ABSTRACT

Preconditioning Mixed Finite Elements
for Tide Models

O. Tate Kernell, Ph.D.

Mentor: Robert C. Kirby, Ph.D.

We describe finite element methods for the linearized rotating shallow water equations which govern tides. Symplectic Euler and Crank-Nicolson time-stepping strategies have good energy preservation properties, which is desirable for tide modeling, but require careful treatment of linear algebra. For symplectic Euler, we have to invert the Raviart-Thomas element mass matrix at every time step. Thus we give estimates for the eigenvalues of these mass matrices. Crank-Nicolson, being fully implicit, has a more complicated system of equations which requires inverting the entire system. For this, we present an effective block preconditioner using parameter-weighted norms in $H(\mathrm{div})$. We give results that are nearly dependent of the given constants. Finally, we provide numerical results that confirm this theory.

Preconditioning Mixed Finite Elements
for Tide Models

by

O. Tate Kernell, B.S., M.S.

A Dissertation

Approved by the Department of Mathematics

Lance L. Littlejohn, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

Robert C. Kirby, Ph.D., Chairperson

Ron Morgan, Ph.D.

Johnny Henderson, Ph.D.

Jeonghun Lee, Ph.D.

Alex Yokochi, Ph.D.

Accepted by the Graduate School
August 2019

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I'd first and foremost like to thank my advisor, Dr. Robert Kirby, for seemingly possessing endless patience throughout all these years. I've enjoyed learning so much from him, and I appreciate how much he has helped me grow as a mathematician.

I would also like to thank my family. I could not have made it this far without their support. No matter how difficult things got, they were always there to lean on.

Lastly, I want to thank my friends who stayed by my side, no matter how busy I became, they made sure that I would occasionally relax and have fun.

CHAPTER ONE

Introduction

Tide modeling is an important component in many areas of scientific research. From coastal flooding and sediment transportation to ocean circulation and deep mixing, the accurate modeling of tides has widespread value in the scientific community. In [15], Garrett and Kunze assert that internal tides play a role in dissipating tidal energy and result in deep ocean mixing. Likewise [29] deliver similar findings concerning interior ocean mixing by winds and tides. Unstructured triangular meshes appear to be useful in modeling the ocean with finite element methods [41]. More specifically, it is enough to use the rotating shallow water equations in order to model tidal forces far from the coasts without including nonlinear advection terms. Often, a parameterized drag term is added to account for friction along the bottom of the ocean [38]. Additionally, these models can increase in complexity by adding more dissipative terms, such as different damping to account for various ocean features, or the global model can be used to create a more advanced regional barotropic tidal model [18, 20]. Many papers have dealt with various aspects of numerical dispersion relations produced from discretizations of the rotation shallow water equations [9, 10, 14, 26, 27, 32, 33, 34].

Cotter and Shipton, in [11], first suggest how mixed finite element methods can be used for the discretization of dynamical cores for numerical weather prediction. They go on to discuss how the correct application of these methods can preserve energy

[11]. These methods, must satisfy the conditions of finite element exterior calculus [2]. Furthermore in [12], Kirby and Cotter study mixed finite element methods for the linearized rotating shallow water equations with linear drag after choosing finite element spaces with a natural discrete Helmholtz decomposition. They go on to prove long-time stability of the system without energy accumulation, along with L^2 error estimates for the linearized momentum and free surface elevation. Their work focuses on linear drag bottom drag (suggested in [25]), even though quadratic drag is more realistic [19, 38]. Likewise, this dissertation will only be concerned with linear damping in order to preserve linearity and allow for easier analysis. Following from that work, they go on to prove the geotryptic state for the same equations but with nonlinear damping using an equivalent second-order formulation [19], again choosing spaces with a natural discrete Helmholtz decomposition. Cotter, Kirby, and Graber analyze the time-dependent attracting solution of barotropic tidal model. This attracting solution of the system is the solution which all solutions converge to as time approaches infinity. Other papers, such as [18], suggest iterative methods can be used to approximate the attracting solution, but we are not concerned with these approaches here.

In this dissertation, our goal is to provide a good preconditioner for discretizations of the linearized rotating shallow-water equations. We accomplish this by extending know effective techniques for the acoustic wave equation to account for additional terms (damping, Coriolis, etc.). Thus, much of our analysis will begin with discretizing the acoustic wave equation. One of the two time discretizations considered in this dissertation, symplectic Euler, has been analyzed in relation to first-order form of

this equation in [23]. Kirby and Kieu demonstrate that the semidiscrete method exactly conserves the system energy and show that symplectic Euler conserves a nearby functional equivalent to the energy. Prior to that publication, Geveci first discussed energy conservation by applying mixed finite element methods to the first order form of the acoustic wave equation in [17]. Likewise, [13, 21] give similar analysis of the second order in time wave equation. In relation to these works, we focus on analyzing the additional terms within $H(\text{div})$ that differentiate the tide model from the acoustic wave equation.

The structure of this dissertation is as follows. In Chapter Two we discuss the mathematical preliminaries required by the remainder of the dissertation. We first outline the finite element method, followed by a discussion of function spaces, especially the Raviart-Thomas space. Then we turn our attention to preconditioning and iterative solvers, specifically how they apply to solving PDE systems. Chapter Three introduces the variants of the acoustic wave and rotating shallow water equations that we will focus on throughout the following chapters. We discuss the symplectic Euler and implicit Crank-Nicolson time stepping methods. We choose these because they both have good energy conserving properties and are reasonably stable (symplectic Euler is stable for small time steps). In Chapter Four we study the mass matrix of the Raviart-Thomas element which is inverted at every time step during the symplectic Euler time stepping method when applied to the tide model. We show how element geometry affects the conditioning of the Raviart-Thomas element mass matrix. Numerical results are included that support our findings. Chapter Five outlines block preconditioning the implicit time stepping method applied to the tide model. Here,

we present a weighted norm that bounds the eigenvalues of the preconditioned system. We provide numerical results to confirm our theory. Chapter Six gives general conclusions and possible future work.

CHAPTER TWO

Preliminaries

This chapter deals with basic finite element method theory, function spaces, the Raviart-Thomas element, and preconditioning that will be required by the rest of the dissertation.

2.1 The Finite Element Method

The finite element method is a numerical method for providing approximate solutions to partial differential equations (PDEs). PDEs are differential equations with functions depending on two or more independent variables (such as x, y, t) and their partial derivatives. Many physics based problems are described by various partial differential equations, including, but not limited to, acoustics, heat transfer, electromagnetics, and fluid flow. Many of these problems are extremely difficult or impossible to solve exactly through standard analytical methods. Thus, we have to turn to numerical methods such as finite element methods to provide approximate solutions. The finite element method excels on unstructured geometries.

The steps of implementing the finite element method are as follows: Determine the weak form of the PDE, discretize the problem and restrict the weak form to a subspace, and finally solve the discrete or algebraic problem. In order to show this, we will briefly outline an example problem. For a more detailed explanation of this example, or the finite element method in general, please refer to [22, 30]. We take a

look at the boundary value problem on $[0,1]$

$$\begin{aligned} -u''(x) &= f(x), \quad \text{for } 0 < x < 1, \\ u(0) &= 0, \\ u(1) &= 0, \end{aligned} \tag{2.1}$$

where f is some continuous function given by the problem and $u' = \frac{du}{dx}$. Clearly this problem has an unique solution u discovered by basic integration.

Before we go further, we want to introduce the linear space H_0^1, where

$$\begin{aligned} H_0^1([0,1]) = \{v : v \text{ is defined on } [0,1] \text{ and } & \int_0^1 v^2 dx < \infty; \\ v' \text{ is defined on } [0,1] \text{ and } & \int_0^1 (v')^2 dx < \infty; \\ v(0) = v(1) = 0\}. & \end{aligned} \tag{2.2}$$

Additionally, some necessary notation follows, so we write

$$(v, w) = \int_0^1 v(x)w(x)dx \tag{2.3}$$

for real-valued, piecewise, continuous bounded functions, which is the $L^2([0,1])$ inner product [22].

We now look to the first step and determine the weak form, also known as the variational form, of the equation (2.1). We know that if u is a solution to (2.1) then u is a solution to the weak form. Therefore, by multiplying (2.1) by a test function v and integrating by parts, we want to find $u \in H_0^1$ such that

$$a(u, v) = (f, v), \quad \forall\, v \in H_0^1, \tag{2.4}$$

where

$$a(u,v) = \int_0^1 u'v'dx \tag{2.5}$$

and

$$(f,v) = \int_0^1 fvdx. \tag{2.6}$$

The next step is to discretize and restrict the weak form to a finite-dimensional subspace. We subdivide our interval $(0,1)$ into "elements" created by letting $0 = x_0 < x_1 < ... < x_N < x_{N+1} = 1$. Each element is then a subinterval defined as $E_i = (x_{i-1}, x_i)$ with length $h_i = x_i - x_{i-1}$ for $i = 1, ..., N+1$. Notice that if we divide the interval into equivalent partitions, $h = h_i$ for every i. Additionally we define

$$h = \max_i h_i. \tag{2.7}$$

This then measures the refinement of the partition of the subspace, which is used in bounding convergence rates. Now, we can define V_h of the space H_0^1 as

$$V_h = \left\{ v : \int_{E_i} v^2 dx < 0 \,\forall\, i; \int_{E_i} (v')^2 dx \,\forall\, i; v(0) = v(1) = 0 \right\}, \tag{2.8}$$

where $v|_E$ is v restricted to an element E. Notice that $V_h \subset H_0^1$.

It's important to also mention how we deal with two dimensional domains. For some domain Ω, we generate a mesh by triangulation. This means we divide up Ω into a set $T_h = \{K_i\}_{i=1}^N$ of non-overlapping triangles K_i, such that

$$\Omega = \bigcup_{K \in T_h} = K_1 \cup K_2 \cup ... \cup K_N, \tag{2.9}$$

where no vertex of one triangle lives on the edge of another triangle [22]. Then, h is defined as the maximum of the longest edge of K_i for any i. The rest follows in a similar manner.

Returning to 1-D, we now define basis functions for V_h, which we call ϕ_j with $j = 1, ..., N$, defined as

$$\phi_j(x_i) = \begin{cases} 1 & \text{if } i = j \\ \\ 0 & \text{if } i \neq j, j = 1, ..., N \end{cases}, \tag{2.10}$$

where ϕ_i is continuous, piecewise, linear. Additionally, at each node x_i we let $\eta_i = v(x_i)$ for $i = 0, ..., N + 1$. Thus we can represent any $v \in V_h$ as

$$v(x) = \sum_{i=1}^{N} \eta_i \phi_i(x) \tag{2.11}$$

for $x \in [0, 1]$. Thus we can reformulate our problem as follows. Find $u_h \in V_h$ such that

$$(u_h', v') = (f, v) \quad \forall\, v \in V_h \tag{2.12}$$

or

$$a(u_h, v) = (f, v) \quad \forall\, v \in V_h. \tag{2.13}$$

If u_h satisfies (2.12) and if we choose test functions as ϕ_i for $i = 1, ..., N$, then we have that

$$(u_h', \phi_i) = (f, \phi_i). \tag{2.14}$$

Additionally,

$$u_h(x) = \sum_{j}^{N} \xi_j \phi_j(x), \text{ where } \xi_j = u_h(x_j). \tag{2.15}$$

with N equations and N unknowns. We can describe this linear system as

$$A\xi = b \tag{2.16}$$

where A is the $N \times N$ matrix with with elements $A_{ij} = (\phi_i', \phi_j')$. Also, we see b is the vector $[b_1, ..., b_N]^T$ where $b_i = (f, \phi_i)$. We call A the stiffness matrix and b the load vector. Notice when we have a uniform mesh, our linear system becomes

$$\frac{1}{h}\begin{bmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & & \vdots \\ 0 & -1 & 2 & -1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 2 & -1 \\ 0 & \dots & \dots & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \\ \vdots \\ \vdots \\ \xi_N \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \\ \vdots \\ \vdots \\ b_N \end{bmatrix}. \tag{2.17}$$

Finally, our last step is simply solving the linear system by applying Jacobi method, which requires a matrix to be diagonally dominant, or Conjugate Gradient, which requires a matrix to be symmetric and positive-definite. If the system is nonsymmetric, GMRES is often the preferred iterative solver.

In this example, we created finite elements on the interval $[0, 1]$. However, we will use more complicated elements throughout this dissertation. The main one we will implement is the Raviart-Thomas (RT) finite element.

2.2 Spaces

Since we will be dealing with a couple of different function spaces throughout this dissertation (including the function space H_0^1 from a previous chapter), we will briefly discuss them here. For a more detailed approach, see [22].

Suppose W is a linear space and L is a bounded linear form on W. Then for all $u, v \in W$ and $\alpha, \beta \in \mathbb{R}$

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v). \tag{2.18}$$

Additionally, a is bilinear on $W \times W$ if $\forall\, v, u, w \in W$ and $\alpha, \beta \in \mathbb{R}$,

$$\begin{aligned} a(\alpha u + \beta v, w) &= \alpha a(u, w) + \beta a(v, w) \\ a(u, \alpha v + \beta w) &= \alpha a(u, v) + \beta a(u, w), \end{aligned} \tag{2.19}$$

and a is symmetric on $W \times W$ if $a(v, w) = a(w, v)$ for all $v, w \in W$. We call a symmetric bilinear form $a(\cdot, \cdot)$ on $W \times W$ an inner product on W if

$$a(w, w) > 0 \,\forall\, w \in W, w \neq 0. \tag{2.20}$$

Using this definition above, we can define a norm $\|\cdot\|_a$ associated with the inner product $a(\cdot, \cdot)$ by $\|w\|_a = (a(w, w))^{\frac{1}{2}}$ for all $w \in W$.

Now that we have these definitions, we can define some various functions spaces that we will need for the following chapters. In general, we will primarily be working with Hilbert spaces. We call a W a Hilbert space if it is a complete inner product space with respect to the norm $\|\cdot\|_W$. For example, let Ω be a bounded domain $\mathbb{R}^2$.

Then, we define the space

$$L^2(\Omega) = \{v : v \text{ is defined on } \Omega \text{ and } \int_\Omega v^2 dx < \infty\}, \tag{2.21}$$

with the inner product $(v, w) = \int_\Omega vw dx$ and v, w defined on Ω and norm $\|v\| = (v, v)^{\frac{1}{2}}$. In this dissertation, $\|\cdot\|$ will represent the L^2 norm and $(\cdot, \cdot)$ will represent the L^2 inner product. Similarly, we can define another Hilbert space by writing

$$H^1(\Omega) = \{v : v \text{ and } v' \text{ belong to } L^2(\Omega)\}. \tag{2.22}$$

This space is equipped with the inner product $(v, w)_{H^1} = \int_\Omega vw + v'w' dx$ and a norm defined in the same way. Note that the space H^1 is fundamental in the analysis and discretization of weak forms for second-order elliptic problems [24]. Now, we can redefine H_0^1 as

$$H_0^1(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \partial\Omega\}. \tag{2.23}$$

H_0^1 shares the same norm and inner product as H^1 and thus is also a Hilbert space. Lastly, we define possibly the most important space for this paper $H(\text{div})$. We have

$$H(\text{div}) = \{v \in (L^2(\Omega))^2; \nabla \cdot v \in L^2(\Omega)\}, \tag{2.24}$$

and it is equipped with the inner product $(u, v)_{H(\text{div})} = (u, v) + (\nabla \cdot u, \nabla \cdot v)$. This space naturally occurs in connections with mixed formulations of second-order elliptic problems [24].

Since we have discussed different types of norms in this section, it is only appropriate to mention the norm $\|A\|_p$ with $p = 1, 2$ for some matrix $A \in \mathbb{R}^{m \times n}$. We turn the reader to [7] for more detailed coverage of this topic. Below, we present the

1-norm and 2-norm:

$$\begin{aligned} \|A\|_1 &= \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|, \\ \|A\|_2 &= \rho(A^TA)^{1/2} = \rho(A^TA)^{1/2} = \max_{1\leq j\leq n} \lambda_j(A^TA)^{1/2}, \end{aligned} \tag{2.25}$$

where λ represents the vector of eigenvalues of A and ρ is the spectral radius. Remember that for a matrix $A \in \mathbb{R}^{n\times n}$, the eigenvalues λ_i and eigenvectors x_i for $i = 1, ..., n$ are

$$Ax_i = \lambda_i x_i. \tag{2.26}$$

Additionally, the spectral radius is defined as

$$\rho(A) = \max_i |\lambda_i|, \tag{2.27}$$

for $i = 1, ..., n$. Thus, for A, we can write the definition of the condition number κ as

$$\kappa(A) = \|A\|\|A^{-1}\| \tag{2.28}$$

for some matrix norm $\|\cdot\|$. If we choose the 2-norm from above, and restrict A to be symmetric and nonsingular, we find that

$$\kappa_2(A) = \|A\|_2\|A^{-1}\|_2 = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}. \tag{2.29}$$

Now we have enough background to define the finite elements used in this dissertation.

2.3 The Raviart-Thomas and L^2 Finite Elements

While we have summarized the application of the finite element method, we have yet to actually give a definition for a finite element. The finite element was first introduced by Ciarlet in [8]. We give the definition from [22] as follows:

Definition 2.3.1. *A finite element is defined to be a triple* (K, P_K, Σ)*, where*

$$\begin{cases} K \text{ is a geometric object, for example a triangle,} \\ P_K \text{ is a finite-dimensional linear space of functions defined on } K, \\ \Sigma \text{ is a set of degrees of freedom,} \end{cases} \tag{2.30}$$

such that a function $v \in P_K$ *is uniquely determined by the degrees of freedom of* Σ*.*

Now that we have the general definition for a finite element, we can define some important finite elements for this dissertation. While there are many finite elements, we will cover the Lagrange element and the Raviart-Thomas element. For further information about these elements or others, reference [24].

The best known finite element is the $\mathscr{P}_1$ Lagrange element, which is an H^1 finite element space. We can define it from [24] as such

$$\begin{aligned} &K \in \{\text{interval, triangle, tetrahedron}\}, \\ &P_K = \mathscr{P}_q(K), \\ &\Sigma = v(x^i), \text{ with } i = 1, ..., n(q), \end{aligned} \tag{2.31}$$

for $q = 1, 2, ...$ and where $\{x^i\}_{i=1}^{n(q)}$ is an enumeration of points in K defined by

$$x = \begin{cases} \frac{i}{q},\ 0 \le i \le q, & K \text{ interval}, \\ \left(\frac{i}{q}, \frac{j}{q}\right),\ 0 \le i + j \le q, & K \text{ triangle}, \\ \left(\frac{i}{q}, \frac{j}{q}, \frac{k}{q}\right),\ 0 \le i + j + k \le q, & K \text{ tetrahedron}. \end{cases} \tag{2.32}$$

Additionally,

$$n(q) = \begin{cases} q+1, & K \text{ interval}, \\ \frac{1}{2}(q+1)(q+2), & K \text{ triangle}, \\ \frac{1}{6}(q+1)(q+2)(q+3), & K \text{ tetrahedron}, \end{cases} \tag{2.33}$$

which is the dimension of the complete polynomials of degree q on K. This element has a mass matrix that is always diagonally dominant. Thus, it is natural to precondition this matrix with a diagonal preconditioner.

Next, we would like to define an element on $H(\text{div})$, the Raviart-Thomas element [31]. Raviart-Thomas elements are commonly used for wave and shallow water equations. Once again, we look to Kirby and Logg [24] for the definition of the RT_q element below

$$K \in \{\text{triangle, tetrahedron}\},$$

$$P_K = [\mathscr{P}_{q-1}(K)]^d + x\mathscr{P}_{q-1}(K),$$

$$\Sigma = \begin{cases} \int_f v \cdot npds, & \text{for a set of basis functions } p \in \mathscr{P}_{q-1}(f) \text{ for each facet } f, \\ \int_K v \cdot pdx, & \text{for a set of basis functions } p \in [\mathscr{P}_{q-2}(K)]^d \text{ for } q \geq 2 \end{cases} \tag{2.34}$$

with dimension

$$n(q) = \begin{cases} q(q+2), & K \text{ triangle}, \\ \frac{1}{2}q(q+1)(q+3), & K \text{ tetrahedron}. \end{cases} \tag{2.35}$$

Note that for any element to be an $H(\text{div})$ finite element, only the normal components must be continuous. An important inequality that will be used in this dissertation

for the Raviart-Thomas space is the inverse assumption, which we state as

$$\exists\, C : \|\nabla \cdot v\|_{H(\mathrm{div})} \leq \frac{C}{h}\|v\|_{L^2} \tag{2.36}$$

for $v \in V_h$. We will assume throughout this dissertation that our meshes allow for the standard inverse assumptions.

2.4 Preconditioning and Iterative Solvers

Another important topic we must discuss is that of preconditioning and iterative solvers. Both of these topics have been outlined in various books and papers, but we reference [5, 40, 7, 39] in particular for this section and direct the reader there for more detailed coverage.

Whenever partial differential equations (PDEs) are discretized (as we will see in the next chapter), generally a large matrix problem is generated with the form

$$Ax = b, \tag{2.37}$$

where $A \in \mathbb{R}^{n\times n}$ and b is a given vector. The goal is to solve for x, which requires the inversion of A. While direct methods, that factorize A into more manageable pieces, work well for fairly small problem, these methods often struggle when A becomes too large. Thus, iterative methods, and more specifically Krylov subspace methods, are a common choice for large-scale problems. However, for some problems, iterative methods can take impractically long to terminate. Preconditioning can be the answer to this issue and provide convergence in acceptable time frames.

Preconditioning is the conversion of a system into a new system with properties that favor the implementation of iterative methods. Some methods, such as Conjugate Gradient, prefer a preconditioner to cluster eigenvalues around 1. Others, such as GMRES, perform better with clustered eigenvalues away from 0.

We then give a preconditioned form of (2.37) as

$$M^{-1}Ax = M^{-1}b \tag{2.38}$$

where M is the preconditioner. Obviously, if we could easily invert A, the perfect preconditioner would be $M = A$, and we would have our solution. However, since inverting A is the problem, we want M to somewhat approximate A in order to create a system that is easier to solve. Clearly, (2.38) has the same solution as (2.37). Similarly, we can also precondition (2.37) from the right by writing

$$AM^{-1}y = b, \quad x = M^{-1}y. \tag{2.39}$$

Benzi, in [5] claims there are two general requirements a good preconditioner M should follow:

- The preconditioned system should be easy to solve.
- The preconditioner should be cheap to construct and apply.

Note that for methods like GMRES, the $M^{-1}A$ is never fully formed.

However, the structure of A matters. If a matrix is diagonally dominant where

$$|A_{ii}| \geq \sum_{i \neq j} |A_i j| \quad \forall\, i, \tag{2.40}$$

the Jacobi method provides fast convergence unless A is ill-conditioned. More often though, Jacobi is simply used as a preconditioner. If a matrix is symmetric positive-definite, where

$$x^T Ax > 0 \quad \forall\, x \in \mathbb{R}^n,\ x \neq 0, \tag{2.41}$$

then the Conjugate Gradient method works well, also as long as A isn't ill-conditioned. If a matrix is nonsymmetric, GMRES is often the first choice. We provide the Jacobi and GMRES methods below, with additional details for other iterative methods in [39].

For the Jacobi method, we require A to be a square matrix. Then, we can write

$$A = D + R, \tag{2.42}$$

where D is the diagonal of A and R is the remainder when the diagonal is removed. We can acquire the solution interatively by

$$x^{(n)} = D^{-1}(b - Rx^{n-1}), \tag{2.43}$$

where x^n is the nth iteration.

We now turn our attention to GMRES, which has a much more extensive discussion. We define the Krylov subspace methods which approximate the solution of (2.37) in the Krylov subspaces

$$\mathscr{K}_m(A, b) = \text{span}\{b, Ab, A^2b, ..., A^{m-1}b\}, \tag{2.44}$$

for $m = 1, 2, ...$, which generate basis vectors. These vectors are columns of a Krylov

matrix $K_m \in \mathbb{R}^{n\times m}$. These methods compute iterates

$$x_m = p(A)b, \quad m = 1, 2, ..., \tag{2.45}$$

which are approximations, with $x_0 = 0$, and p is a polynomial of order less than or equal to $m - 1$.

This dissertation requires the use of GMRES, the generalized minimal residual method [36], which will be outlined here from [39]. First, however, we must mention the Arnoldi iteration [3] in order to explain GMRES. The Arnoldi process converts a matrix A to Hessenberg form H, which is a matrix with zeros either above or below the first subdiagonal. We can write A as

$$AQ = QH. \tag{2.46}$$

Let Q_m be the $n \times m$ matrix whose columns are the first m columns of Q. We assume n is extremely large in this case. Define $\tilde{H}_m$ to be the $(m+1) \times m$ upper-left section of H. Then we get

$$AQ_m = Q_{m+1}\tilde{H}_m. \tag{2.47}$$

We can define an orthonormal basis for $\mathscr{K}_M$ as $\{q_i\}_{i=1}^m$, which are the column vectors of Q_m. Thus, K_m must have a reduced QR factorization

$$K_m = Q_m R_m, \tag{2.48}$$

where $R_m \in \mathbb{R}^{m\times m}$.

Now we can outline GMRES. For convenience, let $x_* = A^{-1}b$ (the exact solution of the system). The strategy of GMRES is simple; at every step m, we approximate

x_* by the vector $x_m \in \mathscr{K}_m$ that minimizes the norm of the residual $r_m = b - Ax_m$. To do this, an Arnoldi iteration is required to generate a sequence of Krylov matrices Q_m with columns $q_1, q_2, ...$ that span the successive Krylov subspaces $\mathscr{K}_m$, giving $x_m = Q_m y$. This provides a least squares problem: find a vector $y \in \mathbb{R}^m$ such that

$$\|AQ_m y - b\|_2 = \text{ minimum } . \tag{2.49}$$

While it may seem this problem has dimensions $n \times m$, in fact it is essentially of dimensions $(m+1) \times m$. We demonstrate this below by applying the Arnoldi iteration to AQ_m to get

$$\|Q_{m+1}\tilde{H}_m y - b\|_2 = \text{ minimum } . \tag{2.50}$$

We now have that bot y and b are in the column space of Q_{m+1}. We can multiply on the left by Q^*_{m+1} to get

$$\|\tilde{H}_m y - Q^*_{m+1} b\|_2 = \text{ minimum } . \tag{2.51}$$

Finally we arrive at

$$\|\tilde{H}_m y - \|b\|_2 e_1\|_2 = \text{ minimum} \tag{2.52}$$

by recognizing that $Q^*_{m+1} b = \|b\|_2 e_1$, with $e_1 = (1, 0, 0, ...)^*$, by the construction of the Krylov matrices $\{Q_m\}$. At step m we solve for y and let $x_m = Q_m y$, thus completing the process.

CHAPTER THREE

The Tide Model

We now turn our attention to the main topic of this dissertation, the tide model. We define these equations on Ω, a two dimensional surface in $\mathbb{R}^2$. Ω can also be curved, but for the purposes of this dissertation, we will focus on the flat case. The tide model is as follows

$$\begin{aligned} u_t + \frac{f}{\epsilon}u^{\perp} + \frac{\beta}{\epsilon^2}\nabla(\eta - \eta') + C(u) &= F \\ \eta_t + \nabla \cdot (Hu) &= 0, \end{aligned} \tag{3.1}$$

where $u : \Omega \to \mathbb{R}^2$ is the velocity field tangent to Ω, $u^{\perp} = (-u_2, u_1)$ is the velocity rotated by $\pi/2$, η is the height of the tide or wave compared to its standard height, $\nabla\eta'$ is the tidal forcing, ϵ is the Rossby number (which is small), f is the Coriolis parameter (which is a sine function), β is the Burger number (which is small), H is the fluid depth at rest, C is the damping, and ∇ and $\nabla\cdot$ are the intrinsic gradient and divergence operators on Ω, respectively [12].

3.1 Wave Equation

In order to build the required theory for the tide model, we will start from the linear acoustic wave equation.

$$\begin{aligned} qu_t + \nabla p &= 0, \\ k^{-1}p_t + \nabla \cdot u &= 0, \end{aligned} \tag{3.2}$$

on some domain $\Omega \times [0,T] \subset \mathbb{R}^d \times \mathbb{R}$ with $d = 2, 3$, with the assumption that Ω is a bounded domain with a polyhedral boundary. The parameter q, the material density, is a measurable function bounded above and below by positive q_* and q^*, respectively. The parameter k is the bulk modulus of compressibility, which we assume is bounded by positive k_* and k^* [23].

Remark 3.1.1. *It is important to note that for L^2 vector-valued, $H(\text{div}) := \{v \in L^2,\ \nabla \cdot (v) \in L^2\}$ and $H_0(\text{div}) := \{v \in H(\text{div}),\ v \cdot \nu|_{\partial\Omega} = 0\}$ where ν is the outward normal vector field on $\partial\Omega$. Also note, L^2_0 is the space L^2 with zero mean.*

For this dissertation we will assume $q = 1 = k$. Additionally, we impose the boundary condition $u \cdot \nu = 0$ on $\partial\Omega$ where ν is the unit outward normal to Ω. We choose initial conditions

$$\begin{aligned} p(x,0) &= p_0(x) \text{ and} \\ u(x,0) &= u_0(x). \end{aligned} \tag{3.3}$$

Converting this system into weak form and integrating by parts gives

$$\begin{aligned} (u_t, v) - (p, \nabla \cdot v) + \underbrace{\langle p, v \cdot \nu \rangle_{\partial\Omega}}_{=0} &= (f, v), \\ (p_t, w) + (\nabla \cdot u, w) &= (g, w), \end{aligned} \tag{3.4}$$

where $u : [0,T] \to V \equiv H_0(\text{div})$ and $p : [0,T] \to W \equiv L^2_0$, along with the initial conditions (3.3). This leads to our final form

$$\begin{aligned} (u_t, v) - (p, \nabla \cdot v) &= (f, v), \\ (p_t, w) + (\nabla \cdot u, w) &= (g, w). \end{aligned} \tag{3.5}$$

3.2 Mixed Formulation

Let $\{T_h\}_h$ be a family of quasiuniform triangulations of Ω [6]. We let $W = L^2(\Omega)$ and V the subspace of $H(\text{div})$ with vanishing normal trace [23]. Additionally, we let V_h be the Raviart-Thomas space of order $r \geq 0$ over each triangulation T_h and W_h the space of discontinuous piecewise polynomials of degree r over T_h [23]. Then, the semidiscrete mixed formulation of (3.5) is to find $u_h : [0, T] \to V_h$ and $p_h : [0, T] \to W_h$ such that

$$\begin{aligned} (u_{h,t}, v_h) - (p_h, \nabla \cdot v_h) &= (f, v_h), \\ (p_{h,t}, w_h) + (\nabla \cdot u_h, w_h) &= (g, w_h), \end{aligned} \tag{3.6}$$

for all $v_h \in V_h$ and $w_h \in W_h$ [23] where $V_h \subset V$ and $W_h \subset W$. Note that Geveci [17] has already provided both existence and uniqueness proofs of the solution of (3.6). Additionally, Geveci showed stability for L^2. We can then partition the time interval $[0, T]$ into time steps $0 \equiv t_0 < t_1 < t_2 < ... < t_N$, where $t_i = i\Delta t$ in order to prepare for time stepping methods. We now take a look at both the Crank-Nicolson and symplectic Euler time stepping methods of this mixed formulation.

3.2.1 Crank-Nicolson Time Discretization

For our implicit method, we apply the Crank-Nicolson method to approximate the solution to the semidiscrete mixed formulation (3.6). We chose Crank-Nicolson primarily because it is exactly energy conserving, but also it provides the benefit of

being absolutely stable. Here, $u_h(t_n) \approx u_h^n \in V_h$ and $p_h(t_n) \approx p_h^n \in W_h$

$$\begin{aligned}\left(\frac{u_h^{n+1}-u_h^n}{\Delta t}, v_h\right) - \left(\frac{p_h^{n+1}+p_h^n}{2}, \nabla \cdot v_h\right) &= \left(f^{n+\frac{1}{2}}, v_h\right), \\ \left(\frac{p_h^{n+1}-p_h^n}{\Delta t}, w_h\right) + \left(\nabla \cdot \frac{u_h^{n+1}+u_h^n}{2}, w_h\right) &= \left(g^{n+\frac{1}{2}}, w_h\right),\end{aligned} \tag{3.7}$$

where $f^{n+\frac{1}{2}} = \frac{f(t_{n+1})+f(t_n)}{2}$ and likewise for g. Letting $f = 0$ and $g = 0$ and multiplying by Δt, we get

$$\begin{aligned}\left(u_h^{n+1}-u_h^n, v_h\right) - \left(\frac{\Delta t}{2}\left(p_h^{n+1}+p_h^n\right), \nabla \cdot v_h\right) &= 0 \\ \left(p_h^{n+1}-p_h^n, w_h\right) + \left(\frac{\Delta t}{2}\nabla \cdot \left(u_h^{n+1}+u_h^n\right), w_h\right) &= 0,\end{aligned} \tag{3.8}$$

Reshuffling terms in (3.8) leads to

$$\begin{aligned}\left(u_h^{n+1}, v_h\right) - \frac{\Delta t}{2}\left(p_h^{n+1}, \nabla \cdot v_h\right) &= \tilde{F}, \\ \left(p_h^{n+1}, w_h\right) + \frac{\Delta t}{2}\left(\nabla \cdot u_h^{n+1}, w_h\right) &= \tilde{G},\end{aligned} \tag{3.9}$$

where

$$\begin{aligned}\tilde{F} &= \left(u_h^n, v_h\right) + \frac{\Delta t}{2}\left(p_h^n, \nabla \cdot v_h\right), \text{ and} \\ \tilde{G} &= \left(p_h^n, w_h\right) - \frac{\Delta t}{2}\left(\nabla \cdot u_h^n, w_h\right).\end{aligned} \tag{3.10}$$

Let $\{\phi_i\}_{i=1}^{|W_h|}$ and $\{\psi_i\}_{i=1}^{|V_h|}$ be bases for W_h and V_h respectively. Then we can define mass matrices

$$\begin{aligned}\tilde{M}_{ij} &= (\phi_j, \phi_i), \\ M_{ij} &= (\psi_j, \psi_i).\end{aligned} \tag{3.11}$$

We can rewrite (3.6) as

$$\begin{bmatrix} M & 0 \\ 0 & \tilde{M} \end{bmatrix} \begin{bmatrix} u_t \\ p_t \end{bmatrix} + \begin{bmatrix} 0 & -D^T \\ D & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{3.12}$$

where

$$D_{ij} = (\nabla \cdot \psi_i, \phi_j), \tag{3.13}$$

is the discrete div operator. In the discretized case, we have a similar structure from (3.9), which we write in matrix form

$$\mathscr{A}_h \begin{bmatrix} u_h^{n+1} \\ \\ p_h^{n+1} \end{bmatrix} = \begin{bmatrix} \tilde{F} \\ \\ \tilde{G} \end{bmatrix}, \tag{3.14}$$

where

$$\mathscr{A}_h = \begin{bmatrix} M & -\frac{\Delta t}{2} D^T \\ \\ \frac{\Delta t}{2} D & \tilde{M} \end{bmatrix}, \tag{3.15}$$

which gives our fully discretized system. To solve this system, we have to invert $\mathscr{A}$ which is nonsymmetric. Thus, we turn to GMRES, which is often applied with a preconditioner P.

We want to show our matrix $\mathscr{A}_h$, multiplied with some preconditioner P is bounded and invertible with respect to a chosen norm. Additionally, we will investigate this question: As the mesh is refined, will the scale of inverse matrices be uniformly bounded in norm? We see that by stripping off the block diagonal we are left with a skew perturbation of a SPD matrix. Unfortunately the bilinear form for (3.14) is not coericive. However, we can still provide an inf-sup condition in order to bound the preconditioned system below in norm. Looking at (3.9), we can view that system

of two bilinear forms as a bilinear form on the Cartesian product, test and trial are pairs.

$$\begin{aligned} a((u,p),(v,w)) &= (u_h^{n+1}, v_h) - k(p_h^{n+1}, \nabla \cdot v_h) + k(\nabla \cdot u_h^{n+1}, w_h) + (p_h^{n+1}, w_h) \\ &= a(U,V), \end{aligned} \tag{3.16}$$

where $U = (u,p)$, $V = (v,w)$, and $k = \frac{\Delta t}{2}$. We see we can substitute U for V and get

$$\begin{aligned} a(U,U) &= (u_h^{n+1}, u_h^{n+1}) - k(p_h^{n+1}, \nabla \cdot u_h^{n+1}) + k(\nabla \cdot u_h^{n+1}, p_h^{n+1}) + (p_h^{n+1}, p_h^{n+1}) \\ &= (u_h^{n+1}, u_h^{n+1}) + (p_h^{n+1}, p_h^{n+1}) \\ &= \|u_h^{n+1}\|^2 + \|p_h^{n+1}\|^2, \end{aligned} \tag{3.17}$$

since the middle two terms cancel each other out. However, this only proves coercivity in $(L^2)^2 \times L^2$, which is not what we need, since the tide model and acoustic wave equations live in $H(\text{div}) \times L^2$. Clearly, further analysis will be needed. By applying Cauchy-Schwarz and the inverse estimate we see

$$\begin{aligned} a(U,V) &\leq \|u_h^{n+1}\|\|v_h\| + \frac{kC_I}{h}\|p_h^{n+1}\|\|v_h\| + \frac{kC_I}{h}\|u_h^{n+1}\|\|w_h\| + \|p_h^{n+1}\|\|w_h\| \\ &\leq \left(2 + \frac{2kC_I}{h}\right)\|U\|\|V\|. \end{aligned} \tag{3.18}$$

Obviously, we care about preconditioning to be able to better control number of iterations. We know that we will have to invert the entire system when solving a Crank-Nicolson time stepping method. Thus, we will discuss how we can designate a weighted norm to bound the eigenvalues of the preconditioned system independently from the parameters.

3.2.2 Symplectic Euler

Similarly, we can choose symplectic Euler as another time stepping method. Here we give the rule, once again with $u_h(t_n) \approx u_h^n \in V_h$ and $p_h(t_n) \approx p_h^n \in W_h$

$$\begin{aligned} \left(\frac{u_h^{n+1} - u_h^n}{\Delta t}, v_h\right) - \left(p_h^{n+1}, \nabla \cdot v_h\right) &= (f^{n+1}, v_h) \\ \left(\frac{p_h^{n+1} - p_h^n}{\Delta t}, w_h\right) + (\nabla \cdot u_h^n, w_h) &= (g^n, w_h), \end{aligned} \tag{3.19}$$

where $f^n = f(t_n)$ and likewise for g. Kirby and Kieu in [23] demonstrated that the fully discrete symplectic Euler method in time exactly conserves an energy functional that is equivalent to the actual energy under a CFL condition. Letting $f = 0$ and $g = 0$ and multiplying by Δt we get

$$\begin{aligned} \left(u_h^{n+1} - u_h^n, v_h\right) - \Delta t \left(p_h^{n+1}, \nabla \cdot v_h\right) &= 0, \\ \left(p_h^{n+1} - p_h^n, w_h\right) + \Delta t \left(\nabla \cdot u_h^n, w_h\right) &= 0. \end{aligned} \tag{3.20}$$

Reshuffling terms in (3.21) leads to

$$\begin{aligned} \left(u_h^{n+1}, v_h\right) - \Delta t \left(p_h^{n+1}, \nabla \cdot v_h\right) &= \tilde{F}, \\ \left(p_h^{n+1}, w_h\right) &= \tilde{G}, \end{aligned} \tag{3.21}$$

where

$$\begin{aligned} \tilde{F} &= \left(-u_h^n, v_h\right), \\ \tilde{G} &= \left(-p_h^n, w_h\right) + \Delta t \left(\nabla \cdot u_h^n, w_h\right). \end{aligned} \tag{3.22}$$

In this case, we have

$$\mathscr{A}_h \begin{bmatrix} u_h^{n+1} \\ \\ p_h^{n+1} \end{bmatrix} = \begin{bmatrix} \tilde{F} \\ \\ \tilde{G}, \end{bmatrix} \tag{3.23}$$

where

$$\mathscr{A}_h = \begin{bmatrix} M & -\Delta t\, D^T \\ 0 & \tilde{M} \end{bmatrix}, \tag{3.24}$$

which gives our semidiscrete symplectic Euler. When solving, we take the second equation in (3.21), and solve by inverting $\tilde{M}$ to get p_h^{n+1}. Next, we substitute that p_h^{n+1} into the first equation and invert M to solve for u_h^{n+1} [23]. Unless $\tilde{M}$ is ill-conditioned, it is easy to invert, since it is symmetric positive-definite. However, M is not so straightforward. We would like to use a diagonal preconditioner on it, but does that make sense? We will explore this more in the next chapter.

3.3 Preconditioning

Returning to the PDE (3.2) with coefficients equal to 1, if we apply Crank-Nicolson in the time derivative without discretizing in space, it becomes

$$\begin{aligned} \frac{u^{n+1}-u^n}{\Delta t} + \nabla\left(\frac{1}{2}\left(p^{n+1}+p^n\right)\right) &= 0, \\ \frac{p^{n+1}-p^n}{\Delta t} + \nabla\cdot\left(\frac{1}{2}\left(u^{n+1}+u^n\right)\right) &= 0, \end{aligned} \tag{3.25}$$

which leads to

$$\begin{aligned} u^{n+1} + \frac{\Delta t}{2}\nabla p^{n+1} &= u^n - \frac{\Delta t}{2}\nabla p^n \\ p^{n+1} + \frac{\Delta t}{2}\nabla\cdot u^{n+1} &= p^n - \frac{\Delta t}{2}\nabla\cdot u^n. \end{aligned} \tag{3.26}$$

Therefore, at each time step we have a discretization of the coefficient operator $\mathscr{A}$, described as

$$\mathscr{A} = \begin{pmatrix} I & k\,\mathrm{grad} \\ k\,\nabla\cdot & I \end{pmatrix} \tag{3.27}$$

where $k = \frac{\Delta t}{2}$.

By discretizing in the finite element space $H_0(\text{div}) \times L^2_0$, as seen in (3.14), we recover our finite dimensional coefficient operator, $\mathscr{A}_h$, defined in (3.15). We claim $\mathscr{A}$ is an isomorphism mapping $H(\text{div}) \times L^2$ onto $H(\text{div})^* \times (L^2)^*$, its dual space. In the view of [40] a common approach to preconditioning is to create an equivalent operator that is easier to invert numerically. Equivalent in this sense means that $\mathscr{B}^{-1}\mathscr{A}$ is a nice operator from the initial space into itself rather than into its dual, where $\mathscr{B}$ is the preconditioner. If $\mathscr{B}^{-1}\mathscr{A}$ is bounded in the Hilbert space, we get mesh independent eigenvalue clustering [28]. Our goal is to find a preconditioner $\mathscr{B}$ which maps $H(\text{div})^* \times L^2$ onto $H(\text{div}) \times L^2$. This preconditioner will be explored below. We can also formulate this problem in an alternative way to be on the space $L^2 \times H^1$. This method is based on the Schur complement, but will not be explored in this dissertation.

From methods described in [28], we want our preconditioner to be a block diagonal operator suggested by the mapping properties of the coefficient operator of the system. The preconditioner for our specific coefficient operator $\mathscr{A}$ utilizes the Riesz map and is derived from the problem's spaces as seen below

$$\mathscr{B} = \begin{bmatrix} \beta\, I - \alpha\, \text{grad}\, \nabla\cdot & 0 \\ 0 & \gamma\, I \end{bmatrix}. \tag{3.28}$$

Here, if $\alpha = \beta = \gamma = 1$, $\mathscr{B}$ is the canonical Riesz map preconditioner that maps the dual space back to our original space. Similarly, the discrete preconditioner is of the

form

$$\mathscr{B}_h = \begin{bmatrix} \alpha(\nabla\cdot,\nabla\cdot) + \beta(\cdot,\cdot) & 0 \\ 0 & \gamma(\cdot,\cdot) \end{bmatrix}. \tag{3.29}$$

For this preconditioner, our goal is to look at the eigenvalues of $\mathscr{B}_h^{-1}\mathscr{A}_h$ as a function of α, β, and γ. We will choose our weighted operator norm so that eigenvalues of the preconditioned system are bounded in the chosen norm. While this is already shown for the wave equation [28], we will prove a theorem about the boundedness of a preconditioner and its inverse for the tide model.

Note that our coefficient operator $\mathscr{A}$ is a bounded map with bounded inverse from $H(\mathrm{div}) \times L^2$ into its dual. We can then premultiply with $\mathscr{B}$, the Riesz map, thus giving that $\mathscr{B}^{-1}\mathscr{A}$ is a bounded operator. Our goal is then to find a ball that bounds the eigenvalues of our operator regardless of the mesh refinement. Additionally, we would like to manipulate α, β, and γ so the ball is also independent of the size of the time step. We will use this on the coefficient operator of the system (3.14) and observe how well it performs. If the preconditioner is easily invertible, and the system has parameter independent bounds, then we have found a good preconditioner for this system.

For explicit time stepping methods, like symplectic Euler, we must invert the mass matrices M and $\tilde{M}$ at every time step to solve. Thus, we are concerned with how we can improve the conditioning of the mass matrix. Since the L^2 mass matrix $\tilde{M}$ is constant up to some multiple dependent on the mesh size, we will concentrate on the Raviart-Thomas mass matrix M. We will prove a theorem in chapter Four on how properties of the mesh determine diagonal dominance.

CHAPTER FOUR

The Mass Matrix

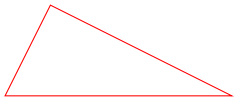
Since we must invert the Raviart-Thomas mass matrix, M, for explicit time stepping methods, we might want to ask when is M diagonally dominant. While we know M is positive definite, diagonal dominance is a stronger condition. Not only that, it also allows us to give eigenvalue bounds through the Gershgorin Circle Theorem [16]. The statement of this theorem follows from Gershgorin,

"Let A be a complex $n \times n$ matrix, with entries a_{ij}. For $i \in \{1, ..., n\}$, let $R_i = \sum_{j \neq i} |a_{ij}|$ be the sum of the absolute values of the non-diagonal entries in the i-th row. Let $D(a_{ii}, R_i)$ be the closed disc centered at a_{ii} with radius R_i. Such a disc is called a Gershgorin disc."

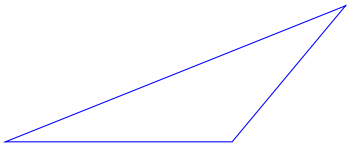
For symplectic Euler, our mass matrices are all symmetric positive-definite. Thus, we have matrices with real eigenvalues in $\mathbb{R}$. Regardless, the Gershgorin Circle Theorem will provide the necessary eigenvalue bounds on $\mathbb{R}^1$ as long as the bounds are greater than zero. Conveniently, P^1 Lagrange finite element mass matrices are always diagonally dominant and in fact are equivalent up to scaling, so this theorem automatically applies. However, the 2-D Raviart-Thomas finite elements in (2.34) are best known for the mixed Poisson equation, but are often used in wave and shallow water equations. These matrices, unfortunately, are not all diagonally dominant. We ask ourselves if it is even reasonable to precondition them with a diagonal preconditioner. The final results of this chapter suggest, in fact, it is not such a bad idea. Since this

is a fairly difficult problem to do in generality, we will look at the lowest order case and try to find the explicit formula for the matrix entries of M_T, the Raviart-Thomas element mass matrix of a triangle T. We hope to determine for which triangles in a mesh this mass matrix is diagonally dominant.

As an example, we provide two triangles below:

$T_1 =$  (4.1)

with $\theta_{min} \approx 26.6°$ and $\theta_{max} = 90.0°$.

$T_2 =$  (4.2)

with $\theta_{min} \approx 21.8°$ and $\theta_{max} \approx 129.8°$. Clearly, T_1 seems like the "nicer" triangle. However, when we determine the diagonal dominance for each row, we receive surprising results with

$$\tilde{\Delta}(M_{T_2}) = \begin{bmatrix} -0.042 \\ -0.042 \\ 0.208 \end{bmatrix}, \tag{4.3}$$

and

$$\tilde{\Delta}(M_{T_2}) = \begin{bmatrix} 0.263 \\ 0.046 \\ 0.015 \end{bmatrix}, \tag{4.4}$$

where $\tilde{\Delta}(M_T)_i = |(M_T)_{ii}| - \sum_{i \neq j} |(M_T)_{ij}|$, $\forall\, i$. Now, we investigate why this occurs.

4.1 Theorem

Let $\mathscr{R}$ as the region bounded by the y-axis and the polar equations

$$\begin{aligned} r^2 - 3 &= r\cos\theta, \\ \frac{r^2+5}{5} &= r\cos\theta, \\ \frac{3+3r^2}{7} &= r\cos\theta, \\ \frac{1+5r^2}{5} &= r\cos\theta, \\ 3r^2 - 1 &= r\cos\theta. \end{aligned} \tag{4.5}$$

Theorem 4.1.1. *Let* $\mathscr{T} = \{conv((0,0),(1,0),(x,y)) : (x,y) \in \mathscr{R}\}$. *Then for any* T, M_T *is diagonally dominant if and only if* $\exists\, T_0 \in \mathscr{T} : T \sim T_0$.

Proof. Now, we must compute a element mass matrix for a triangular Raviart-Thomas element. Let the vertices of T be defined as $v_i = (x_i, y_i)$ for $i = 1, 2, 3$. We define our 2-D RT basis functions ψ_i below by letting

$$\psi_i = \begin{bmatrix} a_i + c_i x \\ b_i + c_i y \end{bmatrix} \text{ for } i = 1, 2, 3. \tag{4.6}$$

Note that these basis functions have the property that for any triangle T, with edge midpoints e_j and normals n_j on each edge for $i, j = 1, 2, 3$,

$$\psi_i(e_j) \cdot n_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \tag{4.7}$$

Then

$$\psi_i(e_j) = \begin{bmatrix} a_i \\ b_i \end{bmatrix} + c_i \begin{bmatrix} e_j^x \\ e_j^y \end{bmatrix} \tag{4.8}$$

and

$$\psi_i(e_j) \cdot n_j = a_i n_j^x + b_i n_j^y + c_i(e_j^x n_j^x + e_j^y + n_j^y). \tag{4.9}$$

We set up a system of equations

$$\begin{bmatrix} n_1^x & n_1^y & e_1^x n_1^x + e_1^y n_1^y \\ n_2^x & n_2^y & e_2^x n_2^x + e_2^y n_2^y \\ n_3^x & n_3^y & e_3^x n_3^x + e_3^y n_3^y \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4.10}$$

and solve to get the coefficients for each ψ_i.

Now by letting $|T|$ be the area of T, we can compute the integral of the dot product of pairs of basis functions,

$$(M_T)_{ij} = \int_T \psi_i \cdot \psi_j dx = \sum_{e \in T} \frac{|T|}{3} \psi_i(e) \cdot \psi_j(e) \tag{4.11}$$

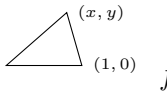
exactly by utilizing the midpoint rule, which exactly integrates quadratics.

This gives the Raviart-Thomas element mass matrix

$$\frac{1}{24|T|} \check{M}_T \in \mathbb{R}^{3\times 3} \tag{4.12}$$

where $1/24|T|$ we can factor from every term and

$$
\begin{aligned}
(\check{M}_T)_{11} &= 3x_1^2 - 3(x_2+x_3)x_1 + x_2^2 + x_3^2 + 3y_1^2 \\
&\quad + y_2^2 + y_3^2 + x_2x_3 + y_2y_3 - 3y_1(y_2+y_3), \\
(\check{M}_T)_{12} &= x_1^2 - 3x_2x_1 + x_2^2 - x_3^2 + y_1^2 + y_2^2 \\
&\quad - y_3^2 + (x_1+x_2)x_3 - 3y_1y_2 + (y_1+y_2)y_3, \\
(\check{M}_T)_{13} &= -x_1^2 + (x_2-3x_3)x_1 - x_2^2 + x_3^2 + y_1^2 \\
&\quad - y_2^2 + y_3^2 + x_2x_3 + y_1y_2 - 3y_1y_3 + y_2y_3, \\
(\check{M}_T)_{21} &= x_1^2 - 3x_2x_1 + x_2^2 - x_3^2 + y_1^2 + y_2^2 \\
&\quad - y_3^2 + (x_1+x_2)x_3 - 3y_1y_2 + (y_1+y_2)y_3, \\
(\check{M}_T)_{22} &= x_1^2 + (x_3-3x_2)x_1 + 3x_2^2 + x_3^2 + y_1^2 \\
&\quad + 3y_2^2 + y_3^2 - 3x_2x_3 - 3y_1y_2 + y_1y_3 - 3y_2y_3, \\
(\check{M}_T)_{23} &= -x_1^2 + (x_2+x_3)x_1 + x_2^2 + x_3^2 - y_1^2 \\
&\quad + y_2^2 + y_3^2 - 3x_2x_3 - 3y_2y_3 + y_1(y_2+y_3), \\
(\check{M}_T)_{31} &= -x_1^2 + (x_2-3x_3)x_1 - x_2^2 + x_3^2 + y_1^2 \\
&\quad - y_2^2 + y_3^2 + x_2x_3 + y_1y_2 - 3y_1y_3 + y_2y_3, \\
(\check{M}_T)_{32} &= -x_1^2 + (x_2+x_3)x_1 + x_2^2 + x_3^2 - y_1^2 \\
&\quad + y_2^2 + y_3^2 - 3x_2x_3 - 3y_2y_3 + y_1(y_2+y_3), \\
(\check{M}_T)_{33} &= x_1^2 + x_2x_1 + x_2^2 + 3x_3^2 + y_1^2 + y_2^2 \\
&\quad + 3y_3^2 - 3(x_1+x_2)x_3 + y_1y_2 - 3(y_1+y_2)y_3.
\end{aligned}
\tag{4.13}
$$

Lemma 4.1.2. *Every triangle is similar to* 

for some (x, y) *in* $Q1$. M_T *is invariant under similarity, meaning* $M_{T_1} = M_{T_2}$ *if* $T_1 \sim T_2$.

Proof. The Piola transform of an $H(\text{div})$ function is:

$$P(u) = \frac{1}{\det J} Ju \circ F^{-1} \tag{4.14}$$

where $F(x) = Jx + b$ is the mapping from triangle T_1 to T_2 [24]. Here, $P(u)$ lives on T_2 and u lives on T_1. The Piola map preserves normal components so it maps the basis functions on one cell to the basis functions on another cell. Writing down the mass matrix for cell 2 (superscripts on ψ indicate which cell they are the RT basis functions for):

$$\begin{aligned} (M_{T_2})_{ij} &= \int_{T_2} \psi_i^2 \cdot \psi_j^2 \\ &= \int_{T_1} P(\psi^1)_i \cdot P(\psi^1)_j \det J dx \end{aligned}, \tag{4.15}$$

which is a change of coordinates that picks up a factor of $\det J$ through the chain rule. Then we have

$$= \int_{T_1} \frac{1}{\det J} (J\psi_i^1)^T (J\psi_j^1) dx \tag{4.16}$$

If $F(x)$ is a rotation, $J^T J = 1$ since the rotation matrix is orthogonal, and J has determinant 1. When $F(x)$ is a translation, then J is the identity. Lastly, if $F(x)$ is a dilation, J is some dilating constant c times the identity which produces a factor of c^2. However, $\det J$ also picks up a c^2 term since it is the square of the area of the triangle. Thus we have finished the proof. □

Next, to reduce the number of possible cases, we fix the vertices $v_1 = (0, 0)$ and $v_2 = (1, 0)$. This allows us to set $v_3 = (x, y)$ and analyze the element matrix on a plane, following our Piola transform proof. Making these substitutions gives our matrix

$$\check{M}_T = \begin{bmatrix} x^2 + x + y^2 + 1 & -x^2 + x - y^2 + 1 & -x^2 - x - y^2 + 1 \\ -x^2 + x - y^2 + 1 & x^2 - 3x + y^2 + 3 & x^2 - 3x + y^2 + 1 \\ -x^2 - x - y^2 + 1 & x^2 - 3x + y^2 + 1 & 3x^2 - 3x + 3y^2 + 1 \end{bmatrix}. \tag{4.17}$$

Since we were able to factor out the triangular element area and diagonal dominance is invariant under constant multiple, our analysis will only be focused on the matrix $\check{M}_T$. We can ignore this factor since it will only scale the eigenvalues. Thus it will have no effect on diagonal dominance. We determine where we will have diagonally dominance by region analysis on our element matrix. For any row, we compute if it is diagonally dominant by subtracting the absolute value of the off diagonal entries from the absolute value of the diagonal entry. Our analysis will be greatly simplified by converting the element matrix to polar coordinates by letting $x = r\cos\theta$ and $y = r\sin\theta$ to give

$$\begin{bmatrix} r^2 + r\cos\theta + 1 & -r^2 + r\cos\theta + 1 & -r^2 - r\cos\theta + 1 \\ -r^2 + r\cos\theta + 1 & r^2 - 3r\cos\theta + 3 & r^2 - 3r\cos\theta + 1 \\ -r^2 - r\cos\theta + 1 & r^2 - 3r\cos\theta + 1 & 3r^2 - 3r\cos\theta + 1 \end{bmatrix}. \tag{4.18}$$

We will require $\cos\theta > 0$ and only allow p_3 to live in the first quadrant. This is possible, since we can relate any triangular element back to this family of reference triangles that has been created. Now we will begin the analysis row by row.

Row 1. We begin by subtracting the absolute values of each off-diagonal element from the diagonal entry to get

$$|1 + r\cos\theta + r^2| - |1 - r^2 + r\cos\theta| - |1 - r^2 - r\cos\theta|. \tag{4.19}$$

Since the diagonal entries are always positive for a symmetric positive definite matrix, we only need to consider four cases for each row

$$\begin{aligned} \text{Case 1:} \quad & 0 < 1 - r^2 + r\cos\theta \text{ and } 0 < 1 - r^2 - r\cos\theta \\ \Rightarrow \quad & \boxed{r^2 < 1 + r\cos\theta} \text{ and } \boxed{r^2 + r\cos\theta < 1}, \end{aligned} \tag{4.20}$$

then

$$\begin{aligned} & 1 + r\cos\theta + r^2 - (1 - r^2 + r\cos\theta) - (1 - r^2 - r\cos\theta) \\ & = -1 + 3r^2 + r\cos\theta > 0, \text{ which is positive if } \boxed{3r^2 + r\cos\theta > 1}. \end{aligned} \tag{4.21}$$

This gives us the region created by the boxed inequalities as show by Case 1 in Figure 4.1. Thus, if (x, y) is in this region, Row 1 has a nonnegative sum. Similar analysis follows with the remaining cases.

$$\begin{aligned} \text{Case 2:} \quad & 1 - r^2 + r\cos\theta < 0 \text{ and } 0 < 1 - r^2 - r\cos\theta \\ \Rightarrow \quad & \boxed{1 + r\cos\theta < r^2} \text{ and } \boxed{r^2 + r\cos\theta < 1}, \end{aligned} \tag{4.22}$$

then

$$\begin{aligned} &1 + r\cos\theta + r^2 + (1 - r^2 + r\cos\theta) - (1 - r^2 - r\cos\theta) \\ &= \boxed{1 + r^2 + 3r\cos\theta > 0}. \end{aligned} \tag{4.23}$$

This case does not correspond to a region in the first quadrant, and thus is not represented on Figure 4.1.

$$\begin{aligned} \text{Case 3:} \quad & 0 < 1 - r^2 + r\cos\theta \text{ and } 1 - r^2 - r\cos\theta < 0 \\ \Rightarrow \quad & \boxed{r^2 < 1 + r\cos\theta} \text{ and } \boxed{1 < r^2 + r\cos\theta}, \end{aligned} \tag{4.24}$$

then

$$\begin{aligned} &1 + r\cos\theta + r^2 - (1 - r^2 + r\cos\theta) + (1 - r^2 - r\cos\theta) \\ &= 1 + r^2 - r\cos\theta > 0, \text{ which is positive if } \boxed{1 + r^2 > r\cos\theta}. \end{aligned} \tag{4.25}$$

The boxed inequalities provide a region for Case 3 as described by Figure 4.1.

$$\begin{aligned} \text{Case 4:} \quad & 1 - r^2 + r\cos\theta < 0 \text{ and } 1 - r^2 - r\cos\theta < 0 \\ \Rightarrow \quad & \boxed{1 + r\cos\theta < r^2} \text{ and } \boxed{1 < r^2 + r\cos\theta}, \end{aligned} \tag{4.26}$$

then

$$\begin{aligned} &1 + r\cos\theta + r^2 + (1 - r^2 + r\cos\theta) + (1 - r^2 - r\cos\theta) \\ &= 3 - r^2 + r\cos\theta > 0, \text{ which is positive if } \boxed{3 + r\cos\theta > r^2}. \end{aligned} \tag{4.27}$$

The boxed inequalities provide a region for Case 4 as described by Figure 4.1.

Taking a union of all of the regions from the first row, we create a new region over which we know the first row will be dominated by the entry on the diagonal. This region is depicted by Figure 4.1.

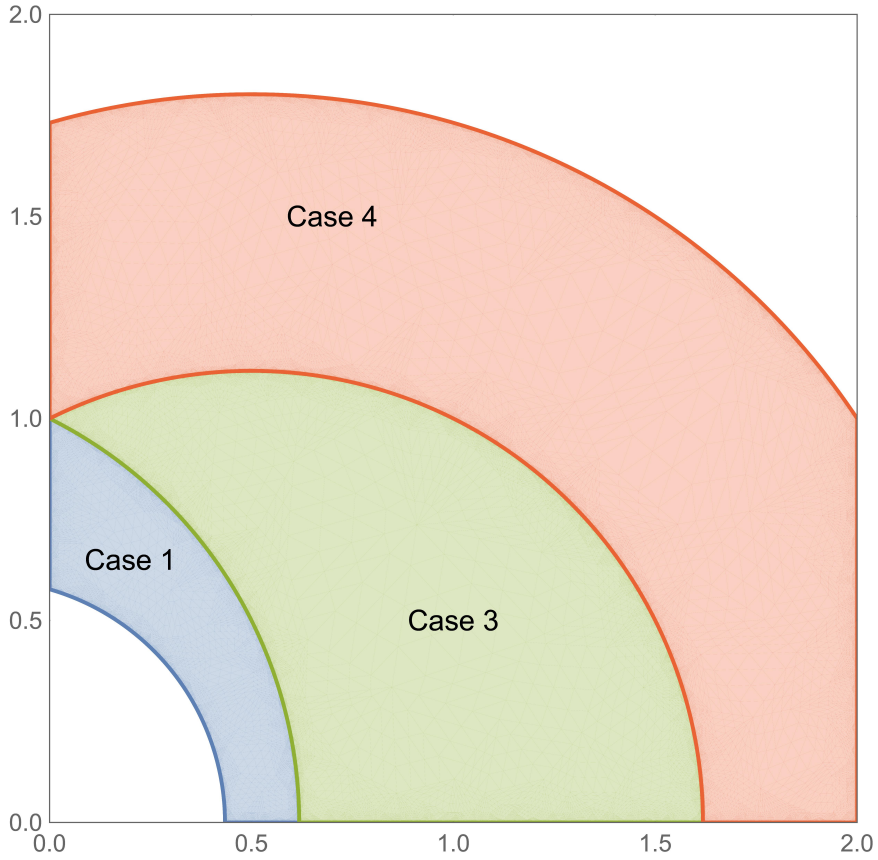


Figure 4.1: Row 1 Final Region. Case 1 refers to inequalities (4.20) and (4.21). Case 3 refers to inequalities (4.24) and (4.25). Case 4 refers to inequalities (4.26) and (4.27).

Row 2. Similarly, we set up our row analysis by subtracting the absolute values of the off-diagonal elements from the diagonal entry to give

$$|3 + r^2 - 3r\cos\theta| - |1 - r^2 + r\cos\theta| - |1 + r^2 - 3r\cos\theta|. \tag{4.28}$$

Once again, we only need to consider four cases since the diagonal entry is always positive.

$$\begin{aligned} \text{Case 1:} \quad & 0 < 1 - r^2 + r\cos\theta \text{ and } 0 < 1 + r^2 - 3r\cos\theta \\ & \Rightarrow \boxed{r^2 < 1 + r\cos\theta} \text{ and } \boxed{3r\cos\theta < 1 + r^2}, \end{aligned} \tag{4.29}$$

then

$$\begin{aligned} & (3 + r^2 - 3r\cos\theta) - (1 - r^2 + r\cos\theta) - (1 + r^2 - 3r\cos\theta) \\ & = 1 + r^2 - r\cos\theta > 0, \text{ which is positive if } \boxed{1 + r^2 > r\cos\theta}. \end{aligned} \tag{4.30}$$

The boxed inequalities provide a region for Case 1 as described by Figure 4.2.

$$\begin{aligned} \text{Case 2:} \quad & 0 < 1 - r^2 + r\cos\theta \text{ and } 1 + r^2 - 3r\cos\theta < 0 \\ & \Rightarrow \boxed{r^2 < 1 + r\cos\theta} \text{ and } \boxed{1 + r^2 < 3r\cos\theta}, \end{aligned} \tag{4.31}$$

then

$$\begin{aligned} & (3 + r^2 - 3r\cos\theta) - (1 - r^2 + r\cos\theta) + (1 + r^2 < 3r\cos\theta) \\ & = 3 + 3r^2 - 7r\cos\theta > 0, \text{ which is positive if } \boxed{3 + 3r^2 > 7r\cos\theta}. \end{aligned} \tag{4.32}$$

The boxed inequalities provide a region for Case 2 as described by Figure 4.2.

$$\begin{aligned} \text{Case 3:} \quad & 1 - r^2 + r\cos\theta < 0 \text{ and } 0 < 1 + r^2 - 3r\cos\theta \\ & \Rightarrow \boxed{1 + r\cos\theta < r^2} \text{ and } \boxed{3r\cos\theta < 1 + r^2}, \end{aligned} \tag{4.33}$$

then

$$\begin{aligned}&(3+r^2-3r\cos\theta)+(1-r^2+r\cos\theta)-(1+r^2-3r\cos\theta)\\&=3-r^2+r\cos\theta>0,\ \text{which is positive if } \boxed{3+r\cos\theta>r^2}.\end{aligned}\tag{4.34}$$

The boxed inequalities provide a region for Case 3 as described by Figure 4.2.

$$\begin{aligned}\text{Case 4:}\quad &1-r^2+r\cos\theta<0 \text{ and } 1+r^2-3r\cos\theta<0\\&\Rightarrow \boxed{1+r\cos\theta<r^2} \text{ and } \boxed{1+r^2<3r\cos\theta},\end{aligned}\tag{4.35}$$

then

$$\begin{aligned}&(3+r^2-3r\cos\theta)+(1-r^2+r\cos\theta)+(1+r^2-3r\cos\theta)\\&=5+r^2-5r\cos\theta>0,\ \text{which is positive if } \boxed{5+r^2>5r\cos\theta}.\end{aligned}\tag{4.36}$$

The boxed inequalities provide a region for Case 4 as described by Figure 4.2.

Taking a union of all of the regions from the second row, we create a new region over which we know the second row will be dominated by the diagonal entry. This region is depicted by Figure 4.2.

Row 3. This row follows in the same manner as the other with

$$|1+3r^2-3r\cos\theta|-|1-r^2-r\cos\theta|-|1+r^2-3r\cos\theta|.\tag{4.37}$$

Once again we will only have to consider four cases.

$$\begin{aligned}\text{Case 1:}\quad &0<1-r^2-r\cos\theta \text{ and } 0<1+r^2-3r\cos\theta\\&\Rightarrow \boxed{r^2+r\cos\theta<1} \text{ and } \boxed{3r\cos\theta<1+r^2},\end{aligned}\tag{4.38}$$

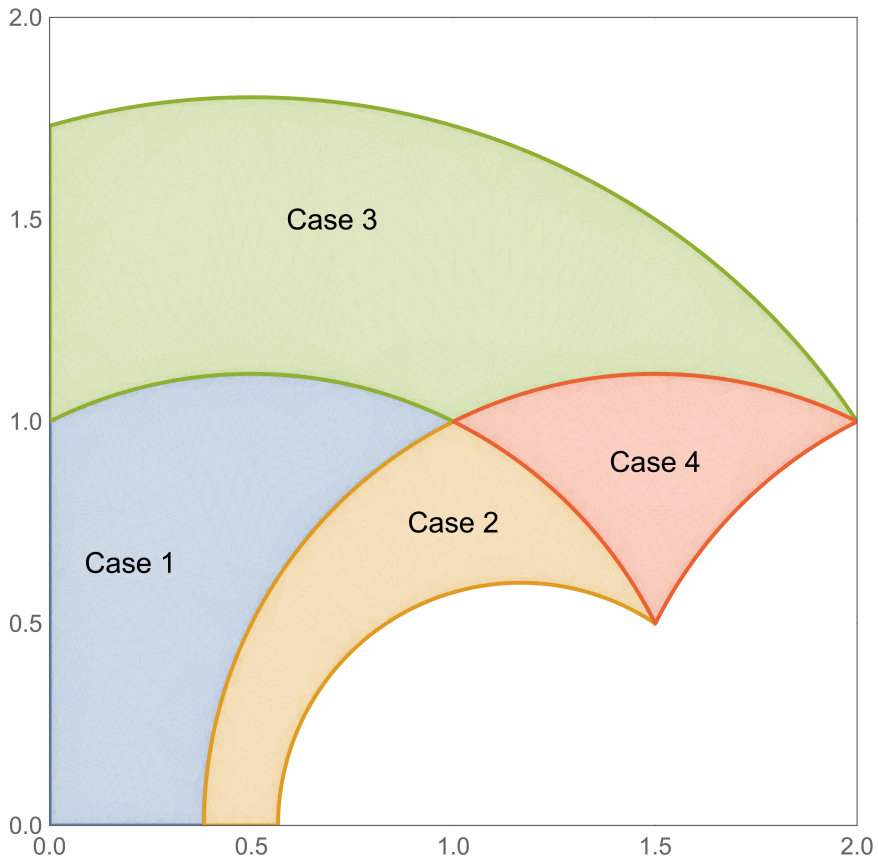


Figure 4.2: Row 2 Final Region. Case 1 refers to inequalities (4.1.1) and (4.30). Case 2 refers to inequalities (4.31) and (4.32). Case 3 refers to inequalities (4.33) and (4.34). Case 4 refers to inequalities (4.35) and (4.36).

then

$$\begin{aligned}&(1+3r^2-3r\cos\theta)-(1-r^2-r\cos\theta)-(1+r^2-3r\cos\theta)\\&=-1+3r^2+r\cos\theta>0,\ \text{which is positive if } \boxed{3r^2+r\cos\theta>1}.\end{aligned}\tag{4.39}$$

The boxed inequalities provide a region for Case 1 as described by Figure 4.3.

$$\begin{aligned}\text{Case 2:}\quad &0<1-r^2-r\cos\theta \text{ and } 1+r^2-3r\cos\theta<0\\&\Rightarrow \boxed{r^2+r\cos\theta<1} \text{ and } \boxed{1+r^2<3r\cos\theta},\end{aligned}\tag{4.40}$$

then

$$\begin{aligned}&(1+3r^2-3r\cos\theta)-(1-r^2-r\cos\theta)+(1+r^2-3r\cos\theta)\\&=1+5r^2-5r\cos\theta>0,\ \text{which is positive if } \boxed{1+5r^2>5r\cos\theta}.\end{aligned}\tag{4.41}$$

The boxed inequalities provide a region for Case 2 as described by Figure 4.3.

$$\begin{aligned}\text{Case 3:}\quad &1-r^2-r\cos\theta<0 \text{ and } 0<1+r^2-3r\cos\theta\\&\Rightarrow \boxed{1<r^2+r\cos\theta} \text{ and } \boxed{3r\cos\theta<1+r^2},\end{aligned}\tag{4.42}$$

then

$$\begin{aligned}&(1+3r^2-3r\cos\theta)+(1-r^2-r\cos\theta)-(1+r^2-3r\cos\theta)\\&=1+r^2-r\cos\theta>0,\ \text{which is positive if } \boxed{1+r^2>r\cos\theta}.\end{aligned}\tag{4.43}$$

The boxed inequalities provide a region for Case 3 as described by Figure 4.3.

$$\begin{aligned}\text{Case 4:}\quad &1-r^2-r\cos\theta<0 \text{ and } 1+r^2-3r\cos\theta<0\\&\Rightarrow \boxed{1<r^2+r\cos\theta} \text{ and } \boxed{1+r^2<3r\cos\theta},\end{aligned}\tag{4.44}$$

then

$$\begin{aligned}&(1+3r^2-3r\cos\theta)+(1-r^2-r\cos\theta)+(1+r^2-3r\cos\theta)\\&=3+3r^2-7r\cos\theta>0,\ \text{which is positive if } \boxed{3+3r^2>7r\cos\theta}.\end{aligned}\tag{4.45}$$

The boxed inequalities provide a region for Case 4 as described by Figure 4.3.

Taking a union of all of the regions from the third row, we create a new region over which we know the third row will be diagonally dominated. This region is depicted by Figure 4.3.

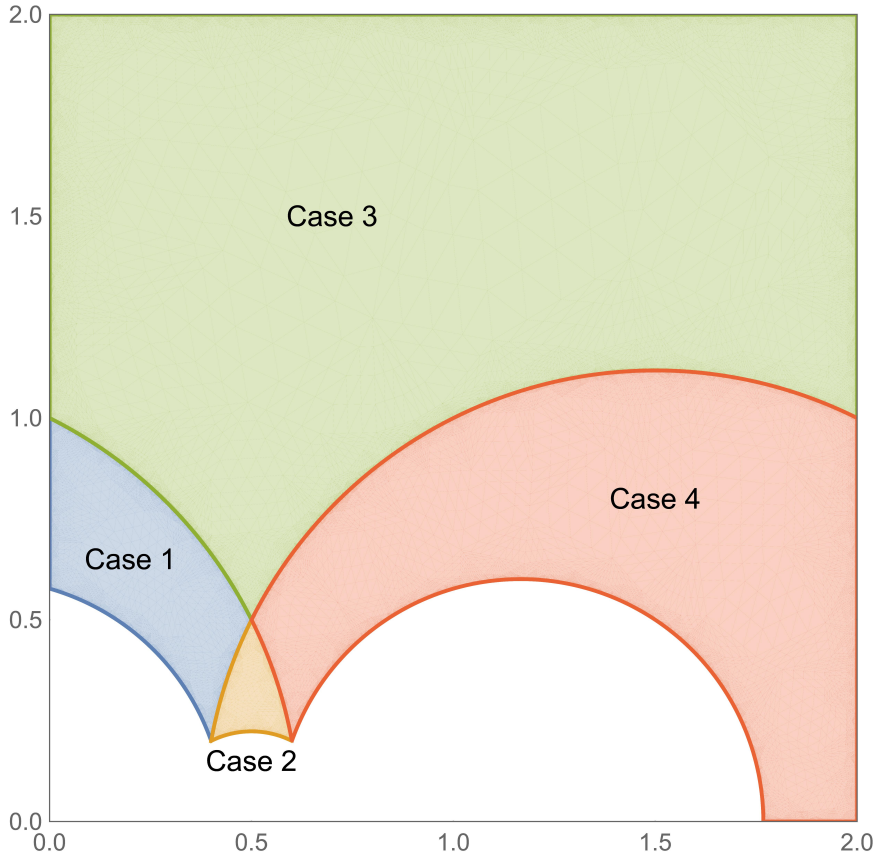


Figure 4.3: Row 3 Final Region. Case 1 refers to inequalities (4.38) and (4.39). Case 2 refers to inequalities (4.40) and (4.41). Case 3 refers to inequalities (4.42) and (4.43). Case 4 refers to inequalities (4.44) and (4.45).

Finally, the region, $\mathscr{R}$, in which all rows give diagonal dominance is obtained by intersecting the regions in figures 4.1, 4.2, and 4.3 to give Figure 4.4. Furthermore, $\mathscr{R}$ is confirmed numerically when we calculate the minimum eigenvalue across a fine grid of points in the first quadrant. We see that the level set following the minimum

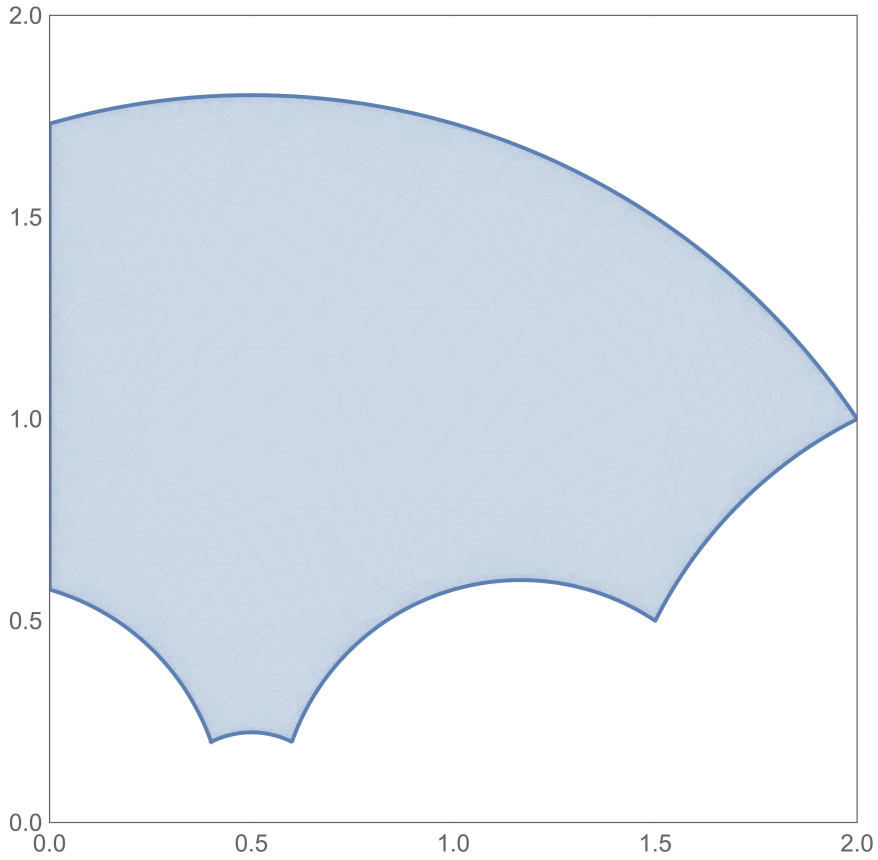


Figure 4.4: Final Region for Diagonally Dominant Raviart-Thomas Element Mass Matrix.

eigenvalue of zero follows along exactly with the boundary of $\mathscr{R}$ in Figure 4.4, as expected. This finalizes our proof.

□

Remark 4.1.1. *It is important to note that the region described in the above theorem is symmetric about the line $x = 0.5$. This is clear by the lemma, seeing as any triangle whose (x, y) vertex where $x \in [0, 0.5]$ is similar to a triangle with $x \in [0.5, 1.0]$. Thus, some of our analysis will only be concerned with portions of the region where $x \geq 0.5$.*

We now return to our two example triangles T_1 and T_2. Based on our theorem, we can see why T_1 does not have a diagonally dominant matrix, but T_2 does in Figure 4.5.

4.2 Minimum Angle Corollaries

Now we can discuss some corollaries that directly describe the necessary and sufficient angle conditions for a Raviart-Thomas triangular element to be strictly diagonally dominant.

4.2.1 Sufficient Angle Condition

Corollary 4.2.1. *If θ_{min} is greater than $\tan^{-1} \frac{\sqrt{13}}{6} \approx 31.0°$ then the Raviart-Thomas triangular element mass matrix is strictly diagonally dominant.*

Proof. In order to determine this angle θ_{min}, we want to find the contour line that traces the minimum angle and fit that within the region. Since we have fixed points at $(0, 0)$ and $(1, 0)$, a portion of each contour line will be a ray emanating from each point. Each ray will represent when θ_{min} is the angle between that ray and

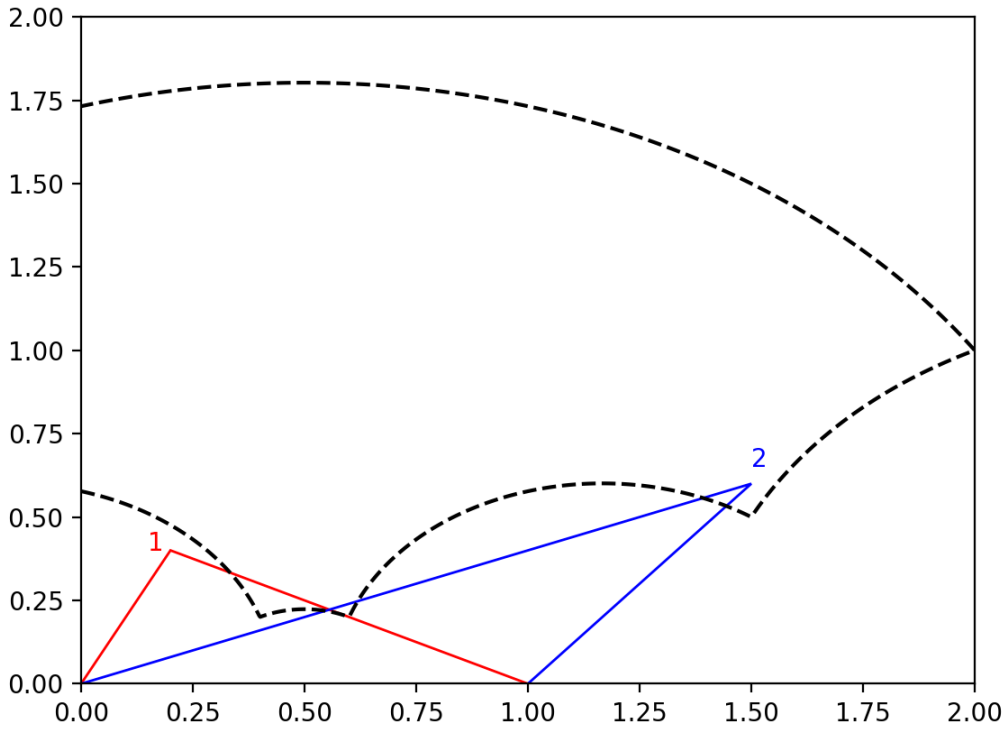


Figure 4.5: Example Triangles: T_1 does not have a diagonally dominant mass matrix because it doesn't lie in $\mathscr{R}$. The opposite is true for T_2.

the side of the triangle along the x-axis. The intersection of those rays will mark the bottom of the region. Additionally, there will be an arc-shaped contour line representing when the top angle is θ_{min}. This forms wedge-shaped contour regions. We have numerically plotted some minimum angle contour lines as an example over the diagonally dominant region shown in 4.6. Thus, we need to find where each ray is tangent $\partial\mathscr{R}$, and if the termination points of the rays and the arc line are contained within or tangent to the region. It is clear from 4.6 that the only portion of the diagonally dominant region we will have to be concerned about are the two immediately left and right of the middle, as well as the top region. We first look at

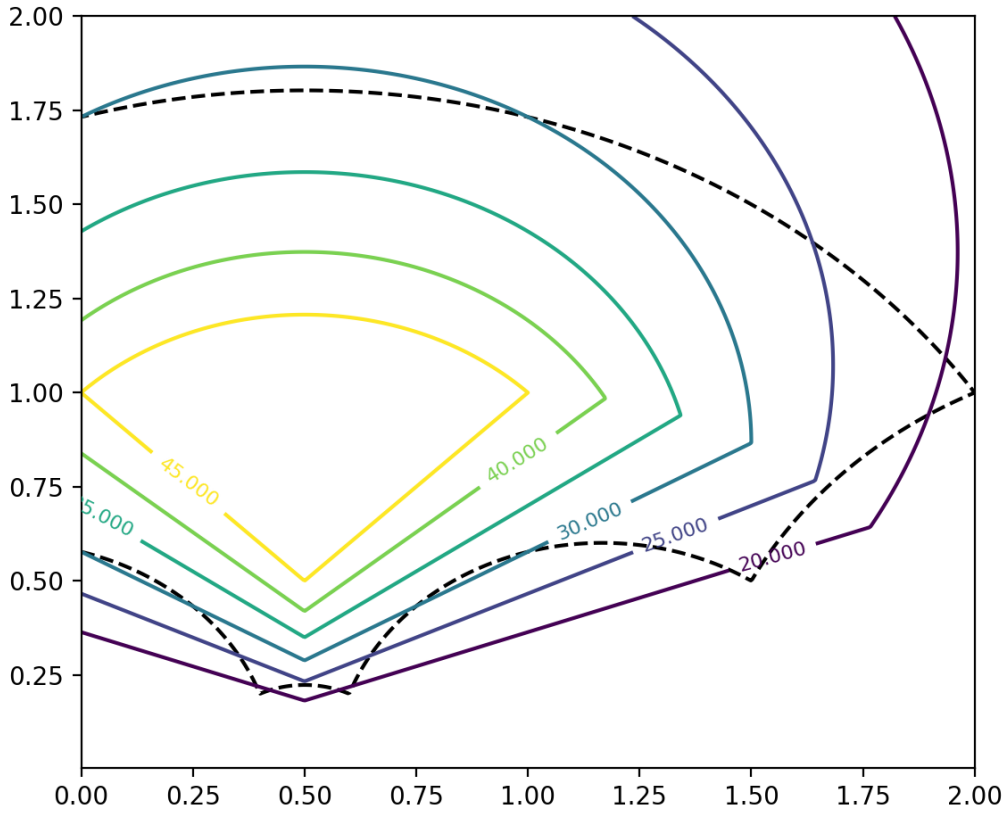


Figure 4.6: Diagonally dominant region represented as dashed line with various contour lines representing families of triangles with the same minimum angle.

the region to the right of center which is represented by the equation

$$3 + 3r^2 = 7r\cos\theta. \tag{4.46}$$

We will turn to basic calculus to uncover this result. First, (4.46) is converted back to Cartesian coordinate form to obtain

$$3 + 3(x^2 + y^2) = 7x. \tag{4.47}$$

Solving for y gives us

$$y = \pm\sqrt{\frac{1}{3}(-3x^2 + 7x - 3)}, \tag{4.48}$$

and we take the positive root since we are concerned with points in quadrant I. Next, we take the derivative with respect to x to get

$$y' = \frac{1}{2}\left[\frac{1}{3}(-3x^2 + 7x - 3)\right]^{-1/2}(-6x + 7) \cdot \frac{1}{3} = m, \tag{4.49}$$

where m represents the slope. Since the ray we are analyzing starts at $(0, 0)$, we also know that

$$m = \frac{y - 0}{x - 0} = \frac{y}{x}. \tag{4.50}$$

Thus we can set (4.49) and (4.50) equal to each other and substitute (4.48) in for y as follows

$$\frac{1}{2}\left[\frac{1}{3}(-3x^2 + 7x - 3)\right]^{-1/2}(-6x + 7) \cdot \frac{1}{3} = \frac{\sqrt{\frac{1}{3}(-3x^2 + 7x - 3)}}{x}, \tag{4.51}$$

$$\Rightarrow \frac{-6x + 7}{6\sqrt{\frac{1}{3}(-3x^2 + 7x - 3)}} = \frac{\sqrt{\frac{1}{3}(-3x^2 + 7x - 3)}}{x}. \tag{4.52}$$

Solving for x gives us

$$x = \frac{6}{7} \Rightarrow y = \frac{\sqrt{13}}{7}. \tag{4.53}$$

Using basic geometry, we construct a right triangle from x and y and solve for the smallest angle, which is

$$\tan\theta = \frac{\frac{\sqrt{13}}{7}}{\frac{6}{7}} = \frac{\sqrt{13}}{6}, \tag{4.54}$$

$$\Rightarrow \theta = \tan^{-1}\left(\frac{\sqrt{13}}{6}\right). \tag{4.55}$$

The same type of computation holds for the ray extending from $(1, 0)$. Lastly, we need to check the top arc. The equation for this section is

$$r^2 = 3 + r\cos\theta, \tag{4.56}$$

which becomes

$$x^2 + y^2 = 3 + x, \tag{4.57}$$

in Cartesian coordinates. Clearly this is an equation for a circle centered at $(0.5, 0)$ with radius $\sqrt{3}$. Since our region is symmetric across the line $x = 0.5$, we simply need to check the point at $x = 0.5$. Plugging this in gives

$$y = \frac{\sqrt{13}}{2}, \tag{4.58}$$

which has been chosen to be positive since we are in quadrant 1. Using a similar geometric process, we have

$$\tan\theta = \frac{\frac{1}{2}}{\frac{\sqrt{13}}{2}}, \tag{4.59}$$

$$\theta = \tan^{-1}\left(\frac{1}{\sqrt{13}}\right). \tag{4.60}$$

However, this only represents half of our total top angle, so we multiply on both sides by 2 to get

$$2\theta = 2\tan^{-1}\left(\frac{1}{\sqrt{13}}\right) = \tan^{-1}\left(\frac{\frac{1}{\sqrt{13}} + \frac{1}{\sqrt{13}}}{1 - \left(\frac{1}{\sqrt{13}}\right)\left(\frac{1}{\sqrt{13}}\right)}\right) \tag{4.61}$$

$$= \tan^{-1}\left(\frac{\frac{2}{\sqrt{13}}}{\frac{12}{13}}\right) = \tan^{-1}\left(\frac{\sqrt{13}}{6}\right). \tag{4.62}$$

This proves our result. □

4.2.2 Necessary Angle Condition

Corollary 4.2.2. *If θ_{min} is less than $\cos^{-1}\frac{3}{\sqrt{10}} \approx 18.4^\circ$ then the Raviart-Thomas triangular element mass matrix is not strictly diagonally dominant.*

Proof. In order to determine the necessary condition for a triangular element to be diagonally dominant, we must determine the domain of each boundary segment in terms of r. Then we must find the minimum angle across all angles of every triangle in the family of triangles represented by each boundary segment. We will analyze each boundary, starting at the top and working clockwise.

The equations for the boundary are as follows:

$$r^2 - 3 = r\cos\theta, \tag{4.63a}$$

$$\frac{r^2+5}{5} = r\cos\theta, \tag{4.63b}$$

$$\frac{3+3r^2}{7} = r\cos\theta, \tag{4.63c}$$

$$\frac{1+5r^2}{5} = r\cos\theta, \tag{4.63d}$$

$$3r^2 - 1 = r\cos\theta. \tag{4.63e}$$

The left bound of the first equation in (4.63a) is found by setting $\theta = \pi/2$ and solving for r. This clearly can only be positive in quadrant 1 and thus our solution is $r = \sqrt{3}$. The right bound is found by setting (4.63a) = (4.63b). Solving gives us the bound $r = \sqrt{5}$. Clearly (4.63b)'s right bound is also $r = \sqrt{5}$. In a similar way, we set (4.63b) = (4.63c) to obtain the left and right bounds of each, respectively. The solution to this equality is $r = \sqrt{5/2}$. Continuing, we set (4.63c) = (4.63d) and

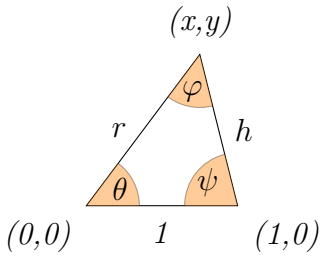


Figure 4.7: Basic triangle within quadrant 1.

(4.63d) = (4.63e) to get boundaries $r = \sqrt{2/5}$ and $r = \sqrt{1/5}$, respectively. Lastly, we set $\theta = \pi/2$ in (4.63e) to get $r = \sqrt{1/3}$. The domains of each boundary segment are then as follows:

$$\begin{aligned} \sqrt{3} \leq r \leq \sqrt{5}, \\ \sqrt{5/2} \leq r \leq \sqrt{5}, \\ \sqrt{2/5} \leq r \leq \sqrt{5/2}, \\ \sqrt{1/5} \leq r \leq \sqrt{2/5}, \\ \sqrt{1/5} \leq r \leq \sqrt{1/3}. \end{aligned} \tag{4.64}$$

Now we determine equations for each angle, θ, ψ, and φ, within a family of triangles along each boundary segment. Figure 4.7 demonstrates the triangle from which we will derive our formulas. For each equation, we will primarily utilize the Law of Cosines to give a general form for side h, which is

$$h = \sqrt{r^2 + 1 - 2r\cos\theta}. \tag{4.65}$$

Similarly, we can discover general equations for both ψ and φ as follows

$$\begin{aligned}\psi &= \cos^{-1}\left(\frac{h^2 - r^2 + 1}{2h}\right), \\ \varphi &= \cos^{-1}\left(\frac{h^2 + r^2 - 1}{2hr}\right).\end{aligned} \tag{4.66}$$

However, this will still give us equations in terms of r and θ when h is substituted. Luckily, notice that we have provided the boundary equations in terms of $r\cos\theta$, which will be substituted into the bold term of h for each boundary segment. Starting with (4.63a), we solve for θ to get

$$\theta = \cos^{-1}\left(\frac{r^2 - 3}{r}\right). \tag{4.67}$$

After substituting (4.63a) into h, ψ and ϕ become

$$\begin{aligned}\psi &= \cos^{-1}\left(\frac{4 - r^2}{\sqrt{7 - r^2}}\right), \\ \varphi &= \cos^{-1}\left(\frac{3}{r\sqrt{7 - r^2}}\right).\end{aligned} \tag{4.68}$$

Lastly, we use basic calculus to minimize each angle function on a domain. We then take the minimum between those three resulting minima. Taking the derivative of each angle equation gives

$$\begin{aligned}\theta' &= \frac{-r^2 - 3}{r^2\sqrt{7 - \frac{9}{r^2} - r^2}}, \\ \psi' &= -\frac{r(-10 + r^2)}{(7 - r^2)^{3/2}\sqrt{r^2 + \frac{9}{r^2 - 7}}}, \\ \varphi' &= \frac{6r^2 - 21}{r(r^2 - 7)\sqrt{-r^4 + 7r^2 - 9}}.\end{aligned} \tag{4.69}$$

Taking the minimum of minima provides $\cos^{-1}\left(\frac{3}{\sqrt{10}}\right)$. We will use the same method for the remaining equations.

For equation (4.63b) we get angle equations

$$\begin{aligned}\theta =& \cos^{-1}\left(\frac{5+r^2}{5r}\right),\\ \psi =& \cos^{-1}\left(\frac{-r^2}{\sqrt{5}\sqrt{3r^2-5}}\right),\\ \varphi =& \cos^{-1}\left(\frac{4r^2-5}{r\sqrt{5}\sqrt{3r^2-5}}\right).\end{aligned} \tag{4.70}$$

Taking the derivative of each gives

$$\begin{aligned}\theta' =& \frac{5-r^2}{r^2\sqrt{15-\frac{25}{r^2}-r^2}},\\ \psi' =& \frac{r(3r^2-10)}{(3r^2-5)^{3/2}\sqrt{\frac{r^4-15r^2+25}{5-3r^2}}},\\ \varphi' =& \frac{25-10r^2}{r^2(3r^2-5)^{3/2}\sqrt{\frac{5}{r^2}+\frac{r^2}{5-3r^2}}}.\end{aligned} \tag{4.71}$$

We find the minima to get $\cos^{-1}\left(\frac{3}{\sqrt{10}}\right)$.

Moving on to (4.63c) we have

$$\begin{aligned}\theta =& \cos^{-1}\left(\frac{3+3r^2}{7r}\right),\\ \psi =& \cos^{-1}\left(\frac{4-3r^2}{\sqrt{7}\sqrt{r^2+1}}\right),\\ \varphi =& \cos^{-1}\left(\frac{4r^2-3}{r\sqrt{7}\sqrt{r^2+1}}\right).\end{aligned} \tag{4.72}$$

We take the derivatives to get

$$\begin{aligned}\theta' =& \frac{3-3r^2}{r\sqrt{-9r^4+31r^2-9}},\\ \psi' =& \frac{r(3r^2+10)}{(r^1+1)^{3/2}\sqrt{40-9r^2-\frac{49}{r^2+1}}},\\ \varphi' =& \frac{-10r^2-3}{r^2(r^1+1)^{3/2}\sqrt{\frac{49}{r^2+1}-\frac{9}{r^2}-9}}.\end{aligned} \tag{4.73}$$

We find the minima to get $\cos^{-1}\left(\frac{3}{\sqrt{10}}\right)$.

Next we have (4.63d) which gives

$$\begin{aligned}\theta =& \cos^{-1}\left(\frac{5r^2+1}{5r}\right),\\ \psi =& \cos^{-1}\left(\frac{4-5r^2}{\sqrt{5}\sqrt{3-5r^2}}\right),\\ \varphi =& \cos^{-1}\left(\frac{-1}{r\sqrt{5}\sqrt{3-5r^2}}\right).\end{aligned} \tag{4.74}$$

We take the derivatives to get

$$\begin{aligned}\theta' =& \frac{1-5r^2}{r\sqrt{r\sqrt{15r^2-1-25r^4}}},\\ \psi' =& \frac{5r(2+5r^2)}{(3-5r^2)^{3/2}\sqrt{5r^2+\frac{1}{5r^2-3}}},\\ \varphi' =& \frac{3-10r^2}{r(5r^2-3)\sqrt{-25r^4+15r^2-1}}.\end{aligned} \tag{4.75}$$

The minimum is then $\cos^{-1}\left(\frac{3}{\sqrt{10}}\right)$.

Finally we consider (4.63e) which has the angle formulas

$$\begin{aligned}\theta =& \cos^{-1}\left(\frac{1-3r^2}{r}\right),\\ \psi =& \cos^{-1}\left(\frac{3r^2}{\sqrt{7r^2-1}}\right),\\ \varphi =& \cos^{-1}\left(\frac{4r^2-1}{r\sqrt{7r^2-1}}\right).\end{aligned} \tag{4.76}$$

Then we get the derivatives below

$$\begin{aligned}\theta' =& \frac{1+3r^2}{r\sqrt{7r^2-1-9r^4}},\\ \psi' =& \frac{r(6-21r^2)}{(7r^2-1)\sqrt{-1+7r^2-9r^4}},\\ \varphi' =& \frac{1-10r^2}{r^2(7r^2-1)^{3/2}\sqrt{\frac{1}{r^2}+\frac{9r^2}{1-7r^2}}}.\end{aligned} \tag{4.77}$$

Once again, our minimum is $\cos^{-1}\left(\frac{3}{\sqrt{10}}\right)$, which completes the proof. □

We can see the contour lines of these minimum angle conditions drawn out in Figure 4.8. It is interesting to notice that the points where the red contour line from the necessary condition intersects the boundary of our diagonally dominant region are in fact all part of the same family of similar triangles. This show that this region is not made up of strictly unique triangle in terms of similarity. More analysis could be done to reduce the region, but that is outside the scope of this dissertation topic.

Furthermore, we would like to bring the reader's attention to another interesting connection regarding the sufficient condition of Corollary 4.2.1. In [37], Shewchuck establishes a 2D triangle mesh generator utilizing Delaunay triangulations. Rupert has shows in [35] that this Delaunay refinement halts for an angle constraint of up to 20.7°. Furthermore, Shewchuck notes that in fact when executing this algorithm, "the algorithm generally halts with an angle constraint of 33.8° , but often fails to terminate given an angle constraint of 33.9°" [37]. We find it intriguing that our necessary minimum angle condition of $\tan^{-1}\left(\frac{\sqrt{13}}{6}\right) \approx 31.0°$ falls close to Shewchuck's observed constraints.

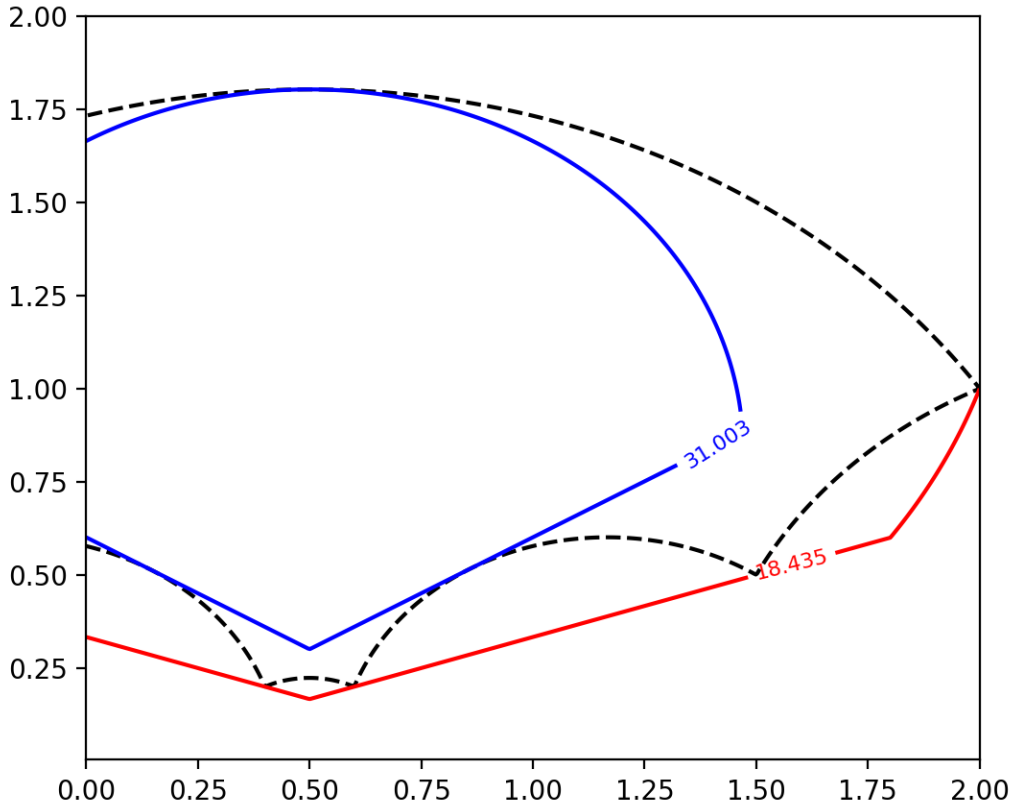


Figure 4.8: Diagonally dominant boundary with necessary (red) and sufficient (blue) angle contour lines. Note that $\tan^{-1}\left(\frac{\sqrt{13}}{6}\right) \approx 31.0027°$ and $\cos^{-1}\left(\frac{3}{\sqrt{10}}\right) \approx 18.4349°$.

4.3 Some Comments on Eigenvalues

We can also consider the condition number of the Raviart-Thomas triangular element in the diagonally dominant region $\mathscr{R}$. We have provided a plot of the condition number with the diagonally dominant boundary in Figure 4.9. However, we could not find much evidence that the boundary of our region and the condition number of the element matrix were directly related. However, the condition number within the region is fairly small, and doesn't blow up anywhere in that region. Additionally

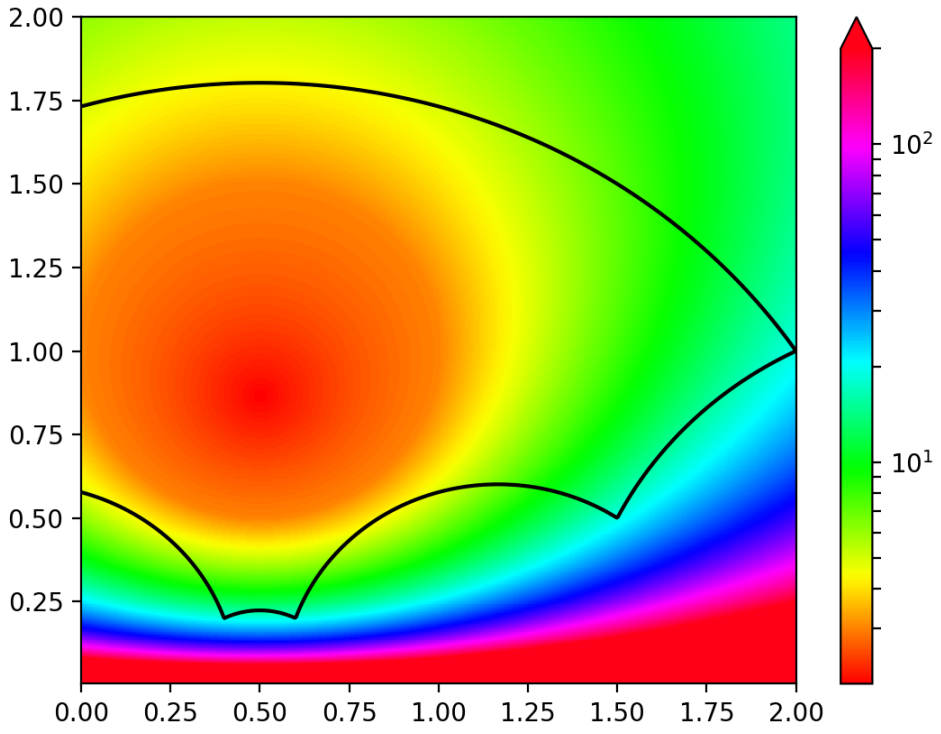


Figure 4.9: Diagonally dominant boundary with the condition number of Raviart-Thomas triangular element matrix as the color gradient in log scale.

we looked at the condition number of the element matrix preconditioned with Jacobi preconditioning. This is shown in Figure 4.10. Since we have been looking at matrices that are strictly diagonally dominant, it was worth considering this preconditioner. It is interesting to note that it did slightly improve the condition number throughout, even in the non-diagonally dominant region.

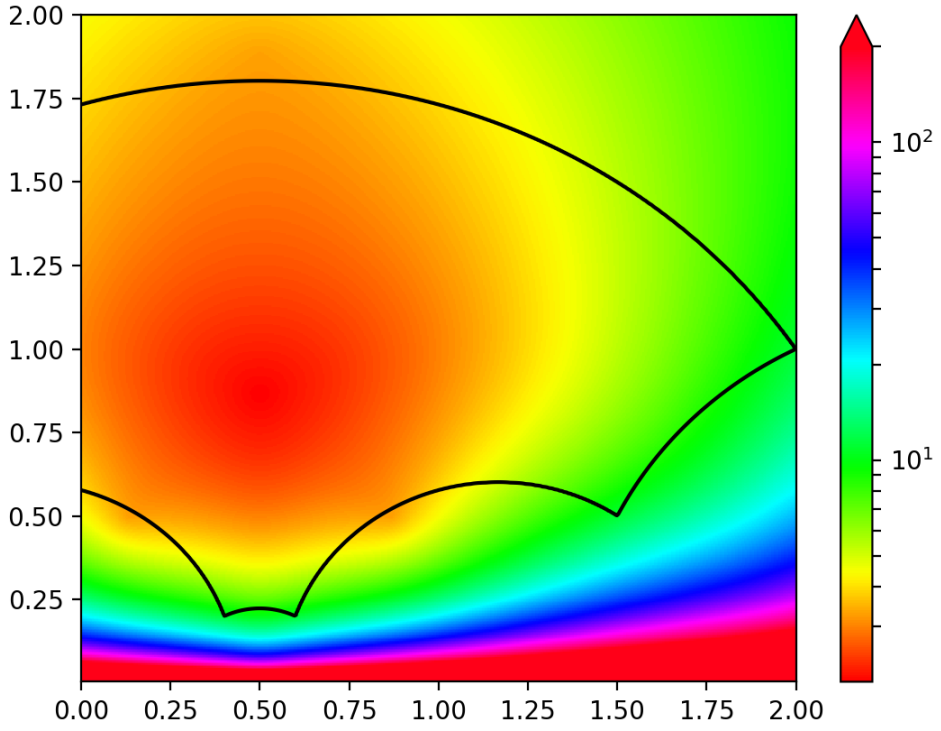


Figure 4.10: Diagonally dominant boundary with the condition number of the Raviart-Thomas triangular element matrix preconditioned with its main diagonal. as the color gradient in log scale.

CHAPTER FIVE

Preconditioning

5.1 The Wave Equation

Before we create a preconditioner for the tide model, we first would like to develop an approach to the wave equation. We start with the mixed formulation of the wave equation described in [1]. Thus, we consider the equations whose solution $(u, p) \in H(\text{div}) \times L^2$ solves

$$\begin{aligned} (u, v) + k(p, \nabla \cdot v) &= 0 \text{ for all } v \in H(\text{div}), \\ k(\nabla \cdot u, q) - (p, q) &= (g, q) \text{ for all } q \in L^2. \end{aligned} \tag{5.1}$$

This formulation gives the differential operator

$$A = \begin{bmatrix} I & -k \operatorname{grad} \\ k \nabla \cdot & -I \end{bmatrix} \tag{5.2}$$

which defines an isomorphism from $H(\text{div}) \times L^2$ onto its dual [1]. Our goal is to give bounds for both A and A^{-1} by providing continuity and inf-sup arguments, respectively. If we equip $H(\text{div})$ with the standard norm, the preconditioned system's condition number will rely heavily on k, the time step. if, we choose to equip $H(\text{div})$ with the norm, however,

$$u \mapsto \left(\|u\|^2 + k^2 \|\nabla \cdot u\|^2 \right)^{1/2}, \tag{5.3}$$

we can prove the preconditioned system has a continuity constant K independent of k. Naturally, this provides the operator norm $\|(u,p)\|_k^2 = \|u\|^2 + k^2\|\nabla \cdot u\|^2 + \|p\|^2$ on $H(\text{div}) \times L^2$. Note that we still use the standard norm for L^2.

Let the bilinear form for A, be

$$a((u,p),(v,q)) = (u,v) + (p, k\nabla \cdot v) + (k\nabla \cdot u, q) - (p,q). \tag{5.4}$$

Then, we show that

$$a((u,p),(v,q)) \leq K\left[\|(u,p)\|_k \|(v,q)\|_k\right]. \tag{5.5}$$

where $K = 2$.

The bilinear form $a(\cdot,\cdot)$ is created by summing the left hand side of (5.1). Then, we can apply the Discrete Cauchy-Schwarz inequality below to get

$$\begin{aligned}
&(u,v) + (p, k\nabla \cdot v) + (k\nabla \cdot u, q) - (p,q) \\
\leq\ & \|u\|\|v\| + \|p\|\|k\nabla \cdot v\| + \|k\nabla \cdot u\|\|q\| + \|p\|\|q\| \\
\leq\ & \sqrt{\|u\|^2 + \|p\|^2 + k^2\|\nabla \cdot u\|^2 + \|p\|^2} \cdot \sqrt{\|v\|^2 + k^2\|\nabla \cdot v\|^2 + \|q\|^2 + \|q\|^2} \\
=\ & 2\|(u,p)\|_k \|(v,q)\|_k,
\end{aligned} \tag{5.6}$$

which proves continuity and confirms that A is bounded in $\|\cdot\|_k$.

Now we look at the inf-sup condition. We want to show

$$\inf_{(u,p)} \sup_{(v,q)} \frac{a((u,p),(v,q))}{\|(u,p)\|_k \|(v,q)\|_k} \geq \alpha. \tag{5.7}$$

Taking the advice in [1], we choose test functions $v = u$ and $q = k\nabla \cdot u - p$ and

substitute them into (5.4). Thus

$$\begin{aligned} a((u,p),(v,q)) &= (u,u) + (p, k\nabla\cdot u) + (k\nabla\cdot u, k\nabla\cdot u - p) - (p, k\nabla\cdot u - p) \\ &= (u,u) + (p, k\nabla\cdot u) + (k\nabla\cdot u, k\nabla\cdot u) - (k\nabla\cdot u, p) - (p, k\nabla\cdot u) + (p,p) \\ &= \|u\|^2 - (p, k\nabla\cdot u) + k^2\|\nabla\cdot u\|^2 + \|p\|^2. \end{aligned} \tag{5.8}$$

Using Young's inequality we get that the RHS of (5.8) is larger than

$$\begin{aligned} &\|u\|^2 + \frac{k^2}{2}\|\nabla\cdot u\|^2 + \frac{1}{2}\|p\|^2 \\ \geq\ &\frac{1}{2}\|(u,p)\|_k^2. \end{aligned} \tag{5.9}$$

Now we can bound v, q in terms of u, p such that $\frac{1}{2}\|(u,p)\|_k \geq \frac{\sqrt{3}}{6}\|(v,q)\|_k$ as shown below,

$$\begin{aligned} \|(v,q)\|_k^2 &= \|v\|^2 + k^2\|\nabla\cdot v\|^2 + \|q\|^2 \\ &= \|u\|^2 + k^2\|\nabla\cdot u\|^2 + \|k\nabla\cdot u - p\|^2 \\ &\leq \|u\|^2 + 3k^2\|\nabla\cdot u\|^2 + 2\|p\|^2 \\ &\leq 3\|(u,p)\|_k^2. \end{aligned} \tag{5.10}$$

Returning to the inf-sup argument, we then have

$$a((u,p),(v,q)) \geq \frac{1}{2}\|(u,p)\|_k^2 \geq \frac{\sqrt{3}}{6}\|(u,p)\|_k\|(v,q)\|_k. \tag{5.11}$$

Thus, we can divide by normed factors on the right hand side and take the supremum with respect to (v,q) and the infimum with respect to (u,p) to prove our result. Therefore, A^{-1} is bounded in the $\|\cdot\|_k$ norm independent of k. Therefore, our preconditioned system has bounded eigenvalues with respect to the k-norm.

5.2 The Tide Model

We now turn our attention back to preconditioning the tide model (3.1). We can use similar techniques to develop a preconditioner for this case. Our goal is to define an inner product so that the norm of the operator and the inverse operator depend as little as possible on the parameters: β, ϵ, C, and f, as well as the time step k.

Below, we present the variational form for one time step for Crank-Nicoloson of the weighted discrete tide model

$$\begin{aligned} \frac{\beta}{\epsilon^2}(\eta, w) + \frac{k\beta}{\epsilon^2}(\nabla \cdot u, w) &= \frac{\beta}{\epsilon^2}(G, w) \,\forall\, w \in L^2 \\ (u, v) + \frac{k}{\epsilon}(fu^{\perp}, v) - \frac{k\beta}{\epsilon^2}(\eta, \nabla \cdot v) + kC(u, v) &= (F, v) \,\forall\, v \in H(\mathrm{div}). \end{aligned} \tag{5.12}$$

Similar to the wave equation, this mixed formulation provides a differential operator A which defines an isomorphism from $H(\mathrm{div}) \times L^2$ onto its dual. We then equip $H(\mathrm{div})$ with the weighted norm

$$u \mapsto \left((1 + Ck) \|u\|^2 + \frac{k^2\beta}{\epsilon^2} \|\nabla \cdot u\|^2 \right)^{1/2} \tag{5.13}$$

and L^2 with the weighted norm

$$p \mapsto \left(\frac{\beta}{\epsilon^2} \|\eta\|^2 \right)^{1/2}. \tag{5.14}$$

We define $|||(\cdot, \cdot)|||$ on $H(\mathrm{div}) \times L^2$ by

$$|||(u, \eta)|||^2 = (1 + Ck) \|u\|^2 + \frac{k^2\beta}{\epsilon^2} \|\nabla \cdot u\|^2 + \frac{\beta}{\epsilon^2} \|\eta\|^2. \tag{5.15}$$

Lastly, we sum the left hand side of (5.12) to get the bilinear form for A represented

by

$$\begin{aligned} &a((u,\eta),(v,w)) = \\ &\frac{\beta}{\epsilon^2}(\eta,w)+\frac{k\beta}{\epsilon^2}(\nabla\cdot u,w)+(u,v)+\frac{k}{\epsilon}(fu^{\perp},v)-\frac{k\beta}{\epsilon^2}(\eta,\nabla\cdot v)+kC(u,v). \end{aligned} \tag{5.16}$$

Proposition 5.2.1. *The operator A is bounded in the norm $|||\cdot|||$ with constant $K_{k,\epsilon} = \max\left\{2, 1+\frac{k}{\epsilon(1+Ck)}\right\}$.*

Proof. We apply the Cauchy-Schwarz method to the discrete tide model. Once again, we are looking to show

$$a((u,\eta),(v,w)) \leq K\left[|||(u,\eta)|||\,|||(v,w)|||\right] \tag{5.17}$$

for any (u,η) and any (v,w). Working from (5.16), we start with the Cauchy-Schwarz inequality, next collect terms, and then redistribute terms in order to apply the Cauchy-Schwarz inequality:

$$a((u,\eta),(v,w)) \tag{5.18}$$

$$\leq \frac{\beta}{\epsilon^2}\|\eta\|\|w\|+\frac{k\beta}{\epsilon^2}\|\nabla\cdot u\|\|w\|+\|u\|\|v\| \tag{5.19}$$

$$+\frac{k}{\epsilon}\|fu^{\perp}\|\|v\|+\frac{k\beta}{\epsilon^2}\|\eta\|\|\nabla\cdot v\|+kC\|u\|\|v\|$$

$$=\left(1+\frac{k}{\epsilon}|f|+kC\right)\|u\|\|v\|+\frac{\beta}{\epsilon^2}\|\eta\|\|w\| \tag{5.20}$$

$$+\frac{k\beta}{\epsilon^2}\|\nabla\cdot u\|\|w\|+\frac{k\beta}{\epsilon^2}\|\eta\|\|\nabla\cdot v\|$$

$$=\sqrt{1+\frac{k}{\epsilon}|f|+kC}\,\|u\|\sqrt{1+\frac{k}{\epsilon}|f|+kC}\,\|v\|+\frac{\sqrt{\beta}}{\epsilon}\|\eta\|\frac{\sqrt{\beta}}{\epsilon}\|w\| \tag{5.21}$$

$$+k\frac{\sqrt{\beta}}{\epsilon}\|\nabla\cdot u\|\frac{\sqrt{\beta}}{\epsilon}\|w\|+\frac{\sqrt{\beta}}{\epsilon}\|\eta\|k\frac{\sqrt{\beta}}{\epsilon}\|\nabla\cdot v\|.$$

Recognizing this as a dot product of $\left(\sqrt{1+\frac{k}{\epsilon}\cdot 1+kC}\,\|u\|, k\frac{\sqrt{\beta}}{\epsilon}\|\nabla\cdot u\|, \frac{\sqrt{\beta}}{\epsilon}\|\eta\|, \frac{\sqrt{\beta}}{\epsilon}\|\eta\|\right)$

and $\left(\sqrt{1+\frac{k}{\epsilon}\cdot 1+kC}\,\|v\|, k\frac{\sqrt{\beta}}{\epsilon}\|\nabla\cdot v\|, \frac{\sqrt{\beta}}{\epsilon}\|w\|, \frac{\sqrt{\beta}}{\epsilon}\|w\|\right)$, the discrete Cauchy-Schwarz inequality gives

$$\leq \sqrt{\left(1+\frac{k}{\epsilon}+kC\right)\ \|u\|^2+\frac{k^2\beta}{\epsilon^2}\|\nabla\cdot u\|^2+2\frac{\beta}{\epsilon^2}\|\eta\|^2} \tag{5.22}$$
$$\cdot\sqrt{\left(1+\frac{k}{\epsilon}+kC\right)\ \|v\|^2+\frac{k^2\beta}{\epsilon^2}\|\nabla\cdot v\|^2+2\frac{\beta}{\epsilon^2}\|w\|^2}$$
$$\leq K_{k,\epsilon}\sqrt{(1+kC)\ \|u\|^2+\frac{k^2\beta}{\epsilon^2}\|\nabla\cdot u\|^2+\frac{\beta}{\epsilon^2}\|\eta\|^2} \tag{5.23}$$
$$\cdot\sqrt{(1+kC)\ \|v\|^2+\frac{k^2\beta}{\epsilon^2}\|\nabla\cdot v\|^2+\frac{\beta}{\epsilon^2}\|w\|^2}$$
$$= K_{k,\epsilon}|||(u,\eta)|||\,|||(v,w)|||, \tag{5.24}$$

where $K_{k,\epsilon}=\max\left\{2, 1+\frac{k}{\epsilon(1+Ck)}\right\}$. We used the fact that $(\cdot)^\perp$ is an isometry, and thus $\|u\|=\|u^\perp\|$. Also, note that since f is a sine function, $|f|\leq 1$. Lastly, the second component of $K_{k,\epsilon}$ was chosen since

$$1+\frac{k}{\epsilon}+Ck=(1+Ck)\left(1+\frac{k}{\epsilon(1+Ck)}\right). \tag{5.25}$$

Since $k<0$, we are independent from all other bounds except ϵ as it goes to 0. □

Proposition 5.2.2. *The operator A^{-1} is bounded in the norm $|||\cdot|||$ with constant $K=\frac{\sqrt{3}}{6}$ and its bound is independent of the parameters ϵ, β, C, and f.*

Proof. Now we would like to show a inf-sup condition [4] of the form

$$\inf_{(u,\eta)}\sup_{(v,w)}\frac{a((u,\eta),(v,w))}{|||(u,\eta)|||\,|||(v,w)|||}\geq\frac{\sqrt{3}}{6}. \tag{5.26}$$

Let u, η be given and let $v, w = u, \eta + k\nabla \cdot u$. Then, in (5.16), we have

$$\begin{aligned}
(\ddagger) &= \frac{\beta}{\epsilon^2}(\eta, \eta + k\nabla \cdot u) + \frac{k\beta}{\epsilon^2}(\nabla \cdot u, \eta + k\nabla \cdot u) \\
&\quad + (u, u) + \frac{k}{\epsilon}(fu^{\perp}, u) - \frac{k\beta}{\epsilon^2}(\eta, \nabla \cdot u) + kC(u, u) \\
&= \frac{\beta}{\epsilon^2}(\eta, \eta) + \frac{k\beta}{\epsilon^2}(\eta, \nabla \cdot u) + \frac{k\beta}{\epsilon^2}(\nabla \cdot u, \eta) + \frac{k^2\beta}{\epsilon^2}(\nabla \cdot u, \nabla \cdot u) \\
&\quad + (u, u) + \frac{k}{\epsilon}(fu^{\perp}, u) - \frac{k\beta}{\epsilon^2}(\eta, \nabla \cdot u) + kC(u, u) \\
&= (1 + Ck)\|u\|^2 + \frac{\beta}{\epsilon^2}\|\eta\|^2 + \frac{k^2\beta}{\epsilon^2}\|\nabla \cdot u\|^2 + \frac{k\beta}{\epsilon^2}(\eta, \nabla \cdot u). \quad (\star)
\end{aligned} \tag{5.27}$$

Remark 5.2.1. *Note that* $(fu^{\perp}, u) = 0$ *since* $u^{\perp}u = 0$ *pointwise almost everywhere.*

Next, Young's inequality gives us

$$\begin{aligned}
(\star) &\geq (1 + Ck)\|u\|^2 + \frac{\beta}{\epsilon^2}\|\eta\|^2 + \frac{k^2\beta}{\epsilon^2}\|\nabla \cdot u\|^2 - \frac{\beta}{2\epsilon^2}\|\eta\|^2 - \frac{k^2\beta}{2\epsilon^2}\|\nabla \cdot u\|^2 \\
&= (1 + Ck)\|u\|^2 + \frac{\beta}{2\epsilon^2}\|\eta\|^2 + \frac{k^2\beta}{2\epsilon^2}\|\nabla \cdot u\|^2 \\
&\geq \frac{1}{2}|||(u, \eta)|||^2.
\end{aligned} \tag{5.28}$$

We have chosen v, q such that $\frac{\sqrt{3}}{3}|||(v, w)||| \leq |||(u, \eta)|||$. So, we can see, that by utilizing the Triangle Inequality and Young's Inequality we have

$$\begin{aligned}
|||(v, q)|||^2 &= (1 + Ck)\|u\|^2 + \frac{k^2\beta}{\epsilon^2}\|\nabla \cdot u\|^2 + \frac{\beta}{\epsilon^2}\|\eta + k\nabla \cdot u\|^2 \\
&\leq (1 + Ck)\|u\|^2 + 3\left(\frac{k^2\beta}{\epsilon^2}\right)\|\nabla \cdot u\|^2 + 2\left(\frac{\beta}{\epsilon^2}\right)\|\eta\|^2 \\
&\leq 3|||(u, p)|||^2,
\end{aligned} \tag{5.29}$$

which clearly validates our inequality above.

Returning to the inf-sup argument, we then have

$$a((u, \eta), (v, w)) \geq \frac{1}{2}|||(u, \eta)|||^2 \geq \frac{\sqrt{3}}{6}|||(u, \eta)|||\,|||(v, w)||| \tag{5.30}$$

Thus, we can divide by the normed factors on the right hand side and take the

supremum with respect to (v, w) and the infimum with respect to (u, η) to prove our result. Therefore, A^{-1} is bounded in $|||\cdot|||$ and is independent of the mentioned parameters. □

Remark 5.2.2. *Note that*

$$\begin{aligned}\frac{\beta}{\epsilon^2}\|\eta + k\nabla\cdot u\|^2 &\leq \frac{\beta}{\epsilon^2}\left[\|\eta\|^2 + 2\|\eta\|\|k\nabla\cdot u\| + \|k\nabla\cdot u\|^2\right] \\ &\leq \frac{\beta}{\epsilon^2}\left[\|\eta\|^2 + 2\left(\frac{\|\eta\|^2}{2} + \frac{\|k\nabla\cdot u\|^2}{2}\right) + \|k\nabla\cdot u\|^2\right] \\ &= 2\left(\frac{\beta}{\epsilon^2}\right)\|\eta\|^2 + 2\left(\frac{k^2\beta}{\epsilon^2}\right)\|\nabla\cdot u\|^2.\end{aligned} \tag{5.31}$$

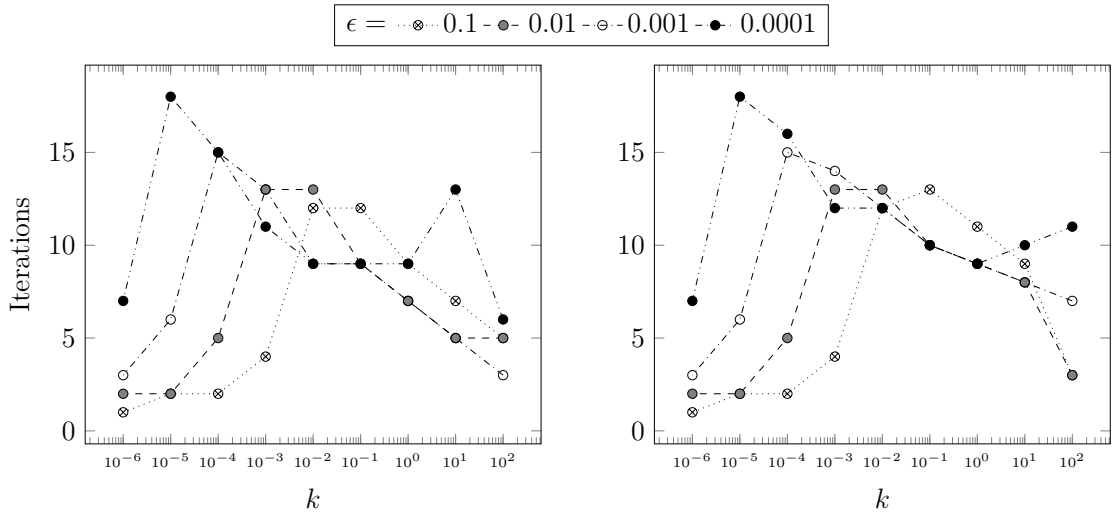
Theorem 5.2.3. *The eigenvalue bounds of the preconditioned system are independent of k, C, and f, but are dependent on ϵ as it goes to 0.*

Proof. The propositions above prove this result. □

Remark 5.2.3. *If ϵ is small, we can pick a k for each ϵ so that $\exists\ k_0$ such that the preconditioner has parameter independent eigenvalue bounds for all $k < k_0$.*

5.3 Numerical Results

Now that we have completed the theoretical analysis, we want to confirm our results numerically. We see that when ϵ is small, the upper bound may pose a problem. By fixing all the parameters, we want to show how badly this term behaves in practice. We compare two different weights on the $\|u\|^2$ term for reference. From Figure 5.1, we see that each plot has a maximum when ϵ is an order higher than k. Clearly larger ϵ can be somewhat bad as ϵ decreases, but overall it is fairly manageable.

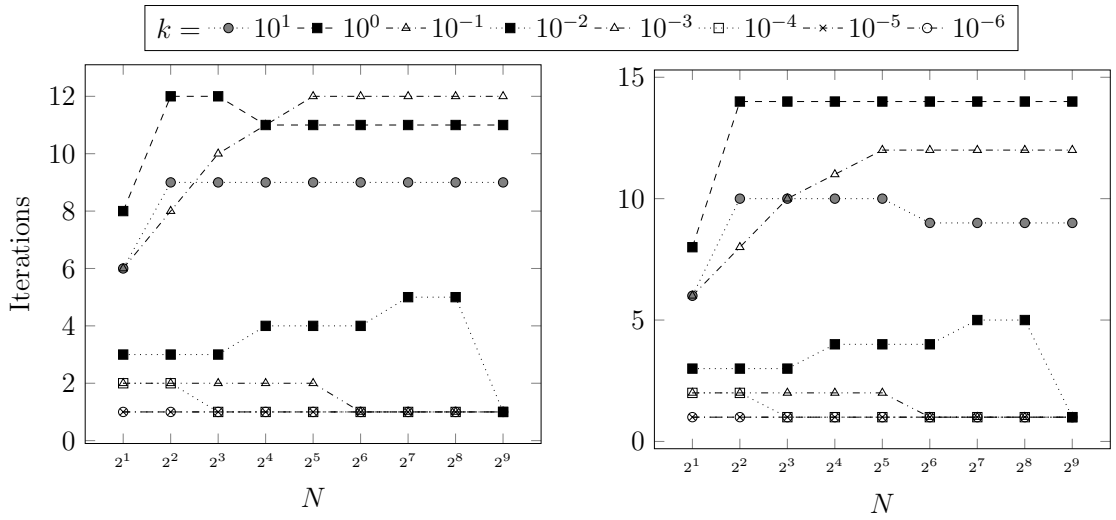


(a) Number of iterations when the $H(\text{div})$ norm weight on $\|u\|^2$ is $1.0+Ck$ for different values of ϵ over time steps k.

(b) Number of iterations when the $H(\text{div})$ norm weight on $\|u\|^2$ is 1.0 for different values of ϵ over time steps k.

Figure 5.1: Varying ϵ over k with all the other parameters fixed.

Based on our theorem, the iteration count for GMRES should be mesh independent for every k. We see that our iteration counts are the worst at intermediate k, but large and small k are much better. Clearly, Figure 5.2 shows good mesh behavior. Some moderate k are more expensive, but overall we have fairly low iteration counts.



(a) Number of iterations when the $H(\text{div})$ norm weight on $\|u\|^2$ is $1.0+Ck$ for different values of ϵ.

(b) Number of iterations when the $H(\text{div})$ norm weight on $\|u\|^2$ is 1.0 for different values of ϵ.

Figure 5.2: Varying k over mesh size N with other parameters held constant.

CHAPTER SIX

Conclusions and Further Research

In this dissertation, we have studied the tide model, providing analysis concerning its preconditioning. We provided a method determine if the element mass matrix for a given Raviart-Thomas element, which is often used in tide modeling, is diagonally dominant. This is extremely helpful when applying an explicit time stepping method, since in order to solve one must invert the mass matrix at every time step. Thus, being able determine which element matrices are diagonally dominant.

Additionally, we used weighted norms to give a parameter independent eigenvalue bound of the preconditioned system. We showed the upper bound was only mildly dependent on the time step and some physical parameters, while the lower bound was totally independent. This analysis is particularly helpful when inverting the entire system of an implicit time stepping method (such as Crank-Nicolson). Future work could be done to quantify different elements, such as quadrilaterals or tetrahedron in the three dimensional case.

A current project is implementing the discrete 2D div operator in Firedrake. Some progress has been made, but not enough to include in this dissertation. Furthermore, we could extend this work with the discrete 3D curl.

APPENDIX

APPENDIX A

More Example Triangles

Here we provide some example triangles that have diagonally dominant mass matrices and non-diagonally dominant mass matrices. The latter is represented by Figure A.1. Below, we provide the element mass matrices for each of these triangles

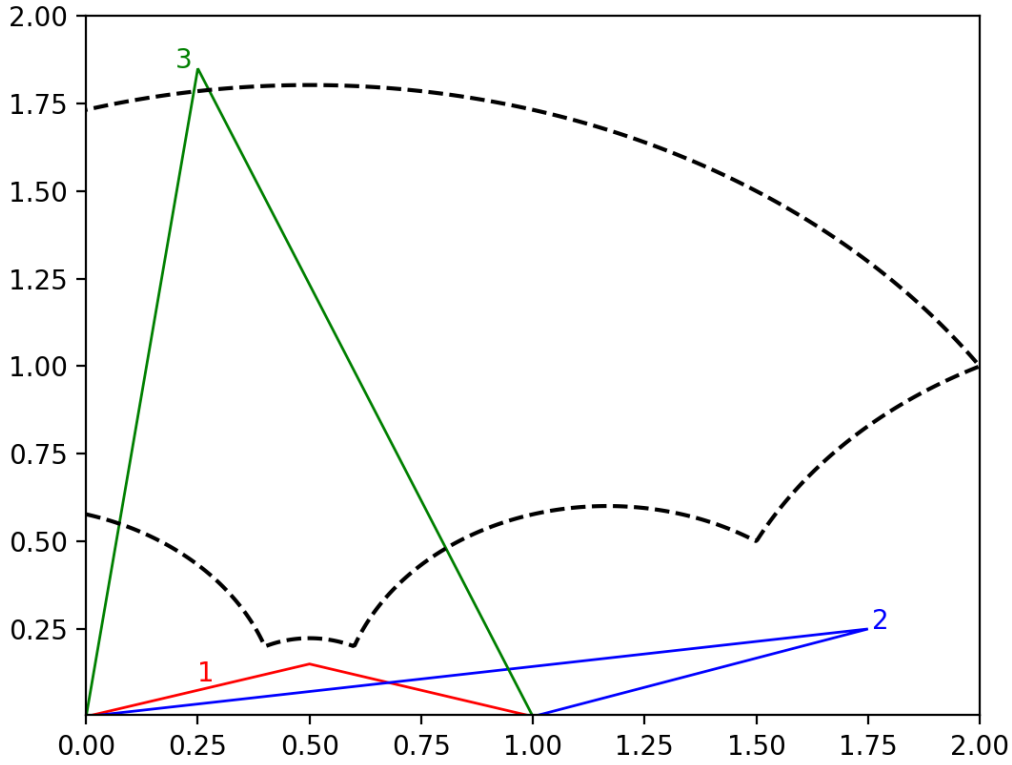


Figure A.1: These three example triangles T_1 (red), T_2 (blue), and T_3 (green) are not diagonally dominant.

with their row sums to demonstrate that every row is not diagonally dominant so the entire element matrix lacks diagonal dominance. The column we list next to each

element mass matrix is created in the same way as above, by subtracting the absolute values of the off diagonal entries from the diagonal entry.

$$T_1 = \begin{bmatrix} 0.98472222 & -0.12638889 & -0.68194444 \\ -0.12638889 & 0.17638889 & -0.12638889 \\ -0.68194444 & -0.12638889 & 0.98472222 \end{bmatrix} \text{ with } \tilde{\Delta}(T_1) = \begin{bmatrix} 0.17638889 \\ -0.07638889 \\ 0.17638889 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 1.95833333 & 1.29166667 & 0.125 \\ 1.29166667 & 1.70833333 & -0.375 \\ 0.125 & -0.375 & 0.29166667 \end{bmatrix} \text{ with } \tilde{\Delta}(T_2) = \begin{bmatrix} 0.54166667 \\ 0.04166667 \\ -0.20833333 \end{bmatrix}$$

$$T_3 = \begin{bmatrix} 0.21328829 & 0.1231982 & 0.10067568 \\ 0.1231982 & 0.48220721 & 0.16824324 \\ 0.10067568 & 0.16824324 & 0.25833333 \end{bmatrix} \text{ with } \tilde{\Delta}(T_3) = \begin{bmatrix} -0.01058559 \\ 0.19076577 \\ -0.01058559 \end{bmatrix}$$

(A.1)

We see that we clearly have different non-diagonally dominant components in each element mass matrix. On the other hand, we follow the same steps in Figure A.2 but with triangles contained in our region. Here we provide the element mass matrices that are clearly diagonally dominant based on the column vectors calculated below. Notice that we chose T_1 to be an equilateral triangle, T_2 to be an acute triangle, and T_3 to be an obtuse triangle.

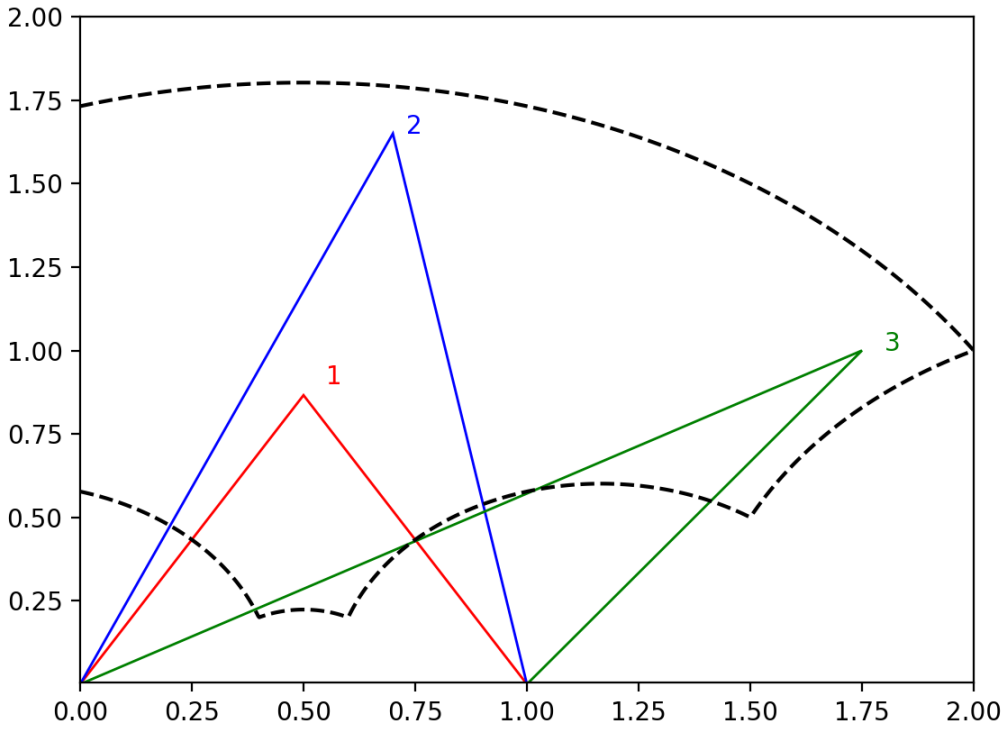


Figure A.2: These three example triangles T_1 (red), T_2 (blue), and T_3 (green) are diagonally dominant.

$$T_1 = \begin{bmatrix} 0.24056261 & 0.04811252 & -0.04811252 \\ 0.04811252 & 0.24056261 & 0.04811252 \\ -0.04811252 & 0.04811252 & 0.24056261 \end{bmatrix} \text{ with } \tilde{\Delta}(T_1) = \begin{bmatrix} 0.14433757 \\ 0.14433757 \\ 0.14433757 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 0.24810606 & 0.14709596 & 0.07638889 \\ 0.14709596 & 0.43118687 & 0.10669192 \\ 0.07638889 & 0.10669192 & 0.20770202 \end{bmatrix} \text{ with } \tilde{\Delta}(T_2) = \begin{bmatrix} 0.02462121 \\ 0.17739899 \\ 0.02462121 \end{bmatrix}$$

$$T_3 = \begin{bmatrix} 0.56770833 & 0.40104167 & 0.109375 \\ 0.40104167 & 0.66145833 & -0.015625 \\ 0.109375 & -0.015625 & 0.15104167 \end{bmatrix} \text{ with } \tilde{\Delta}(T_3) = \begin{bmatrix} 0.05729167 \\ 0.24479167 \\ 0.02604167 \end{bmatrix}$$

(A.2)

Clearly, all these matrices are diagonally dominant. As expected, the equilateral triangle's column elements are equal. Thus, we see these examples confirm our results numerically.

BIBLIOGRAPHY

[1] Douglas N. Arnold, Richard S. Falk, and R. Winther. "Preconditioning in H(div) and Applications". In: *Math. Comput.* 66.219 (July 1997), pp. 957–984. ISSN: 0025-5718. DOI: 10.1090/S0025-5718-97-00826-0. URL: http://dx.doi.org/10.1090/S0025-5718-97-00826-0.

[2] Douglas N. Arnold, Richard S. Falk, and Ragnar Winther. "Finite element exterior calculus, homological techniques, and applications". In: *Acta Numerica* 15 (2006), 1–155. DOI: 10.1017/S0962492906210018.

[3] W. E. Arnoldi. "The principle of minimized iterations in the solution of the matrix eigenvalue problem". In: *Quarterly of Applied Mathematics* 9.1 (1951), pp. 17–29. DOI: 10.1090/qam/42792. URL: https://doi.org/10.1090%2Fqam%2F42792.

[4] Ivo Babuška. "Error-bounds for finite element method". In: *Numerische Mathematik* 16.4 (1971), pp. 322–333. ISSN: 0945-3245. DOI: 10.1007/BF02165003. URL: https://doi.org/10.1007/BF02165003.

[5] Michele Benzi. "Preconditioning Techniques for Large Linear Systems: A Survey". In: *Journal of Computational Physics* 182.2 (2002), pp. 418 –477. ISSN: 0021-9991. DOI: https://doi.org/10.1006/jcph.2002.7176. URL: http://www.sciencedirect.com/science/article/pii/S0021999102971767.

[6] S. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics. Springer New York, 2002. ISBN: 9780387954516. URL: https://books.google.com/books?id=YhPJf_4pu8kC.

[7] Ke Chen. *Matrix preconditioning techniques and applications*. eng. Cambridge monographs on applied and computational mathematics ; 19. Cambridge University Press, 2005. ISBN: 051111558X.

[8] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2002. ISBN: 9780898715149. URL: https://books.google.com/books?id=1PF-WS0Nl9IC.

[9] Richard Comblen et al. "Practical evaluation of five partly discontinuous finite element pairs for the non-conservative shallow water equations". In: *International Journal for Numerical Methods in Fluids* 63.6 (2010), pp. 701–724. DOI: 10.1002/fld.2094. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/fld.2094. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/fld.2094.

[10] C.J. Cotter and D.A. Ham. "Numerical wave propagation for the triangular P1DG–P2 finite element pair". In: *Journal of Computational Physics* 230.8 (2011), pp. 2806 –2820. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2010.12.024. URL: http://www.sciencedirect.com/science/article/pii/S0021999110006984.

[11] C.J. Cotter and J. Shipton. "Mixed finite elements for numerical weather prediction". In: *Journal of Computational Physics* 231.21 (2012), pp. 7076 –7091. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2012.05.020. URL: http://www.sciencedirect.com/science/article/pii/S0021999112002628.

[12] Colin J. Cotter and Robert C. Kirby. "Mixed finite elements for global tide models". In: *Numerische Mathematik* 133.2 (June 2016), pp. 255–277. ISSN: 0945-3245. DOI: 10.1007/s00211-015-0748-z. URL: https://doi.org/10.1007/s00211-015-0748-z.

[13] Lawrence C. Cowsat, Todd F. Dupont, and Mary F. Wheeler. "A priori estimates for mixed finite element methods for the wave equation". In: *Computer Methods in Applied Mechanics and Engineering* 82.1 (1990). Proceedings of the Workshop on Reliability in Computational Mechanics, pp. 205 –222. ISSN: 0045-7825. DOI: https://doi.org/10.1016/0045-7825(90)90165-I. URL: http://www.sciencedirect.com/science/article/pii/004578259090165I.

[14] Sergey Danilov. "On utility of triangular C-grid type discretization for numerical modeling of large-scale ocean flows". In: *Ocean Dynamics* 60.6 (Dec. 2010), pp. 1361–1369. ISSN: 1616-7228. DOI: 10.1007/s10236-010-0339-6. URL: https://doi.org/10.1007/s10236-010-0339-6.

[15] Chris Garrett and Eric Kunze. "Internal Tide Generation in the Deep Ocean". In: *Annual Review of Fluid Mechanics* 39.1 (2007), pp. 57–87. DOI: 10.1146/annurev.fluid.39.050905.110227. eprint: https://doi.org/10.1146/annurev.fluid.39.050905.110227. URL: https://doi.org/10.1146/annurev.fluid.39.050905.110227.

[16] S. Gershgorin. "Über die Abgrenzung der Eigenwerte einer Matrix." Russian. In: *Bull. Acad. Sci. URSS* 1931.6 (1931), pp. 749–754.

[17] Geveci, Tunc. "On the application of mixed finite element methods to the wave equations". In: *ESAIM: M2AN* 22.2 (1988), pp. 243–250. DOI: 10.1051/m2an/1988220202431. URL: https://doi.org/10.1051/m2an/1988220202431.

[18] D. F. Hill et al. "High-resolution numerical modeling of tides in the western Atlantic, Gulf of Mexico, and Caribbean Sea during the Holocene". In: *Journal of Geophysical Research: Oceans* 116.C10 (2011). DOI: 10.1029/2010JC006896. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JC006896. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006896.

[19] Colin J. Cotter, P Graber, and Robert Kirby. "Mixed finite elements for global tide models with nonlinear damping". In: *Numerische Mathematik* (June 2017). DOI: 10.1007/s00211-018-0980-4.

[20] Steven R. Jayne and Louis C. St. Laurent. "Parameterizing tidal dissipation over rough topography". In: *Geophysical Research Letters* 28.5 (2001), pp. 811–814. DOI: 10.1029/2000GL012044. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000GL012044. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL012044.

[21] E. Jenkins, B. Riviaere, and M. Wheeler. "A Priori Error Estimates for Mixed Finite Element Approximations of the Acoustic Wave Equation". In: *SIAM Journal on Numerical Analysis* 40.5 (2002), pp. 1698–1715. DOI: 10.1137/S0036142901388068. eprint: https://doi.org/10.1137/S0036142901388068. URL: https://doi.org/10.1137/S0036142901388068.

[22] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Dover Books on Mathematics Series. Dover Publications, Incorporated, 2012. ISBN: 9780486131597. URL: https://books.google.com/books?id=PYXjyoqy5qMC.

[23] Robert C. Kirby and Thinh Tri Kieu. "Symplectic-mixed finite element approximation of linear acoustic wave equations". In: *Numerische Mathematik* 130.2 (June 2015), pp. 257–291. ISSN: 0945-3245. DOI: 10.1007/s00211-014-0667-4. URL: https://doi.org/10.1007/s00211-014-0667-4.

[24] Robert C. Kirby et al. "Common and unusual finite elements". In: *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book*. Ed. by Anders Logg, Kent-Andre Mardal, and Garth Wells. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 95–119. DOI: 10.1007/978-3-642-23099-8_3. URL: https://doi.org/10.1007/978-3-642-23099-8_3.

[25] H. Lamb. *Hydrodynamics. 6th edition*. C.U.P, 1932. URL: https://books.google.com/books?id=nOHRXwAACAAJ.

[26] D. Le Roux. "Dispersion Relation Analysis of the PNC1 - P1 Finite-Element Pair in Shallow-Water Models". In: *SIAM Journal on Scientific Computing* 27.2 (2005), pp. 394–414. DOI: 10.1137/030602435. eprint: https://doi.org/10.1137/030602435. URL: https://doi.org/10.1137/030602435.

[27] D. Le Roux, V. Rostand, and B. Pouliot. "Analysis of Numerically Induced Oscillations in 2D Finite-Element Shallow-Water Models Part I: Inertia-Gravity Waves". In: *SIAM Journal on Scientific Computing* 29.1 (2007), pp. 331–360. DOI: 10.1137/060650106. eprint: https://doi.org/10.1137/060650106. URL: https://doi.org/10.1137/060650106.

[28] Kent-André Mardal and Ragnar Winther. "Preconditioning discretizations of systems of partial differential equations". In: *Numerical Lin. Alg. with Applic.* 18 (2011), pp. 1–40.

[29] Walter Munk and Carl Wunsch. "Abyssal recipes II: energetics of tidal and wind mixing". In: *Deep Sea Research Part I: Oceanographic Research Papers* 45.12

(1998), pp. 1977 –2010. ISSN: 0967-0637. DOI: https://doi.org/10.1016/S0967-0637(98)00070-3. URL: http://www.sciencedirect.com/science/article/pii/S0967063798000703.

[30] J N. Reddy. *An Introduction to Finite Element Method*. Jan. 2006. ISBN:

[31] 9780072466850. DOI: 10.1115/1.3265687.

[32] P. A. Raviart and Justin M. Thomas. "Primal hybrid finite element methods for 2nd order elliptic equations". In: 1977.

[33] V. Rostand and D. Y. Le Roux. "Raviart–Thomas and Brezzi–Douglas–Marini finite-element approximations of the shallow-water equations". In: *International Journal for Numerical Methods in Fluids* 57.8 (2008), pp. 951–976. DOI: 10.1002/fld.1668. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/fld.1668. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/fld.1668.

[34] Daniel Y. Le Roux. "Spurious inertial oscillations in shallow-water models". In: *Journal of Computational Physics* 231.24 (2012), pp. 7959 –7987. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2012.04.052. URL: http://www.sciencedirect.com/science/article/pii/S0021999112002872.

[35] Daniel Y. Le Roux and Benoit Pouliot. "Analysis of Numerically Induced Oscillations in Two-Dimensional Finite-Element Shallow-Water Models Part II: Free Planetary Waves". In: *SIAM J. Scientific Computing* 30 (2008), pp. 1971–1991.

[36] J. Ruppert. "A Delaunay Refinement Algorithm for Quality 2-Dimensional Mesh Generation". In: *J. Algorithms* 18.3 (May 1995), pp. 548–585. ISSN: 0196-6774. DOI: 10.1006/jagm.1995.1021. URL: http://dx.doi.org/10.1006/jagm.1995.1021.

[37] Y. Saad and M. Schultz. "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems". In: *SIAM Journal on Scientific and Statistical Computing* 7.3 (1986), pp. 856–869. DOI: 10.1137/0907058. eprint: https://doi.org/10.1137/0907058. URL: https://doi.org/10.1137/0907058.

[37] Jonathan Richard Shewchuk. "Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator". In: (1996). Ed. by Ming C. Lin and Dinesh Manocha, pp. 203–222.

[38] D. Stammer et al. "Accuracy assessment of global barotropic ocean tide models". In: *Reviews of Geophysics* 52.3 (2014), pp. 243–282. DOI: 10.1002/2014RG000450. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014RG000450. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014RG000450.

[39] L.N. Trefethen and D. Bau. *Numerical Linear Algebra*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, 1997. ISBN: 9780898713619. URL: https://books.google.com/books?id=bj-Lu6zjWbEC.

[40] A. J. Wathen. "Preconditioning". In: *Acta Numerica* 24 (2015), 329–376. DOI: 10.1017/S0962492915000021.

[41] Hilary Weller et al. "Challenges Facing Adaptive Mesh Modeling of the Atmosphere and Ocean". In: *Bulletin of the American Meteorological Society* 91.1 (2010), pp. 105–108. DOI: 10.1175/2009BAMS2907.1. eprint: https://doi.org/10.1175/2009BAMS2907.1. URL: https://doi.org/10.1175/2009BAMS2907.1.