Finite Element Methods for the Shallow Water Equations

Master's Thesis

Katherine L. McLain

Advisor: Prof. Dr. Volker John
Second Reviewer: Dr. Alfonso Caiazzo

Contents

List of Acronyms and Symbols

The following list defines the acronyms and symbols that will be used throughout this work. The physical units of symbols (where applicable) are included. Units are also presented with the first use of the symbol in the text.

Acronyms

ADCIRC	advanced circulation model
CE	continuity equation
CE-CME	shallow water system pairing the continuity equation with the conservative momentum equation
CE-NCME	shallow water system pairing the continuity equation with the nonconservative momentum equation
CG	continuous Galerkin
CME	conservative momentum equation
DG	discontinuous Galerkin
FEM	finite element method
GWCE	generalized wave continuity equation
GWCE-CME	shallow water system pairing the generalized wave continuity equation with the conservative momentum equation
NCME	nonconservative momentum equation
ParMooN	"Parallel Mathematics and Object-Oriented Numerics" finite element package
SWE	shallow water equations
UTBEST	University of Texas Bays and Estuaries Simulator
WCE	wave continuity equation

Symbols

Λ	rotation rate of the Earth	$\mathrm{rad\,s^{-1}}$
ν	kinematic viscosity	$\mathrm{m^2\,s^{-1}}$

$\overline{\mathbf{T}}$	matrix of stress terms	$\mathrm{N\,m^{-2}}$
ρ	density	$\mathrm{kg\,m^{-3}}$
τ_{bx}, τ_{by}	bottom shear stress	$\mathrm{N\,m^{-2}}$
τ_{ij}	viscous stress	$\mathrm{N\,m^{-2}}$
τ_{sx}, τ_{sy}	surface shear stress	$\mathrm{N\,m^{-2}}$
$\mathbf{b}$	body force	$\mathrm{N\,kg^{-1}}$
$\mathbf{F}_c$	Coriolis force	$\mathrm{m\,s^{-2}}$
$\mathbf{T}$	Cauchy stress tensor	$\mathrm{N\,m^{-2}}$
$\mathbf{v}$	velocity	$\mathrm{m\,s^{-1}}$
θ	geographic latitude	$-$
a	depth of flow	m
c_f	friction coefficient	$-$
f	Coriolis parameter	$\mathrm{s^{-1}}$
F_x, F_y	driving forces	$\mathrm{N\,m^{-2}}$
g	gravitational acceleration	$\mathrm{m\,s^{-2}}$
H	surface elevation	m
p	fluid pressure	$\mathrm{N\,m^{-2}}$
p_a	atmospheric pressure	$\mathrm{N\,m^{-2}}$
t	time	s
u	velocity in x-direction	$\mathrm{m\,s^{-1}}$
v	velocity in y-direction	$\mathrm{m\,s^{-1}}$
w	velocity in z-direction	$\mathrm{m\,s^{-1}}$
x	spatial coordinate	m
y	spatial coordinate	m
z	spatial coordinate	m
z_b	bottom height	m

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The shallow water equations, which are derived from the Navier-Stokes equations, describe flows in domains where the horizontal length scale is significantly greater than the vertical length scale. Simulations of this set of equations have been used extensively in many areas of research, mainly weather and environmental studies, and including, but not limited to, modeling tidal flows, storm surges caused by tropical storms and hurricanes, and atmospheric flows.

Due to the complexity of this system of partial differential equations, it is necessary to use numerical methods to approximate the solution as the analytic solution is impossible to obtain. Many approaches have been suggested in the literature to solve the shallow water equations, including finite difference, finite volume and finite element methods. The finite element method is of particular interest in this work because of the flexibility that it allows in the representation of the complex geometries present in most real-world applications.

The goal of the thesis is to provide an introduction to the literature that examines finite element methods for approximating the solution of the shallow water equations. There are a few notable challenges in determining suitable finite element pairs for the discretization of the shallow water equations, which are discussed in detail, along with techniques that have been developed to address these issues. Due to the breadth of research that has been written on the topic over the last decades, the literature review will be limited to a few select works.

Finite element methods for the incompressible Navier-Stokes equations have been implemented in ParMooN [46], an open source finite element package used to numerically approximate the solutions to elliptic and parabolic partial differential equations. Given that the shallow water equations are derived from the Navier-Stokes equations, an additional aim of this work is to modify the existing code to allow for the implementation of the shallow water equations. This first implementation of the shallow water equations in ParMooN focuses specifically on the 2D inviscid, nonconservative form of the equations and uses finite-element pairs from the literature review. Numerical studies, such as convergence and error analysis, are performed to validate the implementation of the equations.

1.2   Outline

The chapters of this work will adhere to the following structure:

- Chapter 2 provides a basic introduction to the shallow water equations and discusses their various applications. A detailed derivation of the incompressible Navier-Stokes equations is provided, followed by a thorough derivation of the shallow water equations through depth integration. The derivation includes a discussion of appropriate boundary conditions and this section concludes with a brief introduction to various forms of the equations.

- In Chapter 3, the fundamental theorems, definitions, and concepts of the finite element method are presented as a basic introduction to finite element theory. This theoretical portion is followed by a historical overview of the literature discussing finite element methods for the shallow water equations. A few select works are discussed in further detail to provide concrete examples. In conclusion, this section presents *a priori* error estimates for various finite element models of shallow water systems.

- Chapter 4 discusses the implementation of a Galerkin finite element method for the shallow water equations in ParMooN. The boundary conditions, finite element formulation and time discretization of the model are discussed in detail. Visualizations of the solutions of the two examples are presented along with a brief discussion of the results and possible improvements.

- Chapter 5 summarizes the findings of this thesis and discusses possibilities for further avenues of research.

Chapter 2

The Shallow Water Equations

2.1 Background and Applications

The shallow water equations (SWE) are a system of hyperbolic (or parabolic) equations
that describe fluid flow in domains where the vertical length scale H is significantly
smaller than the horizontal length scale L, i.e. $H \ll L$. In these situations, the equations can be simplified by averaging over the depth, allowing the vertical dimension to
be neglected. More specifically, the shallow water equations are derived through depth
integration of the 3D incompressible Navier-Stokes equations, which themselves are derived from the physical laws of conservation of mass and conservation of momentum. An
additional important condition for the domains of shallow water problems is that the
bottom topography does not change too quickly [45].

One should not be deceived by the name of the shallow water equations as it can be misleading as to where these equations can be applied. Contrary to what is suggested by the
name, the equations can be used to describe other flows besides water, e.g. atmospheric
flows. Additionally, the word *shallow* can be misleading as a domain with a high ratio of
horizontal to vertical length scale does not necessarily imply the domain is shallow in the
colloquial sense [25]. This can be seen in modeling the propagation of a tsunami wave.
Although the ocean is a deep body of water, the tsunami waves can build up across the
entire ocean, making the wavelength still significantly greater in comparison. Indeed, this
is a common application of the shallow water equations. While there is not a concrete
definition of what the ratio of the vertical to horizontal length scale in a domain should be
for shallow water theory to apply, suggestions for an upper limit of around $H/L < 10^{-1}$
are given [45].

In addition to the previously mentioned examples, the shallow water equations have a
wide range of other applications. The shallow water equations can be used to model tides
and surges (such as tsunamis or hurricanes) caused by earthquakes or storms. Knowledge
of the tides is especially important in planning the development of coastal regions. Models
of tidal fluctuations can also be useful for those interested in harnessing tidal energy. The
shallow water equations can also be coupled with a transport model, which can be useful
in studying polluted bodies of water and informing environmental remediation efforts.
Furthermore, these models can be used to study acceptable discharge levels necessary
to meet water quality standards, which is important for wastewater management. The
shallow water equations can also be used to model atmospheric and planetary flows and

have been used extensively in weather forecasting. A more comprehensive overview of applications can be found in Chapter 1 of [41].

The derivation of the shallow water equations follows three basic steps, which will be presented in detail in the following section:

1. Derive the incompressible Navier-Stokes equations from the physical laws of conservation of mass and conservation of momentum.

2. Define appropriate free-surface and boundary conditions for the Navier-Stokes equations.

3. Assume hydrostatic pressure and use the specified boundary conditions to integrate the Navier-Stokes equations over depth.

The next section will begin with a basic introduction to the Navier-Stokes equations.

2.2 Derivation of the Incompressible Navier-Stokes Equations

The Navier-Stokes equations are the standard equations for fluid dynamics and are derived from the following two physical properties: the law of conservation of mass and the law of conservation of momentum. The derivation of these equations closely follows the depiction in [10].

Conservation of Mass

The law of conservation of mass states that the rate of change of mass in a volume is equal to the flux of the mass across the boundary. Consider an arbitrary volume $\Omega \subset \mathbb{R}^3$. The above property can be written as the following

$$\frac{d}{dt} \int_\Omega \rho \, dV = - \int_{\partial\Omega} \rho \mathbf{v} \cdot \mathbf{n} \, dA, \tag{2.1}$$

where t is the time (s), ρ is the fluid density (kg/m^3), the vector $\mathbf{v} = (u, v, w)$ is the fluid velocity (m/s), and $\mathbf{n}$ is the outward unit normal vector to the boundary $\partial\Omega$. Applying the divergence theorem to the right-hand side gives

$$\frac{d}{dt} \int_\Omega \rho \, dV = - \int_\Omega \nabla \cdot (\rho \mathbf{v}) \, dV.$$

Assuming that ρ is sufficiently smooth, the integral and derivative on the left-hand side can be interchanged using the Leibniz integral rule, yielding

$$\int_\Omega \frac{\partial}{\partial t} \rho \, dV = - \int_\Omega \nabla \cdot (\rho \mathbf{v}) \, dV.$$

Combining the integrals and rearranging the terms gives

$$\int_\Omega \left(\frac{\partial}{\partial t} \rho + \nabla \cdot (\rho \mathbf{v}) \right) dV = 0.$$

Since Ω is an arbitrary volume, it can be assumed that the integrand is equal to zero, and hence

$$\frac{\partial}{\partial t}\rho + \nabla \cdot (\rho \mathbf{v}) = 0.$$

The above equation, called the continuity equation (CE), is the first of the two Navier-Stokes equations. The derivation of the second Navier-Stokes equation follows a similar process.

Conservation of Momentum

The law of conservation of momentum implies that the rate of change of the momentum in an arbitrary volume Ω is equal to the sum of the flux of the momentum across the boundary $\partial \Omega$, the body forces acting on the volume Ω, and the contact forces acting on the boundary $\partial \Omega$. This property can be written out as the following equation

$$\frac{d}{dt}\int_{\Omega} \rho \mathbf{v} \, dV = -\int_{\partial\Omega} (\rho \mathbf{v})\mathbf{v} \cdot \mathbf{n} \, dA + \int_{\Omega} \rho \mathbf{b} \, dV + \int_{\partial\Omega} \mathbf{T}\mathbf{n} \, dA, \qquad (2.2)$$

where $\mathbf{b}$ is the body force density per unit mass (N/kg) acting on the fluid and $\mathbf{T}$ is the Cauchy stress tensor (N/m^2). These body and contact forces will be discussed in more detail below.

As with the continuity equation, the divergence theorem can be applied to the first and third terms on the right-hand side, which gives

$$\frac{d}{dt}\int_{\Omega} \rho \mathbf{v} \, dV = -\int_{\Omega} \nabla \cdot (\rho \mathbf{v}\mathbf{v}) \, dV + \int_{\Omega} \rho \mathbf{b} \, dV + \int_{\Omega} \nabla \cdot \mathbf{T} \, dV.$$

Under the assumption that $\rho \mathbf{v}$ is sufficiently smooth, the Leibniz integral rule can be applied and allows for the interchange of the integral and derivative, yielding

$$\int_{\Omega} \frac{\partial}{\partial t}(\rho \mathbf{v}) \, dV = -\int_{\Omega} \nabla \cdot (\rho \mathbf{v}\mathbf{v}) \, dV + \int_{\Omega} \rho \mathbf{b} \, dV + \int_{\Omega} \nabla \cdot \mathbf{T} \, dV.$$

Combining the integrals and rearranging the terms gives

$$\int_{\Omega} \left(\frac{\partial}{\partial t}(\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) - \rho \mathbf{b} - \nabla \cdot \mathbf{T} \right) dV = 0.$$

Again, since Ω is an arbitrary volume,

$$\frac{\partial}{\partial t}(\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) - \rho \mathbf{b} - \nabla \cdot \mathbf{T} = 0.$$

This is the second of the Navier-Stokes equations, the so-called momentum equation.

Conservation Laws in Differential Form

In summary, the following differential forms of the equations for conservation of mass and linear momentum have been derived thus far

$$\frac{\partial}{\partial t}\rho - \nabla \cdot (\rho \mathbf{v}) = 0, \qquad (2.3)$$

$$\frac{\partial}{\partial t}(\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v} \mathbf{v}) - \rho \mathbf{b} - \nabla \cdot \mathbf{T} = 0. \tag{2.4}$$

In order to derive the Navier-Stokes equations, it is necessary to next make assumptions about the density ρ, body forces $\mathbf{b}$, and the stress tensor $\mathbf{T}$.

Incompressible Navier-Stokes Equations

The focus of this work will be restricted to the incompressible Navier-Stokes equations, i.e. it is assumed that the density ρ is not dependent on the pressure p. It is important to note that this does not necessarily imply that the density is constant, as the density could still vary for other reasons such as temperature and salinity. In this work, however, any additional dependencies will be assumed to be constant in order to assume a constant density ρ. Based on this assumption, the continuity equation (2.3) can be reduced to

$$\nabla \cdot \mathbf{v} = 0. \tag{2.5}$$

Body and Contact Forces

In fluid dynamics, body forces are external forces that act on the entire fluid uniformly. The origin of the force is distant from the domain of interest and the strength of such forces changes very slowly [33]. The body force term in the momentum equation (2.4) can be rewritten as

$$\rho \mathbf{b} = \rho \left(\mathbf{g} + \mathbf{b}_{\text{others}} \right), \tag{2.6}$$

where $\mathbf{g}$ represents the acceleration due to gravity (m/s^2) and $\mathbf{b}_{\text{others}}$ includes all other body forces. In this derivation, the Coriolis force $\mathbf{F}_c$ (m/s^2) will be the only body force besides gravity taken into consideration.

The Coriolis force can be summarized as a 'pseudo-force' that arises for objects moving on a rotating body and consists of two different effects [35]. First, an object that is moving in the same direction in absolute space will appear to change direction as the Earth rotates. Second, an object on the Earth's surface carries the velocity of the Earth's rotation with it. Each of these two effects causes an acceleration of the magnitude $\Lambda \sin \theta \mathbf{v}$, where Λ is the Earth's rate of rotation and θ is the geographic latitude. Therefore, the combination of these two effects leads to a total acceleration of $2\Lambda \sin \theta \mathbf{v}$ or $f\mathbf{v}$, with the Coriolis parameter $f = 2\Lambda \sin \theta$. The Coriolis force acts at a 90-degree angle to the movement of the object, which leads to the following formulation of the Coriolis force in the Navier-Stokes equations

$$\mathbf{F}_c = f\mathbf{k} \times \mathbf{v},$$

where $\mathbf{k}$ is a normal vector in the vertical direction.

Contact forces can be written in the following form for a Newtonian fluid

$$\mathbf{T} = -p\mathbf{I} + \overline{\mathbf{T}}, \tag{2.7}$$

where p is the pressure of the fluid (N/m^2), $\mathbf{I}$ is the identity tensor, and $\overline{\mathbf{T}}$ is a matrix of stress terms. $\overline{\mathbf{T}}$ consists of nine components and represents the state of stress at a given point:

$$\begin{pmatrix} \tau_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \tau_{yy} & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \tau_{zz} \end{pmatrix}.$$

In this work, τ_{ij} is considered to represent the viscous stress at a given point and can be expressed as

$$\tau_{ij} = \rho\nu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \tag{2.8}$$

where ν is the kinematic viscosity (m^2/s). To simplify the notation, x_j stands for (x, y, z) in the coordinate system and u_j for (u, v, w) of the velocity vector $\mathbf{v}$, where $i, j = 1,2,3$.

Three-Dimensional Navier-Stokes Equations

The inclusion of the above assumptions in the momentum equation (2.4) gives the final form of the 3D incompressible Navier-Stokes equations

$$\nabla \cdot \mathbf{v} = 0, \tag{2.9}$$

$$\partial_t \rho\mathbf{v} + \nabla \cdot (\rho\mathbf{v}\mathbf{v}) - \rho\mathbf{g} + \rho f \mathbf{k} \times \mathbf{v} + \nabla p - \nabla \cdot \overline{\mathbf{T}} = 0, \tag{2.10}$$

where the vector $\mathbf{g} = (0, 0, -g)^\mathsf{T}$. The vector form can be equivalently written out as the following system of equations:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{2.11}$$

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho uw)}{\partial z} - \rho f v + \frac{\partial p}{\partial x} - \frac{\partial \tau_{xx}}{\partial x} - \frac{\partial \tau_{xy}}{\partial y} - \frac{\partial \tau_{xz}}{\partial z} = 0, \tag{2.12}$$

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho uv)}{\partial x} + \frac{\partial(\rho v^2)}{\partial y} + \frac{\partial(\rho vw)}{\partial z} + \rho f u + \frac{\partial p}{\partial y} - \frac{\partial \tau_{xy}}{\partial x} - \frac{\partial \tau_{yy}}{\partial y} - \frac{\partial \tau_{yz}}{\partial z} = 0, \tag{2.13}$$

$$\frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho uw)}{\partial x} + \frac{\partial(\rho vw)}{\partial y} + \frac{\partial(\rho w^2)}{\partial z} + \rho g + \frac{\partial p}{\partial z} - \frac{\partial \tau_{xz}}{\partial x} - \frac{\partial \tau_{yz}}{\partial y} - \frac{\partial \tau_{zz}}{\partial z} = 0. \tag{2.14}$$

Here the gravitational constant is $g \approx 9.8$ m/s^2.

2.3 Derivation of the Shallow Water Equations

The shallow water equations can now be derived from the Navier-Stokes equations by depth integration. The derivation in this section closely follows [10] and [41].

2.3.1 Boundary Conditions

Before deriving the equations, it is first necessary to establish the boundary conditions on the domain. The surface and bottom boundary conditions can be classified into two categories: kinematic and dynamic. Kinematics describe the motion of particles, whereas dynamics describe the physical laws that govern the motion. A discussion of the boundary conditions at the surface and bottom will be presented here.
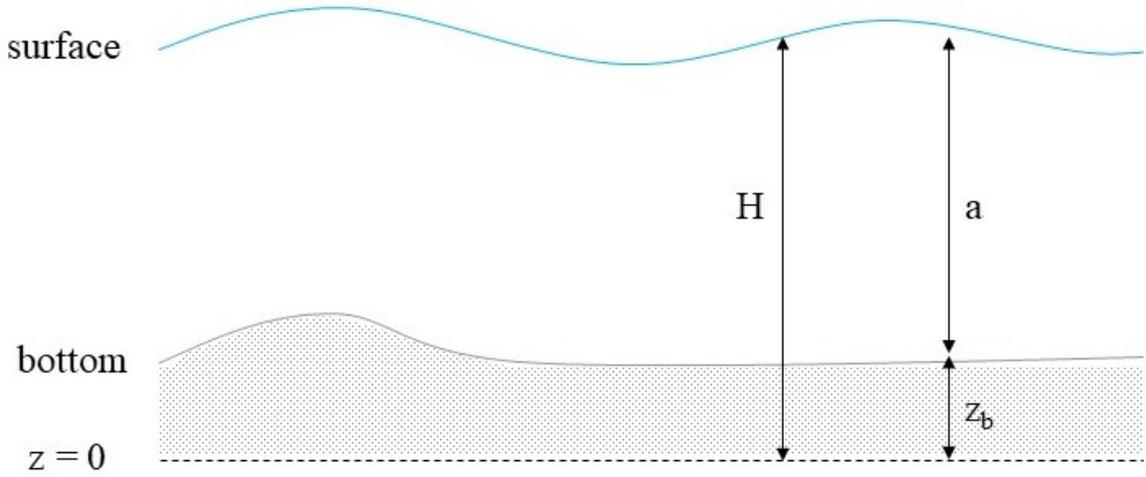
Figure 2.1: Definition of boundaries

Kinematic Boundary Conditions

For the derivation of the shallow water equations, the kinematic boundary conditions state that the fluid particles will not cross the boundaries. For the bottom, this means that the normal velocity component must vanish. In other words, when considering a normal vector to the bottom $\left(\frac{\partial z_b}{\partial x}, \frac{\partial z_b}{\partial y}, -1\right)^{\mathsf{T}}$, the velocity vector $\mathbf{v} = (u, v, w)$ must be perpendicular, i.e. the two vectors must have a dot product of zero. This gives the following equation

$$u\frac{\partial z_b}{\partial x} + v\frac{\partial z_b}{\partial y} - w = 0 \quad \text{at } z = z_b, \tag{2.15}$$

where z_b represents the bottom height measured from a horizontal reference plane (see Figure 2.1).

The surface is more complicated to describe because the surface itself may be moving. A similar equation to (2.15) is derived, with the exception of an additional term compensating for the change in the surface over time

$$\frac{\partial H}{\partial t} + u\frac{\partial H}{\partial x} + v\frac{\partial H}{\partial y} - w = 0 \quad \text{at } z = H, \tag{2.16}$$

where H represents the surface elevation measured from the horizontal reference plane (see Figure 2.1).

Dynamic Boundary Conditions

The dynamic boundary conditions describe the forces acting at the boundaries. At the bottom, there is a so-called "no-slip" condition, which implies $u = v = 0$, i.e. there is no velocity in the x- and y-directions. Additionally, there exist bottom shear stress forces (τ_{bx}, τ_{by}) due to friction

$$\tau_{bx} = \tau_{xx}\frac{\partial z_b}{\partial x} + \tau_{xy}\frac{\partial z_b}{\partial y} + \tau_{xz} \quad \text{at } z = z_b, \tag{2.17}$$

and similarly for the y-direction.

At the surface, a continuity of pressure is assumed, i.e. the pressure of the fluid near the surface is the same as the atmospheric pressure. This assumption yields the following equality

$$p = p_a, \tag{2.18}$$

where p_a represents the atmospheric pressure. As with the bottom, there exist shear stress forces (τ_{sx}, τ_{sy}) acting on the surface in the x- and y-directions due to wind, which can be represented by the following equation

$$\tau_{sx} = -\tau_{xx}\frac{\partial H}{\partial x} - \tau_{xy}\frac{\partial H}{\partial y} + \tau_{xz} \quad \text{at } z = H, \tag{2.19}$$

and similarly for the y-direction.

2.3.2 Hydrostatic Approximation

It is now necessary to examine a central property of shallow water theory called the hydrostatic pressure distribution. For the derivation of the shallow water equations, it is assumed that the pressure is hydrostatic, i.e. given a continuous fluid, the fluid pressure has a linear distribution across the water column [32]. This assumption allows the vertical momentum equation (2.14) to be simplified to the following "state" equation for pressure

$$\frac{\partial p}{\partial z} = -\rho g,$$

as all terms except for the pressure and gravitational acceleration are negligible. Integration of the hydrostatic approximation over depth yields

$$\int_{\bar{z}}^{H} \frac{\partial p}{\partial z}\,dz = \int_{\bar{z}}^{H} -\rho g\,dz,$$

where $\bar{z}$ represents an arbitrary depth. Evaluation of the left-hand side and the substitution of the dynamic surface boundary condition $p = p_a$ at $z = H$ allows the above equation to be rewritten as

$$p = p_a + g\int_{\bar{z}}^{H} \rho\,dz.$$

The assumption that the density ρ is constant over depth allows for further simplification and yields the final pressure estimate

$$p = p_a + \rho g(H - z), \tag{2.20}$$

which can be used to calculate the pressure gradients in the momentum equations. For example, consider the pressure gradient in the x-direction

$$\frac{\partial p}{\partial x} = \frac{\partial(p_a + \rho g(H - z))}{\partial x}.$$

The use of the product rule and linearity yields the following identity, which can be substituted back into the original equations:

$$\frac{\partial p}{\partial x} = \frac{\partial p_a}{\partial x} + g\rho\frac{\partial H}{\partial x}. \tag{2.21}$$

The same process is used to estimate the pressure gradient in the y-direction.

The estimates for the pressure gradient can be inserted into the momentum equations (2.12) and (2.13) and, for simplification, both sides can be divided by the density ρ. This yields the momentum equations thus far:

$$\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}(u^2)+\frac{\partial}{\partial y}(uv)+\frac{\partial}{\partial z}(uw)-fv+\frac{1}{\rho}\frac{\partial p_a}{\partial x}+g\frac{\partial H}{\partial x}-\frac{1}{\rho}\{\frac{\partial\tau_{xx}}{\partial x}+\frac{\partial\tau_{xy}}{\partial y}+\frac{\partial\tau_{xz}}{\partial z}\}=0, \tag{2.22}$$

$$\frac{\partial v}{\partial t}+\frac{\partial}{\partial x}(uv)+\frac{\partial}{\partial y}(v^2)+\frac{\partial}{\partial z}(vw)+fu+\frac{1}{\rho}\frac{\partial p_a}{\partial y}+g\frac{\partial H}{\partial y}-\frac{1}{\rho}\{\frac{\partial\tau_{yx}}{\partial x}+\frac{\partial\tau_{yy}}{\partial y}+\frac{\partial\tau_{yz}}{\partial z}\}=0. \tag{2.23}$$

2.3.3 Depth Integration

The final step in deriving the shallow water equations is the integration of the continuity equation and horizontal momentum equations over the depth $a = H - z_b$. This allows the vertical velocity component to be neglected from the equations.

Continuity Equation

Integration of the continuity equation over the depth gives the following equation

$$0 = \int_{z_b}^{H} \nabla \cdot \mathbf{v}\,dz,$$

which, after evaluating the term containing z, can be simplified to

$$0 = \int_{z_b}^{H} (\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y})\,dz + w|_H - w|_{z_b}.$$

Applying the Leibniz integral rule gives

$$0 = \frac{\partial}{\partial x}\int_{z_b}^{H} u\,dz+\frac{\partial}{\partial y}\int_{z_b}^{H} v\,dz-u|_{z=H}\frac{\partial H}{\partial x}+u|_{z=z_b}\frac{\partial z_b}{\partial x}-v|_{z=H}\frac{\partial H}{\partial y}+v|_{z=z_b}\frac{\partial z_b}{\partial y}+w|_{z=H}-w|_{z=z_b}.$$
$$\tag{2.24}$$

Let the depth-averaged velocities $\bar{u}$ and $\bar{v}$ be defined as

$$\bar{u} = \frac{1}{a}\int_{z_b}^{H} u\,dz \quad\text{and}\quad \bar{v} = \frac{1}{a}\int_{z_b}^{H} v\,dz. \tag{2.25}$$

The insertion of these depth-averaged terms into (2.24) yields

$$0 = \frac{\partial}{\partial x}(a\bar{u}) + \frac{\partial}{\partial y}(a\bar{v}) - u|_H\frac{\partial H}{\partial x} + u|_{z_b}\frac{\partial z_b}{\partial x} - v|_H\frac{\partial H}{\partial y} + v|_{z_b}\frac{\partial z_b}{\partial y} + w|_H - w|_{z_b}.$$

Finally, the kinematic boundary conditions (2.15) and (2.16) allow the above equation to be further simplified and yield the so-called depth-averaged continuity equation

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(a\bar{u}) + \frac{\partial}{\partial y}(a\bar{v}) = 0. \tag{2.26}$$

Momentum Equation

Depth integration of the momentum equation follows a similar process. For the advective term, the surface and bottom terms cancel completely through the boundary conditions, but nonlinear terms of the form

$$\int_{z_b}^{H} uv \, dz = a\bar{u}\bar{v} + \int_{z_b}^{H} (u - \bar{u})(v - \bar{v}) \, dz$$

emerge. The integral on the right-hand side is referred to as the "differential advective term" and accounts for the fact that the average of the product of two functions is not the product of the averages.

Depth integration of the pressure terms and the Coriolis forces is a straightforward calculation. The stress terms are integrated using the Leibniz integral rule and have the following form in the x-momentum equation

$$\int_{z_b}^{H} \left(\frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xz}}{\partial z} \right) dz = \frac{\partial}{\partial x} \int_{z_b}^{H} \tau_{xx} \, dz + \frac{\partial}{\partial y} \int_{z_b}^{H} \tau_{xy} \, dz$$

$$- \left[\tau_{xx} \frac{\partial H}{\partial x} + \tau_{xy} \frac{\partial H}{\partial y} - \tau_{xz} \right]_{z=H} + \left[\tau_{xx} \frac{\partial z_b}{\partial x} + \tau_{xy} \frac{\partial z_b}{\partial y} - \tau_{xz} \right]_{z=z_b},$$

and similarly for the y-momentum equation. The boundary integrals produce the bottom and surface shear stress components described in (2.17) and (2.19). The remaining integrals are absorbed into F_x, F_y defined below.

The resulting depth-averaged momentum equations are

$$\frac{\partial}{\partial t}(a\bar{u}) + \frac{\partial}{\partial x}(a\bar{u}^2) + \frac{\partial}{\partial y}(a\bar{u}\bar{v}) - fa\bar{v} + ga\frac{\partial h}{\partial x} = \frac{1}{\rho}\left[\tau_{sx} - \tau_{bx} + F_x \right], \tag{2.27}$$

$$\frac{\partial}{\partial t}(a\bar{v}) + \frac{\partial}{\partial x}(a\bar{u}\bar{v}) + \frac{\partial}{\partial y}(a\bar{v}^2) + fa\bar{u} + ga\frac{\partial h}{\partial y} = \frac{1}{\rho}\left[\tau_{sy} - \tau_{by} + F_y \right], \tag{2.28}$$

where F_x and F_y are driving forces determined by the data of the problem. These include factors such as atmospheric pressure gradient, wind stress, density gradients, radiation stress, and tidal stresses. For further discussion on the driving forces, see Section 2.6 of [41].

2.4   Shallow Water Equations

Together, the continuity equation (2.26) and momentum equations (2.27) and (2.28) give the full system of shallow water equations:

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(au) + \frac{\partial}{\partial y}(av) = 0, \tag{2.29}$$

16

$$\frac{\partial}{\partial t}(au) + \frac{\partial}{\partial x}(au^2) + \frac{\partial}{\partial y}(auv) - fav + ga\frac{\partial H}{\partial x} = \frac{1}{\rho}\left[\tau_{sx} - \tau_{bx} + F_x\right], \qquad (2.30)$$

$$\frac{\partial}{\partial t}(av) + \frac{\partial}{\partial x}(auv) + \frac{\partial}{\partial y}(av^2) + fau + ga\frac{\partial H}{\partial y} = \frac{1}{\rho}\left[\tau_{sy} - \tau_{by} + F_y\right]. \qquad (2.31)$$

The overlines have been excluded, but it is still assumed that the velocity is depth-averaged.

Forms of the Shallow Water Equations

There are various forms of the shallow water equations and associated terminology that are discussed throughout the literature, a few of which will be briefly introduced here. For more details, refer to Chapter 2 of [22] and Chapter 2 of [41].

Conservative and Nonconservative Momentum Equation   In the literature and numerical studies, it is typically specified whether the conservative or nonconservative form of the momentum equation is used. The conservative form of the momentum equation (CME) is obtained from (2.30) and (2.31) by disregarding lateral stresses and driving forces, as well as assuming the simplest possible expression of the bottom stresses [41]:

$$\frac{\tau_{bx}}{\rho} = c_f u\sqrt{u^2 + v^2},$$

$$\frac{\tau_{by}}{\rho} = c_f v\sqrt{u^2 + v^2},$$

where c_f is a friction coefficient.   The CME reads

$$\partial_t(a\mathbf{v}) + (\mathbf{v}\cdot\nabla)\mathbf{v} + f\mathbf{k}\times(a\mathbf{v}) + ag\nabla H + c_f\|\mathbf{v}\|\mathbf{v} = 0, \qquad (2.32)$$

or equivalently written out:

$$\frac{\partial}{\partial t}(au) + \frac{\partial}{\partial x}(au^2) + \frac{\partial}{\partial y}(auv) - fav + ga\frac{\partial H}{\partial x} + c_f u\sqrt{u^2 + v^2} = 0, \qquad (2.33)$$

$$\frac{\partial}{\partial t}(av) + \frac{\partial}{\partial x}(auv) + \frac{\partial}{\partial y}(av^2) + fau + ga\frac{\partial H}{\partial y} + c_f v\sqrt{u^2 + v^2} = 0. \qquad (2.34)$$

Paired with the continuity equation (2.29), this gives the full system of the shallow water equations, which will be denoted as the CE-CME system.

By performing the differentiations in the momentum equations (2.33) and (2.34) and using the continuity equation (2.29) to cancel out terms, the nonconservative form of the momentum equations (NCME) can be derived:

$$\partial_t\mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} + f\mathbf{k}\times\mathbf{v} + g\nabla H + \frac{c_f}{a}\|\mathbf{v}\|\mathbf{v} = 0, \qquad (2.35)$$

or equivalently written out:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + g\frac{\partial H}{\partial x} + c_f\frac{u}{a}\sqrt{u^2 + v^2} = 0, \qquad (2.36)$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + g\frac{\partial H}{\partial y} + c_f\frac{v}{a}\sqrt{u^2 + v^2} = 0. \qquad (2.37)$$

This is again paired with the continuity equation (2.29) to give the full system of the shallow water equations, which will be denoted as the CE-NCME system. It is important to note that while the conservative and nonconservative forms of the momentum equations are equivalent, this does not hold true for the discretized forms [41].

While the versions of the CME and NCME presented here do not consider lateral stresses and driving forces, there are other versions in the literature that include these forces as well as the viscosity term in equations, e.g. Equation (2.5) in [32]. Therefore, it is important to verify the exact use of the terminology within each individual work. The numerical studies presented in Chapter 4 will use the version of the nonconservative momentum equation presented above.

Dimensionless Shallow Water Equations

Nondimensionalization removes the physical dimensions from the system of equations and yields the so-called dimensionless equations. This process allows for the parametrization of systems with measured units and is particularly useful for mathematical analysis and numerical simulations. A summary of the dimensionless form of the Navier-Stokes equations can be found in Section 2.3 of [21].

To derive the dimensionless shallow water equations, it is first necessary to define the quantities

- $U[\text{m/s}]$: characteristic velocity scale

- $L[\text{m}]$: characteristic length scale

- $H^*[\text{m}]$: characteristic height scale

- $T^*[\text{s}]$: characteristic time scale.

The dimensional variables will be denoted by a prime. Using this notation, the following dimensionless variables can be defined by

$$a = \frac{a'}{H^*}, \quad H = \frac{H'}{H^*}, \quad \mathbf{v} = \frac{\mathbf{v}'}{U}, \quad \mathbf{x} = \frac{\mathbf{x}'}{L}, \quad \text{and} \quad t = \frac{t'}{T^*}.$$

Inserting the transformed variables into the continuity equation (2.29) yields

$$\left(\frac{L}{UT^*}\right)\frac{\partial H}{\partial t} + \nabla \cdot (a\mathbf{v}) = 0.$$

For the momentum equation (2.35), a transform of variables and rescaling gives

$$\left(\frac{L}{UT^*}\right)\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} + \left(\frac{L}{U}\right)(f'\mathbf{k} \times \mathbf{v}) + \left(\frac{H^*}{U^2}\right)g'\nabla H + \left(\frac{L}{H^*}\right)\frac{c'_f}{a}\|\mathbf{v}\|\mathbf{v} = 0.$$

The use of the characteristic time scale $T^* = L/U$ allows the above system to be simplified to

$$\frac{\partial H}{\partial t} + \nabla \cdot (a\mathbf{v}) = 0, \tag{2.38}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} + \left(\frac{L}{U}\right)(f'\mathbf{k} \times \mathbf{v}) + \left(\frac{H^*}{U^2}\right)g'\nabla H + \left(\frac{L}{H^*}\right)\frac{c'_f}{a}\|\mathbf{v}\|\mathbf{v} = 0. \tag{2.39}$$

Let the dimensionless variables f, g, and c_f be defined by

$$f = \frac{f'L}{U}, \quad g = \frac{g'H^*}{U^2}, \quad \text{and } c_f = \frac{c_f'L}{H^*}.$$

Rewriting (2.39) with these variables yields a system equivalent to the dimensional system. Therefore, to arrive at the dimensionless equations, one must only scale the variables g, f, and c_f according to the characteristic scales of the flow problem. The dimensionless form of the equations will be used for the numerical studies in Chapter 4.

Chapter 3

Finite Element Methods for the Shallow Water Equations

3.1 Introduction to Finite Element Methods

For partial differential equations, such as the shallow water equations, an analytic solution may be difficult, even impossible, to calculate. Therefore, it becomes necessary to approximate the solution using numerical methods. The method of focus in this thesis is the so-called finite element method (FEM). This section serves as a basic introduction to the theory of finite element methods and follows closely from Appendix B of [21] and the lecture notes [38].

3.1.1 Overview

The general idea of the finite element method is to look for an approximation of the solution $u(x)$ to a differential equation in a finite-dimensional space V_h with basis $\phi_1, ..., \phi_N$. The approximate solution u_h can then be written in the form

$$u_h(x) = \sum_{i=1}^{N} u_i \phi_i(x).$$

Since the basis $(\phi_i)_{i=1}^{N}$ of V_h is given, the approximate solution u_h is solely determined by the coefficients $u_i \in \mathbb{R}$, which become the unknowns that need to be calculated.

The question becomes how to select an appropriate finite-dimensional space V_h and solve for the coefficients u_i, such that u_h is a good approximation of the solution $u(x)$ of the original differential equation. Either the minimization form or variational (weak) form of the differential equation can be used to derive a linear system of equations to solve for the coefficients u_i. The theory behind these methods will be discussed in the following section.

3.1.2 The Ritz Method and the Galerkin Method

The Ritz method was one of the earliest ideas for approximating the solution to a differential equation in finite-dimensional spaces and involves transforming the original boundary value problem into a minimization problem. The principle behind the Ritz method originates from the Representation Theorem of Riesz.

Representation Theorem of Riesz

Let V be a Hilbert space with an inner product $a(\cdot, \cdot) : V \times V \to \mathbb{R}$ and the norm $\|v\|_V = a(v, v)^{\frac{1}{2}}$. The definition of inner product implies that $a(\cdot, \cdot)$ is symmetric and it can also easily be proven that the inner product is bounded.

Definition 1 (Bounded, Coercive). Let $b(\cdot, \cdot) : V \times V \to \mathbb{R}$ be a bilinear form on the Banach[1] space V. It is *bounded* if

$$|b(u, v)| \leq M \|u\|_V \|v\|_V \quad \forall u, v \in V, \quad M > 0, \tag{3.1}$$

where the constant M is independent of u and v. The bilinear form $b(\cdot, \cdot)$ is considered to be *coercive* if

$$b(u, u) \geq m \|u\|_V^2 \quad \forall u \in V, \quad m > 0, \tag{3.2}$$

where the constant m is independent of u.

Theorem 3.1.1 (Representation Theorem of Riesz). *Let $f \in V'$ be a continuous, linear functional. Then there is a uniquely determined $u \in V$ with*

$$a(u, v) = f(v) \quad \forall v \in V. \tag{3.3}$$

Additionally, u is the unique solution of the variational problem

$$F(v) = \frac{1}{2} a(v, v) - f(v) \to min \quad \forall v \in V. \tag{3.4}$$

Proof. The proof is divided into three parts: first, proving the existence of a solution to the variational problem (3.4), second, showing that the solution of variational problem (3.4) is also the solution of the weak equation (3.3), and third, proving the uniqueness of the solution.

1. Proof of the existence of a solution to the variational problem (3.4)
Since f is a continuous functional, there exists a c such that

$$|f(v)| \leq c \|v\|_V \quad \forall v \in V.$$

Inserting the above inequality into (3.4) yields the following

$$F(v) = \frac{1}{2} \underbrace{a(v, v)}_{\|v\|_V^2} - \underbrace{f(v)}_{\leq c\|v\|_V} \geq \frac{1}{2} \|v\|_V^2 - c\|v\|_V \geq -\frac{1}{2} c^2,$$

where the second inequality comes from the minimum of the quadratic. Hence, $F(\cdot)$ is bounded from below and there exists $d = \inf_{v \in V} F(v)$.
Define a sequence $\{v_k\}_{k \in \mathbb{N}}$ such that $F(v_k) \to d$ as $k \to \infty$. The parallelogram law of Hilbert spaces gives the following equality

$$\|v_k - v_l\|_V^2 + \|v_k + v_l\|_V^2 = 2\|v_k\|_V^2 + 2\|v_l\|_V^2.$$

Rearranging the above equation gives

$$\|v_k - v_l\|_V^2 = 2\|v_k\|_V^2 + 2\|v_l\|_V^2 - \|v_k + v_l\|_V^2.$$

---

[1]Hilbert spaces are by definition also Banach spaces.

Using the linearity of $f(\cdot)$ and the properties of norms, it can be determined that

$$\|v_k - v_l\|_V^2 = 2\|v_k\|_V^2 + 2\|v_l\|_V^2 - 4\|\frac{v_k + v_l}{2}\|_V^2$$

$$= 2\|v_k\|_V^2 + 2\|v_l\|_V^2 - 4\|\frac{v_k + v_l}{2}\|_V^2 + \underbrace{(8f(\frac{v_k + v_l}{2}) - 4f(v_k) - 4f(v_l))}_{=0}.$$

Rearranging the terms above helps to simplify to

$$\|v_k - v_l\|_V^2 = \underbrace{2\|v_k\|_V^2 - 4f(v_k)}_{4F(v_k)} + \underbrace{2\|v_l\|_V^2 - 4f(v_l)}_{4F(v_l)} - \underbrace{4\|\frac{v_k + v_l}{2}\|_V^2 + 8f(\frac{v_k + v_l}{2})}_{8F(\frac{v_k+v_l}{2})}.$$

Since $d \le F(v)$, it can be determined that

$$\|v_k - v_l\|_V^2 \le 4F(v_k) + 4F(v_l) - 8d \to 0 \quad \text{for} \quad k, l \to \infty.$$

Therefore, $\{v_k\}_{k \in \mathbb{N}}$ is a Cauchy sequence. The space V is complete, which implies, by definition, that there exists a limit u of $\{v_k\}_{k \in \mathbb{N}}$ with $u \in V$. Since $F(\cdot)$ is continuous, $F(u) = d$ and u is the solution to the variational problem.

2. Proof that the solution of the variational problem (3.4) is also a solution of the weak equation (3.3)
Define $\Phi(\varepsilon) = F(u + \varepsilon v)$. By definition of $F(\cdot)$,

$$\Phi(\varepsilon) = \frac{1}{2}a(u + \varepsilon v, u + \varepsilon v) - f(u + \varepsilon v).$$

The linearity and symmetry of the inner product $a(\cdot, \cdot)$ allows the above equation to be rewritten as

$$\Phi(\varepsilon) = \frac{1}{2}a(u, u) + \varepsilon a(u, v) + \frac{\varepsilon_2}{2}a(v, v) - f(u) - \varepsilon f(v).$$

If u is the minimum of the variational problem (3.4), $\Phi(\varepsilon)$ must have a local minimum at $\varepsilon = 0$. The necessary condition for a local minimum leads to

$$0 = \Phi'(0) = a(u, v) - f(v) \quad \forall v \in V,$$

and hence proving u is also a solution of the weak equation (3.3).

3. Proof of uniqueness of the solution
It is sufficient to prove the uniqueness of the solution of the weak equation (3.3). If the solution of (3.3) is unique, two solutions to the variational problem (3.4) would be a contradiction. Let u_1 and u_2 be two solutions of (3.3). The difference equation is

$$a(u_1 - u_2, v) = 0 \quad \forall v \in V.$$

The above equation holds for $v = u_1 - u_2$, which implies that $\|u_1 - u_2\|_V = 0$. Hence, $u_1 = u_2$ and the uniqueness of the solution is proven. $\qquad \square$

The Ritz Method

The Ritz method assumes V to have a countable, orthonormal basis. In other words, V is a separable Hilbert space. Following from Parseval's equality, finite-dimensional subspaces $V_1, V_2, \ldots \subset V$ with $\dim V_k = k$ can be found, which satisfy the following property: for each $u \in V$ and each $\varepsilon > 0$, there exists a $K \in \mathbb{N}$ and $u_k \in V_k$ such that

$$\|u - u_k\|_V \leq \varepsilon \quad \forall k \geq K. \tag{3.5}$$

This leads to the definition of the so-called Ritz approximation: Find $u_k \in V_k$ with

$$a(u_k, v_k) = f(v_k) \quad \forall v_k \in V_k. \tag{3.6}$$

It can be proven that the Ritz approximation has a unique solution, as seen in the following lemma.

Lemma 3.1.2 (Existence and Uniqueness of Solution to the Ritz Approximation). *There exists a single solution of the Ritz approximation (3.6).*

Proof. Finite-dimensional subspaces of Hilbert spaces are also Hilbert spaces. Therefore, the Representation Theorem of Riesz can also be applied to the Ritz approximation (3.6), which gives the statement of the lemma. Furthermore, the solution to the Ritz approximation (3.6) also solves a minimization problem on V_k. $\square$

Not only does exactly one solution of the Ritz approximation (3.6) exist, this unique solution is also the best approximation of u in the finite-dimensional subspace V_k. This result is summarized and proved in the following lemma.

Lemma 3.1.3 (Best Approximation Property). *The solution of the Ritz approximation (3.6) is the best approximation of u in V_k. In other words,*

$$\|u - u_k\|_V = \inf_{v_k \in V_k} \|u - v_k\|_V. \tag{3.7}$$

Proof. The property $V_k \subset V$ allows one to use test functions from V_k in the weak equation (3.3). Considering the difference between (3.3) and the Ritz approximation (3.6), one arrives at the Galerkin orthogonality

$$a(u - u_k, v_k) = 0 \quad \forall v_k \in V_k. \tag{3.8}$$

Therefore, the error $u - u_k$ is orthogonal to the space V_k. In other words, u_k is an orthogonal projection of u onto V_k with respect to the inner product of V. Consider an arbitrary element $w_k \in V_k$. By the Galerkin orthogonality (3.8),

$$\|u - u_k\|_V^2 = a(u - u_k, u - \underbrace{(u_k - w_k)}_{v_k}) = a(u - u_k, u - v_k).$$

Using the Cauchy-Schwarz inequality, one can put the following upper bound on the right-hand side

$$a(u - u_k, u - v_k) \leq \|u - u_k\|_V \|u - v_k\|_V.$$

Combining these two equations, it can be determined that

$$\|u - u_k\|_V^2 \leq \|u - u_k\|_V \|u - v_k\|_V.$$

23

Since $w_k \in V_k$ is an arbitrary element, v_k is also arbitrary. If $\|u - u_k\| > 0$, dividing both sides of the above equation by $\|u - u_k\|$ gives the statement of the lemma. If $\|u - u_k\| = 0$, the statement of the lemma is trivial. Hence,

$$\|u - u_k\|_V = \inf_{v_k \in V_k} \|u - v_k\|_V$$

and the statement of the lemma is proven. □

Another important property of numerical methods is convergence. The next theorem proves that the approximation error of the Ritz approximation approaches zero as k tends to infinity.

Theorem 3.1.4 (Convergence of the Ritz Approximation). *The Ritz approximation converges, i.e.*

$$\lim_{k \to \infty} \|u - u_k\|_V = 0. \tag{3.9}$$

Proof. The best approximation property (3.7) above gives

$$\|u - u_k\|_V = \inf_{v_k \in V_k} \|u - u_k\|_V.$$

Additionally, the assumed property of the finite-dimensional subspaces (3.5) states that

$$\|u - u_k\|_V \leq \varepsilon \quad \forall k \geq K.$$

Combining the two equations yields

$$\|u - u_k\|_V = \inf_{v_k \in V_k} \|u - u_k\|_V \leq \varepsilon \quad \forall k \geq K,$$

where $\varepsilon > 0$ and K is dependent on ε. Hence, the Ritz approximation converges. □

The Ritz Method as a Linear System of Equations

An arbitrary basis $\phi_1, ..., \phi_k$ of the space V_k can be used to calculate the approximate solution u_k. Since the Ritz approximation equation (3.6) holds for all $v_k \in V_k$, it must also hold for each basis function ϕ_i. The above assertion holds true because each function v_k can be expressed as a linear combination of basis functions, i.e $v_k(x) = \sum_{i=1}^{k} \alpha_i \phi_i(x)$, and both sides of the equation are linear with respect to the test function. Inserting the above formulation of v_k into the Ritz approximation equation (3.6) yields

$$a(u_k, v_k) = \sum_{i=1}^{k} \alpha_i a(u_k, \phi_i) = \sum_{i=1}^{k} \alpha_i f(\phi_i) = f(v_k).$$

The linearity of the bilinear form allows the coefficient α_i to be pulled out. The above equation is satisfied if $a(u_k, \phi_i) = f(\phi_i)$ for $i = 1, ..., k$. In particular, if (3.6) holds, it holds for each basis function ϕ_i.

An ansatz for the solution u_k can also be expressed as a linear combination of the basis functions

$$u_k(x) = \sum_{j=1}^{k} u_j \phi_j,$$

24

where $u_j \in \mathbb{R}$ are the unknown coefficients. Inserting this ansatz into the Ritz approximation equation (3.6) and using the basis functions as test functions yields

$$\sum_{j=1}^{k} a(u_j \phi_j, \phi_i) = \sum_{j=1}^{k} a(\phi_j, \phi_i) u_j = f(\phi_i)$$

for $i = 1...k$.

This equation can be written as a linear system of equations of the form $A\underline{u} = \underline{f}$ with stiffness matrix $A = (a)_{ij}^{k} = a(\phi_j, \phi_i)_{i,j=1}^{k}$. The right-hand side $f_i = f(\phi_i)$ for $i = 1, ..., k$ is a vector of length k. There exists a one-to-one mapping from the vector $(v_1, ...v_k)^\intercal$ and $v_k = \sum_{i=1}^{k} v_i \phi_i$, which one can use to show that the matrix A is symmetric and positive definite, i.e.

$$A = A^\intercal \iff a(v, w) = a(w, v) \quad \forall v, w \in V_k,$$

$$\underline{x}^\intercal A \underline{x} > 0 \text{ for } \underline{x} \neq \underline{0} \iff a(v, v) > 0 \quad \forall v \in V_k, v \neq 0.$$

The Galerkin Method

The Galerkin method is a more general case of the Ritz method, where the bilinear form $b(\cdot, \cdot)$ is bounded and coercive, but not symmetric. In this case, an equivalent result to the Representation Theorem of Riesz is used, the so-called Lax-Milgram Theorem.

Theorem 3.1.5 (Lax-Milgram Theorem). *Let $b(\cdot, \cdot) : V \times V \to \mathbb{R}$ be a bounded and coercive bilinear form on the Hilbert space V. Then, for each bounded linear functional $f \in V'$, there is exactly one $u \in V$ with*

$$b(u, v) = f(v) \quad \forall v \in V. \tag{3.10}$$

Proof. A complete proof of this theorem can be found in Appendix B.1. of [21]. $\square$

The Galerkin approximation therefore consists in finding $u_k \in V_k$ such that

$$b(u_k, v_k) = f(v_k) \quad \forall v_k \in V_k. \tag{3.11}$$

Lemma 3.1.6 (Lemma of Cea). *Let $b : V \times V \to \mathbb{R}$ be a bounded and coercive linear form on the Hilbert space V and let $f \in V'$ be a bounded linear functional. Let u be the solution of (3.10) and u_k be the solution of (3.11). Then, the following error estimate holds:*

$$\|u - u_k\|_V \leq \frac{M}{m} \inf_{v_k \in V_k} \|u - u_k\|_V, \tag{3.12}$$

where the constants M and m are from the definitions of bounded and coercive bilinear forms given in (3.1) and (3.2).

Proof. From the difference of the weak equation (3.3) and discrete Galerkin approximation (3.11), one arrives at the following error equation

$$b(u - u_k, v_k) = 0 \quad \forall v_k \in V_k.$$

This is also known as Galerkin orthogonality. By the definition of coercive (3.2),

$$\|u - u_k\|_V^2 \leq \frac{1}{m} b(u - u_k, u - u_k).$$

Using the Galerkin orthogonality, the above inequality can be rewritten as

$$\|u - u_k\|_V^2 \leq \frac{1}{m} b(u - u_k, u - v_k) \quad \forall v_k \in V_k.$$

The bilinear form is bounded, therefore one can obtain the following upper bound on the right-hand side

$$\frac{1}{m} b(u - u_k, u - v_k) \leq \frac{M}{m} \|u - u_k\|_V \|u - v_k\|_V \quad \forall v_k \in V_k.$$

Combining the above statements yields

$$\|u - u_k\|_V^2 \leq \frac{M}{m} \|u - u_k\|_V \|u - v_k\|_V \quad \forall v_k \in V_k.$$

Since v_k is an arbitrary vector, the statement of the lemma follows directly. $\qquad \square$

The Lemma of Cea implies that the error of the Galerkin approximation is bounded by a multiple of the best approximation error. Therefore, the best approximation error is important for the study of error estimates of approximations in finite-dimensional spaces.

3.1.3 Finite Element Spaces

This section will give a brief overview of the construction of an appropriate finite element space and follows from Appendix B in [21] and the lecture notes [38].

Finite Elements

A local finite element consists of three components: a mesh cell K, a finite-dimensional space $P(K)$ defined on the mesh cell, and a set of linear functionals defined on $P(K)$. These will be explained more thoroughly in the following definitions.

Definition 2 (Mesh Cell, m-Faces). A *mesh cell* K is a compact polyhedron in $\mathbb{R}^d$ with a non-empty interior, where $d \in \{2, 3\}$. The boundary of the mesh cell ∂K is made up of *m-faces*, which are m-dimensional linear manifolds (i.e. points, line segments, etc.), where $0 \leq m \leq d - 1$.

Definition 3 (Finite-Dimensional Spaces Defined on K). Finite element methods rely on finite-dimensional spaces $P(K) \subset C^s(K)$ that are defined on the mesh cell K, where $s \in \mathbb{N}$. These finite-dimensional subspaces typically consist of polynomials and the dimension is denoted by N_K.

Definition 4 (Linear Functionals Defined on P(K), Nodal Functionals). The linear functionals $\Sigma = \{\Phi_{K,1}, ... \Phi_{K,N_K} : C^s(K) \to \mathbb{R}\}$ defined on P(K) are called *nodal functionals*.

From these definitions, it can be seen that K is the domain on which the finite element is defined and $P(K)$ is the finite-dimensional approximation space. A property called unisolvence is needed to establish that Σ defines a basis of $P(K)$.

Unisolvence and Local Basis

Definition 5 (Unisolvence of P(K) with respect to functionals Σ). A space $P(K)$ is considered *unisolvent with respect to the functionals* Σ if for each N-tuple $\underline{a} = (a_1, ..., a_{N_K})^\intercal$, there exists a unique element $p \in P(K)$ such that

$$\Phi_{K,i}(p) = a_i, \quad 1 \le i \le N_K.$$

If one chooses the Cartesian unit vectors for $\underline{a}$, it can be shown using unisolvence that a set $\{\phi_{K,i}\}_{i=1}^{N_K}$ exists with $\phi_{K,i} \in P(K)$ and

$$\Phi_{K,i}(\phi_{K,j}) = \delta_{ij}, \quad i,j = 1, ..., N_K,$$

where δ_{ij} is the Kronecker delta. This set forms a basis of P(K) and is called the local basis. If an arbitrary basis of P(K) is know, the local basis can be computed by solving a linear system of equations.

Finite Element Space and Global Basis

In order to define global finite element spaces, it is first necessary to decompose the domain of the problem. This leads to the following definition.

Definition 6 (Triangulation, Mesh Cell, Grid, Mesh). A decomposition of the domain Ω into polyhedra K is called a *triangulation T_h*. The polyhedra K are called *mesh cells* and the union of the polyhedra is known as the *grid* or *mesh*.

Definition 7 (Finite Element Space and Global Basis). A function $v(x)$ defined on Ω with $v|_K \in P(K)$ for all mesh cells $K \in T_h$ is called continuous with respect to the functional $\Phi_i : \Omega \to \mathbb{R}$ if

$$\Phi_i(v|_{K_1}) = \Phi_i(v|_{K_2}) \ \ K_1, K_2 \in \chi_i,$$

where χ_i is the union of K of mesh cells that possess a vertex (or a boundary). A *finite element space S* is defined as

$$S = \{v \in L^\infty(\Omega) : v|_K \in P(K) \ \text{and} \ v \ \text{is continuous with respect to} \ \Phi_i, i = 1, ..., N\}.$$

The *global basis* $\{\phi_j\}_{j=1}^N$ of the finite element space S is defined as $\phi_j \in S$, where

$$\Phi_i(\phi_j) = \delta_{ij}, \quad i,j = 1, ..., N.$$

The global basis function coincides with the local basis function on each mesh cell, which implies the uniqueness of the global basis functions. It is important to note, however, that continuity of the global basis functions does not necessarily imply continuity of the finite element functions; this property is dependent on the functionals that define the finite element space.

3.2 Standard Finite Element Methods for the SWE

Whereas the previous section focused on the general theory of the finite element method, the goal of this section is to provide an introduction to finite element research specifically pertaining to the shallow water equations. Research on the topic began in the the 1970s and has grown significantly over the past five decades. Although many articles on the topic are very narrow in scope (i.e., specific domains, forms of the equations, applications, etc.), this section aims to provide a more general overview of the development of finite element research on the shallow water equations. In addition, a selection of papers is discussed in further detail to provide a few concrete examples. Due to the breadth of the literature, this section is by no means comprehensive, especially concerning research in the most recent years.

3.2.1 Finite Element Meshes

The implementation of a finite element method involves many important decisions, which include determining a mesh structure. Due to the complex geometries present in many of the applications of the shallow water equations, unstructured meshes have proven to be a promising option. Unlike structured meshes, which are comprised of uniform elements, unstructured meshes offer more flexibility in size, shape, and orientation and, for this reason, have become standard practice when studying the shallow water equations (see [26] and references therein). A main advantage of an unstructured mesh is that it allows for increased resolution in specific areas of interest, while maintaining a lower resolution in other portions of the domain [17].

3.2.2 Finite Element Pairs

Another important decision when implementing a finite element method is the selection of an appropriate finite element pair for the velocity and surface-elevation spaces. Although the Navier-Stokes equations and shallow water equations have a similar mathematical structure, the behavior of solutions to the shallow water equations is more complex due to the surface elevation-velocity coupling. This leads to a few notable challenges in the selection of an appropriate finite element pair, which are outlined in Le Roux et al. [27] and briefly introduced in this section.

One of these difficulties is adequately approximating the geostrophic balance on unstructured meshes. Almost all large-scale flows, both in the atmosphere and ocean, are in 'geostrophic balance', which is a balance between the pressure gradient and the Coriolis force. The only exception is at the equator where the Coriolis force does not hold [35]. Systems that are imbalanced will go through a process called geostrophic adjustment, in which imbalances will spread throughout the domain until the system reaches a 'balanced' or 'geostrophically-adjusted' state [37]. Due to the geostrophic balance, finite element pairs that have been effectively used to approximate the viscous Stokes and Navier-Stokes equations may not be suitable for shallow water equations. For these sets of equations, the smoothing effect of viscosity and the neglect of the Coriolis force likely conceal the difficulties of suitably representing the geostrophic equilibrium [27].

Spurious noise poses another problem for selecting the function spaces. Noise can result from difficulties approximating the geostrophic balance or can also be attributed to the coupling of the continuity and momentum equations. The severity of the noise is dependent on where the velocity and surface-elevation nodes are placed on the mesh as well as the choice of basis functions, which demonstrates the significance of the choice of finite element pairs [27]. Unlike the Navier-Stokes equations, in which oscillations are only present in the pressure, oscillations are present in *both* surface elevation and velocity in the shallow water equations. This is due to the presence of the extra term $(\frac{\partial H}{\partial t})$ in the continuity equation, which couples oscillations between surface elevation and velocity solutions [43].

A final challenge for the shallow water equations is ensuring that the discrete solution converges. As with the Navier-Stokes equations, an inappropriate choice of approximations spaces for $\mathbf{v}$ and p, or in the case of the shallow water equations, $\mathbf{v}$ and H, can lead to approximations that do not converge as the mesh is refined (see [43] and references therein). Therefore, the selection of a finite element pair is a determining factor in the rate of convergence of a finite element discretization and the existence of spurious modes can be an indication of loss of convergence [27].

Efforts to reduce the effects of spurious noise have utilized three general methods: (1) numerical damping, (2) generation of elements without spurious modes, and (3) modification of the governing equations [42]. These methods will become apparent in the next section, which provides an overview of strategies used to address the aforementioned challenges.

3.2.3 Continuous Galerkin Methods

Early suggestions for noise-suppression techniques included numerical smoothing [2] and enhanced bottom friction [3]. While these methods can be effective in reducing spurious oscillations, they can also have the unintended consequence of damping out physical aspects of the solution in addition to spurious noise [32]. In other words, it can be difficult to select appropriate parameters such that only significant noise is removed without impacting the rest of the spectrum. Ideally, viscous damping should be used in a way that is physically realistic, rather than to suppress numerically-induced problems [27].

Later efforts therefore sought to discover the root cause of the numerical noise and develop methods that address these problems accordingly. These included (1) equal-order elements with variables defined at sets of points staggered in space, (2) vorticity-divergence schemes, (3) mixed-order interpolation, and (4) the wave equation formulation [27]. These approaches will be briefly introduced in the following paragraphs.

Equal-Order Elements on Staggered Grids

Hua and Thomasset [19] found that discretizations with equal-order finite element spaces for the velocity and surface elevation produce small-scale noise. For this reason, Williams [47] concluded that the primitive[2] form of the shallow water equations should be used

---

[2]The term "primitive equations" is used to refer to the standard equations which have not been modified.

with staggered grids[3]. Schoenstadt [37] confirmed in his geostrophic adjustment studies that unstaggered finite element grids of the shallow water equations are generally poor. He proposes that the noise resulting from unstaggered grids is a result of the coupling of the velocity and surface-elevation fields, making specifying both quantities at each grid point an overspecification of the problem.

Vorticity-Divergence Scheme

An alternative approach to the use of staggered grids is the use of the vorticity-divergence scheme, as suggested by Williams [47]. In his analysis, the scheme using the primitive equations with a staggered grid performed comparably to the vorticity-divergence scheme with an unstaggered grid. Correspondingly, Cullen and Hall [9] affirm that some of the effects of staggering can be obtained through using the vorticity-divergence equations instead of the typical velocity components. The vorticity-divergence scheme can be derived by differentiating the primitive momentum equations with respect to the spatial coordinates.

Mixed-Order Interpolation

Several studies extended the work of Taylor and Hood [18] for the Navier-Stokes and employed mixed-order interpolation for the shallow water equations, i.e. different basis functions for the velocity and surface-elevation spaces. These efforts were met with limited success. Walters and Cheng [44], who used linear basis functions to interpolate the surface elevation and quadratic functions for the velocity component, found the scheme to have the following advantage over using the same interpolations functions – increased computation efficiency and increased accuracy in the surface elevation. However, Walters [42] later numerically confirmed that, while the above mixed interpolation scheme can remove noise in the surface elevation, it can still leave considerable noise in the velocity. A combination of piecewise constant and piecewise linear basis functions performed well in a study by Walters and Zienkiewicz [48]. Walters and Carey [43], who studied various element pairs, found that those with a higher ratio of continuity constraints to momentum constraints on each element had more spurious oscillations.

Wave Continuity Equation

One of the most widespread techniques intended to eliminate oscillations without numerical damping is the so-called wave-continuity approach, which was proposed by Lynch and Gray in 1979 [31]. The approach makes use of the wave continuity equation (WCE), a reformulation of the continuity equation,

$$\frac{\partial \mathrm{CE}}{\partial t} - \nabla \cdot (\mathrm{CME}) + \tau(\mathrm{CE}) = 0,$$

which is then paired with either the conservative or nonconservative momentum equation to give the full shallow water system [4]. Here τ is a non-linear friction coefficient, which depends on $\mathbf{x}$ and t. This formulation of the continuity equation is labeled as the wave

---

[3]The term "staggered grid" comes from staggering the variables by using a different element mesh for each variable, e.g., piecewise linear functions defined at the center of finite elements for the velocity field and surface-elevation fields defined at the vertices [9].

form because it is a second-order wave equation in H when linearized [32]. The wave-continuity approach has been used in the development of robust finite-element algorithms and been proven to achieve high computational accuracy and efficiency through several studies (see [32] and references therein).

Generalized Wave Continuity Equation

The wave continuity equation was extended to the generalized wave continuity equation (GWCE) in 1984 by Kinnmark [22]. The key difference from the WCE is that in the GWCE, τ is replaced with a general function, which may be independent of time [4]. Again, the GWCE is paired with the CME or NCME to give the full system. Advantages of this method include: (1) it allows for a weaker coupling between the continuity and momentum equations, (2) it results in symmetric positive definite matrices, and (3) it helps to stabilize the numerical solution (see [13] and references therein). The GWCE approach established the basis for multiple shallow water simulators, most notably, the advanced circulation model (ADCIRC), developed by Luettich and Westerink [30], which has been successfully used in several finite element studies (see [23] and references therein). A major disadvantage of the GWCE approach, however, is that the mass conservation properties and performance of the model depend on a weighting parameter, which must be defined by the user and is difficult to adjust in practice [23].

Finite Element Pair Comparison - Le Roux et al. (1998)

The previous paragraphs sought to provide a general overview of the development of finite element research for the shallow water equations. As a concrete example of a performance comparison for different finite-element pairs, a study by Le Roux et al. [27] is discussed in further detail. The goal of the study was to analyze the performance of nine finite element pairs in their ability to represent geostrophic balance in a noise-free way on both structured and unstructured meshes. The standard nomenclature $P_m - P_n$ is used to describe the finite-element pairs, meaning the velocity and surface elevation are represented by piecewise-defined polynomials of degree m and n respectively.

The nine finite-element pairs under consideration were: $P_1^{NC} - P_0$, P_1 iso $P_2 - P_0$, $P_2 - P_0$, $P_1^{NC} - P_1$, P_1 iso $P_2 - P_1$, $P_2 - P_1$, $P_1 - P_1$, $P_1 - P_0$, and P_1 iso $P_2 - P_{0-3}$. The $P_1^{NC} - P_0$ finite element pair, also known as the Crouzeix-Raviert finite element pair [8], approximates the velocity using piecewise linear functions at the barycenters of the faces of the mesh cells and uses a piecewise constant function to approximate the surface elevation. The NC implies the element is non-conforming, i.e. $V_h \not\subset V$. In the case of the P_1 iso $P_2 - P_0$ and P_1 iso $P_2 - P_1$ finite element pairs, the P_1 in P_1 iso P_2 indicates the velocity is approximated using linear functions, whereas the iso $P2$ implies that the nodal placement is that associated with quadratic elements [27]. The P_1 iso $P_2 - P_{0-3}$ finite element pair was proposed by Gunzburger in 1989 as an alternative possibility to solve the coupled momentum and continuity equations [16]. Rather than just using constant basis functions for the surface elevation, the P_{0-3} indicates that there are three degrees of freedom for the surface elevation, which correspond to three constant and piecewise-discontinuous basis functions [27].

For their analysis, Le Roux et al. [27] combined a semi-Lagrangian approach with finite elements. Unlike Eulerian methods, which monitor fixed points in space, and Lagrangian

schemes, which examine fluid parcels that move with the flow, semi-Lagrangian methods select new fluid parcels at each time step, such that they arrive exactly on the nodes of a regular mesh by the end of every time step [17]. The idea behind semi-Lagrangian schemes is to combine the advantages of the regular meshes of the Eulerian schemes with the increased stability of the Lagrangian methods, as suggested by Staniforth and Cote [39]. While a traditional Eulerian finite element approach requires the use of higher-order elements to accurately model advection, the semi-Lagrangian methods rely on lower-order elements, hence the selection of the above nine pairs.

In their investigation of how the nine finite element pairs performed in representing the geostrophic balance, Le Roux et al. [27] found that the first three pairs ($P_1^{NC} - P_0$, P_1 iso $P_2 - P_0$, and $P_2 - P_0$) performed very poorly and excluded them from further consideration. The next three pairs ($P_1^{NC} - P_1$, P_1 iso $P_2 - P_1$, and $P_2 - P_1$) performed better in comparison, but still exhibited noise, especially when the mesh was distorted. The final three pairs ($P_1 - P_1$, $P_1 - P_0$, and P_1 iso $P_2 - P_{0-3}$) offered the most promising performance in terms of noise and were the least sensitive to mesh distortion. Le Roux et al. [27] attribute the noisy behavior of the first six pairs to the ratio of the degrees of freedom of the velocity and surface-elevation spaces, with a ratio greater than one leading to noise when geostrophically calculating the velocity from the surface elevation.

Due to the poor performance of the first six finite element pairs, only the final three finite element pairs were considered in the remainder of the analysis, which was focused on how the element pairs perform in terms of spurious noise resulting from the coupling of the momentum and continuity equations. Le Roux et al. [27] note the LBB or inf-sup condition as a typical stability criterion of incompressible flows. In their analysis, both the $P_1 - P_1$ and $P_1 - P_0$ pairs, which do not satisfy the inf-sup condition, suffer from spurious surface-elevation nodes. Difficulties with the $P_1 - P_1$ element pair can also be attributed to the coincident placement of the nodes rather than a staggered one, as was discussed earlier in this chapter. The P_1 iso $P_2 - P_{0-3}$ finite element pair, however, which performed almost noise-free in terms of representing the geostrophic balance, does not suffer from the spurious surface-elevation modes as the previous two and has the advantage of satisfying the inf-sup condition. While it does still introduce spurious noise in the velocity, where the surface elevation is 0, Le Roux et al. [27] introduce this pair as a promising compromise for semi-implicit, semi-Lagrangian finite element shallow water models.

Semi-Lagrangian vs. Eulerian Approach - Hanert et al. (2005)

While Le Roux et al. [27] used a semi-Lagrangian and semi-implicit time discretization, the type of fluid description (Eulerian, Lagrangian, or semi-Lagrangian) and time discretization are examples of additional aspects of a finite element scheme that can be modified and would result in a different outcome and performance of the model. Interestingly, Hanert et al. [17] compared a Eulerian and semi-Lagrangian finite element approach to shallow water ocean modeling. Their analysis focused specifically on the $P_1^{NC} - P_1$ finite element pair, following from the work of Hua and Thomasset in 1984 [19], who were the first to study combinations of linear and non-conforming elements to solve the shallow water equations and found the $P_1^{NC} - P_1$ finite element pair to be computationally efficient and perform well in modeling the two-layer shallow water equations.

In their two test problems, Hanert et al. [17] also found the Eulerian $P_1^{NC} - P_1$ model to be a promising candidate for constructing a general ocean circulation model. While the semi-Lagrangian approach has the advantage of being able to use larger time steps, it proved to be ten times more computationally expensive compared to the Eulerian approach as implemented in their code. In addition, diffusion had a small impact on the Eulerian scheme, while it was necessary to add a small amount of Laplacian diffusion to the semi-Lagrangian model in order to obtain an acceptable solution.

3.2.4   Discontinuous Galerkin Methods

The methods discussed thus far, apart from those using the P_1^{NC} finite element, use continuous spaces for approximating the shallow water equations and are therefore classified as continuous Galerkin (CG) methods. In more recent years, discontinuous approximating spaces have become the focus of study for the shallow water equations, in part to address issues with the GWCE model [12, 23]. These methods, which were first introduced in the 1970s, are known as discontinuous Galerkin (DG) methods and combine aspects of both finite element and finite volume methods [15].

In contrast to CG methods, there is no continuity constraint for DG methods, meaning that continuity along the mesh cell edges is not required. Therefore, the approximate solution is a piecewise polynomial and jumps across the cell edges of the mesh may exist. These jumps can be viewed as numerical fluxes at interfaces and are given by surface integrals on the mesh cells. The numerical fluxes are the key difference when comparing CG and DG methods, and in this way, DG methods resemble finite volume methods. DG methods are intrinsically local, and elements only require boundary data from their neighbors, which is another stark contrast to CG methods, which require more information [28]. A complete description of DG methods will not be provided here. For a more comprehensive overview, refer to [7] or [15].

The DG method has many advantages for solving the shallow water equations compared to methods using continuous approximating spaces. These include the ability to model advective flows, including problems with discontinuities such as hydraulic jumps, the ability to use different polynomial orders for different areas of the domain, and the ability to easily apply non-conforming meshes (e.g., meshes with hanging nodes) [12]. Additionally, the DG method fulfills local mass conservation properties. DG methods are also highly parallelizable, which can improve computational time [23].

Although the DG method is favorable in many ways, it is not without its drawbacks, most notably, the larger number of degrees of freedom in comparison with CG methods. For example, a DG method with linear triangular elements has on average six times as many degrees of freedom in comparison with a corresponding CG method. The relatively greater number of degrees of freedom leads to decreased efficiency when employing DG methods [23].

CG and DG Method Comparison - Kubatko et al. (2009)

Kubatko et al. [23] conducted a comparative study of CG and DG models for the shallow water equations that included an analysis of accuracy, convergence rates, serial and parallel run times and efficiency for a series of test problems. The two models compared in their study were the ADCIRC model, introduced in Section 3.2.3, which uses the GWCE and piecewise, linear functions on triangular elements, and what they call the "ADCIRC-DG model", which is detailed in [24].

Overall, the analysis by Kubatko et al. [23] showed the DG model to be preferable to the CG model in most aspects. Although the DG model had a worse computing time on serial machines, the solution errors for the DG model were smaller than those for the CG model, typically by an order of magnitude. A cost vs. accuracy analysis showed that the mesh width required by the CG model to obtain the same errors as the DG model resulted in greater computing time than for the DG model. Additionally, the DG model had consistently better convergence rates. The DG method showed greater parallel efficiency, although this did depend on the problem size and computing resources. Their analysis was limited to piecewise, linear approximations over triangular elements, but a model using a combination of linear and quadrilateral elements could prove the DG model to be even more efficient. The study was also limited to two dimensions.

Coupled DG/CG Method - Dawson et al. (2006)

Similarly, Dawson et al. [12] performed a comparative study of CG and DG models, but also included a novel coupled DG/CG model that uses a DG method for the continuity equation and CG method for the momentum equation. The interest in a coupled model emerged from a desire to reduce the number of degrees of freedom compared to the DG method, while maintaining its high level of accuracy. The traditional CG approach used was the ADCIRC model and the DG approach used was the UTBEST model described in [6]. The study evaluated the error and convergence properties of the coupled model compared with the traditional approaches and also assessed the associated costs. For their test case, they used three different mesh configurations as well as four levels of mesh resolution in order to analyze convergence rates. They analyzed the error and convergence properties on the entire domain and also a fixed portion of the inner domain to eliminate the effects of boundary conditions.

The DG method proved to be the most accurate, while the GWCE model was only slightly more accurate than the coupled DG/CG model. Unlike the other models, the accuracy of the DG method did not experience much change whether the boundary condition effects were included or not. When considering the oscillations, the DG method had the smallest oscillations, and the DG/CG method had the largest. In terms of efficiency, the GWCE continued to be the most efficient, the DG/CG less so, and the DG method remained the costliest.

Dawson et al. [12] conclude that there are advantages and disadvantages of the three models. The GWCE model has the advantage of being the most efficient in terms of run time and also suppresses spurious oscillations relative to other CG methods. However, it also has some significant flaws including the user-selected weighting parameter, mass imbalances, difficulties modeling highly advective flows, and its complex structure. While

the DG method deals with many of these weaknesses and is highly accurate, it can be very costly. The DG/CG was proposed as a novel middle ground between these two methods. While the DG/CG method has the advantage of being mass conserving and is less costly than the DG method, it proved to be less accurate than the DG, more susceptible to oscillations and limited when modeling highly advective conditions.

Concluding Remark

The above studies help to illustrate the considerable number of choices involved in the development and modification of a finite element model and highlight specific challenges associated with the shallow water equations. Each choice has its own advantages and disadvantages, and a compromise must often be reached. The complexity of these models and their corresponding decisions open up many avenues for further research and the body of literature on finite element methods for the shallow water equations continues to develop at present.

3.3   Error Estimates for the Galerkin Method

When studying finite element models, it is important to consider their effectiveness through error analysis. There are two types of error estimates for finite element methods, *a posteriori* and *a priori* estimates, as described in [14].

A posteriori estimates use the computed finite element approximation u_h to calculate an error of the form $\|u - u_h\| \leq \varepsilon$, where ε is a real number. This type of error is useful for determining the actual error of a finite element approximation and can be used for adaptive mesh refinement, a technique that locates areas of the mesh where the error is particularly high and refines the mesh in these specific areas until a tolerable error level is obtained.

The study of *a priori* error estimates is significant because it allows one to determine the order of convergence of a given finite element method, i.e. $\|u - u_h\|$ is $O(h^k)$ in a certain norm $\|\cdot\|$, where h is the mesh size and k is a positive number. This knowledge about how quick the error decreases as the mesh width decreases is an important measure for determining the efficiency of a method. The focus of this section is *a priori* error estimates for the shallow water equations.

3.3.1   *A Priori* Error Estimate Procedure

While the derivations of the *a priori* estimates will not be provided in detail, they tend to adhere to the following standard procedure, which is outlined in Section 3.2 of [32]:

1. Subtract the finite-dimensional weak form of the equations from the equation satisfied by the finite element approximation to obtain the weak form of the error equation.

2. Separate the discretization error $E_\phi = \phi - \phi_h$, where ϕ is the weak solution and ϕ_h the finite element approximation, into the sum of an approximation error (or projection error) $\overline{E}_\phi = \phi - \tilde{\phi}$ and an affine error $\tilde{E}_\phi = \phi_h - \tilde{\phi}$ using a comparison function $\tilde{\phi}$. This approach is chosen because the discretization error is difficult to

35

calculate directly. The triangle inequality can then be used to separate and bound the errors:

$$\|E_\phi\| \leq \|\overline{E}_\phi\| + \|\tilde{E}_\phi\|.$$

The approximation error is either already known or easily determined, leaving the affine error to be estimated.

3. Choose a comparison function $\tilde{\phi}$ such that the approximation error can be estimated from either approximation theory or a related problem. Standard choices for the comparison function include the $\mathcal{L}^2$ projection or the $\mathcal{H}^1$ projection.

4. Choose a test function in the weak form of the error equations that is well-suited for the derivation of the affine error estimate.

5. Add the affine error, once estimated, to the approximation in order to determine an upper bound estimate for the discretization error.

3.3.2 Approximation Theory: Notation and Definitions

In order to present the error analysis results, it is first necessary to introduce relevant notation and theory, which follows from Chapter 3 of [32]. Let T_h be a quasi-uniform triangulation of Ω into elements ω_i, $i = \{1, ..., n_T\}$, where $\mathrm{diam}(\omega_i) = h_i$ and $h = \max_i h_i$.

Definition 8 (Lebesgue Space, Norm). For $p \in [1, \infty]$, the space of functions that are *Lebesgue integrable* on the domain Ω are denoted

$$\mathcal{L}^p(\Omega) = \{f : \int_\Omega |f|^p \, dx < \infty\} \qquad \text{for } 1 \leq p < \infty,$$

$$\mathcal{L}^\infty(\Omega) = \{|f(x)| < \infty \text{ almost everywhere on } \Omega\} \quad \text{for } p = \infty,$$

and equipped with the norms

$$\|f\|_{\mathcal{L}^p(\Omega)} = \left(\int_\Omega |f|^p \, dx \right)^{\frac{1}{p}} \quad \text{for } 1 \leq p < \infty,$$

$$\|f\|_{\mathcal{L}^\infty(\Omega)} = \operatorname*{ess\,sup}_{\mathbf{x} \in \Omega} |f(\mathbf{x})| \quad \text{for } p = \infty.$$

A norm without a subscript is assumed to be the $\mathcal{L}^2$ norm, i.e. $\|f\| = \|f\|_{\mathcal{L}^2(\Omega)}$.

Definition 9 (Sobolev Space, Norm). For $\ell > 0$ and $p \in [1, \infty]$, the *Sobolev space* $\mathcal{W}^{\ell,p}(\Omega)$ is defined as

$$\mathcal{W}^{\ell,p}(\Omega) = \{u \in \mathcal{L}^p(\Omega) \mid D^\alpha u \in \mathcal{L}^p(\Omega) \,\forall\, |\alpha| \leq \ell\},$$

and equipped with the norms

$$\|u\|_{\mathcal{W}^{\ell,p}(\Omega)} = \sum_{|\alpha| \leq \ell} \|D^\alpha u\|_{\mathcal{L}^p(\Omega)} \qquad \text{for } 1 \leq p < \infty,$$

$$\|u\|_{\mathcal{W}^{\ell,\infty}(\Omega)} = \max_{|\alpha| \leq \ell} \|D^\alpha u\|_{\mathcal{L}^\infty(\Omega)} \qquad \text{for } p = \infty.$$

The Sobolev space $\mathcal{H}^\ell$ is defined to be $\mathcal{H}^\ell = \mathcal{W}^{\ell,2}$.

For the discrete time estimate, it is necessary to introduce additional notation, which follows from [32]. Define a discrete temporal subset of $[0, T]$ by

$$J_{\Delta t} = \{t^k \mid t^k \in [0, T],\ t^k = k\Delta t,\ 0 \le k \le N,\ N\Delta t = T,\ \Delta t \ge 0\}.$$

For a space X with norm $\|\cdot\|_X$ and map $\varphi : [0, T] \longrightarrow X$, define the following norms:

$$\|\varphi\|_{\mathcal{L}^2((0,T);X)}^2 = \int_0^T \|\varphi\|_X^2\, dt,$$

$$\|\varphi\|_{\mathcal{L}^\infty((0,T);X)} = \sup_{0 \le t \le T} \|\varphi\|_X,$$

$$\|\varphi\|_{\ell_{\Delta t}^2((0,T);X)}^2 = \sum_{k=0}^N \|\varphi^k\|_X^2 \Delta t,$$

$$\|\varphi\|_{\ell_{\Delta t}^\infty((0,T);X)} = \sup_{0 \le k \le N} \|\varphi^k\|_X,$$

where $\varphi^k = \varphi^k(\mathbf{x}) = \varphi(\mathbf{x}, t^k)$.

Let S_h^0 be a finite-dimensional subspace of $\mathcal{H}_0^1(\Omega)$ defined on T_h that consists of piecewise polynomials less than some degree s_1, such that it also satisfies the standard approximation property

$$\inf_{\zeta \in S_h^0} \|\phi - \zeta\|_{\mathcal{H}^{s_0}(\Omega)} \le K_0 h^{\ell - s_0} \|\phi\|_{\mathcal{H}^\ell(\Omega)},\quad \phi \in \mathcal{H}_0^1(\Omega) \cap \mathcal{H}^\ell(\Omega), \tag{3.13}$$

where s_0 and ℓ are integers, $0 \le s_0 \le \ell \le s_1$, and K_0 is a constant independent of h and ϕ. Additionally, S_h^0 satisfies the following inverse estimate. Define S_h to be a finite-dimensional subspace of $\mathcal{H}^1(\Omega)$ that also satisfies (3.13).

Lemma 3.3.1 (Lemma 3.1 in [32]). *Let $h \in (0, 1]$ and $S_h \in \mathcal{W}^{r,p}(\Omega) \cap \mathcal{W}^{m,q}(\Omega)$, where Ω is a polyhedral domain in $\mathbb{R}^n$, $1 \le p \le \infty, 1 \le q \le \infty$, and $0 \le m \le r$, then there exists a $K_0 = K_0(r, p, q)$ such that $\forall v \in S_h$,*

$$\|v\|_{\mathcal{W}^{r,p}(\Omega)} \le K_0^{m - r + \min(0, \frac{n}{p} - \frac{n}{q})} \|v\|_{\mathcal{W}^{m,q}(\Omega)}. \tag{3.14}$$

3.3.3 *A Priori* Error Estimates

Overall, the literature is lacking in *a priori* error estimates for the shallow equations and this appears to be an area of open research. The results of a few of these limited number of studies are discussed below.

GWCE-CME Shallow Water System

Martinez [32] derived error estimates for the nonlinear, coupled GWCE-CME system of equations, which were also published in Chippada et al.[4, 5]. Let $\mathbf{U} = a\mathbf{v}$. The formulation of the GWCE used in the analysis is defined to be

$$\frac{\partial^2 H}{\partial t^2} + \tau_0 \frac{\partial H}{\partial t} - \nabla \cdot [\nabla \cdot (\frac{\mathbf{U}^2}{a}) + ga\nabla H + \nu\nabla\frac{\partial H}{\partial t} + (\tau_b - \tau_0)\mathbf{U} + f\mathbf{k} \times \mathbf{U} + a\mathcal{F}] = 0, \tag{3.15}$$

where τ_0 is a time-independent positive constant, ν is the viscosity, the bottom friction function is $\tau_b = c_f \frac{\sqrt{u^2 + v^2}}{a}$, and $\mathcal{F} = (-\frac{1}{a}\tau_s + \nabla p_a - g\nabla\mathcal{N})$. Here τ_s is the surface wind stress

and $\mathcal{N}$ is the Newtonian equilibrium tide potential relative to the Earth's elasticity factor.

For the complete shallow water system, the GWCE was coupled with the following formulation of the CME

$$\partial \mathbf{U}_t + \nabla \cdot (\frac{\mathbf{U}^2}{a}) + ga\nabla H - \nu\Delta\mathbf{U} + \tau_b\mathbf{U} + f\mathbf{k} \times \mathbf{U} + a\mathcal{F} = 0. \qquad (3.16)$$

For simplicity, homogeneous Dirichlet boundary conditions were considered.

Continuous Time Estimate The first error estimate was calculated for a continuous time model.

Theorem 3.3.2 (Theorem 4.1 in [32], Theorem 4.4 in [4])*. Let $0 \leq s_0 \leq \ell \leq s_1$. Let $(H, \boldsymbol{U})$ be the solution to the weak form of (3.15) and (3.16) with homogeneous Dirichlet boundary conditions. Let $(H_h, \boldsymbol{U}_h)$ be the Galerkin approximation of $(H, \boldsymbol{U})$. Suppose $H(t) \in \mathcal{H}_0^1(\Omega) \cap \mathcal{H}^l(\Omega) \cap \mathcal{W}^{1,\infty}(\Omega)$ and $\boldsymbol{U}(t) \in \boldsymbol{\mathcal{H}}_0^1(\Omega) \cap \boldsymbol{\mathcal{H}}^l(\Omega) \cap \boldsymbol{\mathcal{W}}^{1,\infty}(\Omega)$ for each t. If $H_h(t) \in S_h(\Omega)$, $\boldsymbol{U}_h(t) \in \boldsymbol{S}_h(\Omega)$ for each t and certain physically reasonable assumptions about the solutions and the data hold (see M1-M7, N1-N3 and D1-D2 in [32]), there exists a constant $\overline{K}$, such that*

$$\|\frac{\partial}{\partial t}(H - H_h)\|_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))} + \|(H - H_h)(\cdot,T)\| + \|H - H_h\|_{\mathcal{L}^2((0,T);\mathcal{H}^1(\Omega))} + \|(\boldsymbol{U} - \boldsymbol{U}_h)(\cdot,T)\|$$

$$+ \|\sqrt{\tau_b}(\boldsymbol{U} - \boldsymbol{U}_h)\|_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))} + \|\nabla\boldsymbol{U} - \nabla\boldsymbol{U}_h\|_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))} \leq \overline{K}h^{l-1}.$$

For h sufficiently small and $s_1 \geq 3$, the sum of the $\|\cdot\|_{\mathcal{L}^\infty((0,T);\mathcal{L}^\infty(\Omega))}$ norms of H_h and $\boldsymbol{U}_h$ are bounded by a constant.

Proof. The complete derivation of this result can be found in [4]. $\qquad \square$

Discrete Time Estimates As an extension of the above continuous-time analysis, Martinez [32] also studied *a priori* error estimates for the GWCE-CME system of equations in discrete time based on an $\mathcal{L}^2$ projection. Their temporal discretization closely followed the scheme in the ADCIRC simulator [30].

Theorem 3.3.3 (Theorem 4.2 in [32], Theorem 4.1 in [5])*. Let $0 \leq s_0 \leq z$, $s_0 \leq \ell \leq s_1$, $0 \leq z < s_1$. Let $(H^k, \boldsymbol{U}^k)$ be the solution to the weak form of (3.15) and (3.16) at time $t = t^k$. Let $(H_h^k, \mathbf{U}_h^k)$ be the Galerkin approximation of $(H^k, \boldsymbol{U}^k)$. Suppose that $H^k \in \mathcal{H}_0^1(\Omega) \cap \mathcal{H}^\ell(\Omega) \cap \mathcal{W}^{1,\infty}(\Omega)$, $\boldsymbol{U}^k \in \boldsymbol{\mathcal{H}}_0^1(\Omega) \cap \boldsymbol{\mathcal{H}}^\ell(\Omega) \cap \boldsymbol{\mathcal{W}}^{1,\infty}(\Omega)$, $H_h^k \in S_h(\Omega)$, and $\boldsymbol{U}_h^k \in \boldsymbol{S}_h(\Omega)$ for each k. If certain physically reasonable and smoothness assumptions about the solutions and the data hold (see A2-A15 in [5]) and h and Δt are sufficiently small, there exists a constant $\overline{K}$ such that*

$$\|H^N - H_h^N\| + \|H - H_h\|_{\ell^2(J_N;\mathcal{H}^1(\Omega))} + \|\boldsymbol{U}^N - \boldsymbol{U}_h^N\| + \|\nabla\boldsymbol{U} - \nabla\boldsymbol{U}_h\|_{\ell^2(J_{N-1};\mathcal{L}^2(\Omega))} \leq \overline{K}(h^{\ell-1} + \Delta t),$$

where $\Delta t \leq \min\{o(h), \kappa_1, \kappa_2\}$. κ_1 and κ_2 can be calculated based on constants defined in the derivation.

Proof. The complete derivation of this result can be found in [5]. $\qquad \square$

Error Estimate for Coupled DG/CG Method

The previous two results were *a priori* estimates for the second-order GWCE-CME system of equations. However, as discussed in Section 3.2, there are disadvantages to replacing the continuity equation with the GWCE and, in recent years, DG methods used with the primitive equations have become more popular as an alternative solution to address spurious oscillations. Dawson and Proft [11] derived an *a priori* error estimate for a coupled DG/CG method. The motivation behind this mixed approach is to use CG methods where the flow is relatively smooth and DG methods where there may be sharp fronts or where local mass conservation is of particular importance. The formulation of the equations used in the analysis is

$$\partial_t H + \nabla \cdot (a\mathbf{v}) = 0,$$

$$\mathbf{v}_t + \mathbf{v} \cdot \nabla \mathbf{v} + \tau_b \mathbf{v} + g\nabla H - \nu \Delta \mathbf{v} = \mathcal{F}.$$

In their scheme, the domain Ω was partitioned into two disjoint subdomains Ω_{DG} and Ω_{CG}, separated by the interface Γ. The triangulation T_h was assumed to be a conforming, quasi-uniform mesh within Ω_{CG}, whereas it could be non-conforming on Ω_{DG}. Within the subdomain Ω_{DG}, the primitive continuity equation was discretized using a DG method and the momentum equation using the interior penalty (NIPG) method [36]. The advection term in the momentum equation was discretized using an upwinding technique introduced by Lesaint and Raviart [29]. Within the subdomain Ω_{CG}, the GWCE and the momentum equation were discretized using CG formulations. Therefore, the solutions are discontinuous in Ω_{DG} and across the interface Γ, and are continuous in Ω_{CG}. The final result of their analysis is summarized in the following theorem.

Theorem 3.3.4 (Theorem 5.2 in [11]). *Given certain stability assumptions for $\boldsymbol{v}$, H, and the initial data, there exists a constant C, such that*

$$\|H - H_h\|_{\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))} + \|\boldsymbol{v} - \boldsymbol{v}_h\|_{\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))} + \|\boldsymbol{v} - \boldsymbol{v}_h\|_{\mathcal{L}^\infty((0,T);\mathcal{H}^1(\Omega))} \leq Ch^k.$$

Proof. The complete derivation of this result can be found in [11]. $\square$

Chapter 4

Numerical Studies

The focus of this chapter is the implementation of a Galerkin finite element method for the shallow water equations in ParMooN [46], which was done through the modification of the existing implementation of the Navier-Stokes equations. The implementation is tested using two examples, one of which uses a prescribed solution to allow for an analysis of the model performance. Visualizations of the solutions were generated using the package ParaView [1]. Prior to a discussion of the test cases, the boundary conditions, finite element formulation, and time discretization of the model are discussed.

4.1   Governing Equations

The Galerkin FEM is implemented for the inviscid, nonconservative form of the shallow water equations discussed in Section 2.4. In the finite element model, it will be assumed that the bottom topography is flat, i.e. $z_b = 0$, and does not change with time. Since $a = H - z_b$, this assumption implies that $H = a$. This yields the following simplified CE-NCME system of equations

$$\partial_t a + \nabla \cdot (a\mathbf{v}) = 0, \tag{4.1}$$

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} + f\mathbf{k} \times \mathbf{v} + g\nabla a + \frac{c_f}{a}\|\mathbf{v}\|\mathbf{v} = 0. \tag{4.2}$$

4.2   Boundary Conditions

While boundary conditions in fluid dynamics typically take either Dirichlet or Neumann boundary conditions, boundary conditions for the shallow water system tend to be more complicated and can take many forms depending on whether boundaries are fixed, moving, or open. A detailed discussion will not be provided here, but can be found in Chapter 3 of [45] and Chapter 5 of [41].

In this thesis, only fixed boundaries will be considered in the numerical simulations. For simplicity, Dirichlet boundary conditions will be assumed for all variables on the boundary. Additionally, a no-slip condition for the velocity is assumed on the boundary, i.e. $u = v = 0$.

4.3 Finite Element Formulation

Weak Formulation

It is assumed that the velocity $\mathbf{v}$ belongs to a subspace V of $\mathcal{H}^1(\Omega) \times \mathcal{H}^1(\Omega)$ and that the depth a is a sufficiently regular scalar function. The weak formulation of (4.1) and (4.2) requires test functions $\boldsymbol{\phi}$ and ψ that belong, respectively, to the same function spaces as $\mathbf{v}$ and a and that are sufficiently regular.

The weak formulation of the problem is obtained in the usual way by multiplying the continuity equation with the test function ψ and the momentum equation with the test function $\boldsymbol{\phi}$, followed by integrating over the domain Ω. For the continuity equation, one obtains

$$(\partial_t a, \psi)_\Omega + (\nabla \cdot (a\mathbf{v}), \psi)_\Omega = 0,$$

where the notation $(\cdot, \cdot)_\Omega$ is used for the $\mathcal{L}^2$-inner product.

Integration by parts is used to shift the derivative from the variables to the test function where possible. For the continuity equation, integration by parts can be applied to the second term

$$\int_\Omega \nabla \cdot (a\mathbf{v})\psi \, dV = -\int_\Omega a\mathbf{v}\nabla\psi \, dV + \int_{\partial\Omega} \psi a\mathbf{v} \cdot \mathbf{n} \, dA.$$

Due to the assumed no-slip boundary condition, the boundary integral vanishes. Therefore, the final expression for the weak form of the continuity equation is

$$(\partial_t a, \psi)_\Omega - (a\mathbf{v}, \nabla\psi)_\Omega = 0. \tag{4.3}$$

A similar process is followed for the momentum equation. Multiplying by test function $\boldsymbol{\phi}$ and integrating over the domain Ω yields

$$(\partial_t \mathbf{v}, \boldsymbol{\phi})_\Omega + ((\mathbf{v} \cdot \nabla)\mathbf{v}, \boldsymbol{\phi})_\Omega + (f\mathbf{k} \times \mathbf{v}, \boldsymbol{\phi})_\Omega + (g\nabla a, \boldsymbol{\phi})_\Omega + (\frac{c_f}{a}\|\mathbf{v}\|\mathbf{v}, \boldsymbol{\phi})_\Omega = 0.$$

The fourth term is then integrated by parts:

$$\int_\Omega g\nabla a\boldsymbol{\phi} \, dV = -g\int_\Omega a\nabla \cdot \boldsymbol{\phi} \, dV + g\int_{\partial\Omega} a\boldsymbol{\phi} \cdot \mathbf{n} \, dA.$$

Here the boundary integral also disappears because of the no-slip condition for the velocity that is assumed on the boundary. Therefore, the final expression of the weak form of the momentum equation is

$$(\partial_t \mathbf{v}, \boldsymbol{\phi})_\Omega + ((\mathbf{v} \cdot \nabla)\mathbf{v}, \boldsymbol{\phi})_\Omega + (f\mathbf{k} \times \mathbf{v}, \boldsymbol{\phi})_\Omega - g(a, \nabla \cdot \boldsymbol{\phi})_\Omega + (\frac{c_f}{a}\|\mathbf{v}\|\mathbf{v}, \boldsymbol{\phi})_\Omega = 0. \tag{4.4}$$

Finite Element Spatial Discretization

Let T_h be a triangulation on the domain Ω. The desired discrete solution for the velocity $\mathbf{v}_h$ is sought in the finite-dimensional velocity subspace V_h, and the discrete solution for the depth a_h in the finite-dimensional subspace Q_h. The pair of velocity and surface-elevation finite element spaces is denoted V_h/Q_h.

Since the spaces V_h and Q_h are finite dimensional, $\mathbf{v}_h$ and a_h can be written as sums of appropriate basis functions, $\boldsymbol{\phi}_j$ and ψ_j respectively, and the discrete approximation is then given in terms of these coefficients. Therefore, $\mathbf{v}_h$ can be expanded to

$$\mathbf{v}_h = \sum_{j=1}^{N_V} \mathbf{v}_j \boldsymbol{\phi}_j,$$

and similarly,

$$a_h = \sum_{j=1}^{N_Q} a_j \psi_j,$$

where $\mathbf{v}_j$ and a_j represent the nodal values of $\mathbf{v}_h$ and a_h and N_V and N_Q are the dimensions of the finite element spaces. These values can be substituted into (4.3) and (4.4) to arrive at the Galerkin finite element spatial discretization. For the continuity equation, this yields

$$(\partial_t a_h, \psi)_\Omega - (a_h \mathbf{v}_h, \nabla \psi)_\Omega = 0, \tag{4.5}$$

and for the momentum equation,

$$(\partial_t \mathbf{v}_h, \boldsymbol{\phi})_\Omega + ((\mathbf{v}_h \cdot \nabla)(\mathbf{v}_h), \boldsymbol{\phi})_\Omega + (f\mathbf{k} \times \mathbf{v}_h, \boldsymbol{\phi})_\Omega - g(a_h, \nabla \cdot \boldsymbol{\phi})_\Omega + (\frac{c_f}{a_h} \|\mathbf{v}_h\| \mathbf{v}_h, \boldsymbol{\phi})_\Omega = 0. \tag{4.6}$$

Temporal Discretization

Let $J_{\Delta t}$ be a discrete temporal subset of $[0, T]$ as described in Section 3.3.2:

$$J_{\Delta t} = \{t^n \mid t^n \in [0, T], \ t^n = n\Delta t, \ 0 \leq n \leq N, \ N\Delta t = T, \ \Delta t \geq 0\}.$$

A time-extrapolated Crank-Nicholson method is used for the time discretization, which can be classified as a semi-explicit scheme or so-called IMEX (implicit-explicit) scheme. IMEX schemes are advantageous in that they avoid solving a nonlinear problem at each discrete time step, which in turn, makes them typically less time-consuming than strictly implicit schemes.

In the method used for the implementation, which follows from [34], the nonlinear terms are quasi-linearized using one term from the current time step and the other term as a linear extrapolation of the previous two time steps. The linearly extrapolated terms, which will be denoted $\mathbf{v}^*$ and a^*, were estimated at the time step $n + \frac{1}{2}$ using a second-order approximation in time as decribed in [34]:

$$\mathbf{v}^* = \mathbf{v}^{n+\frac{1}{2}} = \frac{3}{2}\mathbf{v}^n - \frac{1}{2}\mathbf{v}^{n-1} + \mathcal{O}(\Delta t^2),$$

$$a^* = a^{n+\frac{1}{2}} = \frac{3}{2}a^n - \frac{1}{2}a^{n-1} + \mathcal{O}(\Delta t^2).$$

For example, consider the nonlinear convective term $(\mathbf{v}^{n+1} \cdot \nabla)\mathbf{v}^{n+1}$, appearing in (4.2), at discrete time t^{n+1}. Using the IMEX scheme, the term is replaced by

$$(\mathbf{v}^* \cdot \nabla)\mathbf{v}^{n+1},$$

where $\mathbf{v}^*$ is computed using solutions from previous time steps. Similarly, $\nabla \cdot (a^{n+1}\mathbf{v}^{n+1})$, appearing in (4.1), is replaced by

$$\nabla \cdot (a^{n+1}\mathbf{v}^*),$$

and $\frac{c_f}{a^{n+1}}\|\mathbf{v}^{n+1}\|\mathbf{v}^{n+1}$ in (4.2) by

$$\frac{c_f}{a^*}\|\mathbf{v}^*\|\mathbf{v}^{n+1}.$$

Additionally, the Coriolis term $f\mathbf{k} \times \mathbf{v}^{n+1}$ in (4.1) is replaced by

$$f\mathbf{k} \times \mathbf{v}^*$$

to simplify the system matrix. These time discretizations can be introduced into the continuity equation (4.5) and momentum equation (4.6) to solve for a given time step $n + 1$. For the continuity equation, this yields

$$((a_h^{n+1} - a_h^n), \psi)_\Omega - \frac{\Delta t}{2} \left[(a_h^{n+1}\mathbf{u}_h^*, \nabla\psi)_\Omega + (a_h^n\mathbf{u}_h^*, \nabla\psi)_\Omega \right] = 0,$$

and similarly, for the momentum equation,

$$((\mathbf{u}_h^{n+1} - \mathbf{u}_h^n), \boldsymbol{\phi})_\Omega + \frac{\Delta t}{2} \left[((\mathbf{u}_h^* \cdot \nabla)\mathbf{u}_h^{n+1}, \boldsymbol{\phi})_\Omega + ((\mathbf{u}_h^* \cdot \nabla)\mathbf{u}_h^n, \boldsymbol{\phi})_\Omega \right] + \Delta t(f\mathbf{k} \times \mathbf{u}_h^*, \boldsymbol{\phi})_\Omega$$
$$- \frac{g\Delta t}{2} \left[(a_h^{n+1}, \nabla\psi)_\Omega + (a_h^n, \nabla\psi)_\Omega \right] + \frac{c_f\Delta t}{2a_h^*}\|\mathbf{u}_h^*\| \left[(\mathbf{u}_h^{n+1}, \boldsymbol{\phi})_\Omega + (\mathbf{u}_h^n, \boldsymbol{\phi})_\Omega \right] = 0.$$

4.4 Examples

4.4.1 Analytical Example

In the first example, a prescribed solution is considered for the velocity and depth in order to validate the implementation. The velocity vector is given by the periodic function

$$\mathbf{v}(x, y, t) = \begin{pmatrix} y\sin(\pi t)\sin(\pi x)\cos(\frac{\pi}{2}y) \\ \cos(\pi t)\sin(\pi x)\sin(\pi y) \end{pmatrix},$$

which satisfies the no-slip boundary condition. The depth function is chosen to have the following expression

$$a(x, y, t) = 1 + 0.5\sin(\pi x)\cos(\pi y)\cos(\pi t).$$

The domain of consideration is the unit square $\Omega = (0, 1)^2$ with Dirichlet boundary conditions calculated using the value of the prescribed solution on the boundary $\partial\Omega$. Initial conditions were given by the exact solution at time $t = 0$. The following values were used for the dimensionless variables f, g, and c_f:

$$f = 5.0 \times 10^{-3}, \quad g = 2.5 \times 10^{-4}, \quad \text{and } c_f = 1.0 \times 10^{-2}.$$

These values were calculated based on standard values for the drag coefficient [41] and Coriolis parameter [40] found in the literature. The linear system of equations were solved using a direct solver from the UMFPACK library. Plots of the solution can be seen in Figure 4.1.

(a) $a(x, y)$ (b) $|v(x, y)|$
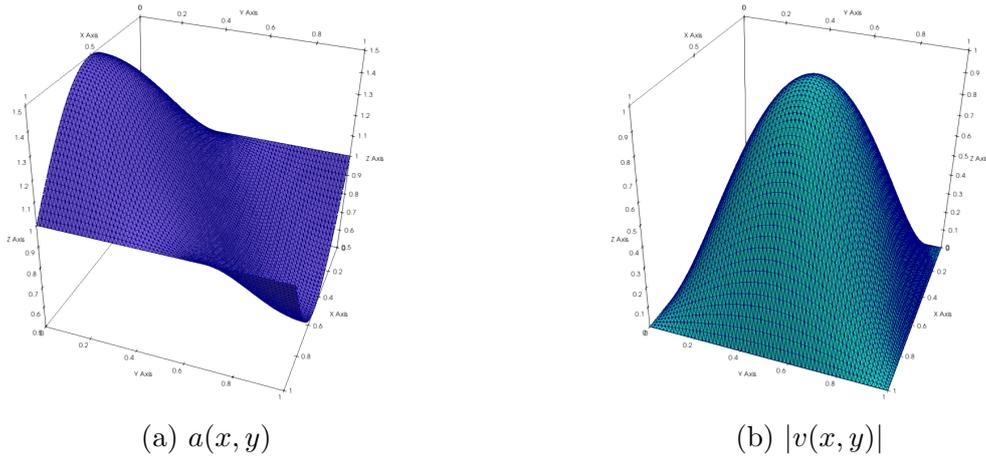
Figure 4.1: Plots of the solution of the analytical example at $t = 0$

In order to validate the implementation and compare the performance of various finite element pairs, convergence plots were generated by considering the $\mathcal{L}^2((0, T); \mathcal{L}^2(\Omega))$ and $\mathcal{L}^\infty((0, T); \mathcal{L}^2(\Omega))$ errors for an increasingly refined mesh [1]. The meshes, which are denoted T_{h_1} to T_{h_4}, were generated by applying successive refinements to an initial mesh decomposing the unit square into two uniform triangles. The number of mesh cells and corresponding mesh width h for the four computational meshes used in the analysis can be seen in Table 4.1. Figure 4.2 shows the first and coarsest computational mesh T_{h_1}.

Mesh	Level of mesh refinement	Mesh width h	Number of mesh cells
T_{h_1}	5	0.0441942	2048
T_{h_2}	6	0.0220971	8192
T_{h_3}	7	0.0110485	32768
T_{h_4}	8	0.00552427	131072

Table 4.1: Computational meshes used in the convergence studies

---

[1] The definition of these norms can be found in Section 3.3.2
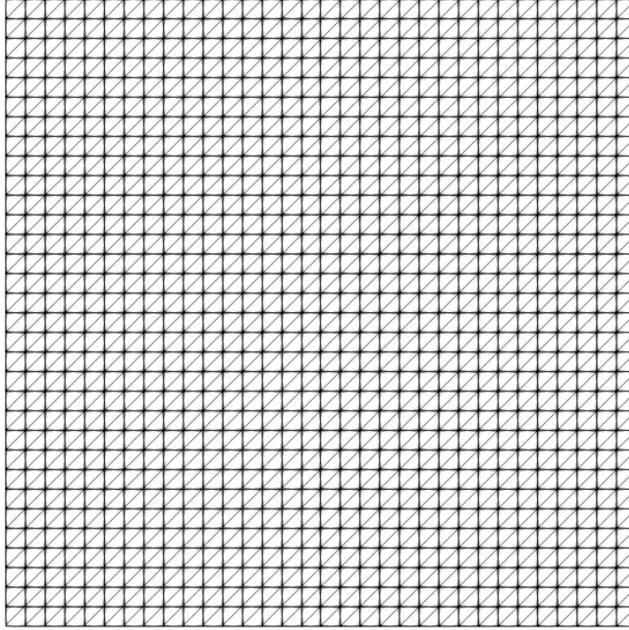
Figure 4.2: Computational mesh T_{h_1} used in the convergence analysis

The simulations were run for a complete period of $t = 2$ with a small time step $\Delta t = 10^{-3}$ to ensure that the error in time becomes negligible to the error in space. The $\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))$ errors estimates were obtained at the final time of the simulation using the output of ParMooN. The $\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))$ errors are the supremum of the $\mathcal{L}^2(\Omega)$ errors over the elapsed time.

The following low-order finite element pairs, which were used in [27], were considered in the analysis: P_2/P_1, P_1/P_1, P_1^{NC}/P_1, P_1/P_0, and P_1^{NC}/P_0. The P_1/P_0 and P_1^{NC}/P_0 finite element pairs did not converge and are excluded from the convergence plots. Convergence analyses for the remaining finite element pairs are visualized in Figure 4.3 and 4.4, where the errors are reported on a logarithmic scale. The plots were made using the Python package Matplotlib [20]. The raw errors can be found in the tables in Appendix A.
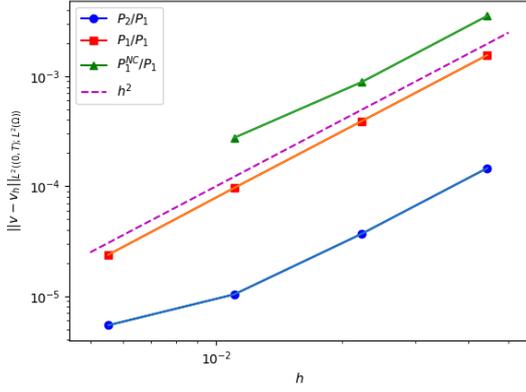
(a) $\|\mathbf{v} - \mathbf{v}_h\|_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))}$
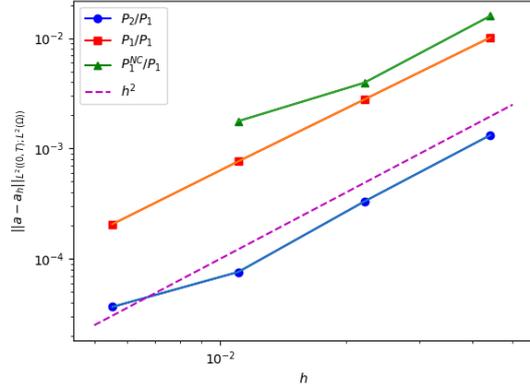
(b) $\|a - a_h\|_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))}$

Figure 4.3: $\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))$ convergence analysis
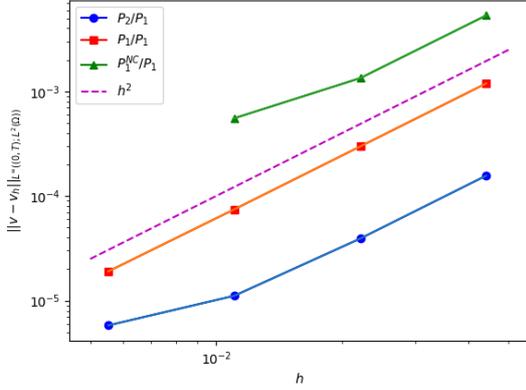
(a) $\|\mathbf{v} - \mathbf{v}_h\|_{\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))}$
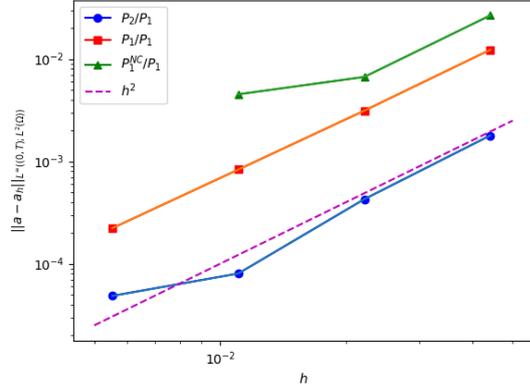
(b) $\|a - a_h\|_{\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))}$

Figure 4.4: $\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))$ convergence analysis

The P_1/P_1 and P_2/P_1 finite element pairs had stable results and exhibited second-order convergence. The slightly increased error for the P_2/P_1 finite element pair on the finest mesh T_{h_4} is likely due to the impact of the temporal error. The P_1^{NC}/P_1 behaved as expected for the first three mesh refinements, but had divergent behavior on the finest mesh T_{h_4}. Therefore, only the results for the first three mesh refinements are reported in the above figures. The cause of the divergent behavior calls for further study. The poor performance of the P_1^{NC}/P_0 finite element pair is consistent with the results of [27] discussed in the literature review. The divergent results for the P_1/P_0 finite element pair, however, stand in contradiction to their results and also require further investigation.

The analysis was limited in that it only considered one set of parameters and perhaps the divergent results in the analysis could be due to a poorly conditioned problem. It would be of interest to compare the performance of the above finite element pairs using other sets of parameters and other test problems. The impact of the size of the time step could also be studied in a systematic way.

4.4.2 "Disappearing" Dam Example

In the second example, a "disappearing" dam is simulated, where the name of the example comes from the initial conditions for the depth, which are at different heights on each half of the domain. The domain of the problem is a square with a side length of 4 km. One half of the domain has a depth of 200 m and the other has a depth of 100 m. The equations were nondimensionalized through the process described in Section 2.4, such that the simulation could be implemented on the unit square $\Omega = (0,1)^2$.

The domain was uniformly triangulated, as in the previous example, and the the Taylor-Hood finite element pair P_2/P_1 [18] was used. Homogeneous Dirichlet boundary conditions are assumed for the velocity on the boundary $\partial\Omega$. The initial condition for the velocity is given by

$$\mathbf{v}(x, y) = 0,$$

and the initial condition for the depth is given by

$$a(x, y) = \begin{cases} 1 & \text{if } x \leq 0.5, \\ 0.5 & \text{if } x > 0.5. \end{cases}$$

The following values were used for the dimensionless variables f, g, and c_f:

$$f = 10^{-3}, \quad g = 1.25 \times 10^{-2}, \quad \text{and } c_f = 2.0 \times 10^{-2}.$$

The results can be visualized in Figure 4.5 and 4.6.

(a) $t = 0$ (b) $t = 5$
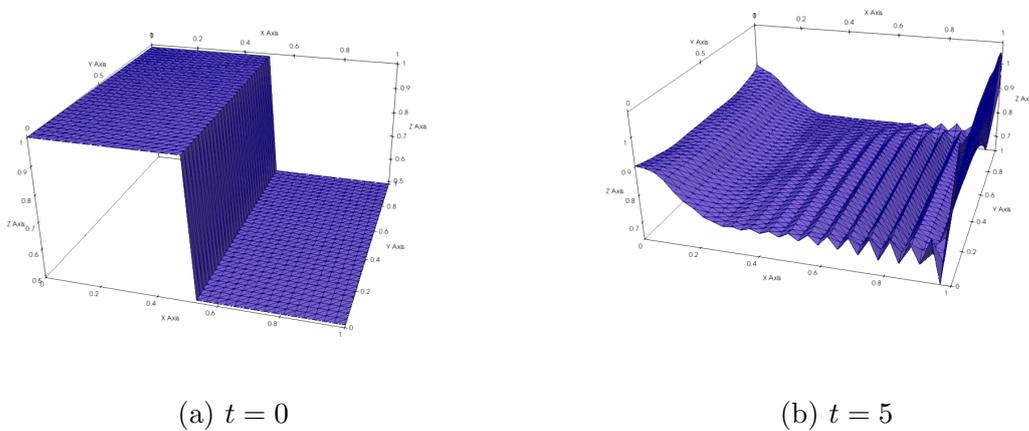
Figure 4.5: Plots of the solution of the "disappearing" dam example: $a(x, y)$
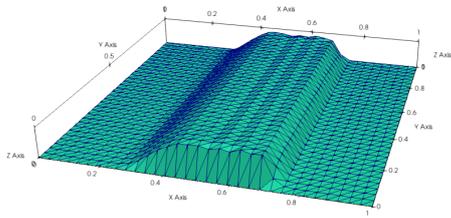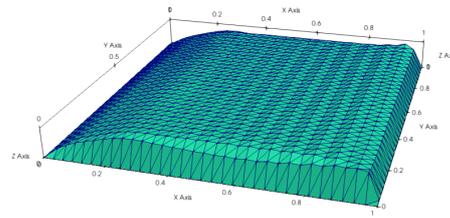
(a) $t = 2$ (b) $t = 5$

Figure 4.6: Plots of the solution of the "disappearing" dam example: $v(x, y)$

Overall, the solution behaved as expected with the higher level flowing to the lower level and the surface slowly equalizing with waves back and forth. The solution suffered from some oscillations in the surface elevation, as can be seen in Figure 4.5, which could possibly be attributed to parameters of the problem.

Chapter 5

Conclusion

5.1 Summary

This thesis provided an introduction to the shallow water equations, starting with a thorough derivation of the system of equations from the Navier-Stokes equations and discussion of the variables. The use of finite element methods to approximate the solution of the shallow water system was of particular interest. An introduction to general finite element theory was provided, followed by a review of the finite element literature pertaining specifically to the shallow water equations. Finally, the shallow water equations were implemented in ParMooN and convergence analyses were performed on various finite element pairs to validate the implementation.

5.2 Outlook

As this thesis only sought to provide an introduction to a very broad field of research, there are many opportunities for further exploration, both mathematically and numerically. The literature review pointed to DG methods being the focus of study for the shallow water equations in recent years. It would be of interest to explore the literature concerning this topic in greater depth and implement such methods in ParMooN.

Also of interest would be the implementation of the GWCE in ParMooN. Of the few studies that derived *a priori* estimates for the SWE, most of them used the GWCE formulation of the CE. A possible next step could be to run simulations using the GWCE in ParMooN and compare the results with the *a priori* error estimates discussed in this work. As the literature was lacking in research on *a priori* estimates, subsequent work could focus attention on the derivation of such estimates for other schemes.

The focus of this work was exclusively on the 2D shallow water equations. Future work could extend the literature review and ParMooN implementations to three dimensions.

Appendix A

Numerical Results

	h_1	h_2	h_3	h_4
P_2/P_1	1.46e−4	3.67e−5	1.04e−5	5.43e−6
P_1/P_1	1.56e−3	3.89e−4	9.66e−5	2.38e−5
P_1^{NC}/P_1	3.54e−3	8.82e−4	2.76e−4	diverged

Table A.1: $\|\mathbf{v} - \mathbf{v}_h\|_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))}$ errors

	h_1	h_2	h_3	h_4
P_2/P_1	1.32e−3	3.31e−4	7.60e−5	3.67e−5
P_1/P_1	1.02e−2	2.78e−3	7.65e−4	2.06e−4
P_1^{NC}/P_1	1.59e−2	3.95e−3	1.77e−3	diverged

Table A.2: $\|a - a_h\|_{\mathcal{L}^2((0,T);\mathcal{L}^2(\Omega))}$ errors

	h_1	h_2	h_3	h_4
P_2/P_1	1.57e−4	3.92e−5	1.11e−5	5.79e−6
P_1/P_1	1.19e−3	2.98e−4	7.46e−5	1.90e−5
P_1^{NC}/P_1	5.33e−3	1.34e−3	5.54e−4	diverged

Table A.3: $\|\mathbf{v} - \mathbf{v}_h\|_{\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))}$ errors

	h_1	h_2	h_3	h_4
P_2/P_1	1.79e−3	4.27e−4	8.03e−5	4.85e−5
P_1/P_1	1.23e−2	3.12e−3	8.32e−4	2.23e−4
P_1^{NC}/P_1	2.66e−2	6.68e−3	4.52e−3	diverged

Table A.4: $\|a - a_h\|_{\mathcal{L}^\infty((0,T);\mathcal{L}^2(\Omega))}$ errors

Bibliography

[1] U. Ayachit, *The ParaView Guide: A Parallel Visualization Application*, Kitware, Incorporated, 2015.

[2] C.A. Brebbia and P.W. Partridge, *Finite Element Models for Circulation Studies*, Mathematical Models for Environmental Problems (1976).

[3] ——, *Finite element simulation of water circulation in the North Sea*, Applied Mathematical Modelling **1** (1976), no. 2, 101–107.

[4] S. Chippada, C.N. Dawson, M.L. Martinez, and M.F. Wheeler, *Finite Element Approximations to the System of Shallow Water Equations, Part I: Continuous Time A Priori Error Estimates*, SIAM Journal on Numerical Analysis **35** (1998), no. 2, 692–711.

[5] ——, *Finite Element Approximations to the System of Shallow Water Equations, Part II: Discrete-Time A Priori Error Estimates*, SIAM Journal on Numerical Analysis **36** (1999), no. 1, 226–250.

[6] S. Chippada, C.N. Dawson, and M.F. Wheeler, *User's Manual for University of Texas Bays and Estuaries Simulator (UTBEST)*, 1997.

[7] B. Cockburn, G.E. Karniadakis, and C. Shu (eds.), *Discontinuous Galerkin Methods: Theory, Computation, and Applications*, Springer, 2000.

[8] M. Crouzeix and P.A. Raviart, *Conforming and nonconforming finite element methods for solving the stationary Stokes equations I*, Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique **7** (1973), 33–75.

[9] M.J.P. Cullen and C.D. Hall, *Forecasting and general circulation results from finite element models*, Quarterly Journal of the Royal Meteorological Society **105** (1979), 571–592.

[10] C. Dawson and C.M. Mirabito, *The Shallow Water Equations*, Institute for Computational Engineering and Sciences University of Texas at Austin, 2008.

[11] C. Dawson and J. Proft, *Coupled Discontinuous and Continuous Galerkin Finite Element Methods for the Depth-Integrated Shallow Water Equations*, Computer Methods in Applied Mechanics and Engineering **193** (2004), no. 3, 289–318.

[12] C. Dawson, J.J. Westerink, J.C. Feyen, and D. Pothina, *Continuous, discontinuous and coupled discontinuous-continuous Galerkin finite element methods for the shallow water equations*, International Journal for Numerical Methods in Fluids **52** (2006), no. 1, 63–88.

[13] C.N. Dawson and M.L. Martinez-Canales, *Finite Element Approximations to the System of Shallow Water Equations, Part III: On the Treatment of Boundary Conditions*, SIAM Journal on Numerical Analysis (1998).

[14] A. Demlow, *Error Estimates for the Finite Element Method.* `https://pi.math.cornell.edu/~demlow/425/chap5.pdf`. Accessed 23 Nov. 2022.

[15] D. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Springer, 2012.

[16] M.D. Gunzburger, *Finite Element Methods for Viscous Incompressible Flows: A Guide to Theory, Practice, and Algorithms*, Academic Press, 1989.

[17] E. Hanert, D.Y. Le Roux, V. Legat, and E. Deleersnijder, *An efficient Eulerian finite element method for the shallow water equations*, Ocean Modelling **10** (2005), no. 1, 115–136.

[18] P. Hood and C. Taylor, *Navier-Stokes equations using mixed interpolation*, Finite element methods in flow problems, 1974, pp. 121–132.

[19] B.L. Hua and F. Thomasset, *A noise-free finite element scheme for the two-layer shallow-water equations*, Tellus A: Dynamic Meteorology and Oceanography **36A** (1984), 157–165.

[20] J.D. Hunter, *Matplotlib: A 2D Graphics Envrionment*, Computing in Science and Engineering (2007), 90–95.

[21] V. John, *Finite Element Methods for Incompressible Flow Problems*, 1st ed., Springer, 2016.

[22] I. Kinnmark, *The Shallow Water Wave Equations: Formulation, Analysis and Application*, Lecture Notes in Engineering, Springer Verlag, 1986.

[23] E.J. Kubatko, S. Bunya, C. Dawson, J.J. Westerink, and C. Mirabito, *A Performance Comparison of Continuous and Discontinuous Finite Element Shallow Water Models*, Journal of Scientific Computing **40** (2009), 315–339.

[24] E.J. Kubatko, J.J. Westerink, and C. Dawson, *hp Discontinuous Galerkin methods for advection dominated problems in shallow water flow*, Computer Methods in Applied Mechanics and Engineering **196** (2006), no. 1, 437–451.

[25] C. Kühbacher, *Shallow Water: Derivation and Applications*, Technische Universität Dortmund, 2009. `https://www.mathematik.tu-dortmund.de/data/pdf/Kuehbacher2009.pdf`. Accessed 21 Apr. 2022.

[26] 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*, SIAM Journal of Scientific Computing **29** (2007), no. 1, 331–360.

[27] D. Le Roux, A. Staniforth, and C.A. Lin, *Finite Elements for Shallow-Water Equation Ocean Models*, Monthly Weather Review **126** (1998), 1931–1951.

[28] R.S. Lehmann, M. Lukáčová-Medvid'ová, B.J.P. Kaus, and A.A. Popov, *Comparison of Continuous and Discontinuous Galerkin Approaches for Variable-Viscosity Stokes Flow*, Zeitschrift für Angewandte Mathematik und Mechanik **96** (2016), 733–746.

[29] P. Lesaint and P.A. Raviart, *On a Finite Element Method for Solving the Neutron Transport Equation*, Mathematical Aspects of Finite Elements in Partial Differential Equations (1974).

[30] Jr. Luettich R.A, J.J. Westerink, and N.W. Scheffner, *ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves, Coasts, and Estuaries, Report 1: Theory and Methodology of ADCIRC-2DDI and ADCIRC-3DL*, Dredging Research Program Technical Report DRP-92-6 (1992).

[31] D.R. Lynch and W.G. Gray, *A wave equation model for finite element tidal computations*, Computers and Fluids **7** (1979), 207–228.

[32] M.L. Martinez, *A Priori Error Estimates of Finite Element Models of Systems of Shallow Water Equations*, Ph.D. Thesis, Rice University, Houston, TX, 1998.

[33] C.C. Mei, *Lecture Notes on Fluid Dynamics*, MIT, 2007. `https://web.mit.edu/1.63/www/Lec-notes/chap1_basics/1-4forces.pdf`. Accessed 12 Aug. 2022.

[34] I.M. Navon, *Finite-element simulation of the shallow-water equations model on a limited-area domain*, Appl. Math. Modelling **3** (1979).

[35] A.W. Otma, *Geophysical Flows*, LibreTexts, 2022.

[36] B. Rivière, M.F. Wheeler, and V. Girault, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I*, Computational Geosciences 3 (1999), 337–360.

[37] A. Schoenstadt, *A Transfer Function Analysis of Numerical Schemes Used to Simulate Geostrophic Adjustment*, Monthly Weather Review **108** (1980), 1248–1259.

[38] E. Sonnendrücker and A. Ratnani, *Advanced Finite Element Methods*, Lecture Notes Wintersemester 2015/2016, Technische Universität München, 2016.

[39] A. Staniforth and J. Cote, *Semi-Lagrangian Integration Schemes for Atmospheric Models - A Review*, Monthly Weather Review **119** (1991), 2206–2223.

[40] R. Stull, *Meteorology for Scientists and Engineers*, 3rd ed., University of British Columbia, 2011.

[41] C.B. Vreugdenhil, *Numerical Methods for Shallow-Water Flow*, Water Science and Technology Library, vol. 13, Kluwer Academic Publishers, 1994.

[42] R.A. Walters, *Numerically induced oscillations in finite-element approximations to the shallow-water equations*, International Journal for Numerical Methods in Fluids **3** (1983), 591–604.

[43] R.A. Walters and G.F. Carey, *Analysis of Spurious Oscillation Modes for the Shallow Water and Navier-Stokes Equations*, Computers and Fluids **11** (1983), no. 1, 51–68.

[44] R.A. Walters and R.T. Cheng, *Accuracy of an Estuarine Hydrodynamic Model Using Smooth Elements*, Water Resources Research **16** (1980), no. 1, 187–195.

[45] T. Weiyan, *Shallow Water Hydrodynamics*, Elsevier Oceanography Series, Elsevier Science Publishers, Amsterdam, 1992.

[46] U. Wilbrandt, C. Bartsch, N. Ahmed, N. Alia, F. Anker, L. Blank, A. Caiazzo, S. Ganesan, S. Giere, G. Matthies, R. Meesala, A. Shamim, J. Venkatesan, and V. John, *ParMooN - A modernized program package based on mapped finite elements*, Computers and Mathematics with Applications **74** (2017), 74–88.

[47] R.T. Williams, *On the Formulation of the Finite-Element Prediction Models*, Monthly Weather Review **109** (1981), 463–466.

[48] R.T. Williams and O.C. Zienkiewicz, *Improved finite element forms for the shallow-water wave equations*, International Journal for Numerical Methods in Fluids **1** (1981), 81–97.