$$\mathcal{A}\mathcal{P}^{-1}u = b, \quad x = \mathcal{P}^{-1}u. \tag{1.8}$$

Split preconditioning can occur when the preconditioner can be separated in simpler parts. One example is if the preconditioner $\mathcal{P}$ has a Cholesky decomposition. In that case $\mathcal{P} = LL^T$ where L is a lower triangular matrix. This type of preconditioning is a combination of the left and right preconditioning methods.

1.5 Review of Some Iterative Methods

The linear system that results from the finite element discretization is often large and sparse. It is well known that direct methods, like Gaussian elimination, require large amounts of storage and many operations to solve a particular system. Although this algorithm can be modified to take advantage of sparsity, the memory and arithmetic requirements may still be too large as the number of unknowns is increased. Therefore, direct methods may not be the most efficient methods for solution, especially when many elements or divisions are used.

For our numerical experiments, we focus primarily on using projection methods which seek to find an approximate solution to the linear system

$$\mathcal{A}x = b. \tag{1.9}$$

For $\mathcal{A} \in \mathbb{R}^{N\times N}$, we will seek to find an approximation, x_m, to the solution $x \in \mathbb{R}^N$ from some subspace $x_0 + \mathcal{K}_m$ of $\mathbb{R}^N$, where x_0 is some initial guess to the solution, $\mathcal{K}_m$ of $\mathbb{R}^N$ is the space of candidate approximations and since $\mathcal{K}_m$ is of size m, then this requires m constraints, which are normally expressed as orthogonality conditions [48]. Often the residual

$$r_m = b - \mathcal{A}x_m$$

is forced to be orthogonal to m linearly independent vectors, which define a subspace $\mathcal{L}_m$, called the subspace of constraints, of dimension m.

Specifically, the methods used are Krylov subspace methods. That is, we set the subspace $\mathcal{K}_m$ to be the following Krylov space

$$\mathcal{K}_m(\mathcal{A}, r_0) = span\left\{r_0, \mathcal{A}r_0, \mathcal{A}^2 r_0, ..., \mathcal{A}^{m-1} r_0\right\}.$$

More information on projection methods and Krylov subspace can be found for example in [48].

1.5.1 Generalized Minimum Residual Method

Many experts in the field of liner algebra now agree that the Generalized Minimum Residual (GMRES) method was first proposed by Marchuk and Kuznetsov [36]. We follow the full formulation presented by Saad [48, 49] in which the choice of $\mathcal{L}_m = \mathcal{A}\mathcal{K}_m$ is designed to minimize the norm of the residual vector over the vectors in $x_0+\mathcal{K}_m$. It is heavily dependent on the Arnoldi procedure [3], which is used to form Arnoldi vectors, $v_1, v_2, ..., v_m$, to make an orthonormal basis of the Krylov space $\mathcal{K}_m(v_1, \mathcal{A})$. In practice, we let v_1 be the normalized residual, $r_0/\|r_0\|_2$, and the rest of the orthonormal basis vectors $v_2, v_3, ...v_m$ are constructed via the Arnoldi−Modified Gram−Schmidt process for stability.

Define $V_m = [v_1, v_2, ...v_m]$. Then any vector x in $\mathcal{K}_m$ can be written as

$$x = x_0 + V_m y$$

where y is some vector in $\mathbb{R}^m$ [48]. Let

$$J(y) = \|b - \mathcal{A}x\|_2 = \|b - \mathcal{A}\left(x_0 + V_m y\right)\|_2$$

GMRES finds

$$y_m = \underset{y}{\operatorname{argmin}}\, J(y).$$

Then the approximate solution is

$$x_m = x_0 + V_m y_m. \tag{1.10}$$

The flexible variant follows the same procedure with a minor change occurring when preconditioners are considered [48]. This is discussed further in the next section.

The implementation of this method is based on the algorithms provided in [4, 49, 48]. A template version of the code for MATLAB or Octave was used from [41]. A restarted variant was also used to control memory requirements. After a predetermined number of iterations, an approximation x_m is found even if the tolerance has not yet been reached. We then use this result as an initial guess, x_0, and begin the method again.

1.5.2 Flexible Generalized Residual Method

To get a flexible version of GMRES (FGMRES), recall that the approximate solution given in (1.10) is a linear combination of vectors v_i. When a preconditioner is applied, this linear combination must also be preconditioned. If a standard GMRES is desired, this preconditioning can be done after the linear combination is formed. However, if the vectors are preconditioned and saved as

$$z_j = \mathcal{P}_j^{-1} v_j, \tag{1.11}$$

then the preconditioner, $\mathcal{P}_j$, does not have to be the same in every step [48]. That is, the solution is written as

$$x_m = x_0 + Z_m y_m,$$

where $Z_m = [z_1, z_2, ...z_m]$ instead of as

$$x_m = x_0 + \mathcal{P}^{-1} V_m y_m,$$

where $V_m = [v_1, v_2, ...v_m]$.

For the purposes of this thesis, FGMRES is generally used as a black box solver and we focus on providing an effective preconditioner. This algorithm can be found, for example, in [49, 48]. Below we directly quote the algorithm for FGMRES given by Saad [48].

Flexible GMRES algorithm

1. Compute $r_0 = b - \mathcal{A}x_0$, $\beta = \|r_0\|_2$, and $v_1 = r_0/\beta$
2. For $j = 1, ..., m$, Do
3. Compute $z_j = \mathcal{P}_j^{-1} v_j$
4. Compute $w = \mathcal{A} z_j$
5. For $i = 1, ..., j$, Do
6. $h_{i,j} := (w, v_i)$
7. $w := w - h_{i,j} v_j$

8. EndDo

9. Compute $h_{j+1,j} = \|w\|_2$ and $v_{j+1} = w/h_{j+1,j}$

10. Define $Z_m := [z_1, ..., z_m]$, $\tilde{H}_m = \{h_{i,j}\}_{1 \leq i \leq j+1; 1 \leq j \leq m}$

11. EndDo

12. Compute $y_m = \text{argmin}_y \|\beta e_1 - \tilde{H}_m y\|^2$ and $x_m = x_0 + Z_m y_m$

13. If satisfied Stop, else set $x_0 \leftarrow x_m$ and go to 1

We focus on finding the preconditioned vectors z_j, which are found in step 3 of the algorithm. Note the preconditioner $\mathcal{P}$ is only used once per FGMRES iteration. As stated above, $\mathcal{P}_j^{-1}$ is not calculated directly. We find z_j as the solution to the system $\mathcal{P}_j z_j = v_j$. This system is often solved with an additional iterative method. Note that the matrix $\mathcal{P}_j$ may be the same for every $j = 1, ..., m$. However, when we find z_j using an iterative method, it is an approximation and each z_j corresponds to a slightly different $\mathcal{P}_j^{-1}$.

1.5.3 Conjugate Gradient Method

The conjugate gradient (CG) method, and its preconditioned version (PCG), is a well known and effective method for solving linear systems (1.9) where $\mathcal{A}$ is symmetric positive definite (SPD) e.g. [48, 4]. At each step of the method, a new approximation

$$x_{j+1} = x_j + \alpha_j p_j$$

is found. In this case, p_j is the so called search direction and α_j is a constant. The next residual, r_{j+1} is then

$$r_{j+1} = r_j - \alpha_j \mathcal{A} p_j$$

and α_j can be found from the requirement that the residuals, r_j and r_{j+1}, are orthogonal. This allows the next search direction to be found as

$$p_{j+1} = r_{j+1} + \beta_j p_j$$

where β_j is found from the requirement that p_{j+1} be orthogonal to $\mathcal{A}p_j$. This is repeated iteratively until the residual is smaller than some predetermined tolerance. More detailed descriptions are available in many texts, for example [48, 4].

The implementation of this method was done using the built in MATLAB and Octave function or the built in hypre function [18, 10]. These implementations allow for preconditioning to be used as well.

1.5.4 Algebraic Multigrid

Algebraic multigrid method was developed as a way to apply the principles of multigrid to a linear system without explicit knowledge of the geometry [17]. The basic ideas of which are smoothing and course grid corrections.

In the geometric case, many standard iterative methods, such as Jacobi and Gauss-Seidel, have a smoothing property. That is, they can eliminate high frequency or oscillatory error very quickly, but are slow to remove low frequency or smooth components of error [9]. It is also known that a good initial guess can speed the convergence of these basic iterative methods, and performing some preliminary iterations on a coarser grid can help provide that better initial guess. So we would like to apply a simple method just until the oscillatory error is removed. Then, find the error and restrict it to a coarser grid. On this coarser grid, the error that was smooth now appears more oscillatory [9].

In the case of only 2 grids, let the exponent h, denote matrices or vectors on the finer grid and the exponent $2h$ denote matrices or vectors on the coarser grid. We perform a predetermined number of relaxation steps on $\mathcal{A}^h x^h = b^h$, then find the residual $r^h = b^h - \mathcal{A}^h x^h$ and restrict it to a coarser grid. We also need to restrict $\mathcal{A}^h$ to the coarser grid. To make this restriction, we will make use of an interpolation matrix $M : \mathbb{R}^{2h} \to \mathbb{R}^h$, which maps coarse grid information to the fine grid, and M^T, which maps from the fine grid to the coarse grid [17]. A common approach used to restrict $\mathcal{A}^h$ to the coarse grid is to use the Galerkin operator, $A^{2h} = M^T A^h M$, which allows the residual to become $r^{2h} = M^T r^h$ [17]. We then solve the smaller system $A^{2h} e^{2h} = r^{2h}$ on the coarse grid and correct our estimate

by adding the interpolated result. That is, $x^h \leftarrow x^h + Mr^{2h}$. We may then perform a few more smoothing steps to $\mathcal{A}^h x^h = b^h$. In practice, there may be several grids or levels. A system is only solved on the coarsest grid.

We are applying these ideas to a linear system where the grid may be unknown, or not present. So we need to make an analogy to smooth error and form a grid. As suggested in [9], we merely consider smooth error to be error not effectively reduced by the relaxation method. Furthermore, we form a grid using the adjacency graph of the matrix.

Determining a good coarse grid is then done using the concept of connection strength. The following definition for connection strength is given in terms of strong dependence and can be found for example in [17]. This definition considers a linear system of the form (1.9) where $a_{i,j}$ is the value in the i^{th} row and j^{th} column of $\mathcal{A}$.

Definition 1. *Given a threshold value* $0 < \theta \leq 1$, *the variable* x_i *strongly depends on the variable* x_j *if*

$$-a_{ij} \geq \theta \max_{k \neq i} \{-a_{ik}.\} \tag{1.12}$$

Following the description in [9], to form a coarse grid consider any fine grid point j that strongly depends on another fine grid point i. Now j is either a coarse grid point or strongly dependent on a coarse grid point. Additionally, the set of coarse grid point should be maximal and have as few as possible coarse grid points that strongly depend on each other. More details on how this is done can be found for example in [9, 17].

1.6 Outline

The remainder of this thesis is organized as follows. In Chapter 2, the details of the specific finite element method used are discussed. This includes a variational formulation and the specific finite element used. Then in Chapter 3, the preconditioning technique is discussed and some numerical results are provided. Finally, in Chapter 4 conclusions and future work are provided.

Chapter 2

Mixed Finite Element Formulation

Recall that we use the mixed finite element method proposed by Yi [60] with the total stress tensor, displacement, flux, and pore pressure, $(\widetilde{\boldsymbol{\sigma}}, \boldsymbol{u}, \boldsymbol{q}, p)$ as its primary unknowns.

2.1 Notation and Spaces

Here we will define some function spaces and their associated norms that will be used to define a mixed variational formulation. We will make use of the Sobolev space, $(H^m(D))^2$, with the associates semi-norm, $|\cdot|_{m,D}$, and norm, $\|\cdot\|_{m,D}$. In the case where $D = \Omega$, the subscript D will be dropped. This space requires a function and all of its partial derivatives of degree m or lower to be square integrable. For example,

$$(H^1(D))^2 = \{v \in L^2(D) : \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y} \in L^2(D)\}.$$

Note that in the case where $m = 0$, $(H^0(D))^2$ is exactly $(L^2(D))^2$. In this special case, the inner product and norm will be denoted by $(\cdot,\cdot)_D$ and $\|\cdot\|_D$ respectively. More specifically, for any $f, g \in L^2(D)$

$$(f, g)_D = \int_D fg\, dx.$$

For simplicity, when $D = \Omega$, the subscript will be omitted. We will also use the notation

$$< f, g >_d = \int_d fg\, ds$$

when d is a line or boundary. In the interest of defining a weak formulation, the following Hilbert spaces will be needed. Let

$$H(\mathrm{div}; \Omega) = \{\boldsymbol{z} \in (L^2(\Omega))^2 : \nabla \cdot \boldsymbol{z} \in L^2(\Omega)\}$$

with the norm $\|\boldsymbol{z}\|_{H(\mathrm{div})} = (\|\boldsymbol{z}\|_0^2 + \|\nabla \cdot \boldsymbol{z}\|_0^2)^{\frac{1}{2}}$ and its subspaces

$$H_{0,\Gamma_f}(\mathrm{div};\Omega) = \{\boldsymbol{z} \in H(\mathrm{div};\Omega) : \boldsymbol{z} \cdot \mathbf{n}|_{\Gamma_f} = 0\}$$

and

$$H_{\Gamma_f}(\mathrm{div};\Omega) = \{\boldsymbol{z} \in H(\mathrm{div};\Omega) : \boldsymbol{z} \cdot \mathbf{n}|_{\Gamma_f} = q_0\}.$$

Additionally, let

$$\boldsymbol{H}(\mathrm{div};\Omega) = \{\boldsymbol{\tau} \in (L^2(\Omega))^{2\times 2} : \nabla \cdot \boldsymbol{\tau} \in (L^2(\Omega))^2\}$$

with the associated norm $\|\boldsymbol{\tau}\|_{\boldsymbol{H}(\mathrm{div})} = (\|\boldsymbol{\tau}\|_0^2 + \|\nabla \cdot \boldsymbol{\tau}\|_0^2)^{\frac{1}{2}}$. When considering symmetric tensors only

$$\boldsymbol{H}^s(\mathrm{div};\Omega) = \{\boldsymbol{\tau} \in \boldsymbol{H}(\mathrm{div};\Omega) : \tau_{ij} = \tau_{ji}, i \leq i, j \leq 2\}$$

with the subspaces

$$\boldsymbol{H}^s_{0,\Gamma_t}(\mathrm{div};\Omega) = \{\boldsymbol{\tau} \in \boldsymbol{H}^s(\mathrm{div};\Omega) : \boldsymbol{\tau} \cdot \mathrm{n}|_{\Gamma_t} = 0\}.$$

and

$$\boldsymbol{H}^s_{\Gamma_t}(\mathrm{div};\Omega) = \{\boldsymbol{\tau} \in \boldsymbol{H}^s(\mathrm{div};\Omega) : \boldsymbol{\tau} \cdot \mathrm{n}|_{\Gamma_t} = \widetilde{\boldsymbol{\sigma}}_0\}.$$

2.2 Mixed Variational Formulation

Biot's consolidation model (1.1) is expanded to its mixed form by using the flux equation (1.4) in (1.1a), and the total stress (1.5) into (1.1b). In order for the formulation to be stable, the following relationship must also be used:

$$\nabla \cdot \boldsymbol{u} = \frac{1}{2\,(\lambda + \mu)} tr(\widetilde{\boldsymbol{\sigma}}) + \frac{\alpha}{\lambda + \mu} p. \tag{2.1}$$

Additionally, let $c_r = \frac{\alpha^2}{\lambda+\mu}$. The resulting system of equations is

$$\frac{\partial}{\partial t}\left((c_o + c_r)\, p + \frac{c_r}{2\alpha} tr(\widetilde{\boldsymbol{\sigma}})\right) - \nabla \cdot (\mathbf{K}\nabla p) = h, \tag{2.2a}$$

$$\mathbf{K}^{-1}\boldsymbol{q} + \nabla p = 0, \tag{2.2b}$$

$$\mathscr{A}\widetilde{\boldsymbol{\sigma}} = \epsilon(\boldsymbol{u}) - \frac{\alpha}{2\,(\lambda+\mu)} p\mathbf{I}, \tag{2.2c}$$

$$-\nabla \cdot \widetilde{\boldsymbol{\sigma}} = \boldsymbol{f}, \tag{2.2d}$$

where $\mathscr{A}$ is a fourth order compliance tensor that is bounded, symmetric positive definite uniformly with respect to $x \in \Omega$ and satisfies

$$\mathscr{A}\boldsymbol{\tau} = \frac{1}{2\mu}\left(\boldsymbol{\tau} - \frac{\lambda}{2(\lambda+\mu)} tr(\boldsymbol{\tau})\mathbf{I}\right). \tag{2.3}$$

Now, for simplicity, let $\Sigma = \boldsymbol{H}^s_{\Gamma_t}(\mathrm{div};\Omega)$, $\mathcal{U} = (L^2(\Omega))^2$, $\mathcal{V} = H_{\Gamma_f}(\mathrm{div};\Omega)$, and $\mathcal{W} = L^2(\Omega)$. The weak formulation then becomes to find $(\widetilde{\boldsymbol{\sigma}}, \boldsymbol{u}, \boldsymbol{q}, p) \in \Sigma \times \mathcal{U} \times \mathcal{V} \times \mathcal{W}$ subject to

$$(c_0 + c_r)\left(\frac{\partial}{\partial t} p, w\right) + \frac{c_r}{2\alpha}\left(\frac{\partial}{\partial t} tr(\widetilde{\boldsymbol{\sigma}}), w\right) + (\nabla \cdot \boldsymbol{q}, w) = (h, w), \qquad \forall w \in L^2(\Omega), \tag{2.4a}$$

$$\left(\mathbf{K}^{-1}\boldsymbol{q}, \mathbf{z}\right) - (p, \nabla \cdot \mathbf{z}) = - < \boldsymbol{z} \cdot \mathbf{n}, p_0 >_{\Gamma_p}, \qquad \forall \mathbf{z} \in H_{0,\Gamma_f}(\mathrm{div};\Omega), \tag{2.4b}$$

$$(\mathscr{A}\widetilde{\boldsymbol{\sigma}}, \boldsymbol{\tau}) + (\boldsymbol{u}, \nabla \cdot \boldsymbol{\tau}) + \frac{c_r}{2\alpha}(p, tr(\boldsymbol{\tau})) = - < \boldsymbol{\tau} \cdot \mathbf{n}, \boldsymbol{u}_0 >_{\Gamma_d}, \qquad \forall \boldsymbol{\tau} \in \boldsymbol{H}^s_{0,\Gamma_t}(\mathrm{div};\Omega), \tag{2.4c}$$

$$(\nabla \cdot \widetilde{\boldsymbol{\sigma}}, \boldsymbol{v}) = -(\boldsymbol{f}, \boldsymbol{v}), \qquad \forall \boldsymbol{v} \in (L^2(\Omega))^2. \tag{2.4d}$$

2.3 Fully-discrete model

To discretize the weak formulation (2.4) we will first need to make use of finite dimensional approximations to the spaces described above. That is, we need subspaces Σ_h, $\mathcal{U}_h$, $\mathcal{V}_h$ and $\mathcal{W}_h$ of Σ, $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ respectively. It is pointed out by Yi [60] that pre-existing stable finite element pairs may be employed. To be stable, the chosen pairs must satisfy the

Ladyzhenskaya-Babuska-Brezzi condition, details about which can be for example in [8]. Specifically, the pair $\Sigma_h \times \mathcal{U}_h$ must be a stable mixed finite element pair based on Hellinger-Reissner formulation for elasticity [26]. The pair $\mathcal{W}_h \times \mathcal{V}_h$ can be any stable mixed finite element pair for the second order elliptic problem [20]. In order to fully discretize the problem, the backward Euler time stepping is applied to (2.2). To simplify the notation, let $\Delta t = \frac{T}{N}$, where N is a positive integer and T is the final time. We assume uniform time steps and let $t^n = n\Delta t$, where n is used to indicate the time step. Now at each time step $t = t^n$, we seek to find $(\widetilde{\boldsymbol{\sigma}}_h^n, \boldsymbol{u}_h^n, \boldsymbol{q}_h^n, p_h^n) \in \Sigma_h \times \mathcal{U}_h \times \mathcal{V}_h \times \mathcal{W}_h$ where

$$(c_0 + c_r)\left(\frac{p_h^n - p_h^{n-1}}{\Delta t}, w\right) + \frac{c_r}{2\alpha}\left(\frac{tr\left(\widetilde{\boldsymbol{\sigma}}_h^n - \widetilde{\boldsymbol{\sigma}}_h^{n-1}\right)}{\Delta t}, w\right) + (\nabla \cdot \boldsymbol{q}_h^n, w) = (h^n, w), \quad \forall w \in \mathcal{W}_h, \tag{2.5a}$$

$$\left(\mathbf{K}^{-1}\boldsymbol{q}_h^n, \mathbf{z}\right) - (p_h^n, \nabla \mathbf{z}) = - < \boldsymbol{z} \cdot \mathbf{n}, p_0 >_{\Gamma_p}, \qquad \forall \mathbf{z} \in \mathcal{V}_h, \tag{2.5b}$$

$$(\mathscr{A}\widetilde{\boldsymbol{\sigma}}_h^n, \boldsymbol{\tau}) + (\boldsymbol{u}_h^n, \nabla \cdot \boldsymbol{\tau}) + \frac{c_r}{2\alpha}(p_h^n, tr(\boldsymbol{\tau})) = - < \boldsymbol{\tau} \cdot \mathbf{n}, \boldsymbol{u}_0 >_{\Gamma_d}, \qquad \forall \boldsymbol{\tau} \in \Sigma_h, \tag{2.5c}$$

$$(\nabla \cdot \widetilde{\boldsymbol{\sigma}}_h^n, \boldsymbol{v}) = -(\boldsymbol{f}^n, \boldsymbol{v}), \qquad \forall \boldsymbol{v} \in \mathcal{U}_h. \tag{2.5d}$$

By writing the solution in terms of its finite element basis functions and rearranging, we are able to write this system in matrix form. Although $\mathcal{W}_h \times \mathcal{V}_h$ and $\Sigma_h \times \mathcal{U}_h$ can be any stable mixed finite elements for the flow and elasticity problems respectively, we have used the Raviart-Thomas space [46] of order 1, and the lowest order space defined by Chen and Wang [12]. Let, $\widetilde{\boldsymbol{\sigma}}_h(t, \boldsymbol{x}) = \Sigma_j \widetilde{\boldsymbol{\sigma}}_j(t)\phi_{\widetilde{\boldsymbol{\sigma}},j}$, $\boldsymbol{u}_h(t, \boldsymbol{x}) = \Sigma_j \boldsymbol{u}_j(t)\phi_{\boldsymbol{u},j}$, $\boldsymbol{q}_h(t, \boldsymbol{x}) = \Sigma_j \boldsymbol{q}_j(t)\phi_{\boldsymbol{q},j}$ and $p_h(t, \boldsymbol{x}) = \Sigma_j p_j(t)\phi_{p,j}$. Let $\{\phi_{g,j}\}$, with $g = \widetilde{\boldsymbol{\sigma}}, \boldsymbol{u}, \boldsymbol{q}, p$, represent the basis functions for each space and let $\{g_j\}$ represent coefficients.

The resulting saddle point system is defined by

$$\mathcal{A}x = b, \tag{2.6}$$

where

$$\mathcal{A} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{C} \end{pmatrix}$$

with

$$\boldsymbol{A} = \begin{pmatrix} C_{pp} & C_{\boldsymbol{\sigma}p}^T \\ C_{\boldsymbol{\sigma}p} & C_{\boldsymbol{\sigma}\boldsymbol{\sigma}} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \Delta t C_{\boldsymbol{q}p} & 0 \\ 0 & C_{\boldsymbol{u}\boldsymbol{\sigma}} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} \Delta t C_{\boldsymbol{q}\boldsymbol{q}} & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$x = \begin{pmatrix} p_h^n \\ \widetilde{\boldsymbol{\sigma}}_h^n \\ \boldsymbol{q}_h^n \\ \boldsymbol{u}_h^n \end{pmatrix}, \quad b = \begin{pmatrix} \Delta t \mathbf{h}_p^n + C_{pp} p_h^{n-1} + C_{\boldsymbol{\sigma}p} \widetilde{\boldsymbol{\sigma}}_h^{n-1} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{F}_{\boldsymbol{u}}^n \end{pmatrix}.$$

In the above equations, $\boldsymbol{h}_p^n$ is a vector whose j^{th} component is $(\boldsymbol{h}_p^n)_j = (h(n\Delta t), \phi_{p,j})$. Similarly, $(\boldsymbol{F}_{\boldsymbol{u}}^n)_j = (\boldsymbol{f}(n\Delta t), \phi_{\boldsymbol{u},j})$. We will be able to prove a number of things about these matrices in the following sections and chapters.

2.4 Approximation Spaces

For the purposes of numerical testing, we have chosen to use a grid of square elements. The basis functions are formed on a reference square $K = [-\frac{h}{2}, \frac{h}{2}] \times [-\frac{h}{2}, \frac{h}{2}]$. The orientations of the edges and the normal vectors $\mathbf{n}_i$ with $i = 1, 2, 3, 4$ for finding the basis functions are defined as shown in Figure 2.1 below.

2.4.1 Mixed Finite Elements for Elasticity

For the elasticity subproblem, we have chosen to use the pair defined by Chen and Wang [12] for $\Sigma_h \times \mathcal{U}_h$. This will require 17 degrees of freedom for stress and 4 degrees of freedom for displacement and can be viewed as a extension of the non-conforming finite elements developed in [29, 59]. We will form basis functions for $\widetilde{\boldsymbol{\sigma}}_h$ and $\boldsymbol{u}$.

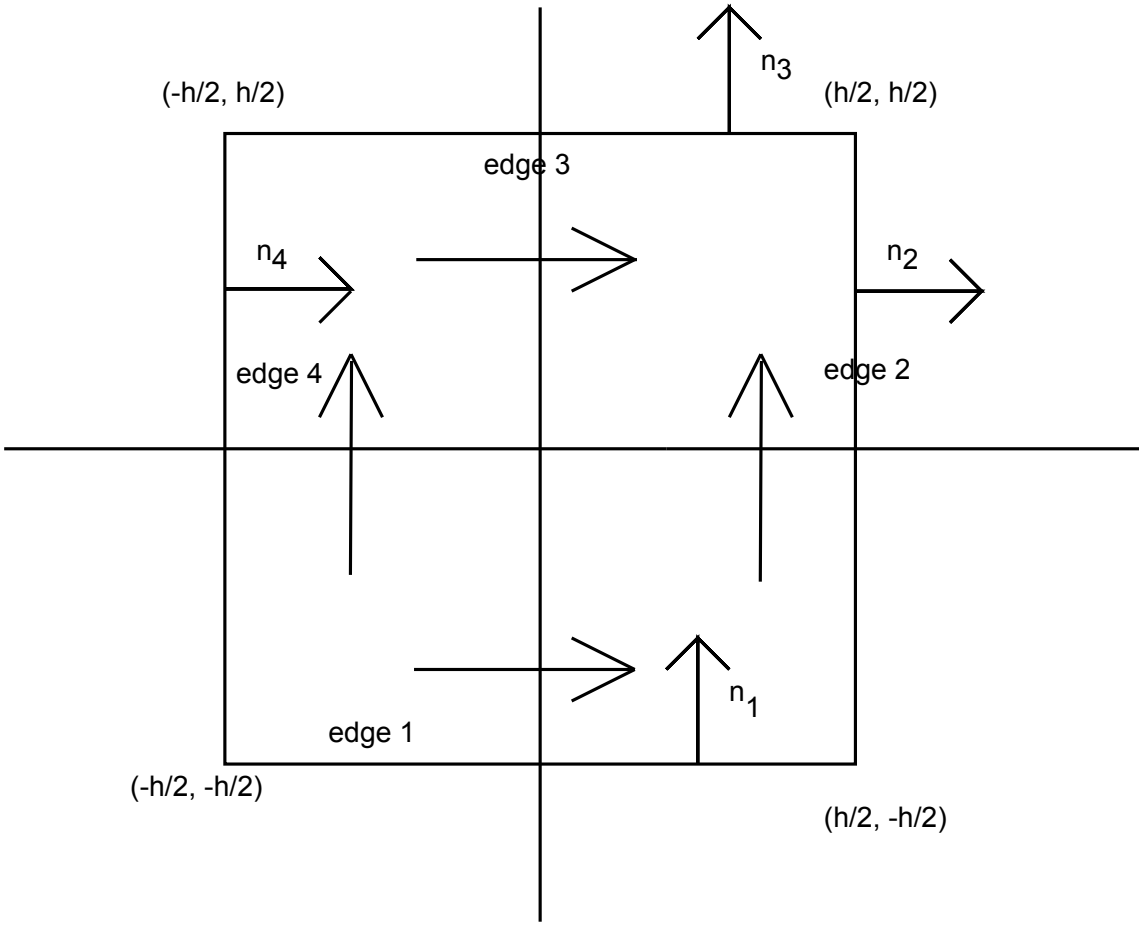

Figure 2.1: The reference element

The basis for the vector $\boldsymbol{u}_h$ is simply $(1,0)$, $(\frac{2(y-y_c)}{h},0)$, $(0,1)$ and $(0,\frac{2(x-x_c)}{h})$, where $c=(x_c,y_c)$ is the center of a particular element. However for a symmetric tensor $\boldsymbol{\tau}\in\Sigma_h$ the basis functions are more complicated. Let $\boldsymbol{\tau}=\begin{pmatrix}\tau_{11} & \tau_{12}\\ \tau_{12} & \tau_{22}\end{pmatrix}$. Recall the reference element, K, shown in Figure 2.1 and let each corner or node be represented by a_i with $i=1,2,3,4$. Let the node a_1 be the node in the lower left corner and proceed numbering the nodes in a counter clockwise fashion. Additionally, $\mathbf{n}_i$ and $\mathbf{t}_i$ will be the normal and tangential vectors along each edge i respectively. Then the degrees of freedom (dofs) defined on K by Chen and Wang [12] are

- the first two moments of $\tau\mathbf{n}_i\cdot\mathbf{n}_i$ on each edge (8 dofs),
- the first moment of $\tau\mathbf{n}_i\cdot\mathbf{t}_i$ on each edge (4 dofs),
- the value of τ_{12} at each node a_i (4 dofs),
- the value of $\int_K \tau_{12}\,dx$ (1 dof).

We will perform convergence studies making use of the norms

$$\|s-s_h\|_{L^\infty(L^2)}=\max_{1\le n\le N}\|s^n-s_h^n\|_0$$

and

$$\|s-s_h\|_{L^2(L^2)}=\left(\Delta t\sum_{n=1}^{N}\|s^n-s_h^n\|_0^2\right)^{\frac{1}{2}}.$$

Based on the analysis of Yi [60] and Chen and Wang [12], we will measure the errors $\|\boldsymbol{u}-\boldsymbol{u}_h\|_{L^\infty(L^2)}$ and $\|\widetilde{\boldsymbol{\sigma}}-\widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$ and the optimal convergence rates, using these norms, for displacement and total stress are 1 and 2 respectively.

2.4.2 Mixed Finite Elements for Flow

In this case, we choose to use the Raviart-Thomas space of index 1 [46] so that the convergence rates for flux and pressure , $\boldsymbol{q}$ and p, will match that of the stress. This chosen

pair for $\mathcal{W}_h \times \mathcal{V}_h$ will require 12 degrees of freedom for flux and 4 degrees of freedom for pressure. Additional details about this pair can be found for example in [32, 27].

As for the elasticity subproblem, the basis functions will be formed on the reference square, K, shown in Figure 2.1. The basis functions for p will be taken as 1, $\frac{2(x-x_c)}{h}$, $\frac{2(y-y_c)}{h}$ and $\frac{2(x-x_c)}{h}\frac{2(y-y_c)}{h}$. For the flux, consider $\boldsymbol{z} \in \mathcal{V}_h$, the degrees of freedom are

- the first two moments of $\boldsymbol{z} \cdot \mathbf{n}_i$ on each edge (8 dofs),
- $\int_K \boldsymbol{z} \cdot \boldsymbol{s}_k \, k = 1, 2, 3, 4$ with $\boldsymbol{s}_1 = (1, 0)$, $\boldsymbol{s}_2 = (0, 1)$, $\boldsymbol{s}_3 = (y, 0)$ and $\boldsymbol{s}_4 = (0, x)$ (4 dofs).

Based on the analysis of Yi [60] and Raviart and Thomas [46], we will measure the errors $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$ and $\|p - p_h\|_{L^\infty(L^2)}$ and the optimal convergence rates, using these norms, for flux and pressure are both 2.

2.5 An Interface Problem

In several of the following numerical experiments, we will attempt to model the case of transport though different layers of media. This will result in discontinuities in some of the physical parameters. At the interface between two layers, certain continuity conditions must link the solutions in the given layers.

For simplicity, consider a domain consisting of two regions, $\Omega = \Omega_1 \cup \Omega_2$, with different properties. These regions are separated by an interface Γ_1. Furthermore, we define the notation of a jump at a point $\psi \in \Gamma_1$

$$[s] = \lim_{x \to \psi^+} s(x) - \lim_{x \to \psi^-} s(x).$$

Certain physical assumptions are required to obtain the interface conditions. First, we will assume that the regions are in perfect hydraulic contact, thus

$$[p] = 0, \quad x \in \Gamma_1.$$

That is, the pressure must be continuous across an interface. Additionally, we will assume that no solid mass is moving across this interface and that the subdomains do not slip with

respect to one another. Formally these assumptions give

$$[\boldsymbol{u}] = 0, \quad x \in \Gamma_1.$$

From mass conservation of the fluid phase, we obtain

$$[\boldsymbol{q} \cdot \mathbf{n}] = 0, \quad x \in \Gamma_1.$$

In other words, the fluid flux across an interface is continuous. Finally, the assumption that total stress should be conserved across the interface gives

$$[\widetilde{\boldsymbol{\sigma}} \cdot \mathbf{n}] = 0, \quad x \in \Gamma_1.$$

These interface conditions can be derived directly from Biot's model, as shown in [23]. This is done by replacing the discontinuity by a thin transition layer in which the properties of the material are allowed to changed smoothly and rapidly. The limit is then taken as this layer's thickness approaches zero.

For these numerical tests, we will use a grid that is fitted to the interface. That is, the interface will fall along the edge of grid elements. In this case, it is a requirement of the mixed finite element space $\mathcal{V}_h$, that the normal component of the fluid flow is continuous [61]. Additionally, for the space defined by Chen and Wang [12], $\widetilde{\boldsymbol{\sigma}} \cdot \mathbf{n}$ is required to be continuous across mesh edges. It is worthwhile to note that these requirements match the continuity requirements of Biot's model.

2.6 Numerical Experiments

2.6.1 Mandel's Problem

Mandel's problem [35] models the case of a 2D sample of unsaturated poroelastic material that is being compressed by two impervious plates. This is used as a test problem because it is one of the few example problems with an analytic solution, see for example [1]. We consider a sample that is $2a$ in height and $2b$ in width with the origin of our axis placed in

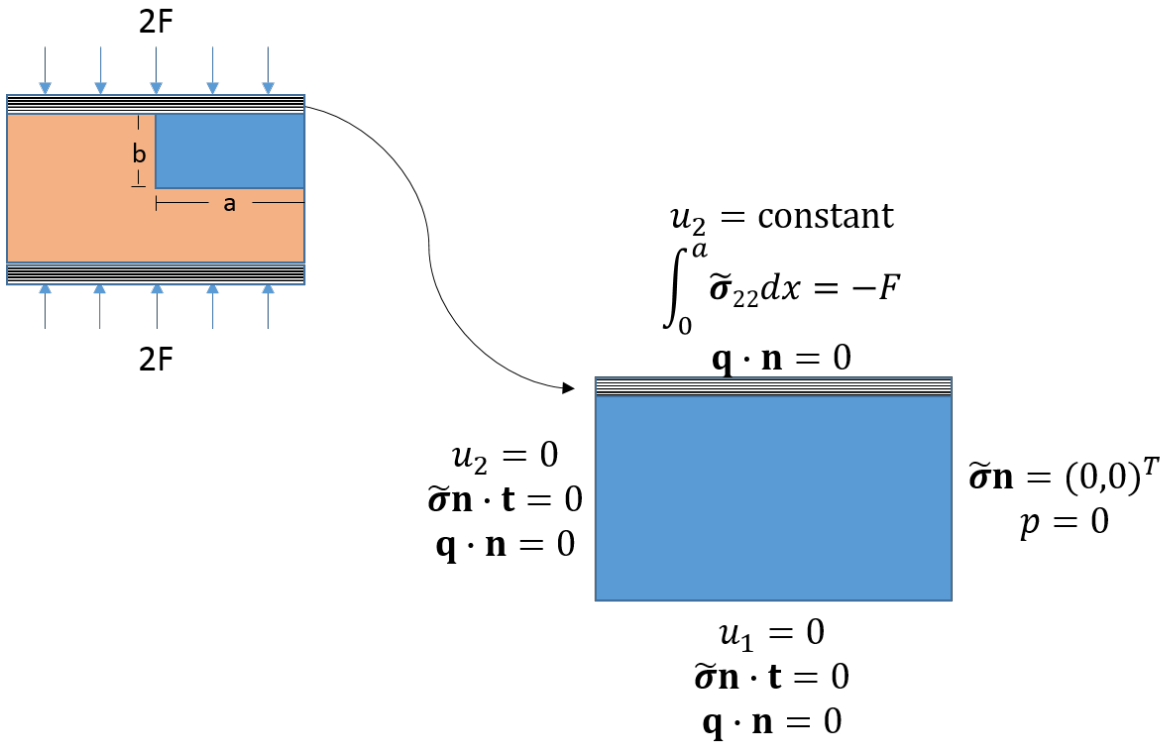

Figure 2.2: The boundary conditions and computational domain for Mandel's problem.

the center of the sample. Then plates are located at $y = \pm b$ and each exerts a force of $2F$, beginning instantaneously at $t = 0^+$. The sample is allowed to drain laterally at $x = \pm a$. The resulting boundary conditions are shown in Figure 2.2. For simplicity in the numerical computations, we will replace the so called impervious plate conditions, $\int_0^a = \widetilde{\boldsymbol{\sigma}}_{22} dx = -F$ and $u_2 =$ constant, with the condition that $u_2 = U_2(b, t)$ on $y = b$ as is done in [37]. Here U is the exact known solution. Additionally, the symmetry of the problem allows us to reduce the computational domain to only one quarter [42].

To provide the analytic solutions to this problem, some notation will be useful. We follow the notation provided in [42]. Let the skeleton bulk modulus K and Poisson's coefficient ν be

$$K = \lambda + \frac{2}{3}\mu$$

and

$$\nu = \frac{3K - 2\mu}{2\left(3K + \mu\right)}$$

respectively. Then the undrained versions of K becomes

$$K_u = K + \frac{\alpha^2}{c_0}.$$

Addtionally, Skempton's coefficient is

$$B = \frac{\alpha}{c_0 K_u},$$

and the diffusivity coefficent is

$$c_f = \frac{1}{c_0} k \frac{K + \frac{4}{3}\mu}{K_u + \frac{4}{3}\mu}$$

where k is the permeability constant. Finally, the undrained version of ν_u is

$$\nu_u = \frac{3\nu + \alpha B\left(1 - 2\nu\right)}{3 - \alpha B\left(1 - 2\nu\right)}$$

Now recall that $\boldsymbol{u} = (u_1, u_2)^T$ and $\widetilde{\boldsymbol{\sigma}} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$. Below, we will provide the

analytic solutions found in [1].

$$p = \frac{2FB(1+\nu_u)}{3a} \sum_{n=1}^{\infty} \frac{\sin\alpha_n}{\alpha_n - \sin\alpha_n \cos\alpha_n} \left(\cos\frac{\alpha_n x}{a} - \cos\alpha_n\right) e^{\frac{-\alpha_n^2 c_f t}{a^2}}$$

$$u_1 = \left(\frac{F\nu}{2\mu a} - \frac{F\nu_u}{\mu a} \sum_{n=1}^{\infty} \frac{\sin\alpha_n \cos\alpha_n}{\alpha_n - \sin\alpha_n \cos\alpha_n} e^{\frac{-\alpha_n^2 c_f t}{a^2}}\right) x + \frac{F}{\mu} \sum_{n=1}^{\infty} \frac{\cos\alpha_n}{\alpha_n - \sin\alpha_n \cos\alpha_n} \sin\frac{\alpha_n x}{a} e^{\frac{-\alpha_n^2 c_f t}{a^2}}$$

$$u_2 = \left(\frac{-F(1-\nu)}{2\mu a} + \frac{F(1-\nu_u)}{\mu a} \sum_{n=1}^{\infty} \frac{\sin\alpha_n \cos\alpha_n}{\alpha_n - \sin\alpha_n \cos\alpha_n} e^{\frac{-\alpha_n^2 c_f t}{a^2}}\right) y$$

$$\sigma_{22} = -\frac{F}{a} - \frac{2F(\nu_u - \nu)}{a(1-\nu)} \sum_{n=1}^{\infty} \frac{\sin\alpha_n}{\alpha_n - \sin\alpha_n \cos\alpha_n} \cos\frac{\alpha_n x}{a} e^{\frac{-\alpha_n^2 c_f t}{a^2}} + \frac{2F}{a} \sum_{n=1}^{\infty} \frac{\sin\alpha_n \cos\alpha_n}{\alpha_n - \sin\alpha_n \cos\alpha_n} e^{\frac{-\alpha_n^2 c_f t}{a^2}}$$

Note that $\sigma_{11} = \sigma_{12} = 0$. In these equations, α_n are the positive solutions to

$$\tan\alpha_n = \frac{1-\nu}{\nu_u - \nu}\alpha_n.$$

The values of α_n are found numerically.

Figure 2.3 shows some solution profiles for a numerical experiment using Mandel's problem on the domain $\Omega = (0,1) \times (0,1)$. For this experiment we used the parameters

$$F = 500, \quad k = 10^{-4}, \quad E = 10^5, \quad \text{and} \quad \nu = 0.2.$$

These figures show the solution at a final time of $T = 1$ and are viewed at a constant $y = 0$. The analytic solutions for the shown variables do not depend on y. The exact solution is shown as a solid black line. The approximate solution, shown as the orange dashed line with square markers, is obtained by using a 16×16 grid of elements and $\Delta t = \frac{1}{128}$. Clearly, the approximation is close to the analytic solution.

Results of convergence studies using Mandel's problem are provided in Tables 2.1 and 2.2. The result show suboptimal convergence rates. Yi [60] points out that this is due to the lack of the required regularity in the solution and its time derivatives. Phillips and Wheeler [42] have shown that $p \in L^2(H^{\frac{3}{2}+\epsilon})$, $0 \leq \epsilon << 1$.

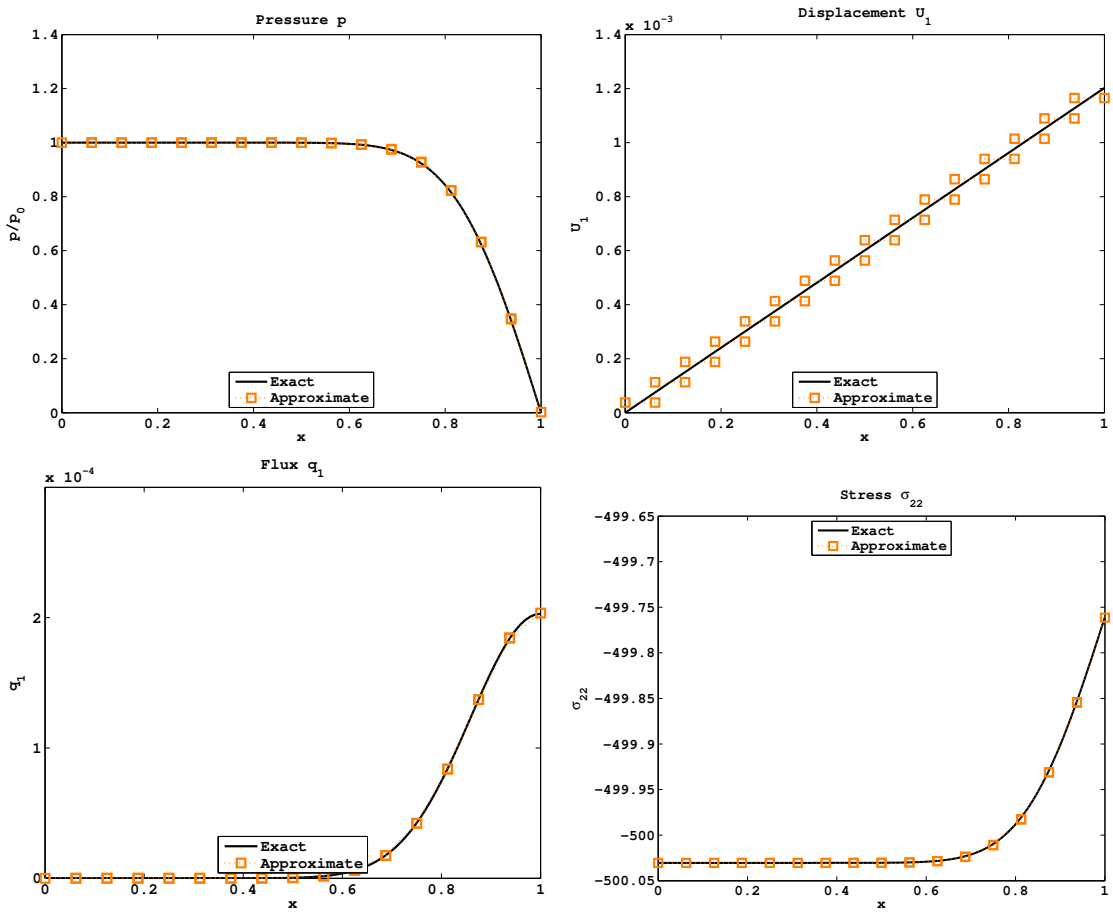

Figure 2.3: Solution profiles for Mandel's problem at $T = 1$.

Table 2.1: Convergence study for Mandel's problem with $\Delta t = \frac{1}{5}h$.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{10}$	1.2935824E-01		7.1390506E-04	
$\frac{1}{4}$	$\frac{1}{20}$	9.2572062E-02	0.48	3.5695126E-04	1.00
$\frac{1}{8}$	$\frac{1}{40}$	6.4813760E-02	0.51	1.7847549E-04	1.00
$\frac{1}{16}$	$\frac{1}{80}$	4.1292532E-02	0.65	8.9237728E-05	1.00
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{10}$	3.4876502E-02		2.7944956E-05	
$\frac{1}{4}$	$\frac{1}{20}$	2.4958516E-02	0.48	1.7485758E-05	0.68
$\frac{1}{8}$	$\frac{1}{40}$	1.7474552E-02	0.51	1.1495493E-05	0.61
$\frac{1}{16}$	$\frac{1}{80}$	1.1132970E-02	0.65	9.1124523E-06	0.34

Table 2.2: Convergence study for Mandel's problem with $\Delta t = 2h^2$.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	1.2682822E-01		7.1390506E-04	
$\frac{1}{4}$	$\frac{1}{8}$	8.9686731E-02	0.50	3.5695126E-04	1.00
$\frac{1}{8}$	$\frac{1}{32}$	6.3417547E-02	0.50	1.7847549E-04	1.00
$\frac{1}{16}$	$\frac{1}{128}$	4.4842773E-02	0.50	8.9237728E-05	1.00
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	3.4194395E-02		2.1271789E-05	
$\frac{1}{4}$	$\frac{1}{8}$	2.4180608E-02	0.50	1.6374353E-05	0.38
$\frac{1}{8}$	$\frac{1}{32}$	1.7098116E-02	0.50	1.1651197E-05	0.49
$\frac{1}{16}$	$\frac{1}{128}$	1.2089601E-02	0.50	8.2464297E-06	0.50

2.6.2 Example 1

In order to perform error analysis, we have modified a three dimensional problem found in [40]. For simplicity, we make up a non-physical example problem. Let the exact solution be

$$p = \frac{t}{k}\sin(\pi x)\sin(\pi y), \tag{2.7}$$

$$u_1 = \frac{t}{\mu}\sin(\pi x)\sin(\pi y), \tag{2.8}$$

$$u_2 = \frac{t}{\lambda + 2\mu}\sin(\pi x)\sin(\pi y). \tag{2.9}$$

Solutions for the other variables, the boundary conditions and the initial conditions are then found from these known solutions. For these experiments, the simple domain $\Omega = (0,1) \times (0,1)$ was chosen. Therefore the boundary conditions are

$$p_0 = 0 \quad \text{and} \quad \boldsymbol{u}_0 = (0,0)^T \quad \text{on} \quad x = 0,1 \quad \& \quad y = 0,1.$$

Additionally, we are able to use

$$p(0) = 0 \quad \text{and} \quad \boldsymbol{u}(0) = (0,0)^T \quad \text{in} \quad \Omega \quad \text{at} \quad t = 0.$$

For simplicity, we use $\lambda = \mu = k = 1$.

For reference, Figure 2.4 shows a few example plots. In the left column of the figure, we show the approximate solution profiles for pressure p and the first term of displacement u_1. These approximations are obtained using a 32×32 spatial grid and $\Delta t = \frac{1}{512}$. For comparison, the exact solution profiles obtained from the solution (2.7). We can see that that approximation profiles are close to the expected profiles.

The results of a convergence study on this example problem are show in Table 2.3. The analytic solution in this case is designed to meet the regularity requirements and we are seeing optimal convergence rates in this case.

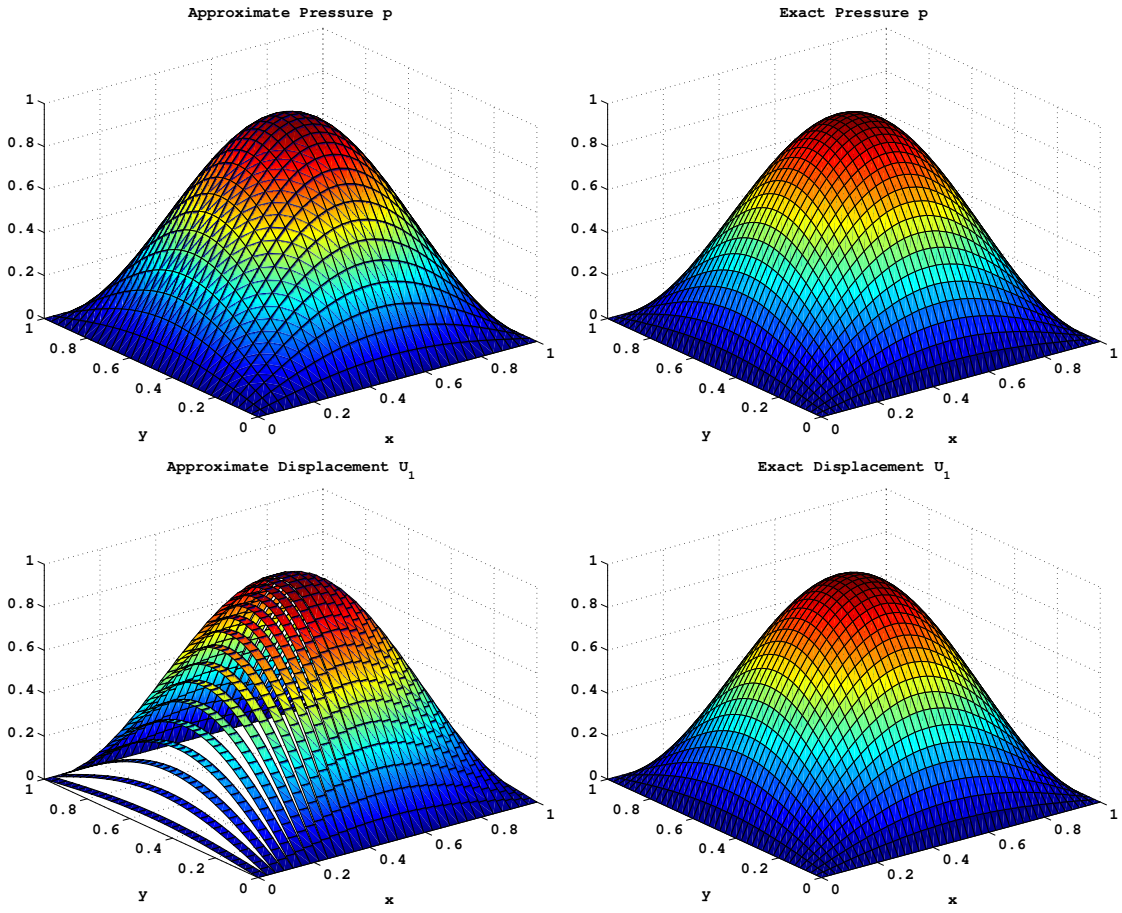

Figure 2.4: Solution profiles for Example 1 at $T = 1$.

Table 2.3: Convergence study for Example 1 with $\Delta t = 2h^2$.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	6.3030484E-02		1.6558077E-01	
$\frac{1}{4}$	$\frac{1}{8}$	1.6177030E-02	1.96	3.3775281E-02	2.29
$\frac{1}{8}$	$\frac{1}{32}$	4.0715700E-03	1.99	7.9803303E-03	2.08
$\frac{1}{16}$	$\frac{1}{128}$	1.0196063E-03	2.00	1.9666771E-03	2.02
$\frac{1}{32}$	$\frac{1}{512}$	2.5500876E-04	2.00	4.8991705E-04	2.00
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	2.4863139E-01		1.3808658E+00	
$\frac{1}{4}$	$\frac{1}{8}$	1.2158968E-01	1.03	3.5881979E-01	1.94
$\frac{1}{8}$	$\frac{1}{32}$	6.0040844E-02	1.02	9.0402656E-02	1.99
$\frac{1}{16}$	$\frac{1}{128}$	2.9911355E-02	1.01	2.2640515E-02	2.00
$\frac{1}{32}$	$\frac{1}{512}$	1.4941564E-02	1.00	5.6625322E-03	2.00

2.6.3 Example 2

Using again the example problems from [40], we constructed a second example that could be tested in a layered media on the domain $\Omega = (0,1) \times (0,1)$. It is constructed so that all of the interface conditions described in Section 2.5 will be satisfied at one interface, $y = 0.5$, regardless of the coefficient values chosen. Let the exact solution be

$$p = \begin{cases} \frac{t^2}{k_1} \sin(\pi x) \sin(\pi y)(y - 0.5), & y < 0.5, \\ \frac{t^2}{k_2} \sin(\pi x) \sin(\pi y)(y - 0.5), & y > 0.5, \end{cases} \tag{2.10}$$

$$u_1 = \begin{cases} \frac{t^2}{\mu_1} x\,(1 - x) \sin(\pi y) \sin(0.5 - y), & y < 0.5, \\ \frac{t^2}{\mu_2} x\,(1 - x) \sin(\pi y) \sin(0.5 - y), & y > 0.5, \end{cases} \tag{2.11}$$

$$u_2 = \begin{cases} \frac{t^2}{\lambda_1 + 2\mu_1} \sin(\pi x) y\,(1 - y) \sin(0.5 - y), & y < 0.5, \\ \frac{t^2}{\lambda_2 + 2\mu_2} \sin(\pi x) y\,(1 - y) \sin(0.5 - y), & y > 0.5. \end{cases} \tag{2.12}$$

Again the boundary and initial conditions are calculated from the known solution.

We will allow the coefficients, λ, μ, and $\mathbf{K}$ to be discontinuous at the interface Γ_1. These are considered to be piecewise constant

$$\mathbf{K}(y) = \begin{cases} k_1 I, & y < 0.5, \\ k_2 I, & y > 0.5, \end{cases} \quad \lambda(y) = \begin{cases} \lambda_1, & y < 0.5, \\ \lambda_2, & y > 0.5, \end{cases} \quad \mu(y) = \begin{cases} \mu_1, & y < 0.5, \\ \mu_2, & y > 0.5. \end{cases} \tag{2.13}$$

To numerically verify that the presence of a layer is not impacting the convergence, we performed a convergence study for several different layers. First, we use the case of no layer (i.e. $k_1 = k_2 = 1$, $\lambda_1 = \lambda_2 = 1$ and $\mu_1 = \mu_2 = 1$). The results of this trial are recorded in Table 2.4. The expected convergence rates are observed for each variable. We have also tested the case where the coefficients vary by several orders of magnitude. Table 2.5 shows results from the case where

$$k_1 = 1, \quad k_2 = 10^3, \quad \lambda_1 = 1, \quad \lambda_2 = 10^3, \quad \mu_1 = 1, \quad \text{and} \quad \mu_2 = 10^3.$$

Figure 2.5 shows a solution profile corresponding to this case. Since it can be difficult to compare three dimensional graphs, we have fixed $x = 0.5$ and shown the resulting curves

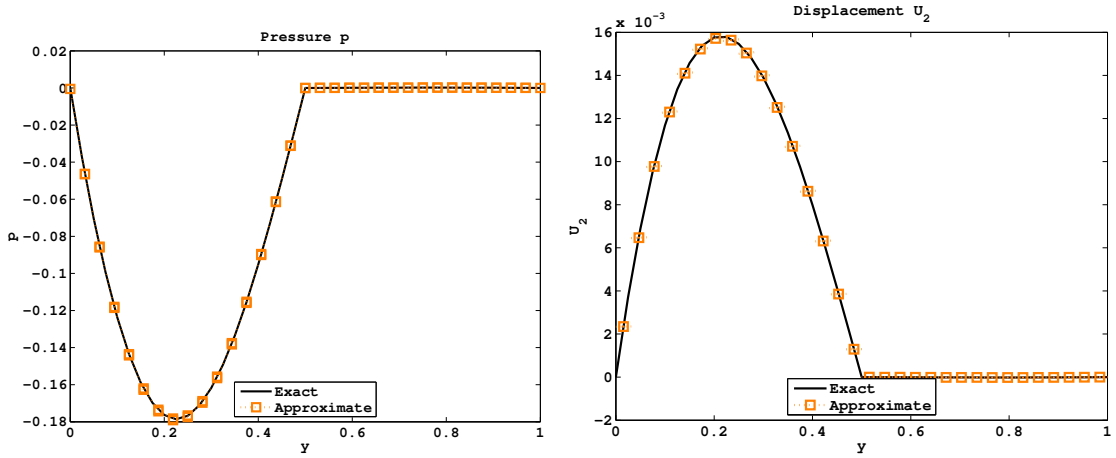

Figure 2.5: Solution profiles for Example 2 at $T = 1$. Here $x = 0.5$ is fixed and the difference in the material parameters of layers is 3 orders of magnitude.

Table 2.4: Convergence study for Example 2 in the case of no layer.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	3.9058672E-02		1.0172006E-01	
$\frac{1}{4}$	$\frac{1}{8}$	9.4472596E-03	2.05	1.6877536E-02	2.59
$\frac{1}{8}$	$\frac{1}{32}$	2.3591107E-03	2.00	3.7369149E-03	2.18
$\frac{1}{16}$	$\frac{1}{128}$	5.8981375E-04	2.00	9.0477807E-04	2.05
$\frac{1}{32}$	$\frac{1}{512}$	1.4745866E-04	2.00	2.2436515E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	1.4754977E-02		1.6824911E-01	
$\frac{1}{4}$	$\frac{1}{8}$	6.7733696E-03	1.12	4.1147780E-02	2.03
$\frac{1}{8}$	$\frac{1}{32}$	3.2707531E-03	1.05	1.0330064E-02	1.99
$\frac{1}{16}$	$\frac{1}{128}$	1.6182793E-03	1.02	2.5863199E-03	2.00
$\frac{1}{32}$	$\frac{1}{512}$	8.0690765E-04	1.00	6.4681799E-04	2.00

Table 2.5: Convergence study for Example 2 in the case of 3 orders of magnitude difference in the coefficients between the layers.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	2.7567381E-02		1.0091838E-01	
$\frac{1}{4}$	$\frac{1}{8}$	6.6746193E-03	2.05	1.6801630E-02	2.59
$\frac{1}{8}$	$\frac{1}{32}$	1.6671570E-03	2.00	3.7229061E-03	2.17
$\frac{1}{16}$	$\frac{1}{128}$	4.1684354E-04	2.00	9.0153093E-04	2.05
$\frac{1}{32}$	$\frac{1}{512}$	1.0421643E-04	2.00	2.2356821E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	9.6238698E-03		1.7274341E-01	
$\frac{1}{4}$	$\frac{1}{8}$	4.6302871E-03	1.06	4.0653090E-02	2.09
$\frac{1}{8}$	$\frac{1}{32}$	2.2905633E-03	1.02	1.0089306E-02	2.01
$\frac{1}{16}$	$\frac{1}{128}$	1.1414428E-03	1.00	2.5199626E-03	2.00
$\frac{1}{32}$	$\frac{1}{512}$	5.7021091E-04	1.00	6.2990499E-04	2.00

as a function of y. We can see that the approximate solutions, shown as orange squares, match the analytic solutions, shown as black lines, quite well.

Next, Table 2.6 shows the case where

$$k_1 = 1, \quad k_2 = 10^6, \quad \lambda_1 = 1, \quad \lambda_2 = 10^6, \quad \mu_1 = 1, \quad \text{and} \quad \mu_2 = 10^6.$$

Finally, in Table 2.7 the coefficients

$$k_1 = 1, \quad k_2 = 10^9, \quad \lambda_1 = 1, \quad \lambda_2 = 10^9, \quad \mu_1 = 1, \quad \text{and} \quad \mu_2 = 10^9$$

are used. These tables show that the presence of large discontinuities between layers is not impacting the expected convergence rates. In fact, with jumps this large, there is little difference in the errors obtained.

Table 2.6: Convergence study for Example 2 in the case of 6 orders of magnitude difference in the coefficients between the layers.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	2.7567369E-02		1.0091765E-01	
$\frac{1}{4}$	$\frac{1}{8}$	6.6746149E-03	2.05	1.6801614E-02	2.59
$\frac{1}{8}$	$\frac{1}{32}$	1.6671557E-03	2.00	3.7229089E-03	2.17
$\frac{1}{16}$	$\frac{1}{128}$	4.1684321E-04	2.00	9.0153201E-04	2.05
$\frac{1}{32}$	$\frac{1}{512}$	1.0421635E-04	2.00	2.2356850E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	9.6240619E-03		1.7289930E-01	
$\frac{1}{4}$	$\frac{1}{8}$	4.6302326E-03	1.06	4.0671747E-02	2.09
$\frac{1}{8}$	$\frac{1}{32}$	2.2905526E-03	1.02	1.0092750E-02	2.01
$\frac{1}{16}$	$\frac{1}{128}$	1.1414410E-03	1.00	2.5207607E-03	2.00
$\frac{1}{32}$	$\frac{1}{512}$	5.7021046E-04	1.00	6.3010109E-04	2.00

Table 2.7: Convergence study for Example 2 in the case of 9 orders of magnitude difference in the coefficients between the layers.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	2.7567369E-02		1.0091765E-01	
$\frac{1}{4}$	$\frac{1}{8}$	6.6746149E-03	2.05	1.6801614E-02	2.59
$\frac{1}{8}$	$\frac{1}{32}$	1.6671557E-03	2.00	3.7229089E-03	2.17
$\frac{1}{16}$	$\frac{1}{128}$	4.1684321E-04	2.00	9.0153201E-04	2.05
$\frac{1}{32}$	$\frac{1}{512}$	1.0421635E-04	2.00	2.2356850E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	9.6240619E-03		1.7289930E-01	
$\frac{1}{4}$	$\frac{1}{8}$	4.6302326E-03	1.06	4.0671747E-02	2.09
$\frac{1}{8}$	$\frac{1}{32}$	2.2905526E-03	1.02	1.0092750E-02	2.01
$\frac{1}{16}$	$\frac{1}{128}$	1.1414410E-03	1.00	2.5207607E-03	2.00
$\frac{1}{32}$	$\frac{1}{512}$	5.7021046E-04	1.00	6.3010109E-04	2.00

2.6.4 Example 3

We study the problem of a 3 layered material in which the middle layer has high permeability. The following example was constructed to provide an analytic solution in this case.

Let the exact solution be

$$p = \begin{cases} \frac{t^2}{k_1}\sin(\pi x)\sin(\pi y)(y-0.25)(y-.5), & y < 0.25, \\ \frac{t^2}{k_2}\sin(\pi x)\sin(\pi y)(y-0.25)(y-0.75), & 0.25 < y < 0.75, \\ \frac{t^2}{k_1}\sin(\pi x)\sin(\pi y)(y-0.25)(y-0.5), & y > 0.75, \end{cases} \tag{2.14}$$

$$u_1 = \begin{cases} \frac{t^2}{\mu_1}x\,(1-x)\sin(\pi y)\sin(y-0.25)\sin(y-0.75), & y < 0.25, \\ \frac{t^2}{\mu_2}x\,(1-x)\sin(\pi y)\sin(y-0.25)\sin(y-0.75), & 0.25 < y < 0.75, \\ \frac{t^2}{\mu_1}x\,(1-x)\sin(\pi y)\sin(y-0.25)\sin(y-0.75), & y > 0.75 \end{cases} \tag{2.15}$$

$$u_2 = \begin{cases} \frac{t^2}{\lambda_1+2\mu_1}\sin(\pi x)y\,(1-y)\sin(y-0.25)\sin(y-0.75), & y < 0.25, \\ \frac{t^2}{\lambda_2+2\mu_2}\sin(\pi x)y\,(1-y)\sin(y-0.25)\sin(y-0.75), & 0.25 < y < 0.75. \\ \frac{t^2}{\lambda_1+2\mu_1}\sin(\pi x)y\,(1-y)\sin(y-0.25)\sin(y-0.75), & y > 0.75 \end{cases} \tag{2.16}$$

Again, the boundary and initial conditions are calculated from the known solution. The following coefficients are considered to be piecewise constant

$$\mathbf{K}(y) = \begin{cases} k_1 I, & y < 0.25, \\ k_2 I, & 0.25 < y < 0.75, \\ k_1 I, & y > 0.75, \end{cases} \quad \lambda(y) = \begin{cases} \lambda_1, & y < 0.25, \\ \lambda_2, & 0.25 < y < 0.75, \\ \lambda_1, & y > 0.75, \end{cases} \quad \mu(y) = \begin{cases} \mu_1, & y < 0.25, \\ \mu_2, & 0.25 < y < 0.75, \\ \mu_1, & y > 0.75. \end{cases}$$

Although the analytic solution allows for λ and μ to be discontinuous as well, for the following trials, we have fixed $\lambda = \lambda_1 = \lambda_2 = 114$ and $\mu = \mu_1 = \mu_2 = 455$ and only allowed the permeability to be discontinuous. Tables 2.8, 2.9, 2.10, and 2.11 show the results of repeating the above experiment in Section 2.6.3. Specifically, for Table 2.8 $k_1 = k_2 = 1$, for Table 2.9 $k_1 = 1$ while $k_2 = 10^3$, for Table 2.10 $k_1 = 1$ while $k_2 = 10^6$, and for Table 2.11 $k_1 = 1$ while $k_2 = 10^9$. Again, we see no impact from the layers on the convergence rates.

Table 2.8: Convergence study for Example 3 in the case of no layer.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	9.2125189E-03		2.7432543E-02	
$\frac{1}{4}$	$\frac{1}{8}$	4.9042540E-03	0.91	8.6129464E-03	1.67
$\frac{1}{8}$	$\frac{1}{32}$	1.2897998E-03	1.93	1.9666668E-03	2.13
$\frac{1}{16}$	$\frac{1}{128}$	3.2602246E-04	1.98	4.7859270E-04	2.04
$\frac{1}{32}$	$\frac{1}{512}$	8.1723954E-05	2.00	1.1881390E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	7.2594480E-06		2.8951839E-02	
$\frac{1}{4}$	$\frac{1}{8}$	4.1426361E-06	0.81	1.7201475E-02	0.75
$\frac{1}{8}$	$\frac{1}{32}$	1.7153350E-06	1.27	4.5089651E-03	1.93
$\frac{1}{16}$	$\frac{1}{128}$	7.9113673E-07	1.12	1.1392310E-03	1.98
$\frac{1}{32}$	$\frac{1}{512}$	3.8622815E-07	1.03	2.8555589E-04	2.00

Table 2.9: Convergence study for Example 3 in the case of 3 orders of magnitude difference in the coefficients between the layers.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	1.5547416E-02		1.1343120E-01	
$\frac{1}{4}$	$\frac{1}{8}$	4.5096889E-03	1.79	8.6180482E-03	3.72
$\frac{1}{8}$	$\frac{1}{32}$	1.1831511E-03	1.93	1.9667944E-03	2.13
$\frac{1}{16}$	$\frac{1}{128}$	2.9864064E-04	1.99	4.7859254E-04	2.04
$\frac{1}{32}$	$\frac{1}{512}$	7.4830982E-05	2.00	1.1881252E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	7.2205634E-06		3.3157451E-02	
$\frac{1}{4}$	$\frac{1}{8}$	4.0656882E-06	0.83	1.5656991E-02	1.08
$\frac{1}{8}$	$\frac{1}{32}$	1.6971082E-06	1.26	4.1281714E-03	1.92
$\frac{1}{16}$	$\frac{1}{128}$	7.8842253E-07	1.10	1.0438830E-03	1.98
$\frac{1}{32}$	$\frac{1}{512}$	3.8587278E-07	1.03	2.6170475E-04	2.00

Table 2.10: Convergence study for Example 3 in the case of 6 orders of magnitude difference in the coefficients between the layers.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	1.5572147E-02		1.1364515E-01	
$\frac{1}{4}$	$\frac{1}{8}$	4.5096864E-03	1.79	8.6180698E-03	3.72
$\frac{1}{8}$	$\frac{1}{32}$	1.1831510E-03	1.93	1.9667947E-03	2.13
$\frac{1}{16}$	$\frac{1}{128}$	2.9864061E-04	1.99	4.7859256E-04	2.04
$\frac{1}{32}$	$\frac{1}{512}$	7.4830975E-05	2.00	1.1881252E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	7.2208197E-06		3.3176293E-02	
$\frac{1}{4}$	$\frac{1}{8}$	4.0656227E-06	0.83	1.5655845E-02	1.08
$\frac{1}{8}$	$\frac{1}{32}$	1.6970932E-06	1.26	4.1279248E-03	1.92
$\frac{1}{16}$	$\frac{1}{128}$	7.8842035E-07	1.10	1.0438235E-03	1.98
$\frac{1}{32}$	$\frac{1}{512}$	3.8587250E-07	1.03	2.6169001E-04	2.00

Table 2.11: Convergence study for Example 3 in the case of 9 orders of magnitude difference in the coefficients between the layers.

h	Δt	$\|p - p_h\|_{L^\infty(L^2)}$	Order	$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{L^2(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	1.5572171E-02		1.1364536E-01	
$\frac{1}{4}$	$\frac{1}{8}$	4.5096864E-03	1.79	8.6180698E-03	3.72
$\frac{1}{8}$	$\frac{1}{32}$	1.1831510E-03	1.93	1.9667947E-03	2.13
$\frac{1}{16}$	$\frac{1}{128}$	2.9864061E-04	1.99	4.7859256E-04	2.04
$\frac{1}{32}$	$\frac{1}{512}$	7.4830975E-05	2.00	1.1881252E-04	2.01
h	Δt	$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^\infty(L^2)}$	Order	$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{L^\infty(L^2)}$	Order
$\frac{1}{2}$	$\frac{1}{2}$	7.2208200E-06		3.3176312E-02	
$\frac{1}{4}$	$\frac{1}{8}$	4.0656226E-06	0.83	1.5655843E-02	1.08
$\frac{1}{8}$	$\frac{1}{32}$	1.6970932E-06	1.26	4.1279245E-03	1.92
$\frac{1}{16}$	$\frac{1}{128}$	7.8842035E-07	1.11	1.0438234E-03	1.98
$\frac{1}{32}$	$\frac{1}{512}$	3.8587250E-07	1.03	2.6169000E-04	2.00

Chapter 3

Preconditioning

3.1 Block Preconditioning Technique

We make use of a preconditioning technique known as block preconditioning. Assume that our linear system will have the form $\mathcal{A}x = b$ with

$$\mathcal{A} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{C} \end{pmatrix} \tag{3.1}$$

where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is symmetric positive semi-definite, $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ with $m < n$, and $\boldsymbol{C} \in \mathbb{R}^{m\times m}$ is symmetric positive semi-definite. One way to construct a preconditioner $\mathcal{P}$ for this system is to form $\mathcal{P}^{-1}$ so that the preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}$ has a low-degree minimal polynomial, or equivalently has only a few eigenvalues [39]. If $\boldsymbol{A}$ is also non-singular or $\boldsymbol{A}$ is symmetric positive definite (SPD) [28], then known optimal block diagonal preconditioner is based on the Schur complement [6]. We considered ideal, block diagonal preconditioners of the form

$$\mathcal{P} = \begin{pmatrix} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{S} \end{pmatrix}, \tag{3.2}$$

where S is the negative of the Schur complement, i.e., $\boldsymbol{S} = \boldsymbol{C} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$. For simplicity, we start with the case where $\boldsymbol{C} = \boldsymbol{0}$.

Proposition 1. *Preconditioners of the form*

$$\mathcal{P} = \begin{pmatrix} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T \end{pmatrix} \tag{3.3}$$

when applied to the saddle point system $\mathcal{A}x = b$ with

$$\mathcal{A} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{pmatrix}.$$

yield a system with at most 3 distinct eigenvalues, see equivalently [6, 39].

Proof. Consider the eigenvalue problem

$$\mathcal{P}^{-1}\mathcal{A}x = \gamma x \quad \forall x \neq \boldsymbol{0} \in \mathbb{R}^{(n+m)\times(n+m)}.$$

This leads to the generalized eigenvalue problem

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \gamma \begin{pmatrix} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{3.4}$$

From the first row, we have

$$\boldsymbol{A}x_1 + \boldsymbol{B}^T x_2 = \gamma \boldsymbol{A}x_1.$$

If $\gamma \neq 1$, this gives

$$x_1 = \frac{1}{\gamma - 1}\boldsymbol{A}^{-1}\boldsymbol{B}^T x_2. \tag{3.5}$$

Plugging (3.5) into the second equation from (3.4) gives

$$\frac{1}{\gamma - 1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T x_2 = \gamma \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T x_2. \tag{3.6}$$

Therefore,

$$1 = \gamma\left(\gamma - 1\right). \tag{3.7}$$

Now the system can have at most 3 distinct eigenvalues. These distinct values are $\frac{1\pm\sqrt{5}}{2}$, from the solution of (3.7), or 1, which was excluded from the calculation. □

If this exact preconditioner is used, then any Krylov subspace iterative method that makes use of optimality conditions, such as GMRES and FGMRES, will converge in at most 3 iterations [39].

For the problems under consideration, $\boldsymbol{C}$ may not be $\boldsymbol{0}$, but it is known by construction to be small. Furthermore, $\boldsymbol{C}$ gets smaller as Δt gets smaller. It will be useful in this case

to that $\boldsymbol{C}$ is not only symmetric positive semi-definite [60], but is also SPD. To show this we will make use of the bilinear form associated with the matrix $\boldsymbol{C}$.

Bilinear forms, associated with matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ will be useful for the following proofs. First let

$$\boldsymbol{V} = \mathcal{W}_h \times \Sigma_h, \quad \boldsymbol{Q} = \mathcal{U}_h \times \mathcal{V}_h,$$

with the associated norms

$$\|(\omega, \boldsymbol{\tau})\|_{\boldsymbol{V}} = \|\omega\|_0 + \|\boldsymbol{\tau}\|_{\boldsymbol{H}(div)}, \quad \forall (\omega, \boldsymbol{\tau}) \in \boldsymbol{V},$$

and

$$\|(\boldsymbol{v}, \boldsymbol{z})\|_{\boldsymbol{Q}} = \|\boldsymbol{v}\|_0 + \|\boldsymbol{z}\|_{H(div)}, \quad \forall (\boldsymbol{v}, \boldsymbol{z}) \in \boldsymbol{Q}.$$

The bilinear form $\varphi_{\boldsymbol{C}}(\cdot, \cdot)$ defined on $\boldsymbol{Q} \times \boldsymbol{Q}$ is

$$\varphi_{\boldsymbol{C}}((\boldsymbol{u}_h, \boldsymbol{q}_h), (\boldsymbol{v}, \boldsymbol{z})) = \left(\mathbf{K}^{-1}\boldsymbol{q}_h, \boldsymbol{z}\right) \tag{3.8}$$

Lemma 1. *The bilinear form $\varphi_{\boldsymbol{C}}$ is positive definite, i.e. for any $(\boldsymbol{v}, \boldsymbol{z}) \neq (\mathbf{0}, \mathbf{0}) \in Q$*

$$\varphi_{\boldsymbol{C}}((\boldsymbol{v}, \boldsymbol{z}), (\boldsymbol{v}, \boldsymbol{z})) > 0$$

.

Proof. This proof is straight forward. □

Also note that we are assuming that the matrix $\mathcal{A}$ is invertible. Therefore, it is well known that 0 is not an eigenvalue of $\mathcal{A}$ since the system $\mathcal{A}x = 0$ has only the solution $x = \mathbf{0}$. For this particular problem, Yi [60] shows that $\mathcal{A}$ is invertible.

Proposition 2. *Assume $\mathcal{A}$ is the invertible matrix described in* (2.6)*. Then, preconditioners of the form*

$$\mathcal{P} = \begin{pmatrix} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{S} \end{pmatrix} \tag{3.9}$$

with $S = \boldsymbol{C} + \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$*, when applied to the saddle point system* $\mathcal{A}x = b$ *with*

$$\mathcal{A} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{C} \end{pmatrix},$$

will yield a system with eigenvalues distinct from 0.

Proof. As in the proof to Proposition 1, we consider the generalized eigenvalue problem

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & -\boldsymbol{C} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \gamma \begin{pmatrix} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{S} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{3.10}$$

The first row has not changed, so (3.5) is still valid. Plugging (3.5) into the second equation from (3.10) gives

$$\frac{1}{\gamma - 1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T x_2 - \boldsymbol{C}x_2 = \gamma \boldsymbol{S}x_2.$$

After some rearrangement we have

$$\frac{1 - \gamma(\gamma - 1)}{\gamma}\boldsymbol{S}x_2 = \boldsymbol{C}x_2.$$

Then, multiplying both sides by x_2^T gives

$$\frac{1 - \gamma(\gamma - 1)}{\gamma}x_2^T\boldsymbol{S}x_2 = x_2^T\boldsymbol{C}x_2.$$

The matrix $\boldsymbol{A}$, and therefore $\boldsymbol{A}^{-1}$, is SPD, so the product $\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$ is positive semi-definite [28]. Therefore matrix $\boldsymbol{S}$ is SPD since it is the sum of an SPD and a positive semi-definite matrix. Therefore, $x_2^T\boldsymbol{S}x_2 > 0$ and we have

$$\frac{1 - \gamma(\gamma - 1)}{\gamma} = \frac{x_2^T\boldsymbol{C}x_2}{x_2^T\boldsymbol{S}x_2}.$$

Note that $x_2^T\boldsymbol{S}x_2 \geq x_2^T\boldsymbol{C}x_2$, so

$$0 < \frac{1 - \gamma(\gamma - 1)}{\gamma} \leq 1.$$

If $\gamma > 0$, then

$$0 < -\gamma^2 + \gamma + 1 \quad \text{and} \quad -\gamma^2 + 1 \leq 0.$$

Therefore,

$$1 \leq \gamma < \frac{1}{2} + \sqrt{\frac{5}{4}}.$$

If $\gamma < 0$, then

$$-\gamma^2 + \gamma + 1 < 0 \quad \text{and} \quad 0 \leq -\gamma^2 + 1.$$

Therefore,

$$-1 \leq \gamma \leq \frac{1}{2} - \sqrt{\frac{5}{4}}.$$

□

Because the eigenvalues are separated from 0, we still expect the preconditioner, $\mathcal{P}$, to speed the convergence of GMRES significantly [55].

In general, we will be using FGMRES to solve the system. This algorithm only requires that a vector, v, be preconditioned via $\tilde{\mathcal{P}}^{-1}v$ once per iteration. Furthermore, this method has the advantage of allowing us to slightly change the preconditioner in each step. However, since the preconditioner is being changed at each step, we will need to save a preconditioned vector in each iteration [48].

For preconditioners of the form (3.2) to be valid for our problem, we will need $\boldsymbol{A}$ to be SPD. Under certain circumstances, $\boldsymbol{A}$ is SPD. We will adhere to those circumstances in our example problems. Rather than prove that $\boldsymbol{A}$ is SPD directly, we will use the bilinear form $\varphi_{\boldsymbol{A}}(\cdot,\cdot)$ defined on $\boldsymbol{V} \times \boldsymbol{V}$

$$\varphi_{\boldsymbol{A}}((p_h, \boldsymbol{\sigma}_h), (\omega, \boldsymbol{\tau})) = (c_o + c_r)(p_h, \omega) + (\not{\mathcal{A}}\boldsymbol{\sigma}_h, \boldsymbol{\tau}) + \frac{c_r}{2\alpha}(tr(\boldsymbol{\sigma}_h), \omega) + \frac{c_r}{2\alpha}(p_h, tr(\boldsymbol{\tau})). \quad (3.11)$$

Lemma 2. *The bilinear form $\varphi_{\boldsymbol{A}}$ is positive definite, i.e. for any $(\omega, \boldsymbol{\tau}) \neq (0, \mathbf{0}) \in V$*

$$\varphi_{\boldsymbol{A}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) > 0$$

if $\lambda + \mu > 0$ and $c_0 > 0$.

Proof. Recall (3.11) and (2.3):

$$\begin{aligned}
&\varphi_{\boldsymbol{A}}\left((\omega,\boldsymbol{\tau}),(\omega,\boldsymbol{\tau})\right)\\
&\quad=(c_o+c_r)(\omega,\omega)+(\mathscr{A}\boldsymbol{\tau},\boldsymbol{\tau})+\frac{c_r}{\alpha}(tr(\boldsymbol{\tau}),\omega)\\
&\quad=\left(c_o+\frac{\alpha^2}{\mu+\lambda}\right)\|\omega\|^2+\frac{1}{2\mu}\|\boldsymbol{\tau}\|^2-\frac{\lambda}{4\mu(\mu+\lambda)}\|\tau_{11}+\tau_{22}\|^2+\frac{\alpha}{\mu+\lambda}(\tau_{11}+\tau_{22},\omega).
\end{aligned}$$

By the Triangle inequality and Young's inequalities

$$\|\tau_{11}+\tau_{22}\|^2\le 2\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right).$$

Also, by Young's inequality,

$$\begin{aligned}
\alpha(\tau_{11}+\tau_{22},\omega)&\le\frac{\|\tau_{11}+\tau_{22}\|^2}{2\epsilon}+\frac{\epsilon}{2}\alpha^2\|\omega\|^2\\
&\le\frac{1}{\epsilon}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right)+\frac{\epsilon}{2}\alpha^2\|\omega\|^2
\end{aligned}$$

for some $\epsilon>0$. Therefore,

$$\begin{aligned}
&\varphi_{\boldsymbol{A}}\left((\omega,\boldsymbol{\tau}),(\omega,\boldsymbol{\tau})\right)\\
&\quad\ge\left(c_o+\frac{\alpha^2}{\mu+\lambda}\right)\|\omega\|^2+\frac{1}{2\mu}\left(\|\tau_{11}\|^2+2\|\tau_{12}\|^2+\|\tau_{22}\|^2\right)-\frac{\lambda}{2\mu(\mu+\lambda)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right)\\
&\qquad-\frac{1}{\epsilon\,(\mu+\lambda)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right)-\frac{\alpha^2\epsilon}{2\,(\mu+\lambda)}\|\omega\|^2\\
&\quad=\left(c_o+\frac{\alpha^2(2-\epsilon)}{2\,(\mu+\lambda)}\right)\|\omega\|^2+\frac{\epsilon-2}{2\epsilon\,(\lambda+\mu)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right)+\frac{1}{\mu}\|\tau_{12}\|^2
\end{aligned}$$

To ensure that $\varphi_{\boldsymbol{A}}\left((\omega,\boldsymbol{\tau}),(\omega,\boldsymbol{\tau})\right)=0$ only if $\omega=0$ and $\boldsymbol{\tau}=\mathbf{0}$, we must have

$$c_o+\frac{\alpha^2(2-\epsilon)}{2\,(\lambda+\mu)}>0$$

and

$$\frac{\epsilon-2}{2\epsilon\,(\lambda+\mu)}>0.$$

Therefore, if $c_0>0$ and $\lambda+\mu>0$, we can guarantee the above condition by choosing ϵ such that

$$2<\epsilon<2+\frac{2c_0\,(\mu+\lambda)}{\alpha^2}. \tag{3.12}$$

□

For the following experiments, we will adhere to the case when $\boldsymbol{A}$ is SPD. Therefore the use of the Schur complement preconditioner (3.2) is justified.

3.2 Spectrally Equivalent Preconditioner

Although using the ideal preconditioner, $\mathcal{P}$, should result in fast convergence of the FGMRES method, it may be expensive to invert. For practical applications, we use an approximation to, $\mathcal{P}$, denoted $\tilde{\mathcal{P}}$. Since preconditioning with $\mathcal{P}^{-1}$ would require finding $\boldsymbol{A}^{-1}$ and $\boldsymbol{S}^{-1}$, we seek to replace $\boldsymbol{A}$ by a matrix that is inexpensive to invert. Toh et. al [54] have explored a few ways to approximate the Schur complement when using it as a preconditioner. However, we use only the diagonal entries of $\boldsymbol{A}$. From now on, the notation, $\boldsymbol{D_Q}$, will be used to denote the diagonal of some matrix $\boldsymbol{Q}$. The preconditioner becomes

$$\tilde{\mathcal{P}} = \begin{pmatrix} \boldsymbol{D_A} & 0 \\ 0 & \tilde{\boldsymbol{S}} \end{pmatrix} \tag{3.13}$$

with $\tilde{\boldsymbol{S}} = \boldsymbol{C} + \boldsymbol{B}\boldsymbol{D}_{\boldsymbol{A}}^{-1}\boldsymbol{B}^T$. As long as $\boldsymbol{D_A}$ is a relatively good approximation to $\boldsymbol{A}$, FGMRES should converge quickly. Additionally, this approximation has the advantage of making the first block, $\tilde{\mathcal{P}}_{1,1}$, inexpensive to apply. We will attempt to use a spectrally equivalent approximation.

Definition 2. *If $\mathcal{B}$ is a symmetric positive definite matrix arising from the approximation of a partial differential operator on a mesh of size h, then $\mathcal{B}$ is spectrally equaivalent to a matrix $\mathcal{C}$ of equal dimension if*

$$C_1 x^T \mathcal{C} x \leq x^T \mathcal{B} x \leq C_2 x^T \mathcal{C} x \quad \forall x \neq \mathbf{0} \in \mathbb{R}^n,$$

where $C_1, C_2 \in \mathbb{R}$, $0 < C_1 < C_2$, and both C_1 and C_2 are independent of h [33, 16]. Spectral equivalence is denoted by $\mathcal{B} \sim \mathcal{C}$.

Note that in some cases, we will allow the the material parameters λ, μ, and $\boldsymbol{K}$ to be only piecewise constants. However, the constrained specific storage coefficient on Ω is

constant and we maintain $c_0 > 0$ and $\lambda + \mu > 0$ thought out the region of interest. We will show that spectral equivalence holds for these cases.

Since $\boldsymbol{A}$ is SPD, the ideal, block diagonal preconditioner is viable. However to apply the preconditioner (3.13), we will need to show that $\boldsymbol{A}$ is spectrally equivalent to $\boldsymbol{D_A}$. This is denoted

$$\boldsymbol{A} = \begin{pmatrix} C_{pp} & C_{\boldsymbol{\sigma}p}^T \\ C_{\boldsymbol{\sigma}p} & C_{\boldsymbol{\sigma\sigma}} \end{pmatrix} \sim \begin{pmatrix} \boldsymbol{D}_{C_{pp}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} \end{pmatrix}.$$

The proof will be done in several steps. First, it will be shown that $\boldsymbol{A} \sim \boldsymbol{D}$ where

$$\boldsymbol{D} = \begin{pmatrix} C_{pp} & \mathbf{0} \\ \mathbf{0} & C_{\sigma\sigma} \end{pmatrix}.$$

Then we will show that $\begin{pmatrix} C_{pp} & \mathbf{0} \\ \mathbf{0} & C_{\sigma\sigma} \end{pmatrix} \sim \begin{pmatrix} \boldsymbol{D}_{C_{pp}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} \end{pmatrix}$ by showing that $C_{pp} \sim \boldsymbol{D}_{C_{pp}}$ and $C_{\sigma\sigma} \sim \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}}$.

For the first step, rather than show $\boldsymbol{A} \sim \boldsymbol{D}$ directly, the corresponding bilinear forms similar to that of (3.11) will be used. The bilinear form associated with $\boldsymbol{D}$ will be

$$\varphi_{\mathbf{D}}\left((p_h, \boldsymbol{\sigma}_h), (\omega, \boldsymbol{\tau})\right) = (c_o + c_r)(p_h, \omega) + (\mathscr{A}\boldsymbol{\sigma}_h, \boldsymbol{\tau}). \tag{3.14}$$

Lemma 3. *Using the conditions of Lemma 2 and assuming that λ and μ are strictly constants in Ω, the bilinear forms $\varphi_{\boldsymbol{A}}$ and $\varphi_{\boldsymbol{D}}$ are spectrally equivalent, i.e. for any $(\omega, \boldsymbol{\tau}) \in \boldsymbol{V}$*

$$C_1\varphi_{\boldsymbol{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq \varphi_{\boldsymbol{A}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq C_2\varphi_{\boldsymbol{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau}))$$

for some $0 < C_1 < C_2$.

Proof. Let $\boldsymbol{\tau} = \begin{pmatrix} \tau_{11} & \tau_{12} \\ \tau_{12} & \tau_{22} \end{pmatrix}$. Recall (2.3). Then

$$\mathscr{A}\tau : \tau = \frac{1}{2\mu}(\tau : \tau) - \frac{\lambda}{4\mu(\mu + \lambda)}(tr(\tau))^2$$

Or equivalently,

$$\mathscr{A}\tau : \tau = \frac{1}{2\mu}\left(\tau_{11}^2 + 2\tau_{12}^2 + \tau_{22}^2\right) - \frac{\lambda}{4\mu(\mu + \lambda)}(\tau_{11} + \tau_{22})^2.$$

After expansion and combining like terms, we have

$$\mathcal{A}\tau : \tau = \frac{2\mu+\lambda}{4\mu(\mu+\lambda)}\left(\tau_{11}^2+\tau_{22}^2\right) - \frac{\lambda}{4\mu(\mu+\lambda)}\left(2\tau_{11}\tau_{22}\right) + (\frac{1}{\mu}\tau_{12}^2).$$

We regroup the terms to get

$$\mathscr{A}\tau : \tau = \frac{1}{2(\mu+\lambda)}\left(\tau_{11}^2+\tau_{22}^2\right) + (\frac{1}{\mu}\tau_{12}^2) + \frac{\lambda}{4\mu(\mu+\lambda)}\left(\tau_{11}^2 - 2\tau_{11}\tau_{22}+\tau_{22}^2\right).$$

Thus

$$\mathscr{A}\tau : \tau = \frac{1}{2(\mu+\lambda)}\left(\tau_{11}^2+\tau_{22}^2\right) + (\frac{1}{\mu}\tau_{12}^2) + \frac{\lambda}{4\mu(\mu+\lambda)}\left(\tau_{11}-\tau_{22}\right)^2.$$

After integration, we have

$$(\mathscr{A}\boldsymbol{\tau},\boldsymbol{\tau}) = \frac{1}{2(\lambda+\mu)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right) + \frac{1}{\mu}\|\tau_{12}\|^2 + \frac{\lambda}{4\mu\left(\lambda+\mu\right)}\|\tau_{11}-\tau_{22}\|^2.$$

Therefore,

$$\varphi_{\mathbf{D}}\left((\omega,\boldsymbol{\tau}),(\omega,\boldsymbol{\tau})\right) = \left(c_o+c_r\right)\|\omega\|^2 + \frac{1}{2(\lambda+\mu)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right) + \frac{1}{\mu}\|\tau_{12}\|^2 + \frac{\lambda}{4\mu\left(\lambda+\mu\right)}\|\tau_{11}-\tau_{22}\|^2.$$

Recall also from Lemma 2

$$\frac{\alpha}{\lambda+\mu}\left(\tau_{11}+\tau_{22},\omega\right) \le \frac{1}{\epsilon\left(\lambda+\mu\right)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right) + \frac{\epsilon c_r}{2}\|\omega\|^2. \tag{3.15}$$

Let $\epsilon = 2$. Then

$$\begin{aligned}
&\varphi_{\boldsymbol{A}}\left((\omega,\boldsymbol{\tau}),(\omega,\boldsymbol{\tau})\right)\\
&\quad\le \left(c_o+2c_r\right)\|\omega\|^2 + \frac{1}{2(\lambda+\mu)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right) + \frac{1}{\mu}\|\tau_{12}\|^2 + \frac{\lambda}{4\mu\left(\lambda+\mu\right)}\|\tau_{11}-\tau_{22}\|^2
\end{aligned}$$

Then let $C_2 \ge \frac{c_0+2c_r}{c_0+c_r} > 1$. Also,

$$\begin{aligned}
&\varphi_{\boldsymbol{A}}\left((\omega,\boldsymbol{\tau}),(\omega,\boldsymbol{\tau})\right)\\
&\quad\ge \left(c_o+c_r\right)\|\omega\|^2 + \frac{1}{2(\lambda+\mu)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right) + \frac{1}{\mu}\|\tau_{12}\|^2 + \frac{\lambda}{4\mu\left(\lambda+\mu\right)}\|\tau_{11}-\tau_{22}\|^2\\
&\qquad - \frac{\epsilon c_r}{2}\|\omega\|^2 - \frac{1}{\epsilon(\lambda+\mu)}\left(\|\tau_{11}\|^2+\|\tau_{22}\|^2\right)
\end{aligned}$$

We must choose a C_1 such that

$$c_0 + (1 - \frac{\epsilon}{2})c_r > C_1 (c_0 + c_r) > 0$$

and

$$\frac{1}{2(\lambda+\mu)} - \frac{1}{\epsilon(\lambda+\mu)} > C_1 \frac{1}{2(\lambda+\mu)} > 0.$$

After some algebra, we require

$$\frac{c_0 + (1 - \frac{\epsilon}{2})c_r}{c_0 + c_r} > C_1 > 0$$

and

$$1 - \frac{2}{\epsilon} > C_1 > 0$$

These can be satisfied as long as

$$2 < \epsilon < 2 + 2\frac{c_0}{c_r},$$

so chose $\epsilon = 2 + \frac{c_0}{c_r}$. Then let

$$C_1 = \min\{\frac{c_0}{2(c_0 + c_r)}, 1 - \frac{2}{2 + \frac{c_0}{c_r}}\}$$

□

Corollary 1. *Under the conditions of Lemma 2 and assuming that λ and μ are piecewise constants in a finite number of connected regions, the bilinear forms $\varphi_{\boldsymbol{A}}$ and $\varphi_{\boldsymbol{D}}$ are spectrally equivalent, i.e. for any $(\omega, \boldsymbol{\tau}) \in \boldsymbol{V}$*

$$C_a \varphi_{\boldsymbol{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq \varphi_{\boldsymbol{A}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq C_b \varphi_{\boldsymbol{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau}))$$

for some $0 < C_a < C_b$.

Proof. Let the domain Ω be divided into N regions, Ω_i with $i = 1, 2..., N$. Then in a region, Ω_n, $\lambda = \lambda_n$ and $\mu = \mu_n$. Lemma 3 implies

$$C_{1_n} \varphi_{\mathbf{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq \varphi_{\mathbf{A}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq C_{2_n} \varphi_{\mathbf{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau}))$$

where

$$C_{1_n} = \min\{\frac{c_0}{2(c_0 + c_{r_n})}, 1 - \frac{2}{2 + \frac{c_0}{c_{r_n}}}\}$$

and

$$C_{2_n} = \frac{c_0 + 2c_{r_n}}{c_0 + c_{r_n}}$$

with $c_{r_n} = \frac{\alpha^2}{\lambda_n + \mu_n}$. Let $C_a = \min_n\{C_{1_n}\}$ and $C_b = \max_n\{C_{2_n}\}$ for $n = 1, 2, 3, ..., N$. □

Now for the second part, we will need to show that $\begin{pmatrix} C_{pp} & \mathbf{0} \\ \mathbf{0} & C_{\sigma\sigma} \end{pmatrix} \sim \begin{pmatrix} \boldsymbol{D}_{C_{pp}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} \end{pmatrix}$. We will make use of the following Lemma.

Lemma 4. *Let $\boldsymbol{M}_1$ be an $n \times n$ matrix and let $\boldsymbol{M}_2$ be an $m \times m$ matrix. Then*

$$\begin{pmatrix} \boldsymbol{M}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M}_2 \end{pmatrix} \sim \begin{pmatrix} \boldsymbol{D}_{M_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_{M_2} \end{pmatrix}$$

if $\boldsymbol{M}_1 \sim \boldsymbol{D}_{M_1}$ and $\boldsymbol{M}_2 \sim \boldsymbol{D}_{M_2}$.

Proof. The proof is straight forward. □

Now we will show that $C_{\boldsymbol{\sigma\sigma}} \sim \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}}$ and $C_{pp} \sim \boldsymbol{D}_{C_{pp}}$. This will be done by making an observation on local elements, then making an extension to the global system, as is done in superelement analysis e.g. [19]. The extension of these local results into a global result relies heavily on the following lemma.

Lemma 5. *Let $a_i > 0$ and $c_i > 0 \quad \forall i \in \mathbb{N}^+$. Then*

$$\frac{\sum_{i=1}^{N} a_i}{\sum_{i=1}^{N} c_i} \leq \max_{1 \leq i \leq N} \left\{ \frac{a_i}{c_i} \right\}$$

and

$$\min_{1 \leq i \leq N} \left\{ \frac{a_i}{c_i} \right\} \leq \frac{\sum_{i=1}^{N} a_i}{\sum_{i=1}^{N} c_i}$$

for any $N \in \mathbb{N}^+$ [34].

Proof. Let's start with the case where N=2. Consider

$$\begin{aligned}\frac{a_1+a_2}{c_1+c_2}-\frac{a_1}{c_1} &= \frac{c_1\left(a_1+a_2\right)-a_1\left(c_1+c_2\right)}{c_1\left(c_1+c_2\right)} \\ &= \frac{c_1a_2-a_1c_2}{c_1\left(c_1+c_2\right)}\end{aligned}$$

and

$$\begin{aligned}\frac{a_1+a_2}{c_1+c_2}-\frac{a_2}{c_2} &= \frac{c_2\left(a_1+a_2\right)-a_2\left(c_1+c_2\right)}{c_2\left(c_1+c_2\right)} \\ &= \frac{a_1c_2-a_2c_1}{c_2\left(c_1+c_2\right)} \\ &= \frac{-\left(c_1a_2-a_1c_2\right)}{c_2\left(c_1+c_2\right)}.\end{aligned}$$

The denominators $c_1\left(c_1+c_2\right)$ and $c_2\left(c_1+c_2\right)$ are positive, but the numerators have opposite sign. Therefore, either both numerators are equal to 0, or one one of them must be negative. If $c_1a_2-a_1c_2=0$ then

$$\frac{a_1+a_2}{c_1+c_2}=\frac{a_1}{c_1}$$

and

$$\frac{a_1+a_2}{c_1+c_2}=\frac{a_2}{c_2}.$$

If $c_1a_2-a_1c_2<0$, then

$$\frac{a_1+a_2}{c_1+c_2}<\frac{a_1}{c_1}.$$

Finally, if $c_1a_2-a_1c_2>0$, then

$$\frac{a_1+a_2}{c_1+c_2}<\frac{a_2}{c_2}.$$

So, in any case

$$\frac{a_1+a_2}{c_1+c_2}\leq\max\left\{\frac{a_1}{c_1},\frac{a_2}{c_2}\right\}.$$

Now, assume that for $N=n$

$$\frac{\sum_{i=1}^{n}a_i}{\sum_{i=1}^{n}c_i}\leq\max_{1\leq i\leq n}\left\{\frac{a_i}{c_i}\right\}$$

and

$$A_n = \sum_{i=1}^{n} a_i > 0$$

and

$$C_n = \sum_{i=1}^{n} c_i > 0.$$

From the above described case for $N = 2$,

$$\frac{A_n + a_{n+1}}{C_n + c_{n+1}} < \max\left\{\frac{A_n}{C_n}, \frac{a_{n+1}}{c_{n+1}}\right\}.$$

On the other hand, from the assumption

$$\frac{A_n}{C_n} \leq \max_{1\leq i\leq n}\left\{\frac{a_i}{c_i}\right\},$$

So, by induction,

$$\frac{\sum_{i=1}^{n+1} a_i}{\sum_{i=1}^{n+1} c_i} \leq \max_{1\leq i\leq n+1}\left\{\frac{a_i}{c_i}\right\}$$

Similarly, we can prove that

$$\min_{1\leq i\leq N}\left\{\frac{a_i}{c_i}\right\} \leq \frac{\sum_{i=1}^{N} a_i}{\sum_{i=1}^{N} c_i}$$

by changing the direction of the inequalities and considering the minimum instead of the maximum. □

For the following proofs it is important to note that the basis functions are formed in such a way that $C_{\boldsymbol{\sigma\sigma}}$ is scaled with h^2. That is $C^K_{\boldsymbol{\sigma\sigma}} = h^2 C^{\hat{K}}_{\boldsymbol{\sigma\sigma}}$. The super script K is used to denote formation on a local element and the superscript $\hat{K}$ is used to denote formation on a reference element, $\hat{K} = [-h/2, h/2] \times [-h/2, h/2]$. Similarly $C^K_{pp} = h^2 C^{\hat{K}}_{pp}$.

Another important detail for the following proof is how the global matrices are assembled from the local ones. We perform assembly with the aid of a connectivity matrix as described

in [32]. The connectivity matrix $\mathcal{V}$ stores the relationship between local numbering and global numbering. For example, the global number for the 3^{rd} local degree of freedom on the 5^{th} element is stored in row 5 and column 3 of the matrix $\mathcal{V}$, denoted by $\mathcal{V}_{5,3}$. In general, $\mathcal{V}_{i,j}$ is the global number for the j^{th} degree of freedom in the i^{th} element, where $1 \leq i \leq N$ with N is the number of elements and $1 \leq j \leq M$ with M is the number of local degrees of freedom. The number of local degrees of freedom varies depending on the variable being discussed. For example, for stress $\widetilde{\boldsymbol{\sigma}}$, $M = 17$ and for flux $\boldsymbol{q}$, $M = 12$.

Lemma 6. *$C_{\boldsymbol{\sigma\sigma}}$ is spectrally equivalent to $\boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}}$ i.e.*

$$C_1 x^T \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} x \leq x^T C_{\boldsymbol{\sigma\sigma}} x \leq C_2 x^T \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} x.$$

Proof. The global matrices $C_{\boldsymbol{\sigma\sigma}}$ and $\boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}}$ are assembled by using a connectivity matrix $\mathcal{V}$ and the local matrices $C^K_{\boldsymbol{\sigma\sigma}} \in \mathbb{R}^{17\times 17}$ and $\boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} \in \mathbb{R}^{17\times 17}$. Let $j = \mathcal{V}_{k,i}$ with $1 \leq k \leq N$ where N is the number of elements and $1 \leq i \leq 17$. Then let x^k be a local version of the vector x:

$$(x^k)_i = x_j.$$

Now we write

$$x^T C_{\boldsymbol{\sigma\sigma}} x = \sum_{k=1}^{N} (x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k$$

and

$$x^T \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} x = \sum_{k=1}^{N} (x^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} x^k.$$

Consider the fraction

$$\frac{x^T C_{\boldsymbol{\sigma\sigma}} x}{x^T \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} x} = \frac{\displaystyle\sum_{k=1}^{N} (x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k}{\displaystyle\sum_{k=1}^{N} (x^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} x^k}.$$

Note that since $C^K_{\boldsymbol{\sigma\sigma}}$ is SPD, for any $1 \leq k \leq N$, $(x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k = 0$ only if $x_k = 0$. The same is true of $\boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}}$. So these zeros can be excluded from the summations when considering the

fraction above. The remaining terms $(x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k$ and $(x^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} x^k$ are strictly positive. Now by Lemma 5,

$$\min_{1\le k\le N} \frac{(x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k}{(x^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} x^k} \le \frac{x^T C_{\boldsymbol{\sigma\sigma}} x}{x^T \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} x} \le \max_{1\le k\le N} \frac{(x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k}{(x^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} x^k}.$$

Since x can be any vector, x^k can be any vector in $\mathbb{R}^{17\times 17}$. However,

$$\min_{1\le k\le N} \frac{(x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k}{(x^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} x^k} = \gamma$$

will occur at some vector, call it w^k. So consider $\frac{(w^k)^T C^K_{\boldsymbol{\sigma\sigma}} w^k}{(w^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} w^k} = \gamma$ or equivalently

$$C^K_{\boldsymbol{\sigma\sigma}} w^k = \gamma \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} w^k$$

Moreover, we can use the local matrices one the reference square when we recall that $C^K_{\boldsymbol{\sigma\sigma}} = h^2 C^{\hat{K}}_{\boldsymbol{\sigma\sigma}}$. We find γ by solving the generalized eigenvalue problem

$$h^2 C^{\hat{K}}_{\boldsymbol{\sigma\sigma}} w^k = \gamma h^2 \boldsymbol{D}^{\hat{K}}_{C_{\boldsymbol{\sigma\sigma}}} w^k$$

There will be 17 eigenvalue solutions, γ_i with $1 \le i \le 17$, to this problem. A similar argument can be made concerning

$$\max_{1\le k\le N} \frac{(x^k)^T C^K_{\boldsymbol{\sigma\sigma}} x^k}{(x^k)^T \boldsymbol{D}^K_{C_{\boldsymbol{\sigma\sigma}}} x^k}.$$

Therefore, we choose

$$C_1 \le \min_{1\le i\le 17} \gamma_i \quad \text{and} \quad C_2 \ge \max_{1\le i\le 17} \gamma_i.$$

Note that the numeric values of the eigenvalues, γ_i, will depend on the chosen parameters λ and μ. In the case that we allow these parameters to be only piecewise constants, we will simply choose the largest and smallest of the values found from solving all generalized eigenvalue on all regions where λ and μ are constant. □

The particular basis functions chosen for pressure result in a matrix C_{pp} that is diagonal. However, if this were not the case, the line of reasoning above can be followed for C_{pp} and $\boldsymbol{D}_{C_{pp}}$, respectively. The eigenvalues of $C^{\hat{K}}_{pp}$ depend on the values of c_0 and c_r.

To form an estimate on the constants needed for the spectral equivalence equations, consider the following.

Lemma 7. *If* $\boldsymbol{M}_1 \sim \boldsymbol{M}_2 \sim \boldsymbol{M}_3$, *then* $\boldsymbol{M}_1 \sim \boldsymbol{M}_3$.

Proof. The proof is straightforward. □

To show how the coefficients of spectral equivalence can depend on the magnitude of these discontinuities in the parameters λ and μ, consider the following example. First consider the case where $\lambda = \mu = 1000$ thoughout the region of interest. Then, from the generalized eigenvalue problem in Lemma 6,

$$0.135x^T \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} x \leq x^T C_{\boldsymbol{\sigma\sigma}} x \leq 3.01 x^T \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}} x.$$

Furthermore, if we take $c_0 = .1$, from Lemma 3

$$0.498\varphi_{\mathbf{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq \varphi_{\boldsymbol{A}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})) \leq 1.005\varphi_{\mathbf{D}}((\omega, \boldsymbol{\tau}), (\omega, \boldsymbol{\tau})).$$

So, $\boldsymbol{A} \sim \boldsymbol{D}_{\boldsymbol{A}}$ with the coefficients $C_1 = 0.067$ and $C_2 = 3.03$. However, if we allow a single discontinuity and 3 orders of magnitude difference in λ and μ, similar to the example in [40], the coefficients will change. Specifically, let the coefficients have the form of (2.13) and

$$k_1 = 1, \quad k_2 = 10^3, \quad \lambda_1 = 1, \quad \lambda_2 = 10^3, \quad \mu_1 = 1, \quad \text{and} \quad \mu_2 = 10^3.$$

Then $\boldsymbol{A} \sim \boldsymbol{D}_{\boldsymbol{A}}$ with the coefficients $C_1 = 0.0112$ and $C_2 = 5.52$.

3.3 Application

Recall that in order to precondition in FGMRES, we need to produce one preconditioned vector. If the inverse of the precondition $\tilde{\mathcal{P}}^{-1}$ were inexpensive to find explicitly, this could be done with a simple matrix vector multiplication. However, the preconditioner $\tilde{\mathcal{P}}$, described above, is still expensive to invert because the matrix $\tilde{\boldsymbol{S}}$ may still be expensive to invert. However, since we are using the flexible variant, we can consider the preconditioning step, $\tilde{\mathcal{P}}^{-1}v$, to be the result of a set of calculations rather than calculating $\tilde{\mathcal{P}}^{-1}$ explicitly. Let $v = (v_1, v_2)^T$. The first block of the preconditioner is a diagonal matrix and its inversion

is inexpensive. We will focus on the inversion of the second block. That is, we seek to find an approximation of $\tilde{\boldsymbol{S}}^{-1}v_2$ via some iterative approximation method.

The preconditioned vector $\tilde{\boldsymbol{S}}^{-1}v_2$ can be found as the solution of a linear system. That is we attempt to find an approximation to y in $\tilde{\boldsymbol{S}}y = v_2$. Note that $\tilde{\boldsymbol{S}}$ is also a block diagonal matrix.

$$\tilde{\boldsymbol{S}} = \boldsymbol{C} + \boldsymbol{B}\boldsymbol{D}_{\boldsymbol{A}}^{-1}\boldsymbol{B}^T \tag{3.16}$$

$$= \begin{pmatrix} \tilde{\boldsymbol{S}}_2 & 0 \\ 0 & \tilde{\boldsymbol{S}}_3 \end{pmatrix}, \tag{3.17}$$

where $\tilde{\boldsymbol{S}}_2 = \Delta t C_{\boldsymbol{qq}} + \Delta t^2 C_{\boldsymbol{qp}} \boldsymbol{D}_{C_{pp}}^{-1} C_{\boldsymbol{qp}}^T$ and $\tilde{\boldsymbol{S}}_3 = C_{\boldsymbol{u\sigma}} \boldsymbol{D}_{C_{\boldsymbol{\sigma\sigma}}}^{-1} C_{\boldsymbol{u\sigma}}^T$. Therefore, we can break the vector v_2 into two parts, $v_2 = (x_2, x_3)^T$. Then we find $\tilde{\boldsymbol{S}}_2^{-1}x_2$ and $\tilde{\boldsymbol{S}}_3^{-1}x_3$, by solving the systems

$$\tilde{\boldsymbol{S}}_2 y_2 = x_2 \tag{3.18}$$

and

$$\tilde{\boldsymbol{S}}_3 y_3 = x_3. \tag{3.19}$$

To solve the systems (3.18) and (3.19), we have used PCG with AMG as a preconditioner. This is done using the example codes given in HYPRE [10] and is treated essentially as a black box solver. Both systems are solved to a tolerance of 10^{-2} as we have observed that there is little benefit to solving these with a higher tolerance. After some testing with AMG, we have chosen to use 1 iteration or cycle of AMG with 1 relaxation sweep of the built in hybrid symmetric Gauss-Seidel smoother [10] for the preconditioner to PCG. Note that AMG forms a set of progressively coarser grids which takes some additional time. We have used the built in Falgout Coarsening [10].

3.4 Numerical Experiments

3.4.1 Example 1

We first attempt the described preconditioner in conjunction with example problem 1. Recall that the analytic solution is given in (2.7). Also, this example does not allow discontinuous coefficients. For the purposes of this experiment, we solve the linear system for only one time step. In order to keep the two terms in $\tilde{\boldsymbol{S}}_2$ in the same ratio, we reduce h by a factor of 2 and Δt is reduced by a factor of 4 in each trial.

To form a basis of comparison, we first attempted to solve the linear system using a direct method. We were unable to solve systems in which h was less than $\frac{1}{32}$ due to a lack of memory. We also attempted to solve the system using the simple diagonal preconditioner $\boldsymbol{D}_{\mathcal{A}}$ and GMRES. In this case, GMRES, restarted after 20 iterations, did not converge in 20,000 iterations when $h < \frac{1}{8}$. Finally, we applied the preconditioner (3.13) with FGMRES restarted after 20 iterations. The results of this experiment are shown in Table 3.1. It is shown that there is only a mild increase in the number of FGMRES iterations required to solve the system. Furthermore, the numbers of iterations required, per FGMRES loop, to apply the preconditioners $\tilde{\boldsymbol{S}}_2^{-1}$ and $\tilde{\boldsymbol{S}}_3^{-1}$ are roughly constant.

3.4.2 Example 2

Again, to form a basis of comparison, we attempted to solve the system using the simple diagonal preconditioner $\boldsymbol{D}_{\mathcal{A}}$ and GMRES. We used the case of no layer as described in Section 2.6.3. In this case, GMRES, restarted after 20 iterations, did not converge in 20,000 iterations for even the largest $h = \frac{1}{2}$.

We next apply the preconditioner (3.13) to the example problem whose analytic solution is (2.10). We vary the coefficients as described in Section 2.6.3 and use FGMRES restatred after 20 iterations. The results of which are shown in Tables 3.1, 3.3, and 3.4. Although the number of FGMRES iterations required to solve the system is not increasing greatly

Table 3.1: Preconditioning analysis with Example 1. The systems $\tilde{\boldsymbol{S}}_2 y_2 = x_2$ and $\tilde{\boldsymbol{S}}_3 y_3 = x_3$ are solved to a tolerance of 10^{-2}.

h	Δt	FGRMRES	$\tilde{\boldsymbol{S}}_2$ Iterations		$\tilde{\boldsymbol{S}}_3$ Iterations	
			total	per loop	total	per loop
$\frac{1}{2}$	$\frac{1}{2}$	80	669	8.36	148	1.85
$\frac{1}{4}$	$\frac{1}{8}$	94	719	7.65	190	2.02
$\frac{1}{8}$	$\frac{1}{32}$	100	813	8.13	296	2.96
$\frac{1}{16}$	$\frac{1}{128}$	110	1025	9.32	416	3.78
$\frac{1}{32}$	$\frac{1}{512}$	127	1153	9.08	567	4.46
$\frac{1}{64}$	$\frac{1}{2048}$	166	1596	9.61	831	5.01
$\frac{1}{128}$	$\frac{1}{8192}$	189	1840	9.74	1061	5.61
$\frac{1}{256}$	$\frac{1}{32768}$	203	1690	8.33	1241	6.11

as the difference between the coefficients is becoming greater, the number of iterations of PCG required to apply the preconditioner $\tilde{\boldsymbol{S}}_2$ appears to have a large dependence on the difference in the coefficients.

In the papers by Haga et.al. [24] and Dillon et.al. [15], the Lam*é* coefficients are considered constants and only the permeability $\mathbf{K}$ is allowed to be discontinuous. Along this line, we take $\lambda = 114$, $\mu = 455$, and $\mathbf{K}(y) = \begin{cases} k_0 I, & y < \zeta, \\ E k_0 I, & y > \zeta, \end{cases}$. The results of this trial, when we allow $k_0 = 10^9$ are shown in the Figure 3.1. It can be seen that when there is no layer present, $E = 1$, and when there is a large jump in coefficents, $E = 10^{-10}$, the number of iterations required are very similar. However, when the jump falls in between, the number of iterations required is larger.

Table 3.2: Preconditioning analysis with Example 2 in the case of no layer. The systems $\tilde{\boldsymbol{S}}_2 y_2 = x_2$ and $\tilde{\boldsymbol{S}}_3 y_3 = x_3$ are solved to a tolerance of 10^{-2}.

h	Δt	FGRMRES	$\tilde{\boldsymbol{S}}_2$ Iterations		$\tilde{\boldsymbol{S}}_3$ Iterations	
			total	per loop	total	per loop
$\frac{1}{2}$	$\frac{1}{2}$	91	815	8.96	162	1.78
$\frac{1}{4}$	$\frac{1}{8}$	95	647	6.81	196	2.06
$\frac{1}{8}$	$\frac{1}{32}$	103	824	8.00	311	3.02
$\frac{1}{16}$	$\frac{1}{128}$	109	937	8.60	432	3.96
$\frac{1}{32}$	$\frac{1}{512}$	138	1258	9.12	646	4.68
$\frac{1}{64}$	$\frac{1}{2048}$	164	1488	9.07	855	5.21
$\frac{1}{128}$	$\frac{1}{8192}$	192	1713	8.92	1096	5.71

Table 3.3: Preconditioning analysis with Example 2 in the case of a 3 orders of magnitude layer. The systems $\tilde{\boldsymbol{S}}_2 y_2 = x_2$ and $\tilde{\boldsymbol{S}}_3 y_3 = x_3$ are solved to a tolerance of 10^{-2}.

h	Δt	FGRMRES	$\tilde{\boldsymbol{S}}_2$ Iterations		$\tilde{\boldsymbol{S}}_3$ Iterations	
			total	per loop	total	per loop
$\frac{1}{2}$	$\frac{1}{2}$	98	3354	34.22	188	1.92
$\frac{1}{4}$	$\frac{1}{8}$	110	11622	105.65	230	2.09
$\frac{1}{8}$	$\frac{1}{32}$	113	25026	221.47	362	3.20
$\frac{1}{16}$	$\frac{1}{128}$	123	28678	233.15	498	4.05
$\frac{1}{32}$	$\frac{1}{512}$	150	46963	313.09	681	4.54
$\frac{1}{64}$	$\frac{1}{2048}$	194	58237	300.19	983	5.07
$\frac{1}{128}$	$\frac{1}{8192}$	198	55582	280.72	1082	5.46

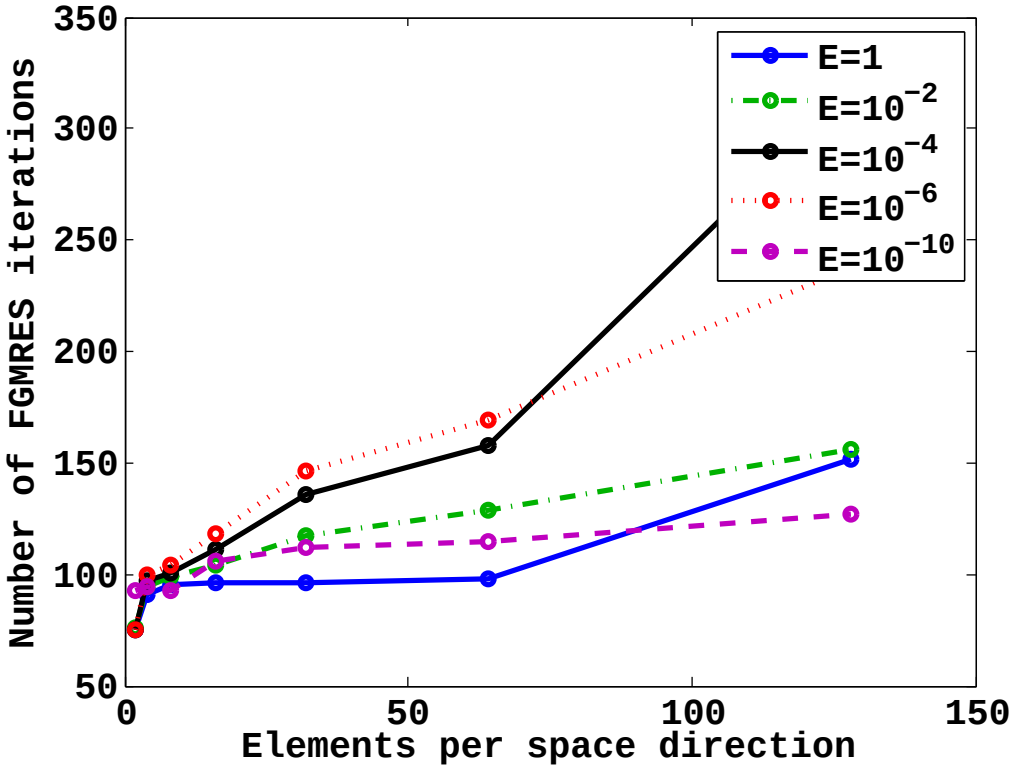

Figure 3.1: Iterations required to solve the system when only the permeability constant is allowed to be discontinuous. The parameter E indicates the amount of discontinuity.

Table 3.4: Preconditioning analysis with Example 2 in the case of a 6 orders of magnitude layer. The systems $\tilde{\boldsymbol{S}}_2 y_2 = x_2$ and $\tilde{\boldsymbol{S}}_3 y_3 = x_3$ are solved to a tolerance of 10^{-2}.

h	Δt	FGRMRES	$\tilde{\boldsymbol{S}}_2$ Iterations		$\tilde{\boldsymbol{S}}_3$ Iterations	
			total	per loop	total	per loop
$\frac{1}{2}$	$\frac{1}{2}$	107	3937	36.79	200	1.87
$\frac{1}{4}$	$\frac{1}{8}$	116	13332	114.93	250	2.16
$\frac{1}{8}$	$\frac{1}{32}$	113	24634	218.00	369	3.27
$\frac{1}{16}$	$\frac{1}{128}$	119	33915	285.00	482	4.05
$\frac{1}{32}$	$\frac{1}{512}$	156	64344	412.46	707	4.53
$\frac{1}{64}$	$\frac{1}{2048}$	192	173662	904.49	968	5.04
$\frac{1}{128}$	$\frac{1}{8192}$	241	303786	1260.52	1300	5.39

Chapter 4

Conclusions and Future Work

4.1 Conclusions

We have attempted to solve the linear system arising from a mixed finite element discretization of Biot's equations. Since this system is large and sparse, we have attempted to use FGMRES with a preconditioner based on an approximation to the Schur complement. To test the validity of this preconditioner, we have attempted to solve several test problems.

First we tested the case of a single homogeneous medium. We use an example problem designed to prove an analytic solution. A convergence study shows the expected convergence rates for all variables. In this case, experimental results indicate only a slight increase in the number of FGMRES iterations required for solution as the number of unknowns is increased. We have used PCG, preconditioned with AMG to apply these preconditioners. The tolerance specified to PCG is 10^{-2} as there did not appear to be a reduction in the number of FGMRES iterations required when a higher tolerance was used. When the grid size h and the time step Δt are varied at corresponding rates, we are able to apply the preconditioner in a small number of iterations that does not appear to depend heavily on the number of unknowns.

Secondly, we consider a problem in a heterogeneous medium. We allow the material parameters to be discontinuous along one interface that is aligned with the finite element grid. Again, an example problem, with analytic solution (2.10), was constructed to meet all of the interface requirements described in Section 2.5. In the case where we allow all of the parameters to be discontinuous, we see that the number of FGMRES iterations required does not increase significantly as the discontinuity is allowed to be greater.

However when attempting to apply the $\tilde{\boldsymbol{S}}_2$ part of the preconditioner, we see a larger jump in the number PCG iterations required. In the case of a moderate jump, 3 orders of magnitude difference, the number of PCG iterations is much larger than those required with continuous coefficients. However, the number of PCG iterations does not appear to be increasing as the number of elements is increased. As the jump in the coefficients gets larger, the number of PCG iterations required does appear to be increasing with the number of elements.

Ideally, we would like to alter the preconditioner $\tilde{\boldsymbol{S}}_2$ or its application so that it can be applied in a roughly constant number of iterations regardless of the size of the discontinuity or the number of elements used. Thus far, attempts have been made to alter the preconditioner using only diagonal parts. Specifically, we attempted to use $\tilde{\boldsymbol{S}}_2 = diag(\Delta t C_{\boldsymbol{qq}} + \Delta t^2 C_{\boldsymbol{qp}} \boldsymbol{D}_{C_{pp}}^{-1} C_{\boldsymbol{qp}}^T)$ and $\tilde{\boldsymbol{S}}_2 = \Delta t C_{\boldsymbol{qq}} + diag(\Delta t^2 C_{\boldsymbol{qp}} \boldsymbol{D}_{C_{pp}}^{-1} C_{\boldsymbol{qp}}^T)$ as attempted by Toh et.al [54] and Haga et. al. [24] respectively. Although both of these approximations reduced the number of PCG iterations required to apply the preconditioner, both of these less accurate approximations greatly increased the number of FGMRES iterations required to solve the system, and introduced a more dramatic dependence on the number of elements used.

4.2 Future Work

In the future, we would like to attempt to find a better way to apply the preconditioner $\tilde{\boldsymbol{S}}_2$. As stated above, we have attempted to apply this preconditioner using AMG as a preconditioner to the CG method. We may be able to improve on this application method by constructing a better preconditioner, using other smoothing methods or by using alternative forms of AMG. Some work on this topic has been done for example in [2, 22]. Thus far, attempts to apply these ideas have not yielded positive results.

We would also like to extend this work to other model equations. For example, we would like to consider a model that allows for wave propagation though the porous media,

or a model for flow over a porous media.

References

[1] Y. Abousleiman, A.H.-D. Cheng, L. Cui, E. Detournay, and J.-C. Roegiers. Mandel's problem revisited. *Geotechnique*, 46(2):187–195, 1996.

[2] D. N. Arnold, R. S. Falk, and R. Winther. Preconditioning in $H(\mathrm{div})$ and applications. *Math. Comp.*, 66(219):957–984, 1997.

[3] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. Appl. Math*, 9(1):17–29, 1951.

[4] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition*. SIAM, Philadelphia, PA, 1994.

[5] J. Bear and Y. Bachmat. *Introduction to Modelling Phenomena of Transport in Porous Media*. Kluwer Academic, Dordrecht, 1991.

[6] M. Benzi and A. J. Wathen. Some preconditioning techniques for saddle point problems. In *Model order reduction: theory, research aspects and applications*, pages 195–211. Springer, 2008.

[7] M. A. Biot. General theory of three-dimensional consolidation. *Journal of applied physics*, 12(2):155–164, 1941.

[8] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*. Springer, 1991.

[9] W. L. Briggs, S. F McCormick, et al. *A Multigrid Tutorial*. SIAM, 2000.

[10] Lawrence Livermore National Laboratory Center for Applied Scientific Computing. hypre users' guide. https://computation-rnd.llnl.gov/linear_solvers/software.php, 2012. [Online; accessed 4-November-2014].

[11] K. Chen. *Matrix Preconditioning Techniques and Applications*. Number 19. Cambridge University Press, 2005.

[12] S.-C. Chen and Y.-N. Wang. Conforming rectangular mixed finite elements for elasticity. *Journal of Scientific Computing*, 47(1):93–108, 2011.

[13] Y. K. Cheung and L. G. Tham. Numerical solutions for Biot's consolidation of layered soil. *Journal of Engineering Mechanics*, 109(3):669–679, 1983.

[14] S. C. Cowin. Bone poroelasticity. *Journal of Biomechanics*, 32(3):217–238, 1999.

[15] G. Dillion, V. E. Howle, and R. C. Kirby. Block preconditioners for Biots equations. https://bearspace.baylor.edu/Robert_Kirby/www/papers/DHKCopper2014.pdf, submitted to SIAM Journal of Scientific Computing. [Online; accessed 20-November-2014].

[16] H. C. Elman, D. J. Silvester, and A. J. Wathen. *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics: with Applications in Incompressible Fluid Dynamics*. Oxford University Press, 2005.

[17] R. D. Falgout. An introduction to algebraic multigrid computing. *Computing in Science Engineering*, 8(6):24–33, Nov 2006.

[18] R. D. Falgout and U. M. Yang. hypre: A library of high performance preconditioners. In *Computational ScienceICCS 2002*, pages 632–641. Springer, 2002.

[19] C. A. Felippa. Chapter 10 superelements and global-local analysis. http://www.colorado.edu/engineering/cas/courses.d/IFEM.d/IFEM.Ch10.d/IFEM.Ch10.pdf, 2014. [Online; accessed 17-November-2014].

[20] M. Fortin and F. Brezzi. *Mixed and hybrid finite element methods*. Springer, 1991.

[21] D. Galloway and F. S. Riley. San Joaquin Valley, California. Largest human alteration of the Earths surface. *Land subsidence in the United States US Geological Survey Circular*, 1182:23–34, 1999.

[22] F. J. Gaspar, J. L. Gracia, F. J. Lisbona, and C. W. Oosterlee. Distributive smoothers in multigrid for problems with dominating grad–div operators. *Numerical linear algebra with applications*, 15(8):661–683, 2008.

[23] B. Gurevich and M. Schoenberg. Interface conditions for Biots equations of poroelasticity. *The Journal of the Acoustical Society of America*, 105(5):2585–2589, 1999.

[24] J. B. Haga, H. Osnes, and H. P. Langtangen. Efficient block preconditioners for the coupled equations of pressure and deformation in highly discontinuous media. *International Journal for Numerical and Analytical Methods in Geomechanics*, 35(13):1466–1482, 2011.

[25] J. B. Haga, H. Osnes, and H. P. Langtangen. On the causes of pressure oscillations in low-permeable and low-compressible porous media. *International Journal for Numerical and Analytical Methods in Geomechanics*, 36(12):1507–1522, 2012.

[26] E. Hellinger. Die allgemeinen ansätze der mechanik der kontinua. enc. d. math. *Encyklopädie der mathematischen Wissenschaften*, 30:602–694, 1914.

[27] R. H. W. Hoppe. Chapter 7 mixed finite element methods. `http://www.math.uh.edu/~rohop/spring_11/downloads/Chapter7.pdf`, 2011. [Online; accessed 28-August-2013].

[28] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge university press, 2012.

[29] J. Hu and Z. Shi. Lower order rectangular nonconforming mixed finite elements for plane elasticity. *SIAM Journal on Numerical Analysis*, 46(1):88–102, 2008.

[30] J. A. Hudson, O. Stephansson, J. Andersson, C.-F. Tsang, and L. Jing. Coupled T–H–M issues relating to radioactive waste repository design and performance. *International Journal of Rock Mechanics and Mining Sciences*, 38(1):143–161, 2001.

[31] R. W. Lewis and B. A. Schrefler. *The finite element method in the static and dynamic deformation and consolidation of porous media*. John Wiley, 1998.

[32] J. Li and Y.T. Chen. *Computational partial differential equations using MATLAB*, volume 17. Chapman & Hall/CRC, 2008.

[33] K. Lipnikov. Numerical methods for the Biot model in poroelasticity, 2002.

[34] K. Lipnikov. Discussion, Summer 2013.

[35] J. Mandel. Consolidation des sols (étude mathématique)*. *Geotechnique*, 3(7):287–299, 1953.

[36] G. I. Marchuk and Kuznetsov Yu. A. On optimal iteration processes. *Soviet Math. Dokl.*, 9(4):1041–1045, 1968.

[37] A. Mikelić, B. Wang, and M. F. Wheeler. Numerical convergence study of iterative coupling for coupled flow and geomechanics. *Comput. Geosci.*, 18(3-4):325–341, 2014.

[38] J. P. Morris, W. W. McNab, S. K. Carroll, Y. Hao, W. Foxall, and J. L. Wagoner. Injection and reservoir hazard management: the role of injection-induced mechanical deformation and geochemical alteration at in salah co2 storage project: status report quarter end, june 2009. Technical report, LLNL Technical Report, doi: 10.2172/964517, 2009.

[39] M. F. Murphy, G. H. Golub, and A. J. Wathen. A note on preconditioning for indefinite linear systems. *SIAM Journal on Scientific Computing*, 21(6):1969–1972, 2000.

[40] A. Naumovich. *Efficient numerical methods for the Biot poroelasticity system in multilayered domains*. Vom fachbereich mathematik, Der Universitat Kaiserslautern, Zur verleihung des akademischen grades, 2007.

[41] Univ. of Tennessee and Oak Ridge National Laboratory. Iterative template routine. `http://www.netlib.org/templates/matlab/`, 1993. [Online; accessed 29-July-2013].

[42] P. J. Phillips and M. F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity I: the continuous in time case. *Computational Geosciences*, 11(2):131–144, 2007.

[43] P. J. Phillips and M. F. Wheeler. A coupling of mixed and discontinuous Galerkin finite-element methods for poroelasticity. *Computational Geosciences*, 12(4):417–435, 2008.

[44] PhillipJoseph Phillips and MaryF. Wheeler. Overcoming the problem of locking in linear elasticity and poroelasticity: an heuristic approach. *Computational Geosciences*, 13(1):5–12, 2009.

[45] R. Rajapakse. Stress analysis of borehole in poroelastic medium. *Journal of engineering mechanics*, 119(6):1205–1227, 1993.

[46] P.-A. Raviart and J.-M. Thomas. A mixed finite element method for 2-nd order elliptic problems. In *Mathematical aspects of finite element methods*, pages 292–315. Springer, 1977.

[47] T. Roose, P. A. Netti, L. L. Munn, Y. Boucher, and R. K. Jain. Solid stress generated by spheroid growth estimated using a linear poroelasticity model. *Microvascular research*, 66(3):204–212, 2003.

[48] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[49] Y. Saad and M. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986.

[50] A. Settari, D. A. Walters, G. A. Behie, et al. Use of coupled reservoir and geomechanical modelling for integrated reservoir analysis and management. *Journal of Canadian Petroleum Technology*, 40(12), 2001.

[51] R. E. Showalter. Diffusion in poro-elastic media. *Journal of Mathematical Analysis and Applications*, 251(1):310–340, 2000.

[52] X.-W. Tang and K. Onitsuka. Consolidation of double-layered ground with vertical drains. *International Journal for Numerical and Analytical Methods in Geomechanics*, 25(14):1449–1465, 2001.

[53] K. Terzaghi. *Theoretical Soil Mechanics*. J. Wiley, 1943.

[54] K.-C. Toh, K.-K. Phoon, and S.-H. Chan. Block preconditioners for symmetric indefinite linear systems. *International Journal for Numerical Methods in Engineering*, 60(8):1361–1381, 2004.

[55] L. N. Trefethen and D. Bau III. *Numerical Linear Algebra*, volume 50. SIAM, 1997.

[56] A. Verruijt. Theory and problems of poroelasticity. `http://geo.verruijt.net/software/PoroElasticity2013.pdf`, 2014. [Online; accessed 17-November-2014].

[57] J. Wan. *Stabilized Finite Element Methods for Coupled Geomechanics and Multiphase Flow*. PhD thesis, Stanford University, 2002.

[58] H. F. Wang. *Theory of Linear Poroelasticity with Applications to Geomechanics and Hydrogeology*. Princeton University Press, 2000.

[59] S.-Y. Yi. A new nonconforming mixed finite element method for linear elasticity. *Mathematical Models and Methods in Applied Sciences*, 16(07):979–999, 2006.

[60] S.-Y. Yi. Convergence analysis of a new mixed finite element method for Biot's consolidation model. *Numer. Methods Partial Differential Equations*, 30(4):1189–1210, 2014.

[61] C. Zhang and H. Chen. Mixed finite element method with interface-fitted grids for the interface problems. In *Information Science and Technology (ICIST), 2011 International Conference on*, pages 251–254. IEEE, 2011.

Curriculum Vitae

Maranda Lee Bean was born in Sioux City, Iowa. The only daughter of Barbara Bean, she graduated from Franklin High School, El Paso, Texas, in the spring of 2003. She then entered The New Mexico Institute of Mining and Technology and, in the spring of 2007, earned a Bachelor of Science degree in Physics. She spent several years working as an online tutor.

In the fall of 2010, she entered Graduate School at The University of Texas at El Paso. She became a teaching assistant for the Mathematics department in January of 2011. In the fall of 2012, she earned a Master of Science degree in Mathematics. Then in the Spring of 2013, she entered the Computational Science PhD program at The University of Texas at El Paso as a research assistant supported a National Science Foundation Grant (DMS 1217123). In 2014, she published *An Immersed Interface Method for a 1D Poroelasticity Problem with Discontinuous Coefficients* in the *Journal of Computational and Applied Mathematics* with her mentor, Dr. Son-Young Yi.

Permanent address: 7117 Orizaba Ave.
El Paso, Texas 79912