Preface

The fiftieth year of the program was dedicated to **Nonlinear Waves**, a topic with many applications in geophysical fluid dynamics. The principal lectures were given jointly by Roger Grimshaw and Harvey Segur and between them they covered material drawn from fundamental theory, fluid experiments, asymptotics, and reaching all the way to detailed applications. These lectures set the scene for the rest of the summer, with subsequent daily lectures by staff and visitors on a wide range of topics in GFD. It was a challenge for the fellows and lecturers to provide a consistent set of lecture notes for such a wide-ranging lecture course, but not least due to the valiant efforts of Pascale Garaud, who coordinated the write-up and proof-read all the notes, we are very pleased with the final outcome contained in these pages.

This year's group of eleven international GFD fellows was as diverse as one could get in terms of gender, origin, and race, but all were unified in their desire to apply their fundamental knowledge of fluid dynamics to challenging problems in the real world. Their projects covered a huge range of physical topics and at the end of the summer each student presented his or her work in a one-hour lecture. As always, these projects are the heart of the research and education aspects of our summer study.

During the summer we also had a couple of special events, namely the public Sears Lecture given by Geoff Spedding from USC on animal propulsion entitled "Flight at small scales", and a one-day special workshop and celebration in honour of Lou Howard's 80^{th} birthday entitled "Nonlinear excursions". These well-attended events brought together old and new friends and a good time was had by all.

This summer also had its share of unscheduled surprises, but the enthusiasm and tenacity of our fellows was able to overcome all, be it rained-out softball training sessions or advisors that had to disappear abruptly to attend to robbed houses at the other end of the country. At the end it all came together, and the final days were marked by a happy reunion of everyone and a graceful farewell party, despite a somewhat disappointing outcome of the Fellow-Staff softball match.

Special thanks go to a number of people who served to make the program flow smoothly. Jeanne Fleming, Kathy Ponti and Penny Foster all helped run the Walsh Cottage office and performed exceptional administrative work for the many visits and lectures. Jeanne Fleming was responsible for assembling and preparing this volume. Janet Fields helped with all aspects of the academic programs for the fellows and as usual did a superb job.

We extend our sincere thanks to the National Science Foundation for support for this program under Grant OCE-0824636 and to the Office of Naval Research, Processes and Prediction Division, Physical Oceanography Program, under Grant N00014-09-10844.

Oliver Bühler and Karl Helfrich,
Co-Directors

TABLE OF CONTENTS

PREFACE i

TABLE OF CONTENTS ii

PARTICIPANTS v

LECTURE SCHEDULE ix

PRINCIPAL LECTURES
Roger Grimshaw, Loughborough University and Harvey Segur, University of Colorado

Lecture 1
Introduction to linear and non-linear waves
Harvey Segur 1

Lecture 2
Concepts from linear theory
Harvey Segur 7

Lecture 3
Introduction to non-linear waves
Roger Grimshaw 17

Lecture 4
Zhakharov formulation of water waves
Harvey Segur 23

Lecture 5
Waves in shallow water, part 1: the theory
Harvey Segur 35

Lecture 6
Derivation of the KdV equation for surface and internal waves
Roger Grimshaw 52

Lecture 7
Oceanographic applications
Harvey Segur 62

Lecture 8
The shallow- water equations
Harvey Segur 71

Lecture 9
Nonlinear waves in a variable medium
Roger Grimshaw 80

Lecture 10
Whitham modulation theory
Roger Grimshaw 91

Lecture 11
Internal solitary waves in the ocean
Roger Grimshaw 99

Lecture 12
Transcritical flow over an obstacle
Roger Grimshaw 110

Lecture 13
Triad (or 3-wave) resonances
Harvey Segur 123

Lecture 14
Waves on deep water, I
Harvey Segur 134

Lecture 15
Waves on deep water, II
Harvey Segur 146

Lecture 16
Solitary waves
Roger Grimshaw 158

Lecture 17
Generalized solitary waves
Roger Grimshaw 167

Lecture 18
Wave-mean flow interaction, part I
Roger Grimshaw 178

Lecture 19
Wave-mean flow interaction, part II: General Theory
Roger Grimshaw 193

Lecture 20
The explosive instability due to 3-wave or 4-wave mixing
Harvey Segur 201

Lecture 21
Potpourri
Harvey Segur 208

FELLOWS' REPORTS

Resonant triad interactions on an extended equatorial ß-plane
Michael Bates, University of South Wales 217

Degradation of the internal tide over long bumpy topography
Erinna Chen, University of California, Santa Cruz 248

Quantum vortices in a glass of Bose-Einstein condensate
Nicholas Grisouard, Université Joseph Fourier, Grenoble 269

Ray theory of nonlinear water waves
Andong He, Penn State University 290

The derivation and application of convective pattern equations
Yiping Ma, University of California, Berkeley 308

Variation of the eddy diffusivity across jets in the southern ocean
Ali Mashayekhi, University of Toronto.....330

Flow-destabilized seiche modes with a movable dam
Hélène Scolan, Ecole Normale Supérieure de Lyon.....358

Nonlinear shelf waves in a rotating annulus
Andrew Stewart, University of Oxford.....373

Nonlinear peristaltic waves: a bitter pill to swallow
Daisuke Takagi, University of Cambridge 403

Linear stability of Su-Gardner solutions to small two-dimensional perturbations
Adrienne Traxler, University of California, Santa Cruz.....428

Laboratory experiments on two coalescing axisymmetric turbulent plumes in a rotating fluid
Hiroki Yamamoto, Kyoto University 442

2009 GFD FELLOWS, STAFF AND VISITORS

Fellows

Michael Bates	University of New South Wales
Erinna Chen	University of California, Santa Cruz
Nicolas Grisouard	Université Joseph Fourier - Grenoble
Andong He	Pennsylvania State University
Yiping Ma	University of California, Berkeley
Alireza Mashayekhi	University of Toronto
Hélène Scolan	Ecole Normale Supérieure de Lyon
Andrew Stewart	University of Oxford
Daisuke Takagi	University of Cambridge
Adrienne Traxler	University of California, Santa Cruz
Hiroki Yamamoto	Kyoto University

Staff and Visitors

Triantaphyllos Akylas	Massachusetts Institute of Technology
James Anderson	Stevens Institute of Technology
Hassan Aref	Virginia Tech
Alexander Balk	University of Utah
Neil Balmforth	University of British Columbia
Andrew Belmonte	Pennsylvania State University
Oliver Buhler	New York University
Friedrich Busse	University of Bayreuth
Colm-cille Caufield	University of Cambridge
Claudia Cenedese	Woods Hole Oceanographic Institution
Paola Cessi	University of California, San Diego
Eric Chassignet	Florida State University
Greg Chini	University of New Hampshire
Predrag Cvitanovic	Georgia Institute of Technology
Charles Doering	University of Michigan
Gavin Esler	University College London
Alexey Fedorov	Yale University
Raffaele Ferrari	Massachusetts Institute of Technology
Pascale Garaud	University of California, Santa Cruz
Johannes Gemmrich	University of Victoria
Jerry Gollub	Haverford College
Roger Grimshaw	Loughborough University
Karl Helfrich	Woods Hole Oceanographic Institution
Miranda Holmes-Cerfon	New York University
Louis Howard	Massachusetts Institute of Technology
Edward Johnson	University College London
Joseph Keller	Stanford University
Nancy Kopell	Boston University

Kevin Lamb	University of Waterloo
Yuri Lvov	Rensselaer Polytechnic Institute
Norman Lebovitz	University of Chicago
Willem Malkus	Massachusetts Institute of Technology
Philip Morrison	University of Texas, Austin
Carline Muller	Massachusetts Institute of Technology
Anatoly Odulo	Suffolk University
Lev Ostrovsky	NOAA
Nathan Paldor	Hebrew University of Jerusalem
Thomas Peacock	Massachusetts Institute of Technology
Joseph Pedlosky	Woods Hole Oceanographic Institution
Lawrence Pratt	Woods Hole Oceanographic Institution
Themis Sapsis	Massachusetts Institute of Technology
Harvey Segur	University of Colorado
Geoffrey Spedding	University of Southern California
Edward Spiegel	Columbia University
Ravi Srinivasan	Brown University
Melvin Stern	Florida State University
Louis St. Laurent	Woods Hole Oceanographic Institution
Bruce Sutherland	University of Alberta
Esteban Tabak	New York University
John Taylor	Massachusetts Institute of Technology
Jean-Luc Thiffeault	University of Wisconsin, Madison
George Veronis	Yale University
Patrick Weidman	University of Colorado
Andrew Wells	University of Cambridge
John Wettlaufer	Yale University
Brian White	University of North Carolina, Chapel Hill
John Whitehead	Woods Hole Oceanographic Institution
Carl Wunsch	Massachusetts Institute of Technology
William Young	University of California, San Diego
Yuan-Nan Young	New Jersey Institute of Technology
Jie Yu	North Carolina State University

Top row from left to right: John Taylor, Alexander Balk, John Marshall, Roger Grimshaw, Hiroki Yamamoto, Joseph Pedlosky, Michael Bates, Adrienne Traxler, Ted Johnson, Hélène Scolan, Alireza Mashayekhi, unknown, unknown, Yiping Ma

Second row from left to right: Ed Spiegel, Neil Balmforth, Raffaele Ferrari, Joe Keller, Claudia Cenedese, George Veronis, Daisuke Takagi, James Anderson, Andrew Stewart, Eric Chassignet, Erinna Chen, Xujing Jia Davis, Oliver Bühler (standing)

First row from left to right: Nicholas Grisouard, Karl Helfrich, Harvey Segur, Andong He, Pascale Garaud

Not pictured: Triantaphyllos Akylas, Hassan Aref, Andrew Belmonte, Friedrich Busse, Colm-cille Caufield, Paola Cessi, Greg Chini, Predrag Cvitanovic, Charles Doering, James Esler, Alexey Fedorov, Johannes Gemmrich, Jerry Gollub, Miranda Holmes-Cerfon, Louis Howard, Kevin Lamb, Norman Lebovitz, Willem Malkus, Philip Morrison, Carline Muller, Anatoly Odulo, Lev Ostrovsky, Nathan Paldor, Geoffrey Spedding, Ravi Srinivasan, Louis St. Laurent, Bruce Sutherland, Esteban Tabak, Jean-Luc Thiffeault, Patrick Weidman, Andrew Wells, John Wettlaufer, John Whitehead, William Young, Yuan-Nan Young

Principal Lecturers

Roger Grimshaw

Harvey Segur

GFD Lecture Schedule

Week 1

Monday, June 15

Lecture 1

Introduction to linear and non-linear waves
Harvey Segur, University of Colorado

Lecture 2

Concepts from linear theory
Harvey Segur, University of Colorado

Tuesday, June 16

Lecture 3

Introduction to non-linear waves
Roger Grimshaw, Loughborough University

Lecture 4

Zhakharov formulation of water waves
Harvey Segur, University of Colorado

Wednesday, June 17

Lecture 5

Waves in shallow water, part 1: the theory
Harvey Segur, University of Colorado

Lecture 6

Derivation of the KdV equation for surface and external waves
Roger Grimshaw, Loughborough University

Afternoon lecture

What can thermal convection teach us about the nature of turbulence?
Friedrich Busse, University of Bayreuth

Thursday, June 18

Lecture 7

Oceanographic applications
Harvey Segur, University of Colorado

Lecture 8

The shallow water equations
Harvey Segur, University of Colorado

Afternoon lecture

The extra invariant for Rossby wave dynamics
Alexander Balk, University of Utah

Friday, June 19

Lecture 9

Non-linear waves in a variable medium
Roger Grimshaw, Loughborough University

Lecture 10

Whitham modulation theory
Roger Grimshaw, Loughborough University

Week 2

Monday, June 22

Lecture 11

Internal solitary waves in the ocean
Roger Grimshaw, Loughborough University

Lecture 12

Transcritical flow over an obstacle
Roger Grimshaw, Loughborough University

Tuesday, June 23

Lecture 13

Triad (or 3-wave) resonances
Harvey Segur, University of Colorado

Lecture 14

Waves on deep water, I
Harvey Segur, University of Colorado

Wednesday, June 24

Lecture 15

Waves on deep water, II
Harvey Segur, University of Colorado

Lecture 16

Solitary waves
Roger Grimshaw, Loughborough University

Thursday, June 25

Lecture 17

Generalized solitary waves
Roger Grimshaw, Loughborough University

Lecture 18

Wave-mean flow interaction, part I
Roger Grimshaw, Loughborough University

Friday, June 26

Lecture 19

Wave-mean flow interaction part II: general theory
Roger Grimshaw, Loughborough University

Lecture 20

The explosive instability due to 3-wave or 4-wave mixing
Harvey Segur, University of Colorado

Lecture 21

Potpourri
Harvey Segur, University of Colorado

Week 3

Monday, June 29

Waves and vortices in fluids and superfluids
Oliver Buhler, Courant Institute

Tuesday, June 30

Resonant forcing of gravity-capillary lumps in deep water
Triantaphyllos Akylas, MIT

Wednesday, July 1

Internal wave representation in OGCMs
Eric Chassignet, Florida State University

Thursday, July 2

Near resonant internal wave triads
Kevin Lamb, University of Waterloo

Hydraulics miscellany: Rossby waves and 2D internal waves
Ted Johnson, University College London

Friday, July 3

HOLIDAY – INSTITUTION CLOSED

Week 4

Monday, July 6

Mixing in the southern ocean
Raffaele Ferrari, MIT

Tuesday, July 7

Fluid dynamics in stellar interiors
Pascale Garaud, University of California, Santa Cruz

Wednesday, July 8

Internal solitary waves with trapped cores
Karl Helfrich, WHOI

Thursday, July 9

Dynamics of inertial particles in fluid flows
Themis Sapsis, MIT

Friday, July 10

Dynamics and stability of interacting buoyant currents
Claudia Cenedese, WHOI

Week 5

Monday, July 13

Large amplitude non-linear internal waves in the South China Sea
Lou St. Laurent, WHOI

Tuesday, July 14

The influence of horizontal current shear on salt fingers
Melvin Stern, Florida State University

Wednesday, July 15

Extreme waves-should one expect the unexpected?
Johannes Gemmrich, University of Victoria

Trapped waves in a baroclinic ocean in mid-latitudes
Nathan Paldor, Hebrew University of Jerusalem

Thursday, July 16

Aspects of stratified shear instability
Neil Balmforth, University of British Columbia

Friday, July 17

Nonlinear Excursions

Rather more can be said: the science of Lou Howard
Oliver Buhler, Courant Institute

Rip current as an instability
Jie Yu, North Carolina State University

Lou, me and science
Nancy Kopell, Boston University

Heat transport by turbulent convection
Charlie Doering, University of Michigan

Week 6

Monday, July 20

A tale of two jets: one liquid, the other granular
Patrick Weidman, University of Colorado

Tuesday, July 21

Rigorous bounds revisited: A prediction for the optimal stratification for turbulent mixing
Colm Caulfield, University of Cambridge

Wednesday, July 22

No lecture scheduled

Thursday, July 23

Propulsion through diffusion
Tom Peacock, MIT

Friday, July 24

A discontinuous Galerkin method for the Vlasov-Poisson equation
Phil Morrison, University of Texas at Austin

Week 7

Monday, July 27

Modeling of "genuinely strong" internal waves
Lev Ostrovsky, NOAA

Stability of relative equilibria of three vortices
Hassan Aref, Technical University of Denmark and Virginia Tech

Tuesday, July 28

Turbulent convection in the presence of density fronts
John Taylor, MIT

Wednesday, July 29

Generation of internal waves by gravity currents in two-layer and continuous stratification
Brian White, University of North Carolina, Chapel Hill

Thursday, July 30

Beyond Ray theory for internal waves: nonlinear effects
Bruce Sutherland, University of Alberta

Diffusion of swimming cells, flagella as coupled oscillators, and advective micro-mixing
Jerry Gollub, Haverford College

Friday, July 31

Stability of the Atlantic meridional overturning circulation in a zonally-averaged ocean model
Alexey Fedorov, Yale University

Week 8

Monday, August 3

No lecture scheduled

Tuesday, August 4

No lecture scheduled

Wednesday, August 5

Internal waves and wave turbulence
Yuri Lvov, Rensselaer Polytechnic Institute

Thursday, August 6

Public lecture

Flight at small scales
Geoff Spedding, Dept. of Aerospace and Mechanical Engineering,
University of Southern California

Friday, August 7

10:30 AM

Sinking amid bubbles
Andrew Belmonte, Penn State University

Week 9 – No lectures scheduled this week

Monday, August 10

Tuesday, August 11

Wednesday, August 12

Thursday, August 13

Friday, August 14

Week 10 – Fellows' Presentations

Tuesday, August 18

Flow-destabilized seiche modes with a movable dam
Hélène Scolan, Ecole Normale Supérieure de Lyon

Nonlinear peristaltic waves: a bitter pill to swallow
Daisuke Takagi, University of Cambridge

Degradation of the internal tide over long bumpy topography
Erinna Chen, University of California, Santa Cruz

Resonant triad interactions on an extended equatorial β-plane
Michael Bates, University of South Wales

Wednesday, August 19

Quantum vortices in a glass of Bose-Einstein condensate
Nicholas Grisouard, Université Joseph Fourier, Grenoble

Variation of the eddy diffusivity across jets in the southern ocean
Alireza Mashayekhi, University of Toronto

Laboratory experiment on two coalescing axisymmetric turbulent plumes in a rotating fluid
Hiroki Yamamoto, Kyoto University

Nonlinear shelf waves in a rotating annulus
Andrew Stewart, University of Oxford

Thursday, August 20

Ray theory of nonlinear water waves
Andong He, Penn State University

Linear stability of Su-Gardner solutions to small two-dimensional perturbations
Adrienne Traxler, University of California, Santa Cruz

The derivation and application of convection pattern equations
Yiping Ma, University of California, Berkeley

Lecture 1: Introduction to Linear and Non-Linear Waves

Lecturer: Harvey Segur. Write-up: Michael Bates

June 15, 2009

1 Introduction to Water Waves

1.1 Motivation and Basic Properties

There are many types of waves. Here, we will be paying particular attention to water waves as they provide a concrete, physical example of a dynamical system that exhibits many of the mathematical concepts that have been developed in recent years, including,

- linear stability,
- nonlinear stability,
- solitons,
- complete integrability,
- chaos,
- sensitive dependence on initial data,
- singularities,
- blow-up in finite time, and
- deterministic versus probabilistic models.

Water waves evolve on a timescale that humans can naturally relate to (as opposed to, say, optical waves). Furthermore, since we are hosted by the Woods Hole Oceanographic Institution, water waves are a fitting phenomenon to examine.

For the purposes of these notes, we shall pay particular attention to surface water waves, that is, those that you would feel if you were in a boat on the surface, or that you might observe if you were at the beach.

Surface water waves have their maximum displacement at the surface (decreasing exponentially with depth) and are approximately periodic. Furthermore, they propagate with little dissipation [1], that is, they lose very little energy to the surrounding environment as they propagate. As a result, they can propagate very long distances. Surface water waves

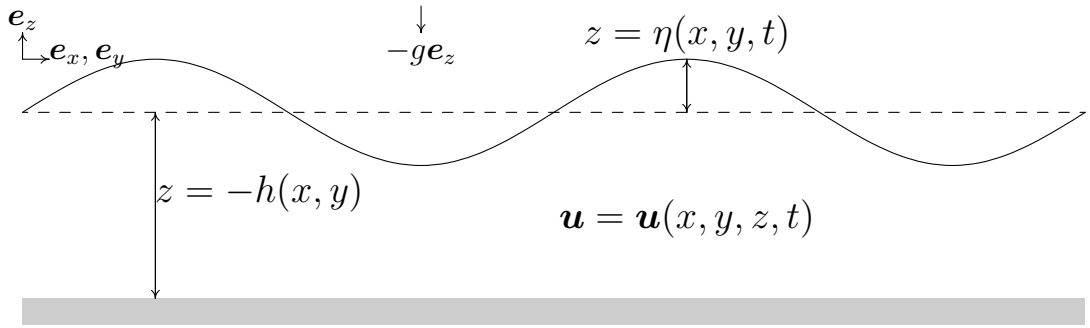


Figure 1: The shaded region indicates the solid earth boundary, the dashed line indicates mean sea level ($z = 0$) and the solid line represents the free surface. h is the distance from mean sea level to the ocean floor, and $\eta(x, y, t)$ is the deviation of the free surface from mean sea level and g is acceleration due to gravity.

are also dispersive, in that waves of different wavelengths travel at different speeds.

There are a number of other oceanic waves that will be ignored in this discussion, including,

- sound waves (or pressure waves) which arise from the compressibility of seawater. Pressure wave speeds are $O(1500\,\mathrm{ms}^{-1})$, which is much greater than, say, the 2004 Tsunami which propagated at less than $200\,\mathrm{ms}^{-1}$,
- internal waves (see Lectures 6 and 11), which arise due to variations in fluid density, have periods in the order of hours (recall surface waves have periods of order seconds),
- inertial waves (which include Rossby Waves), which are due to the rotation of the Earth and have periods greater than 12 h.

1.2 Derivation of the Governing Equations

Following Stokes[2], we derive some equations representing the evolution of the displacement of the sea surface height, $\eta = \eta(x, y, t)$, from a rest position at $z = 0$ and for the velocity throughout the fluid domain, $\boldsymbol{u} = \boldsymbol{u}(x, y, z, t)$, where $-h(x, y) < z < \eta(x, y, t)$ (h is the depth of the ocean floor), with $t > 0$ and $\eta + h > 0$.

Here (x, y, z) denote fixed, laboratory coordinates with z indicating the vertical coordinate and x and y denoting horizontal coordinates. This is illustrated in Figure 1. We can denote the location of a fixed fluid parcel as $\boldsymbol{x}(t) = \{x(t), y(t), z(t)\}$ and the velocity of a fluid parcel,

$$\boldsymbol{u}(\boldsymbol{x}, t) = (u, v, w) = \left(\frac{\mathrm{D}x}{\mathrm{D}t}, \frac{\mathrm{D}y}{\mathrm{D}t}, \frac{\mathrm{D}z}{\mathrm{D}t}\right), \tag{1}$$

where $\frac{\mathrm{D}}{\mathrm{D}t}$ is the material derivative, or in other words, the derivative that follows the fluid parcel in a Lagrangian perspective of fluid motion. The material derivative may be written

in Eulerian form as,

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}. \tag{2}$$

For instance, the material derivative for the z component of velocity is,

$$\frac{\mathrm{D}w}{\mathrm{D}t} = \frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}. \tag{3}$$

To set about deriving equations governing η and $\boldsymbol{u}$, we assume that the water (again, following Stokes [2]),

- is incompressible (which precludes sound waves),
- has uniform density (which precludes inertial waves),
- is not affected by the rotation of the Earth (which precludes inertial waves),
- is inviscid,
- and finally that gravity is uniform and is anti-parallel with $\boldsymbol{e}_z$.

Note that the theory can be expanded to include all of the neglected effects (Lectures 6 and 11 analyse the case of internal waves, while Lecture 15 discusses the effect of viscous damping on the fluid motion).

In addition, we assume that the flow is irrotational. This assumption does not change the governing equations but restricts the types of initial conditions considered, and therefore the types of solutions obtained.

We begin with mass conservation, noting that the net mass flux into an arbitrary volume V must be zero by incompressibility, so that

$$\iint_{\partial V} \rho \boldsymbol{u} \cdot \hat{\mathbf{n}} \mathrm{d}s = 0 \tag{4}$$

where ∂V is the surface bounding the volume, $\mathrm{d}s$ an infinitesimal area on that surface, and $\hat{\mathbf{n}}$ the unit normal vector to the surface. Using the divergence theorem, and the fact that we have assumed a uniform density, $\rho =$ constant, we may rewrite equation (4) as

$$\rho \iiint_V \nabla \cdot \boldsymbol{u} \, \mathrm{d}V = 0. \tag{5}$$

This expression is valid for all choices of V. We can thus say that the fluid velocity is divergence free

$$\nabla \cdot \boldsymbol{u} = 0. \tag{6}$$

We now examine the assumption that the flow is irrotational. The vorticity of the fluid is defined by

$$\boldsymbol{\omega} = \nabla \times \boldsymbol{u}. \tag{7}$$

If we assume that the vorticity is zero, then we can define a velocity potential $\phi = \phi(\boldsymbol{x}, t)$ such that

$$\boldsymbol{u} = \nabla \phi. \tag{8}$$

Recall from vector calculus that the curl of a gradient is identically zero, thus, if the vorticity of the flow is non-zero, then the velocity potential does not exist. Combining the definition for the velocity potential with the conservation (divergence) equation shows that ϕ is a solution of Laplace's equation

$$\nabla^2\phi = 0. \tag{9}$$

We have not yet discussed the validity of the irrotational assumption. To investigate this, we first state the Navier-Stokes equations,

$$\rho\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \nabla p + \rho g\boldsymbol{e}_z = \mu\nabla^2\boldsymbol{u}. \tag{10}$$

If we neglect viscosity by setting $\mu = 0$, we regain Euler's equation,

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \frac{\nabla p}{\rho} + g\boldsymbol{e}_z = 0. \tag{11}$$

If we then take the curl of Euler's equation, we find an expression for the time evolution of vorticity,

$$\frac{\mathrm{D}\boldsymbol{\omega}}{\mathrm{D}t} = \frac{\partial\boldsymbol{\omega}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{\omega} = \boldsymbol{\omega}\cdot\nabla\boldsymbol{u} \tag{12}$$

showing that if the vorticity is zero at $t = 0$, it remains so for all subsequent time.

We now examine Bernoulli's Law. We first rewrite the material derivative as

$$\frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} = \frac{\partial\boldsymbol{u}}{\partial t} + \frac{1}{2}\nabla|\boldsymbol{u}|^2 - \boldsymbol{u}\times\boldsymbol{\omega}\,, \tag{13}$$

where we have made use of equation (2) and of the vector identity $\boldsymbol{u}\cdot\nabla\boldsymbol{u} = \frac{1}{2}\nabla|\boldsymbol{u}|^2 - \boldsymbol{u}\times\nabla\times\boldsymbol{u}$. Since we have assumed that the initial vorticity is zero and we have shown that it shall remain so, the third term on the right is identically zero. As a result, Euler's equation becomes,

$$\nabla\left\{\frac{\partial\phi}{\partial t} + \frac{1}{2}|\nabla\phi|^2 + \frac{p}{\rho} + gz\right\} = 0. \tag{14}$$

Integrating this expression gives

$$\frac{\partial\phi}{\partial t} + \frac{1}{2}|\nabla\phi|^2 + \frac{p}{\rho} + gz = F(t)\,, \tag{15}$$

where $F(t)$ is an additive constant. Note that since the velocity potential is defined up to an arbitrary additive function of time, we may absorb $F(t)$ into $\frac{\partial\phi}{\partial t}$. As such, we may set $F(t) = 0$ in (15) without loss of generality, and thus recover the well-known Bernoulli's Law.

We assume that the bottom boundary is impermeable, and thus, enforce a "no-normal flow" boundary condition,

$$\boldsymbol{u}\cdot\nabla\{z + h(x,y)\} = 0 \quad \text{at } z = -h(x,y)\,, \tag{16}$$

where $\nabla\{z + h(x, y)\}$ is the normal vector to the bottom surface. Using the definition of the velocity potential, equation (8), we obtain,

$$\frac{\partial \phi}{\partial z}\frac{\partial z}{\partial z} + \frac{\partial \phi}{\partial x}\frac{\partial h}{\partial x} + \frac{\partial \phi}{\partial y}\frac{\partial h}{\partial y} = 0$$
$$\frac{\partial \phi}{\partial z} + \nabla\phi \cdot \nabla h = 0\,. \tag{17}$$

On the surface, $z = \eta(x, y, t)$, we require the continuity of the pressure field p. Just above the surface, there are two contribution to pressure: a pressure due to the weight of the atmosphere and a pressure given by the surface tension, which conceptually acts like an elastic membrane stretched over the surface of the water. The surface tension is evaluated from the Young-Laplace equation. Hence,

$$p = p_{\text{air}} - \sigma\nabla \cdot \hat{\mathbf{n}} \quad \text{at } z = \eta(x, y, t)\,, \tag{18}$$

where σ is a constant, with units Nm^{-1}, and $\hat{\mathbf{n}}$ is the surface normal unit vector. From here on, we assume that $p_{\text{air}} = 0$, again without loss of generality. Note that we are ignoring the effects of wind. Vector calculus tells us that

$$\hat{\mathbf{n}} = \frac{\nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}\,. \tag{19}$$

Using (18) in Bernoulli's Law (15) gives us the dynamic boundary condition on the free surface,

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}|\nabla\phi|^2 + g\eta = \frac{\sigma}{\rho}\nabla \cdot \left\{\frac{\nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}\right\}. \tag{20}$$

We obtain a kinematic boundary condition by assuming that a material element on the free surface stays on the free surface,

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(z(t) - \eta(x, y, t)\right) = 0. \tag{21}$$

Noting that $\frac{\mathrm{D}z}{\mathrm{D}t} = w = \partial_z\phi$ and $\frac{\mathrm{D}\eta}{\mathrm{D}t} = \partial_t\eta + \boldsymbol{u} \cdot \nabla\eta$, we obtain

$$\frac{\partial \eta}{\partial t} + \nabla\phi \cdot \nabla\eta = \frac{\partial \phi}{\partial z}\,, \tag{22}$$

when evaluated at $z = \eta$.

To summarize, the governing equations are

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}|\nabla\phi|^2 + g\eta = \frac{\sigma}{\rho}\nabla \cdot \left\{\frac{\nabla\eta}{\sqrt{1 + |\nabla\eta|^2}}\right\} \qquad \text{on} \quad z = \eta(x, y, t)$$
$$\frac{\partial \eta}{\partial t} + \nabla\phi \cdot \nabla\eta = \frac{\partial \phi}{\partial z} \qquad \text{on} \quad z = \eta(x, y, t)$$
$$\nabla^2\phi = 0 \qquad -h(x, y) < z < \eta(x, y, t)$$
$$\frac{\partial \phi}{\partial z} + \nabla\phi \cdot \nabla h = 0 \qquad \text{on} \quad z = -h(x, y).$$

The first equation relates the evolution of the velocity potential for a material element on the surface to the restoring force of gravity and the surface tension. The second equation describes the kinematic evolution of the free surface. The third equation is the continuity equation, where we have assumed that the fluid is incompressible and irrotational. The last equation is a statement that we do not allow any flow across the impermeable bottom boundary.

References

[1] F. E. Snodgrass, G. W. Groves, K. F. Hasselmann, G. R. Miller, W. H. Munk, and W. H. Powers, *Propagation of ocean swell across the pacific*, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences., 259 (1966), pp. 431–497.

[2] G. G. Stokes, *On the theory of oscillatory waves*, Transactions of the Cambridge Philosophical Society, 8 (1847), pp. 441–473.

Lecture 2: Concepts from Linear Theory

Lecturer: Harvey Segur. Write-up: Alireza Mashayekhi and Michael Bates

June 15, 2009

1 Introduction

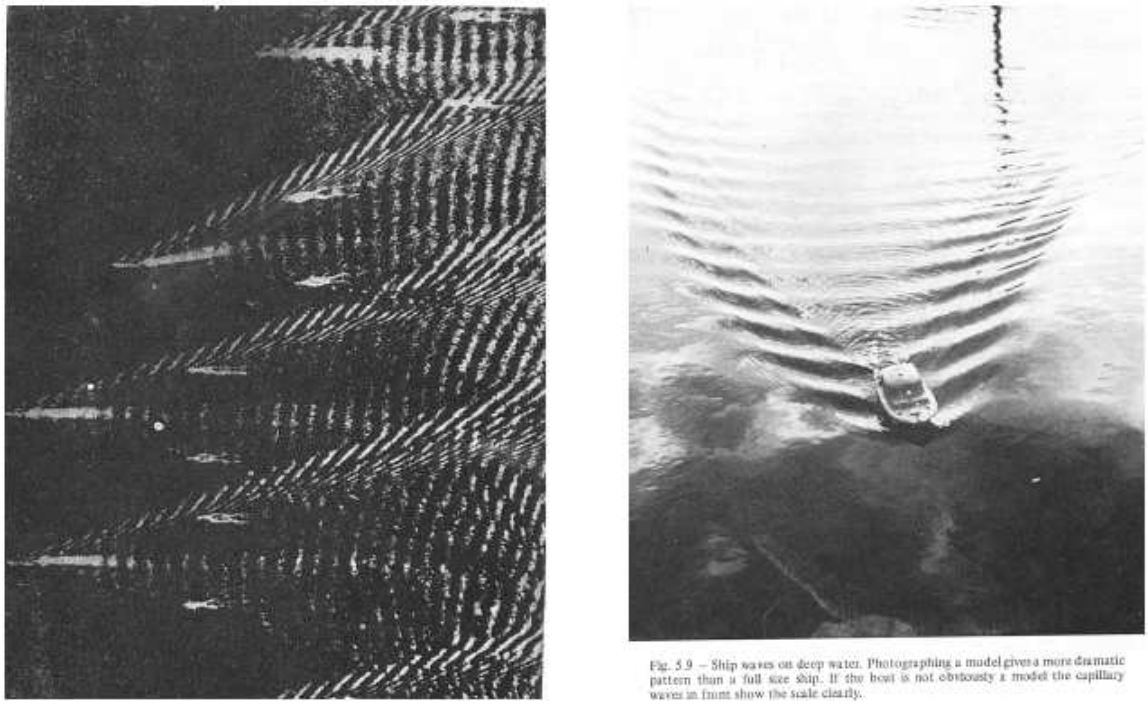


Fig. 5.9 – Ship waves on deep water. Photographing a model gives a more dramatic pattern than a full size ship. If the boat is not obviously a model the capillary waves in front show the scale clearly.

Figure 1: Ship waves from WW II battleships shown on the left (from *Water Waves* by Stoker [3]) and a toy boat on the right from (*Introduction to Water Waves* by Crapper [1]). Kelvin's (1887) method [2] of stationary phase predicts both.

Figure 1 shows the wave patterns observed in the wake of ships and boats, which are among the few kinds of waves that are well-explained by linear theory. In this lecture, we study linear theory by starting from the nonlinear equations of water waves derived in Lecture 1, linearizing them, and exploring some fundamental concepts which emerge from the linearized equations such as the phase and group velocities and particle orbits. The predictions from linear theory also help distinguish between cases where surface tension effects are negligible (gravity waves) and where they dominate (capillary waves). Different

characteristics of shallow and deep water waves (such as particle paths in each case) are also pointed out.

2 Nonlinear equations of motion

Starting from nonlinear equations of motion for an irrotational flow with no wind forcing derived in Lecture 1, we know that

$$\partial_t \eta + \nabla\phi \cdot \nabla\eta = \partial_z \phi, \tag{1}$$

$$\partial_t \phi + \frac{1}{2}|\nabla\phi|^2 + g\eta = \frac{\sigma}{\rho}\nabla \cdot \left(\frac{\nabla\eta}{\sqrt{1+|\nabla\eta|^2}} \right), \tag{2}$$

have to be satisfied at the surface, defined as $z = \eta(x, y, t)$. The first relation is the kinematic boundary condition while the second one is the Bernoulli's law expressing the dynamic boundary condition at the surface. For the interior flow, the Laplace equation

$$\nabla^2\phi = 0 \qquad\qquad -h(x,y) < z < \eta(x,y,t), \tag{3}$$

needs to be satisfied and the bottom boundary condition is

$$\partial_z\phi + \nabla\phi \cdot \nabla h = 0, \qquad\qquad \text{on} \quad z = -h(x,y). \tag{4}$$

For small amplitude waves, the nonlinear terms in (1) and (2) can be neglected to linearize the problem around the state of rest. Further simplifications can be achieved by arguing that to first order the boundary conditions can be evaluated at $z = 0$ rather than $z = \eta$. This can be justified by expanding (1) and (2) in a Taylor series around $z = 0$ and keeping linear terms only. For example the $\partial_z\phi$ in (1) yields

$$\frac{\partial\phi}{\partial z}\,|_{z=\eta} = \frac{\partial\phi}{\partial z}\,|_{z=0} + \eta\frac{\partial^2\phi}{\partial z^2}\,|_{z=0} + \ldots \simeq \frac{\partial\phi}{\partial z}\,|_{z=0}\,. \tag{5}$$

Applying these simplifications and assuming a flat bottom (h=const) leads to the following set of linearized equations:

$$\partial_t\eta = \partial_z\phi, \qquad\qquad \text{on} \quad z = 0, \tag{6}$$

$$\partial_t\phi + g\eta = \frac{\sigma}{\rho}\nabla^2\eta \qquad\qquad \text{on} \quad z = 0, \tag{7}$$

$$\nabla^2\phi = 0 \qquad\qquad -h < z < 0, \tag{8}$$

and

$$\partial_z\phi = 0, \qquad\qquad \text{on} \quad z = -h. \tag{9}$$

The goal is to solve (8) for the interior flow subject to the conditions (6), (7), and (9). As a warm-up problem, let us first find solutions to (8) assuming that ϕ is known on the upper boundary. For example, if

$$\phi(x, y, 0, t) = a \sin(kx), \tag{10}$$

at the top boundary and (9) at the bottom boundary, the solution is

$$\phi(x, y, z, t) = a \frac{\cosh(k(z + h))}{\cosh(kh)} \sin(kx). \tag{11}$$

If the top boundary condition is changed to

$$\phi(x, y, 0, t) = a \sin(kx) \cos(ly), \tag{12}$$

then

$$\phi(x, y, z, t) = a \frac{\cosh(\kappa(z + h))}{\cosh(\kappa h)} \sin(kx) \cos(ly), \tag{13}$$

where $\kappa^2 = k^2 + l^2$.

For simplicity, the problem can be rewritten in complex variables. Hence if the function ϕ takes the form

$$\phi(x, y, 0, t) = \Re\{a \; e^{ikx+ily}\} \qquad \text{on} \qquad z = 0. \tag{14}$$

then

$$\phi(x, y, z, t) = \Re \left\{ a \frac{\cosh(\kappa(z + h))}{\cosh(\kappa h)} \; e^{ikx+iky} \right\}. \tag{15}$$

in the whole domain. Finally, note that if ϕ is time-dependent on the boundary, it can be expressed as a linear combination of components in the form of (14), so that the solution will be the equivalent linear combination of the respective solutions (15):

$$\phi(x, y, z, t) = \Re \left\{ \iiint \Phi(k, l, \omega) \frac{\cosh(\kappa(z + h))}{\cosh(\kappa h)} \; e^{ikx+ily-i\omega t} \mathrm{d}k \, \mathrm{d}l \, \mathrm{d}\omega \right\}, \tag{16}$$

leading to an equivalent expression for η:

$$\eta(x, y, t) = \Re \left\{ \iiint H(k, l, \omega) \; e^{ikx+ily-i\omega t} \mathrm{d}k \, \mathrm{d}l \, \mathrm{d}\omega \right\}. \tag{17}$$

Next, (16) and (17) can be used to satisfy the boundary conditions (6) and (7) at the surface to find the linearized dispersion relation for the frequency ω:

$$\omega^2 = \left(g + \frac{\sigma}{\rho} \kappa^2 \right) [\kappa \cdot \tanh(\kappa h)], \qquad \kappa^2 = k^2 + l^2. \tag{18}$$

Since the expression obtained for ω depends on the wave number, it is called a "dispersion" relation. If a system of linear evolution equations has a dispersion relation, then that

relation encodes all of the information about wave propagation from the partial differential equations (but not from their boundary or initial conditions).

From this point on, we restrict our attention to 2-D motions. For a linearized 2-D case ($\partial_y \equiv 0$), the dispersion relation (18) reduces to

$$\omega^2 = \left(gh + \frac{\sigma}{\rho}k^2 h\right)\left[\frac{\tanh(kh)}{kh}\right] \cdot k^2, \tag{19}$$

and we may choose,

$$\omega(k) = \sqrt{gh + \frac{\sigma}{\rho}k^2 h} \cdot \sqrt{\frac{\tanh(kh)}{kh}} \cdot k. \tag{20}$$

The general solution of the 2-D linear problem can then be written in the form of a sum over all components like those in (16) and (17):

$$\begin{aligned}\phi(x,z,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_-(k)\frac{\cosh(k(z+h))}{\cosh(kh)}e^{i(kx-\omega(k)t)}\mathrm{d}k + \\ \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_+(k)\frac{\cosh(k(z+h))}{\cosh(kh)}e^{i(kx+\omega(k)t)}\mathrm{d}k,\end{aligned} \tag{21}$$

and

$$\begin{aligned}\eta(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{i\omega(k)}{g+\frac{\sigma}{\rho}k^2}\Phi_-(k)e^{i(kx-\omega(k)t)}\mathrm{d}k - \\ \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{i\omega(k)}{g+\frac{\sigma}{\rho}k^2}\Phi_+(k)e^{i(kx+\omega(k)t)}\mathrm{d}k.\end{aligned} \tag{22}$$

The first integrals in (21) and (22) correspond to waves traveling to the right and the second integrals correspond to leftward-travelling waves.

3 Phase velocity and group velocity

The "phase velocity" for each wavenumber k in (22) is defined by

$$c_p(k) = \frac{w(k)}{k} \geq 0, \tag{23}$$

where $\omega(k)$ was defined by the dispersion relation (20). The phase speed is the velocity with which the crest of each wave travels. If the phase speed varies with k, each wave travels with its own velocity and they progressively disperse. In contrast if all waves travel with the same velocity, then the system is called "non-dispersive". Sound waves are examples of non-dispersive waves.

Another velocity can be defined which is called the "group velocity":

$$c_g = \frac{\mathrm{d}w}{\mathrm{d}k}. \tag{24}$$

This is the velocity with which the energy of the waves propagates and is not necessarily equal to the wave's phase velocity. To see this, consider adding two waves with slightly different wavenumbers $k, k+\delta k$ $(\delta k \ll k)$ and frequencies $\omega(k), \omega(k+\delta k) \approx \omega(k) + \frac{d\omega}{dk} \cdot \delta k$. The combination of the two waves has the form

$$\eta(x,t) = \sin\{kx - \omega t\} + \sin\{(k+\delta k)x - (\omega + c_g \delta k)t\}$$

$$= 2\sin\{(k + \frac{\delta k}{2})x - (\omega + \frac{c_g \delta k}{2}t\} \cdot \cos\left\{\frac{\delta k}{2}x - \frac{c_g \delta k}{2}t\right\} \tag{25}$$

$$\Rightarrow \quad \eta(x,t) \approx 2\sin\{k(x - c_p t)\} \cdot \cos\left\{\frac{\delta k}{2}(x - c_g t)\right\}. \tag{26}$$

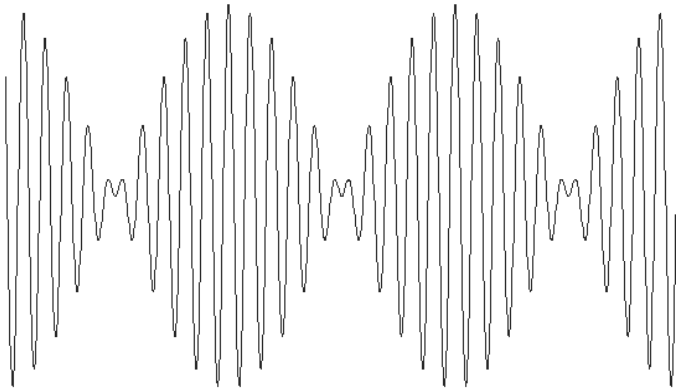


Figure 2: Formation of a wave group by linear combination of two waves

The $\sin(k(x-c_p t))$ term corresponds to the fast oscillations (shown in Figure 2) moving with the phase speed $c_p = \omega/k$. The amplitude of this wave is modulated by a slowly varying function $\cos\{\frac{\delta k}{2}(x - c_g t)\}$ which has a longer wavelength $4\pi/\delta k$ and propagates at a speed $c_g = d\omega/dk$. Multiplication of a rapidly varying and a slowly varying wave in (25) leads to periodic wave groups (Figure 2). The individual wave components travel with the phase speed c_p while the envelope of the wave group travels with the group speed c_g. For a wave with a phase speed larger than the group speed, the wave crests appear from behind the wave envelope at the nodal point, progress forward, and disappear at the front nodal point. For $c_p < c_g$ the reverse is true. Thus, following the crest of the largest wave in a wave envelope shows that it only dominates the group for a limited time due to the slow modulation. For this reason, a surfer would not be able to enjoy riding a large wave for a long time unless the phase and group velocities were almost the same. In this scenario, the large wave almost holds its shape as it travels with the group. Since no energy can pass through the nodal points, the wave energy can only travel with the group velocity and not the phase velocity.

4 Gravity waves

In most oceanographic applications, surface tension is negligible. In this section, we therefore restrict our attention to the case where the surface tension does not play an important

role, so that equation (20) reduces to

$$\omega(k) = \sqrt{gh} \cdot \sqrt{\frac{\tanh(kh)}{kh}} \cdot k, \tag{27}$$

which is the dispersion relation for "gravity waves". Figure 3 shows the plot of the frequency ω versus k. The slope of the straight line connecting a point on the curve and the origin is the phase velocity ω/k while the slope of the tangent to the curve at that point is the group velocity $\mathrm{d}\omega/\mathrm{d}k$. The bottom panel shows the phase velocity and the group velocity curves. The peak of the phase speed curve corresponds to the fastest moving wave. For gravity waves, the longest waves travel fastest and with the maximum speed of $\sqrt{gh}$. This can be obtained from (27) by letting $kh \to 0$ or by calculating the slope of the dispersion curve in the $kh \to 0$ limit. It is important to note that the maxima of the phase and group velocity curves do not necessarily coincide.

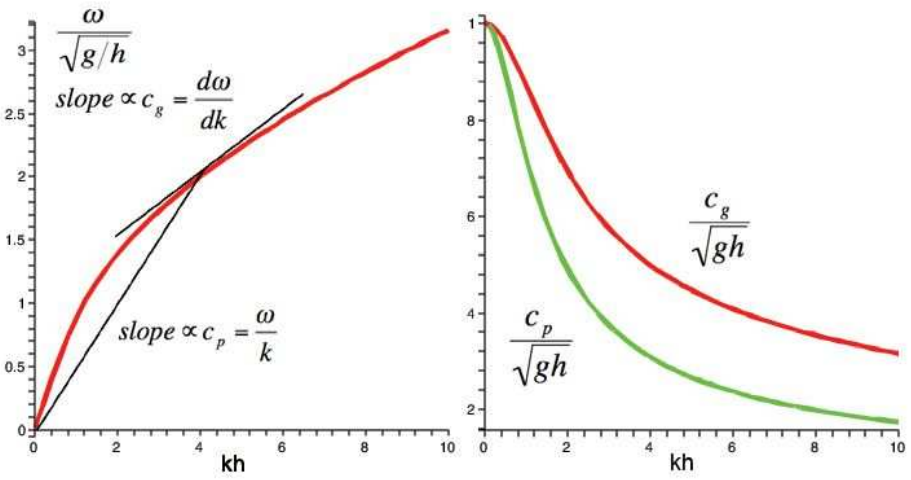


Figure 3: (left) Determining the group and phase velocities from the dispersion curve. (right) Phase and group velocity.

To have an idea of the maximum phase speed of gravity waves in nature, the maximum wave speed in the bay of Bengal (a potential tsunami region with a depth of 3500 m) is 185 m/s (ie. 670 km/hr=415 mi/hr). The deepest point in the ocean (near Guam) is around 11,000m deep which is deeper than the height of Mt. Everest (8848 m). The fastest gravity wave in the ocean can have a velocity of 328 m/s (1182 km/hr = 734 mi/hr). Although this is very close to speed of sound in the air, it is considerably slower than the speed of sound in water which is 1450 m/s (at 10^oC).

Finally, note that two limiting cases can be considered to simplify the dispersion relation (27). For $h/\lambda \gg 1$ (or $hk \gg 1$), the dispersion relation and its corresponding phase velocity reduce to

$$\omega = \sqrt{gk}, \tag{28}$$

$$c_p = \frac{\omega}{k} = \sqrt{\frac{g}{k}}, \tag{29}$$

since $\tanh(x) \to 1$ for $x \to \infty$. This is called the "Deep Water" approximation and implies that deep-water waves are dispersive. It should be noted that kh needs not be very large for this approximation to hold as $\tanh(kh) = 0.94138$ for only $kh = 1.75$ and $\tanh(kh) = 0.999993$ for $kh = 2\pi$. In the "Shallow Water" limit on the other hand, we have $kh \ll 1$. Since $\tanh(x) \to x$ as $x \to 0$, the dispersion relation and phase velocity relations for gravity waves take the form of

$$\omega = \sqrt{gh} \cdot k \tag{30}$$

$$c_p = c_g = \sqrt{gh}, \tag{31}$$

which implies that shallow water waves are approximately non-dispersive and all travel with the same speed at a similar depth. This makes waves in shallow enough waters good candidates for surfing, a fact well-known by surfers.

5 Waves with surface tension

The dispersion relation (20) includes the surface tension and can be used to find the general expression for the phase speed:

$$c_p = \sqrt{\left(\frac{g}{k} + \frac{\sigma k}{\rho}\right) \tanh(kh)} = \sqrt{\left(\frac{g\lambda}{2\pi} + \frac{2\pi\sigma}{\rho\lambda}\right) \tanh(\frac{2\pi h}{\lambda})}, \tag{32}$$

where $\lambda = 2\pi/k$ is the wavelength. Figure 4 shows the plot of the phase speed versus the wavelength to demonstrate the effect of surface tension. As seen in equation (32), surface tension increases the phase speed for all wavenumbers, a fact which can easily be interpreted on physical grounds since tension increases the restoring force on the free surface. However, Figure 4 shows that this effect is only noticeable for short waves ($\lambda < \lambda_{\min}$). For deep water waves($\tanh(kh) \simeq 1$), the phase velocity takes the form

$$c_p = \frac{\omega}{k} = \frac{\sqrt{g + \frac{\sigma}{\rho}k^2}}{\sqrt{|k|}} \tag{33}$$

and the wavelength of the slowest wave ($c_{\min}$) can be obtained by setting $\mathrm{d}c_p/\mathrm{d}\lambda = 0$ yielding

$$c_{\min} = \left[\frac{4g\sigma}{\rho}\right]^{1/4} \quad \text{at} \quad \lambda_{\min} = 2\pi\sqrt{\frac{\sigma}{\rho g}}. \tag{34}$$

For air-water interface at 20^oC, $c_{\min}$ and $\lambda_{\min}$ take the values of 23.2 cm/s and 1.73 cm respectively. As the figure shows, this implies that waves with wavelengths of a few centimeters or less are affected by surface tension. These waves are called "capillary" waves and are often referred to as "ripples". For wavelengths of few millimeters surface tension is the dominating factor with the effect of gravity being negligible. For waves longer than a few centimeters gravity plays the dominant role. As mentioned earlier and also shown on the figure, gravity waves have a maximum speed of $\sqrt{gh}$. As the figure shows, for every

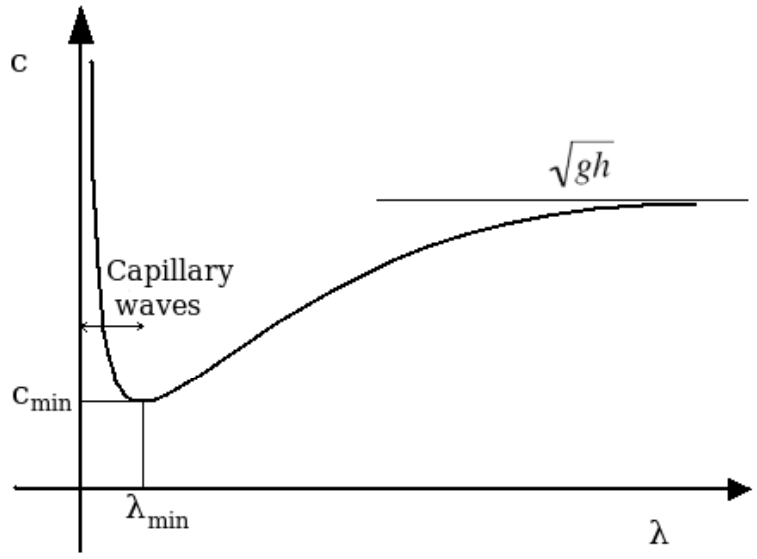


Figure 4: Phase velocity vs wavelength for surface gravity waves.

(long) gravity wave, there is a (short) capillary wave with the same phase speed. Remote sensing of ocean waves depends on this fact. One can also plot the graph of the group velocity versus the wavelength and observe that the group velocity also has a minimum, at a wavelength slightly different from that of the phase velocity.

It should also be noted that there are capillary waves which can travel faster than the fastest gravity waves although they are very small and rapidly dissipate in the presence of viscosity. Another point shown on the figure is that for gravity waves, long waves travel faster than short waves while the opposite is true for capillary waves. To show this, Figure 5 demonstrates the propagation of gravity waves (left) and capillary waves (right) in deep water. Clearly, long gravity waves travel faster at the outer front of the wave trains in the left panel while the shorter capillary waves travel faster and form the front in the right panel. Since the effect of surface tension on capillary waves guarantees a minimum group speed, the innermost waves seen in the right panel of Figure 5 (darkest ring) are the ones which travel with the minimum possible group velocity and form the inner boundary of the wave pattern. This leaves a flat interior region.

6 Particle paths

In this section, we investigate fluid particle paths for a wave traveling to the right with a constant phase speed $c_p(k)$. Assuming the rest position of the particle to be (x_0, z_0), and its instantaneous position to be $(x = x_0 + x_1, z = z_0 + z_1)$, the equations of motion for the

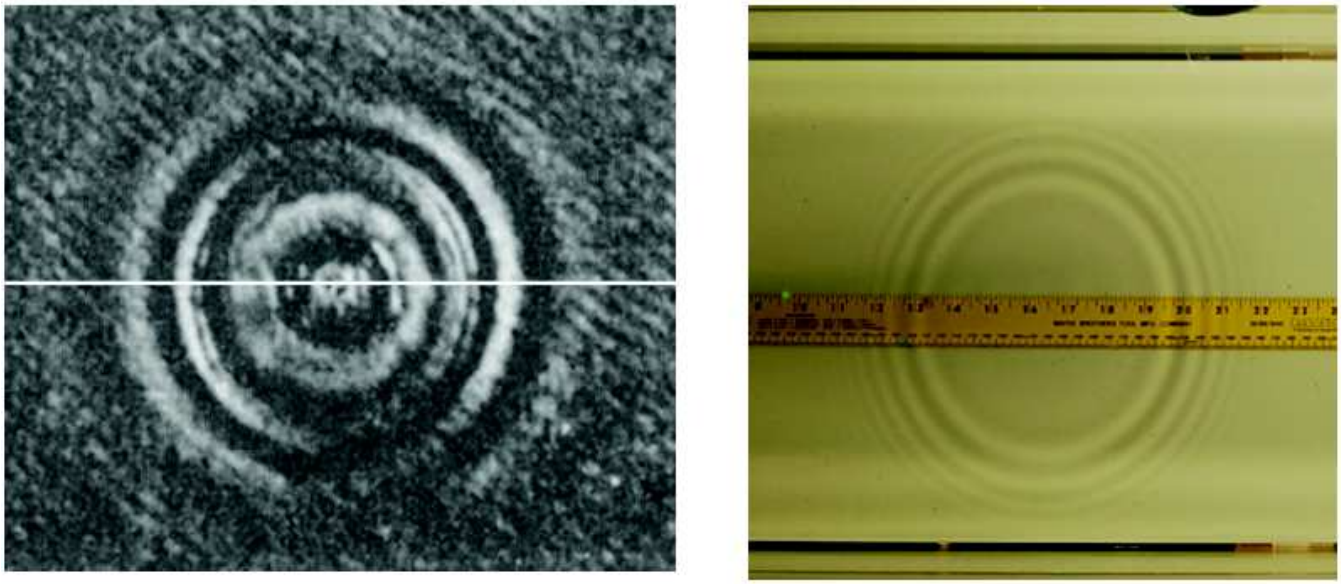


Figure 5: Propagation of surface gravity waves on the left (from *Water Waves* by Stoker [3]) and capillary waves on the right (Courtesy of Diane Henderson).

particle are:

$$\frac{\mathrm{D}x}{\mathrm{D}t} = \frac{\mathrm{D}x_1}{\mathrm{D}t} = u(x,z,t) = a\cos(kx - \omega(k)t)\frac{\cosh(k(z+h))}{\cosh(kh)}, \tag{35}$$

$$\frac{\mathrm{D}z}{\mathrm{D}t} = \frac{\mathrm{D}z_1}{\mathrm{D}t} = w(x,z,t) = a\sin(kx - \omega(k)t)\frac{\sinh(k(z+h))}{\cosh(kh)}. \tag{36}$$

Note that these two equations describe the motion of the fluid particle based on a velocity field obtained from linear theory. The approximate particle paths can be obtained using these ODEs by making some additional simplification. For small-amplitude waves, one can assume the particle displacement (x_1, z_1) away from its neutral position (x_0, z_0) to be small. The particle velocity is then the same as that of the neutral position to linear order so that

$$\frac{\mathrm{D}x_1}{\mathrm{D}t} = u(x,z,t) = a\cos(kx_0 - \omega(k)t)\frac{\cosh(k(z_0+h))}{\cosh(kh)}, \tag{37}$$

$$\frac{\mathrm{D}z_1}{\mathrm{D}t} = w(x,z,t) = a\sin(kx_0 - \omega(k)t)\frac{\sinh(k(z_0+h))}{\cosh(kh)}. \tag{38}$$

Integrating in time we get

$$x_1 = x(t) - x_0 \approx -\frac{a}{\omega}\sin(kx_0 - \omega(k)t)\frac{\cosh(k(z_0+h))}{\cosh(kh)}, \tag{39}$$

$$z_1 = z(t) - z_0 \approx \frac{a}{\omega}\cos(kx_0 - \omega(k)t)\frac{\sinh(k(z_0+h))}{\cosh(kh)}. \tag{40}$$

The above two relations can be combined to give

$$\frac{(x(t)-x_0)^2}{\cosh^2(k(z_0+h))}+\frac{(z(t)-z_0)^2}{\sinh^2(k(z_0+h))}=\frac{a^2}{\omega^2} \tag{41}$$

which indicates that the particle orbits are ellipses to the first order. It should be noted that for deep water waves the orbits are nearly circular as $\cosh(k(z_0+h)) \approx \sinh(k(z_0+h))$ for $kh \gg 1$. Meanwhile, for shallow water waves the orbits become thin ellipses with their major axes almost independent of the depth. Figure 6 shows the motion of marked fluid particles over one wave period. Orbits are nearly circular near the top of this flow, progressively becoming thin ellipses towards the bottom boundary. Stokes [4] showed that at second order, there is a slow drift near the free surface (Stokes drift). The drift is visible in Figure 6 by inspecting some of the larger orbits near the top.

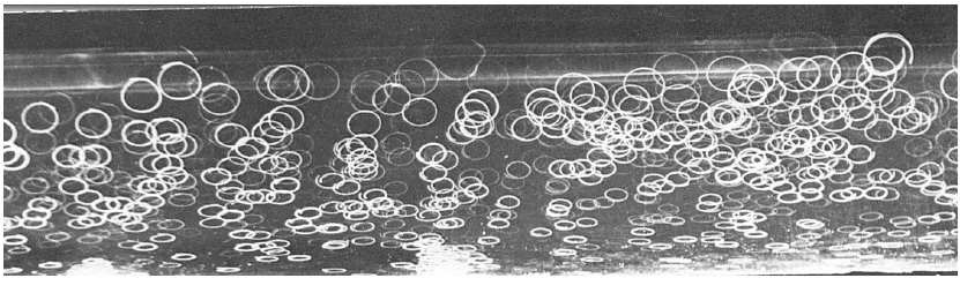


Figure 6: Motion of marked fluid particles over one wave period From [5].

References

[1] G. Crapper, *Introduction to water waves*, Ellis Horwood Ltd., (1984).

[2] L. W. T. Kelvin, Phil Mag., 4 (1887), pp. 252–255.

[3] J. Stoker, *Water waves: The mathematical theory with applications*, John Wiley & Sons, Inc., (1992).

[4] G. G. Stokes, *On the theory of oscillatory waves*, Transactions of the Cambridge Philosophical Society, 8 (1847), pp. 441–473.

[5] F. Wallet, A & Ruellan, Houille Blanche, 5 (1950), pp. 483–489.

Lecture 3: Introduction to Non-Linear Waves

Lecturer: Roger Grimshaw, Write-up: Alireza Mashayekhi

June 15, 2009

1 Introduction

The aim of this lecture is to introduce briefly the various kinds of nonlinear equations which have been proposed as models of water waves. These equations are presented here on physical grounds, and are derived more formally in following lectures. We begin by summarizing the properties of linear waves, and under which limits the linear approximations are thought to break down. Then, we investigate these limits and discuss basic nonlinear water wave equations.

1.1 Linear water waves

A one-dimensional linear wave can be represented by Fourier components

$$u = \Re\left\{A\exp(ikx - i\omega t)\right\}, \tag{1}$$

where k is the wavenumber, ω is the frequency, and A is the amplitude. Both ω and A may be functions of k. The linear wave dynamics are determined by the dispersion relation

$$\omega = \omega(k), \tag{2}$$

the form of which depends on the circumstances. In the instance of surface water waves where surface tension is negligible, we saw in Lecture 2 that the dispersion relation is

$$\omega^2 = gk\tanh(kh) \tag{3}$$

where g is gravity and h is the still water depth. Note that there are two branches of the dispersion relation $\pm\omega$, corresponding to leftward and rightward traveling waves respectively. Depending on the physical system considered there may be any number of branches of solution. For stable waves, ω is real for all real valued k.

We also saw that there are two important velocities when considering waves,

$$c_p = \frac{\omega}{k} \tag{4}$$

$$c_g = \frac{\mathrm{d}\omega}{\mathrm{d}k}, \tag{5}$$

where c_p is the phase velocity and c_g is the group velocity. For dispersive waves, these two velocities are not the same. The phase velocity is the speed at which the phase of the wave propagates (for instance, a wave crest). Meanwhile, the wave energy propagates at the group velocity. The energy, E, of each Fourier component is usually given by the form,

$$E = F(k)|A|^2. \tag{6}$$

As will be seen in subsequent lectures, in the case of water waves the energy is given by $E = g|A|^2/2$.

As time evolves in a linear dispersive system, each Fourier component propagates at its own phase velocity and thus, a group of waves of mixed k disperses. Meanwhile, the effect of non-linearities becomes important, typically leading to three possible scenarios depending on the waves considered:

1. long waves: for $k \to 0$, the dispersion relation is only weakly dispersive as $\omega = c_0 k + O(k^3)$ (see equation 3). This dispersion is comparable with the weakly nonlinear effects in modulating the amplitude of the wave.

2. wave packets: when the wave energy is concentrated about a finite wavenumber, k_0 say, dispersion is also weak and the wave group propagates with approximately a uniform group velocity. Again, the weak dispersion is comparable with the weak nonlinearity and modulates the amplitude of the wave group.

3. resonant wave interaction: due to non-linearities, two linear waves of wavenumber k_1 and k_2 may interact to form another wave $k_0 = k_1 + k_2$. In the instance where there is resonance with $\omega_0 \approx \omega_1 + \omega_2$, this can potentially be a strong effect in amplifying and/or modulating the waves. This third scenario is discussed in other lectures.

1.2 The Korteweg-de Vries Equation

The Korteweg-de Vries (KdV) equation [6] is used to consider the weakly nonlinear, weakly dispersive behavior of the long wave case discussed in the previous section, that is, when $k \to 0$. If we use a Taylor expansion about $kh = 0$ of the dispersion relation (3), and retain the first two non-trivial terms only, we get an approximate dispersion relation with error $O(k^5)$,

$$\omega = c_0 k - \beta k^3. \tag{7}$$

where c_0 is the limit of both phase and group velocity as $k \to 0$. Identifying $-i\omega$ with $\partial/\partial t$, and ik with $\partial/\partial x$ for each Fourier component, we deduce that the evolution equation for u is:

$$u_t + c_0 u_x + \beta u_{xxx} = 0. \tag{8}$$

In the long-wave approximation the dominant terms are the first two, showing that the wave nearly propagates with constant velocity c_0 except for the weak dispersion (the third term).

The original system being nonlinear, we expect that the dispersion term is balanced by another weak, but this time nonlinear term. The exact identification of the correct nonlinear term requires a formal study of the original equations, which will be done in the following lectures. For the moment, it is sufficient to say that a term of the form $\mu u u_x$ suitably balances the dispersive term, hence introducing

$$u_t + c_0 u_x + \mu u u_x + \beta u_{xxx} = 0, \tag{9}$$

which is the KdV equation. We can simplify the equation by performing the transform $x \to x - c_0 t$, which puts the observer in a reference frame moving with velocity c_0, in which case

$$u_t + \mu u u_x + \beta u_{xxx} = 0. \tag{10}$$

Equation (10) is an integrable equation, a fact established in the 1960's. The principle solution of the KdV equation are solitons. Solitons are solitary waves, that is, they are isolated, steadily propagating pulses given by,

$$u = a\mathrm{sech}^2(\gamma(x - Vt)), \tag{11}$$

where $V = \mu a/3 = 4\beta\gamma^2$ is the soliton velocity in the moving frame.

Solitons form a one-parameter family of solutions, parameterized for example by their amplitude a. The speed, V, is proportional to the amplitude and is positive when $\beta > 0$. Conversely, V is negative when $\beta < 0$. The soliton wavenumber γ is proportional to the square root of a. As such, large-amplitude waves are thinner and travel faster. Note that solitons are waves of elevation when $\mu\beta > 0$ and of depression when $\mu\beta < 0$.

A consequence of integrability means that the initial-value problem is solvable, with methods such as the Inverse Scattering Transform (IST) for a localized initial condition (see Lecture 5). The generic outcome the initial value problem is a finite number of solitons propagating in the positive x direction and some dispersing radiation propagating in the negative x direction (when $\mu\beta > 0$).

1.3 Nonlinear Schrödinger equation

To deal with the nonlinearity associated with the wave envelopes mentioned in section 1.1 we assume that the solution is a narrow-band wave packet, where the wave energy in Fourier space is concentrated around a dominant wavenumber k_0. The dispersion relation $\omega = \omega(k)$ can then be approximated for $k \approx k_0$ by

$$\omega - \omega_0 = c_{g0}(k - k_0) + \delta(k - k_0)^2, \tag{12}$$

where $\omega_0 = \omega(k_0)$, $c_{g0} = c_g(k_0)$ and $\delta = c_{gk}(k_0)/2$, and we recall that $c_g(k) = \mathrm{d}\omega/\mathrm{d}k$, so that $c_{gk} = \omega_{kk}$. This translates to an evolution equation for the wave amplitude

$$i(A_t + c_{g0}A_x) + \delta A_{xx} = 0, \tag{13}$$

where $u = \Re[A\exp(ikx - i\omega t)]$. Here it is assumed that the envelope function $A(x,t)$ is slowly-varying with respect to the carrier phase $kx - \omega t$. The dominant term is $A_t + c_{g0}A_x \approx$

0, showing that the wave envelope propagates with the group velocity c_{g0}, modified by the effect of weak dispersion due to the term A_{xx}. This equation is well-known in quantum mechanics as the Schrödinger equation. As for the KdV equation, the small dispersion effect introduced in (13) needs to be balanced by nonlinearity. In this case, the lowest possible nonlinearity has to be a cubic term of the form $\nu|A|^2A$, for some constant ν. This form can be roughly argued on the grounds that the associated phase for all the terms in (13) should be the same, $e^{ikx-i\omega t}$. By contrast, a quadratic term for example would have phase $e^{2ikx-2i\omega t}$ or no phase at all (e.g. as in the $|A|^2$ term). A proper derivation of the NLS is presented later in this lecture series.

Thus the model evolution equation for the wave envelope is the *nonlinear* Schrödinger equation (NLS), expressed here in the reference frame moving with speed c_{g0} (as before, transform $x \to x - c_{g0}t$),

$$iA_t + \nu|A|^2A + \delta A_{xx} = 0. \tag{14}$$

Like the KdV equation, (14) is a valid model for many physical systems, including notably water waves and nonlinear optics, a result first realized in the late 1960's. Remarkably, and again like the KdV equation, it is also an integrable equation through the IST, first established by Zakharov and collaborators in 1972 ([9]). Equation (14) also has soliton solutions, and the single soliton or solitary wave solution has the form

$$A(x,t) = a\ \mathrm{sech}(\gamma(x - Vt))\ \exp(iKx - i\Omega t), \tag{15}$$

$$\gamma^2 = \frac{\nu a^2}{2\delta},\ \ V = 2\delta K,\ \text{and}\ \Omega = \delta(K^2 - \gamma^2). \tag{16}$$

This solution forms a two-parameter family, the parameters being the amplitude, a, and the "chirp" wavenumber K; however, K amounts to a perturbation of the carrier wavenumber k to $k + K$, $|K| \ll |k|$, and so can be removed by a gauge transformation. Note that this soliton solution exists only when $\delta\nu > 0$ which is the so-called focusing case.

2 Higher space dimensions

2.1 The 2D dispersion relation

In two space dimensions the wavenumber becomes a vector $\mathbf{k} = (k, l)$ and the dispersion relation is then in the form of

$$\omega = \omega(\mathbf{k}) = \omega(k, l), \tag{17}$$

where the wave phase is $\mathbf{k}\cdot\mathbf{x} - \omega t = kx + ly - \omega t$. The phase velocity is the vector $\mathbf{c} = \omega\mathbf{k}/\boldsymbol{\kappa}^{\mathbf{2}}$, where $\kappa = |\mathbf{k}|$. The group velocity becomes the vector

$$\mathbf{c}_g = \nabla_{\mathbf{k}} \cdot \omega = \left(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial l}\right). \tag{18}$$

Note that in general, the group velocity and the phase velocity differ in both magnitude and direction.

For water waves the dispersion relation is

$$\omega(k,l) = g\kappa \tanh \kappa h. \tag{19}$$

This is an example of an isotropic medium, where the wave frequency depends only on the wavenumber magnitude, and not its direction. In this case the group velocity is parallel to the wavenumber $\mathbf{k}$, and hence parallel to the phase velocity, with a magnitude $c_g = \mathrm{d}\omega/\mathrm{d}\kappa$.

2.2 Kadomtsev-Petviashvili equation

The Kadomtsev-Petviashvili equation (KP equation [5]) is the two-dimensional extension of the KdV equation for isotropic systems. In the reference frame moving with the linear long-wave phase speed c_0 aligned with the x-direction, the KP equation is

$$(u_t + \mu u u_x + \beta u_{xxx}) + \frac{c_0}{2} u_{yy} = 0. \tag{20}$$

Equation (20) assumes that there is weak diffraction in the y-direction, that is $\partial/\partial y \ll \partial/\partial x$. The linear terms can be deduced from the linear dispersion relation $\omega = \omega(\kappa)$, $\kappa = (k^2 + l^2)^{1/2}$, where it is assumed that $l^2 \ll k^2$. In the long-wave limit we then have $\kappa \approx k + l^2/2k$, so that

$$\omega \approx c_0 k - \beta k^3 + \frac{c_0\, l^2}{2k} \ldots . \tag{21}$$

Identifying, as in Section 1.2, $-i\omega$ with $\partial/\partial t$, $ik \sim \partial/\partial x$, and $il \sim \partial/\partial y$, we see that (20) follows. When $c_0\beta > 0$, the system is referred to as the "KPII" equation, and it can be shown that the solitary wave solution (11) is stable to transverse disturbances. This is the case for water waves. On the other hand if $\beta c_0 < 0$, (20) is the "KPI" equation for which the 1D solitary wave is unstable; instead this equation supports fully 2D "lump" solitons. Like the KdV equation, both KPI and KPII are integrable equations.

2.3 Benney-Roskes equation

Finally, for systems with an isotropic dispersion relation, the NLS equation can also be extended to two dimensions. In the reference frame moving with the group velocity c_{g0} aligned with the x-direction

$$iA_t + \nu|A^2|A + \delta A_{xx} + \delta_1 A_{yy} + QA = 0. \tag{22}$$

where Q is a functional of the amplitude $A(x,t)$. This equation deserves a little discussion. The linear term in (22) can be found by expanding the dispersion relation as in the one-dimensional case (12), so that for $k \approx k_0 \approx l \approx 0$,

$$\omega - \omega_0 = c_{g0}(k - k_0) + \delta(k - k_0)^2 + \delta_1 l^2, \tag{23}$$

where, as before $\delta = \omega_{kk}(k_0, 0)/2 = c_{gk}(k_0, 0)/2$ and $\delta_1 = \omega_{ll}(k_0, 0) = c_{g0}/2k_0$.

The final term QA on the other hand arises from the 2D extension of the cubic term. The quantity Q depends on $|A|^2$ only, a quantity which does not oscillate, and is therefore

typically called the "wave-induced mean flow". The term QA is then interpreted as the effect of mean flows, generated by the nonlinear wave stresses, back on the wave-packet amplitude $A(x,t)$. The precise form of Q depends on the particular physical system being considered. For water waves, where ${c_0}^2 = gh$, it has been shown that

$$\left(1 - \frac{c_{g0}^2}{c_0^2}\right) Q_{xx} + Q_{yy} + \nu_1 |A|^2_{yy} = 0. \tag{24}$$

Note that if we set the $y-$derivatives to 0 in this equation then $Q_{xx} = 0$, and the Q term vanishes thus recovering the NLS equation.

The resulting system (22, 24) form the Benney-Roskes equations [1], also known as the Davey-Stewartson equations. Note that for water waves $\delta < 0$, $\delta_1 > 0$ and $c_{g0} < c_0$, so that (22) is hyperbolic, but (24) is elliptic.

References

[1] D. Benney and G. Roskes, *Wave instabilities*, Stud. Appl. Math, (1969).

[2] P. Drazin and J. R.S., *Solitons: an introduction*, Cambridge University Press, London, (1989).

[3] R. Grimshaw, *Nonlinear waves in uids: Recent advances and modern applications*, CISM Courses and Lectures,Springer, Wien New York, (2005), p. 196.

[4] ——, *Solitary waves in fluids*, Advances in Fluid Mechanics, WIT press, UK, 47 (2007), p. 288.

[5] B. Kadomtsev and Petviashvil, *On the stability of solitary waves in weakly dispersive media*, Sov. Phys. Dokl, 15 (1970), pp. 539–541.

[6] D. J. Korteweg and G. de Vries, *On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves*, Philosophical Magazine, 39 (1895), pp. 422–443.

[7] A. Scott, *Nonlinear science: Emergence and dynamics of coherent structures*, Oxford University Press Inc., New York., (1999).

[8] G. Whitham, *Linear and nonlinear waves*, Wiley, New York, (1974).

[9] V. E. Zakharov and A. B. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, Soviet Journal of Experimental and Theoretical Physics, 34 (1972), p. 62.

Lecture 4: Zhakharov Formulation of Water Waves

Lecturer: Harvey Segur. Write-up: Andrew Stewart

June 16, 2009

Hamiltonian mechanics is an alternative formulation of the laws of classical mechanics that was developed by William Rowan Hamilton in the early 19^{th}century. It is desirable to express mathematical systems in Hamiltonian form, as this brings us one step closer to determining whether they are completely integrable. We first present a brief overview of Hamiltonian systems, then demonstrate that the problem of inviscid surface water waves presented in the previous lectures is in fact Hamiltonian. We conclude by discussing the consequences of this result in terms of integrability.

1 Review of Hamiltonian systems

We begin with a brief introduction to Hamiltonian mechanics and the properties of Hamiltonian systems. The interested reader is referred to [1] for a more comprehensive discussion of this topic.

1.1 Formal definition

A system of $2N$ first-order ordinary differential equations is said to be Hamiltonian if there exist N pairs of coordinates in the phase space,

$$\{p_j(t), q_j(t)\}, \quad j = 1, 2, \ldots, N, \tag{1}$$

and a real-valued Hamiltonian function

$$H(\mathbf{p}(t), \mathbf{q}(t), t),$$

where $\mathbf{p}(t) = (p_1(t), \ldots, p_N(t))^T$ and $\mathbf{q}(t) = (q_1(t), \ldots, q_N(t))^T$, such that the original equations expressed in this coordinate system are:

$$\dot{q}_j = \frac{\mathrm{d}q_j}{\mathrm{d}t} = \frac{\partial H}{\partial p_j}, \quad \dot{p}_j = \frac{\mathrm{d}p_j}{\mathrm{d}t} = -\frac{\partial H}{\partial q_j}, \quad j = 1, \ldots, N. \tag{2}$$

While not all systems of $2N$ equations are Hamiltonian, many commonly known examples are. For example, consider the equation of a nonlinear oscillator,

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \omega^2\theta + \alpha\theta^3 = 0, \tag{3}$$

where $\omega, \alpha \in \mathbb{R}$. This equation describes for example the gravity-driven evolution of the angle θ of a pendulum with the downward vertical, for small angles $|\theta| \ll 1$. Indeed, for an undamped pendulum of unit length, Newton's second law states that

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = -g\sin\theta = -g\left(\theta - \tfrac{1}{6}\theta^3\right) + \mathrm{O}\left(\theta^5\right), \tag{4}$$

which reduces to (3) for small angles (i.e. neglecting terms of $\mathrm{O}\left(\theta^5\right)$) and setting $\omega^2 = g$ and $\alpha = -g/6$. Choosing $q(t) = \theta(t)$ and $p(t) = \dot{\theta}(t)$, we now express (3) as the following system of first-order ordinary differential equations,

$$\dot{p} = -\omega^2 q - \alpha q^3, \quad \dot{q} = p. \tag{5}$$

We find that an appropriate choice of H is

$$H\left(p(t), q(t), t\right) = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega^2 q^2 + \tfrac{1}{4}\alpha q^4, \tag{6}$$

and it is readily verified that the Hamiltonian equations (2) yield the equations of motion (5) upon evaluation of the partial derivatives of H.

1.2 Properties of Hamiltonian systems

The Hamiltonian is frequently, but not always, the total energy of the system. In the case of the nonlinear oscillator H is exactly the constant total energy E, which may be seen by multiplying (3) by $\dot{\theta}$ and integrating,

$$E = \frac{1}{2}\dot{\theta}^2 + \frac{g}{2}\theta^2 - \frac{g}{24}\theta^4 = H(\dot{\theta}, \theta, t).$$

An essential property of a Hamiltonian system is that the 'flow' of the coordinates (1) in the phase space is volume-preserving. Consider a system of M time-dependent variables $x_j(t)$, $j = 1, \ldots, M$ governed by M first-order ordinary differential equations,

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = v_j\left(\mathbf{x}, t\right), \quad j = 1, \ldots, M. \tag{7}$$

The vector $\mathbf{x}(t) = \left(x_1(t), \ldots, x_M(t)\right)^T$ may be thought of as the coordinates of a "fluid" particle in M-dimensional phase space, and $\mathbf{v}(\mathbf{x}, t) = \left(v_1(\mathbf{x}, t), \ldots, v_M(\mathbf{x}, t)\right)^T$ as the fluid velocity vector. The fluid is "incompressible", so volume is preserved if

$$\nabla \cdot \mathbf{v} = 0. \tag{8}$$

Now suppose that $M = 2N$ and that (7) is a Hamiltonian system satisfying (2) with $p_j = x_{2j-1}$ and $q_j = x_{2j}$ for $j = 1, \ldots, N$. Then

$$\begin{aligned}\nabla \cdot \mathbf{v} = \sum_{j=1}^{M} \frac{\partial v_j}{\partial x_j} = \sum_{j=1}^{M} \frac{\partial}{\partial x_j}\left(\frac{\mathrm{d}x_j}{\mathrm{d}t}\right) &= \sum_{j=1}^{N} \frac{\partial}{\partial p_j}\left(\frac{\mathrm{d}p_j}{\mathrm{d}t}\right) + \frac{\partial}{\partial q_j}\left(\frac{\mathrm{d}q_j}{\mathrm{d}t}\right) \\ &= \sum_{j=1}^{N} \frac{\partial}{\partial p_j}\left(-\frac{\partial H}{\partial q_j}\right) + \frac{\partial}{\partial q_j}\left(\frac{\partial H}{\partial p_j}\right) = 0.\end{aligned}$$

The final equality above requires that H has continuous 2^{nd} partial derivatives with respect to p_j and q_j. Thus every sufficiently continuous Hamiltonian system meets the "incompressibility" condition and thereby preserves volume. Note that (8) is a necessary, but not sufficient, condition for a system to be Hamiltonian. A more comprehensive discussion of this property may be found in [1], page 69.

1.3 Extension to continuous variables

We now extend Hamiltonian mechanics to systems that depend on continuous variables, such that the discrete $p_j(t)$ and $q_j(t)$ are replaced by functions $p(x,t)$ and $q(x,t)$. Consider the following nonlinear wave equation with periodic boundary conditions,

$$\theta_{tt} = c^2\theta_{xx} - \omega^2\theta - \alpha\theta^3, \tag{9}$$

where $\theta = \theta(x,t)$ and θ_t denotes partial differentiation with respect to t. We proceed as we did with the nonlinear oscillator, first seeking an energy equation. Multiplying (9) by θ_t and integrating with respect to x over the domain considered yields

$$\int \left\{\theta_t\theta_{tt} + \omega^2\theta\theta_t + \alpha\theta^3\theta_t + c^2\theta_x\theta_{xt} - c^2\left(\theta_t\theta_x\right)_x\right\}\mathrm{d}x = 0,$$

where the second derivative in x has been split to allow integration by parts. The final term in the integrand may be integrated exactly with respect to x, and disappears due to the periodic boundary conditions. The remaining terms then form an exact derivative with respect to t, which leads to the following energy equation,

$$E\left[\theta_t,\theta,t\right] = \int \left\{\tfrac{1}{2}(\theta_t)^2 + \tfrac{1}{2}c^2(\theta_x)^2 + \tfrac{1}{2}\omega^2\theta^2 + \tfrac{1}{4}\alpha\theta^4\right\}\mathrm{d}x, \quad \frac{\mathrm{d}E}{\mathrm{d}t} = 0,$$

where E is now a functional of θ_t, θ and t. Following the method of Section 1.1, we guess that a suitable Hamiltonian is $H = E$, and choose

$$p(x,t) = \theta_t(x,t), \quad q(x,t) = \theta(x,t).$$

The new functions p and q are called the *conjugate variables*. The Hamiltonian functional of the system is then

$$H\left[p,q,t\right] = \int \left\{\tfrac{1}{2}p^2 + \tfrac{1}{2}c^2(q_x)^2 + \tfrac{1}{2}\omega^2q^2 + \tfrac{1}{4}\alpha q^4\right\}\mathrm{d}x. \tag{10}$$

If (10) is indeed the correct Hamiltonian for this system, then we expect to recover the nonlinear wave equation (9) via a set of equations analogous to (2), namely

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\delta H}{\delta p}, \quad \frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\delta H}{\delta q}, \tag{11}$$

where $\delta/\delta p$ denotes a *variational derivative* with respect to p. For a small change δp of the function p in the functional H, the variational derivative $\delta H/\delta p$ is defined as the coefficient of δp in the leading-order contribution to the integrand in the following expression:

$$H\left[p+\delta p,q,t\right] - H\left[p,q,t\right] = \int \left\{\left(\frac{\delta H}{\delta p}\right)\delta p + \mathrm{O}\left((\delta p)^2\right)\right\}\mathrm{d}x.$$

In the case of the nonlinear wave equation (10) it is straightforward to show that

$$\frac{\delta H}{\delta p} = p = \frac{\mathrm{d}q}{\mathrm{d}t},$$

but $\delta H/\delta q$ requires more work. Taking a small variation δq, we find that the integrand of δH contains a derivative of δq,

$$H\left[p, q+\delta q, t\right] - H\left[p, q, t\right] = \int \left\{c^2 q_x \left(\delta q\right)_x + (\omega^2 q + \alpha q^3)\delta q + \mathrm{O}\left((\delta q)^2\right)\right\} \mathrm{d}x.$$

In order to express the integrand as a sum of terms proportional to powers of δq, we must integrate the first term by parts,

$$\int \left\{c^2 q_x (\delta q)_x\right\} \mathrm{d}x = \int \left\{c^2 \left(q_x \delta q\right)_x - c^2 q_{xx}\delta q\right\} \mathrm{d}x = -\int \left\{c^2 q_{xx}\delta q\right\} \mathrm{d}x,$$

where the final equality follows by noting that once integrated, the boundary term vanishes. Indeed, for the perturbed functions $p+\delta p$ and $q+\delta q$ to satisfy the boundary conditions of the original problem, it is sufficient to require that the perturbations (both δq and δp) are zero on the boundary. Hence,

$$\frac{\delta H}{\delta q} = -c^2 q_{xx} + \omega^2 q + \alpha q^3 = -\frac{\mathrm{d}p}{\mathrm{d}t},$$

and so H is a Hamiltonian for the nonlinear wave equation.

2 Water waves as a Hamiltonian system

We now turn our attention to the equations of inviscid, incompressible, irrotational water waves propagating on a free fluid surface. Many attempts were made to prove that this system is Hamiltonian, but it was Vladimir Zakharov who finally published the Hamiltonian structure in 1968[2]. The full details of the following can be found, albeit in a compacted form, in Zakharov's paper.

2.1 The inviscid water wave problem

Consider an incompressible, irrotational, inviscid fluid of constant density ρ with velocity $\mathbf{u}(x, y, z, t)$ and pressure $p(x, y, z, t)$. The fluid lies above a rigid, impermeable lower boundary $z = -h(x, y)$, and has a single-valued surface at $z = \eta(x, y, t)$ that is subject to surface tension (with coefficient σ). In all that follows, we restrict our study to a finite-size domain with periodic boundary conditions, although a similar proof can be derived for an infinite domain with ϕ and η going to 0 at $\pm\infty$. This set-up is illustrated in Figure 1. In the first lecture of this series we showed that this system is governed by the following set of

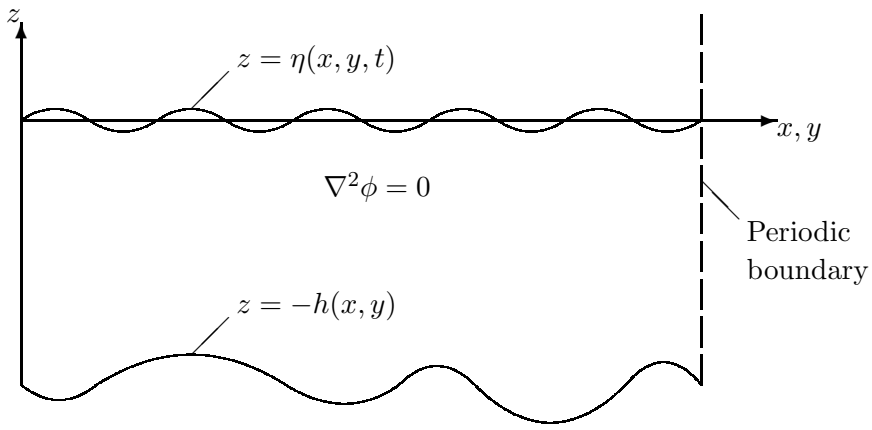


Figure 1: Capillary-gravity waves on the surface of an inviscid, incompressible, irrotational fluid of constant density.

equations and boundary conditions,

$$\frac{\partial \eta}{\partial t} + \nabla\phi \cdot \nabla\eta = \frac{\partial \phi}{\partial z} \qquad \text{on } z = \eta(x,y,t), \tag{12a}$$

$$\frac{\partial \phi}{\partial t} + \tfrac{1}{2}\left|\nabla\phi\right|^2 + g\eta = \frac{\sigma}{\rho}\nabla \cdot \left\{ \frac{\nabla\eta}{\sqrt{1+\left|\nabla\eta\right|^2}} \right\} \qquad \text{on } z = \eta(x,y,t), \tag{12b}$$

$$\nabla^2\phi = 0 \qquad -h(x,y) < z < \eta(x,y,t), \tag{12c}$$

$$\frac{\partial \phi}{\partial z} + \nabla\phi \cdot \nabla h = 0 \qquad \text{on } z = -h(x,y), \tag{12d}$$

where $\phi(x,y,z,t)$ is the velocity potential defined as $\nabla\phi = \mathbf{u}$. The Hamiltonian structure for this system is developed entirely in terms of the surface variables $\eta(x,y,t)$ and a new function ψ,

$$\psi(x,y,t) = \phi(x,y,\eta,t),$$

the velocity potential at the free surface. This approach is plausible because although the entire system is time-dependent, the time-derivatives only appear in the boundary conditions at the free surface. That is, if at a given time $t = t_0$, $\eta(x,y,t_0)$ and $\psi(x,y,t_0)$ are known, then the solution in the rest of the fluid domain is determined uniquely by the boundary conditions and Laplace's equation (12c). The functions η and ψ are therefore plausible conjugate variables for the Hamiltonian.

We obtain an energy equation for the inviscid water wave equations by multiplying (12c) by $\partial\phi/\partial t$ and integrating over $-h(x,y) \le z \le \eta(x,y,t)$,

$$\int_{-h}^{\eta} \phi_t \nabla^2\phi \, \mathrm{d}z = 0.$$

Integrating by parts, we apply Leibniz's rule and the boundary conditions (12a), (12b) and (12d) to obtain the following conservation law for the energy,

$$\frac{\partial E}{\partial t} + \nabla_{H} \cdot \mathbf{F} = 0, \tag{13}$$

where

$$E(x,y,t) = \int_{-h}^{\eta} \tfrac{1}{2}|\nabla\phi|^2 \mathrm{d}z + \tfrac{1}{2}g\eta^2 + \frac{\sigma}{\rho}\left(\sqrt{1+|\nabla\eta|^2}-1\right),$$
$$\mathbf{F}(x,y,t) = -\int_{-h}^{\eta} \phi_t \nabla_H \phi \,\mathrm{d}z - \frac{\sigma}{\rho}\frac{\eta_t \nabla\eta}{\sqrt{1+|\nabla\eta|^2}}, \tag{14}$$

are the energy flux and energy density respectively, and $\nabla_H = (\partial/\partial x, \partial/\partial y, 0)^T$ is the two-dimensional gradient vector. Note that this derivation takes a little work, and that an additional constant was added to E to set $E = 0$ if both ϕ and η are identically 0.

We obtain the expression for the total conserved energy of the system by integrating (13) over the horizontal domain $R \subset \mathbb{R}^2$,

$$\frac{\partial}{\partial t}\iint_R E \,\mathrm{d}x\,\mathrm{d}y + \int_{\partial R} \mathbf{F}\cdot\hat{\mathbf{n}}\,\mathrm{d}s = 0,$$

where we have applied Green's theorem to express the second term as an integral over the boundary ∂R. Recalling that the system has periodic boundary conditions, the integral over ∂R vanishes, showing that the total energy of the system is constant:

$$\iint_R E \,\mathrm{d}x\,\mathrm{d}y = \text{constant}. \tag{15}$$

2.2 Hamiltonian structure

We now show that the inviscid water wave problem is Hamiltonian, and that the correct Hamiltonian is simply the total energy given by (15). We define

$$H[\eta,\psi] = \iint_R \mathrm{d}x\,\mathrm{d}y \left\{\int_{-h}^{\eta} \tfrac{1}{2}|\nabla\phi|^2\mathrm{d}z + \tfrac{1}{2}g\eta^2 + \frac{\sigma}{\rho}\left(\sqrt{1+|\nabla\eta|^2}-1\right)\right\}, \tag{16}$$

and seek to show that the Hamilton equations,

$$\frac{\partial\eta}{\partial t} = \frac{\delta H}{\delta\psi}, \quad \frac{\partial\psi}{\partial t} = -\frac{\delta H}{\delta\eta}, \tag{17}$$

are equivalent to the boundary conditions at the free surface, (12a) and (12b).

As the Hamiltonian (16) is a functional of η and ψ, it is practical to express the boundary conditions at $z = \eta$ in terms of the surface variables. It is also desirable that their form should be similar to that of the Hamilton equations (17). We therefore rearrange (12a) and (12b) in terms of η, ψ and the component of the velocity normal to the free surface $\hat{\mathbf{n}}_\eta \cdot \nabla\phi$ (where $\hat{\mathbf{n}}_\eta$ is the vector normal to the $z = \eta$ surface). Let

$$F(x,y,z,t) = z - \eta(x,y,t),$$

such that $F(x,y,z,t) = 0$ defines the free surface. The gradient vector ∇F is then normal to the surface, so we may write the normal unit vector as

$$\hat{\mathbf{n}}_\eta = \frac{\nabla F}{|\nabla F|} = \frac{(-\eta_x, -\eta_y, 1)^T}{\sqrt{1+|\nabla\eta|^2}}.$$

The velocity normal to the surface is then

$$\nabla\phi \cdot \hat{\mathbf{n}}_\eta = \frac{-\nabla\phi \cdot \nabla\eta + \phi_z}{\sqrt{1+|\nabla\eta|^2}}, \tag{18}$$

Substituting (12a) in the above and rearranging leads to the following alternative form for the kinematic boundary condition,

$$\frac{\partial \eta}{\partial t} = \sqrt{1+|\nabla\eta|^2}\left[\hat{\mathbf{n}}_\eta \cdot \nabla\phi\right]_{z=\eta} \tag{19}$$

Having expressed for $\partial\eta/\partial t$ purely in terms of surface variables, we now seek a similar expression for the time derivative of ψ, our other conjugate variable. We first apply the chain rule for differentiation as follows,

$$\frac{\partial \psi}{\partial t} = \left.\frac{\partial \phi}{\partial t}\right|_{z=\eta} + \left.\frac{\partial \eta}{\partial t}\,\frac{\partial \phi}{\partial z}\right|_{z=\eta} = \left.\frac{\partial \phi}{\partial t}\right|_{z=\eta} + \left.\frac{\partial \phi}{\partial z}\right|_{z=\eta}\left[\frac{\partial \phi}{\partial z} - \nabla\phi\cdot\nabla\eta\right]_{z=\eta},$$

where the final equality follows from (12a). Substituting this into (12b) leads to the following expression for the dynamic boundary condition,

$$\begin{aligned}\frac{\partial \psi}{\partial t} + \frac{1}{2}&\left[\left(\frac{\partial \phi}{\partial x}\right)^2 + \left(\frac{\partial \phi}{\partial y}\right)^2 - \left(\frac{\partial \phi}{\partial z}\right)^2\right]_{z=\eta} \\ &+ \left[\frac{\partial \phi}{\partial z}\left(\nabla\phi\cdot\nabla\eta\right)\right]_{z=\eta} + g\eta - \frac{\sigma}{\rho}\nabla\cdot\left\{\frac{\nabla\eta}{\sqrt{1+|\nabla\eta|^2}}\right\} = 0.\end{aligned} \tag{20}$$

Although we are unable to write $\partial\psi/\partial t$ purely in terms of surface variables, we shall see that in fact we are still able to obtain this equation from the Hamiltonian.

Our task is now to show that the Hamilton equations (17) are equivalent to the rewritten kinematic and dynamic boundary conditions at the free surface, (19) and (20). Let us first consider variations of the Hamiltonian with respect to η. Remembering that ϕ implicitly depends on η, we find that

$$\begin{aligned}H\left[\eta+\delta\eta,\psi\right] - H\left[\eta,\psi\right] = \iint_R \mathrm{d}x\mathrm{d}y\Bigg\{&\int_\eta^{\eta+\delta\eta} \tfrac{1}{2}\left|\nabla(\phi+\delta\phi)\right|^2 \mathrm{d}z \\ &+ \int_{-h}^{\eta} \nabla\phi\cdot\nabla\delta\phi\,\mathrm{d}z + g\eta\,\delta\eta + \frac{\sigma}{\rho}\frac{\nabla\eta\cdot\nabla\delta\eta}{\sqrt{1+|\nabla\eta|^2}}\Bigg\} + \mathrm{O}\left((\delta\eta)^2\right).\end{aligned} \tag{21}$$

Here we have expanded the expressions in the integrand in the small variations $\delta\eta$ and $\delta\phi$, retaining only first-order terms. We may resolve the first term in the integrand of (21) by noting that, by the definition of integration,

$$\int_\eta^{\eta+\delta\eta} \tfrac{1}{2}\left|\nabla(\phi+\delta\phi)\right|^2 \mathrm{d}z = \left[\tfrac{1}{2}\left|\nabla(\phi+\delta\phi)\right|^2\right]_{z=\eta}\delta\eta + \mathrm{O}\left((\delta\eta)^2\right) = \left.\tfrac{1}{2}\left|\nabla\phi\right|^2\right|_{z=\eta}\delta\eta + \mathrm{O}\left((\delta\eta)^2\right)$$

for infinitesimally small variations $\delta\eta$. In the final term of (21) we may integrate by parts and apply Green's theorem to obtain the following,

$$\iint_R \mathrm{d}x\mathrm{d}y \int_{-h}^{\eta} \frac{\sigma}{\rho} \frac{\nabla\eta\cdot\nabla\delta\eta}{\sqrt{1+|\nabla\eta|^2}}\mathrm{d}z = \int_{\partial V} \frac{\sigma}{\rho}\frac{\delta\eta\nabla\eta}{\sqrt{1+|\nabla\eta|^2}}\cdot\hat{\mathbf{n}}\,\mathrm{d}s - \iiint_V \nabla\cdot\left\{\frac{\sigma}{\rho}\frac{\nabla\eta}{\sqrt{1+|\nabla\eta|^2}}\right\}\delta\eta\,\mathrm{d}V.$$

where V denotes the volume bounded by R in the horizontal direction, and by $-h$ and η in the vertical direction. Here the first term vanishes due to the periodic boundary conditions, leaving only the second term to contribute to δH.

The second term of (21) requires more work. We note first that

$$\iiint_V \nabla\phi\cdot\nabla\delta\phi\,\mathrm{d}V = \iiint_V \nabla\cdot(\delta\phi\nabla\phi)\,\mathrm{d}V$$

because $\nabla^2\phi = 0$ everywhere in the fluid from (12c). We then apply the divergence theorem to this integral to transform it into an integral over the fluid boundaries ∂V,

$$\iiint_V \nabla\cdot(\delta\phi\,\nabla\phi)\,\mathrm{d}V = \iint_{\partial V}\delta\phi\,\nabla\phi\cdot\hat{\mathbf{n}}\,\mathrm{d}s = \iint \delta\phi\,\left[\nabla\phi\cdot\hat{\mathbf{n}}_\eta\right]_{z=\eta}\mathrm{d}s_\eta. \tag{22}$$

where $\mathrm{d}s$ denotes an infinitesimal area of the fluid boundary ∂V, while $\mathrm{d}s_\eta$ denotes an infinitesimal area on the free surface only. The final equality follows because the component of the velocity normal to the surface vanishes at $z=-h$ (12d), and because the lateral boundary terms cancel out in this periodic domain, so that in fact we are left with an integral over the free surface $z=\eta$. Noting that we may expand variations in ϕ due to variations in η at the free surface as

$$\delta\phi = \left.\frac{\partial\phi}{\partial z}\right|_{z=\eta}\delta\eta \quad \text{on } z=\eta,$$

we transform the integral on the free surface to an integral over x and y only, using the following expression for an infinitesimal area,

$$\iint \mathrm{d}s_\eta = \iint_R \sqrt{1+|\nabla\eta|^2}\,\mathrm{d}x\mathrm{d}y. \tag{23}$$

and our expression for the normal velocity at the free surface (18). We thus obtain

$$\iint \delta\phi\,\left[\hat{\mathbf{n}}_\eta\cdot\nabla\phi\right]_{z=\eta}\mathrm{d}s_\eta = \iint_R \mathrm{d}x\mathrm{d}y\left[\frac{\partial\phi}{\partial z}\delta\eta\left(\nabla\phi\cdot\nabla\eta - \frac{\partial\phi}{\partial z}\right)\right]_{z=\eta}.$$

Collecting together the above results yields the following expression for the variation of the Hamiltonian with respect to η,

$$\delta H = \iint_R \mathrm{d}x\mathrm{d}y\left\{\left[\tfrac{1}{2}|\nabla\phi|^2 + \frac{\partial\phi}{\partial z}\left(\nabla\phi\cdot\nabla\eta - \frac{\partial\phi}{\partial z}\right)\right]_{z=\eta} + g\eta - \frac{\sigma}{\rho}\nabla\cdot\left[\frac{\nabla\eta}{\sqrt{1+|\nabla\eta|^2}}\right]\right\}\delta\eta. \tag{24}$$

Thus, the Hamilton equation

$$\frac{\partial \psi}{\partial t} = -\frac{\delta H}{\delta \eta}$$

is exactly equivalent to the dynamic boundary condition on the free surface (20).

Let us now consider variations of the Hamiltonian with respect to ψ. The form of (16) is such that only the first term in the integrand (the kinetic energy) depends on ψ,

$$H[\eta, \psi + \delta\psi] - H[\eta, \psi] = \delta \iiint_V \tfrac{1}{2}|\nabla\phi|^2 \mathrm{d}V.$$

However, the dependence of this term on ψ is not initially clear, so we proceed by noting that

$$\nabla\phi \cdot \nabla\phi = \nabla \cdot (\phi \nabla\phi),$$

because $\nabla^2\phi = 0$ everywhere in the fluid from (12c). We use this to rewrite our expression for the kinetic energy and apply the divergence theorem, which yields the following surface integral,

$$\iiint_V \tfrac{1}{2}|\nabla\phi|^2 \mathrm{d}z = \iint_{\partial V} \tfrac{1}{2}\phi \nabla\phi \cdot \hat{\mathbf{n}} \, \mathrm{d}s$$

where again ∂V is the surface bounding the fluid and $\mathrm{d}s$ is an infinitesimal area of that surface. By a similar argument as was used to derive equation (22), the surface integral vanishes on the horizontal and bottom boundaries, leaving only an integral over the free surface. Transforming back to an integral over the horizontal domain using (23) yields

$$\iint \tfrac{1}{2}\phi \nabla\phi \cdot \hat{\mathbf{n}} \, \mathrm{d}s_\eta = \iint_R \mathrm{d}x \, \mathrm{d}y \; \tfrac{1}{2}\psi \left[\nabla\phi \cdot \hat{\mathbf{n}}_\eta\right]_{z=\eta} \sqrt{1 + |\nabla\eta|^2}. \tag{25}$$

All that now remains is to relate $\left[\nabla\phi \cdot \hat{\mathbf{n}}_\eta\right]_{z=\eta}$ to ψ. Using a Dirichlet-to-Neumann map, it can be shown that there exists a symmetric Green's function of two variables $G(x, y; \mu, \nu)$ such that

$$\left[\nabla\phi \cdot \hat{\mathbf{n}}_\eta\right]_{z=\eta} = \iint \psi(\mu, \nu, t) G(x, y; \mu, \nu) \, \mathrm{d}s_\eta = \iint_R \mathrm{d}\mu \, \mathrm{d}\nu \; \psi(\mu, \nu, t) G(x, y; \mu, \nu) S(\mu, \nu, t), \tag{26}$$

where we have written $S(x, y, t) = \sqrt{1 + |\nabla\eta(x, y, t)|^2}$ for convenience. We may now find $\delta H / \delta\psi$ by substituting this into (25) as follows,

$$\begin{aligned}
&\delta \iint_R \mathrm{d}x \, \mathrm{d}y \; \tfrac{1}{2}\psi \left[\nabla\phi \cdot \hat{\mathbf{n}}_\eta\right]_{z=\eta} S(x, y, t) \\
&= \delta \iint_R \mathrm{d}x \, \mathrm{d}y \; \tfrac{1}{2}\psi(x, y, t) S(x, y, t) \iint_R \mathrm{d}\mu \, \mathrm{d}\nu \; \psi(\mu, \nu, t) G(x, y; \mu, \nu) S(\mu, \nu, t).
\end{aligned}$$

Taking variations with respect to ψ and neglecting terms of $\mathrm{O}\left((\delta\psi)^2\right)$ yields

$$\begin{aligned}
\tfrac{1}{2} \iint_R \mathrm{d}x \, \mathrm{d}y \; S(x, y, t) \iint_R \mathrm{d}\mu \, \mathrm{d}\nu \; & G(x, y; \mu, \nu) S(\mu, \nu, t) \\
& \times \left[\psi(\mu, \nu, t)\delta\psi(x, y, t) + \psi(x, y, t)\delta\psi(\mu, \nu, t)\right]
\end{aligned}$$

Finally, we use the property that the Green's function is symmetric, so we may rewrite this as,

$$\iint_R \mathrm{d}x\,\mathrm{d}y\ S(x,y,t)\delta\psi(x,y,t)\iint_R \mathrm{d}\mu\,\mathrm{d}\nu\ \psi(\mu,\nu,t)G(x,y;\mu,\nu)S(\mu,\nu,t)$$
$$= \iint_R \mathrm{d}x\,\mathrm{d}y\ \sqrt{1+|\nabla\eta|^2}\left[\nabla\phi\cdot\hat{\mathbf{n}}_\eta\right]_{z=\eta} S(x,y,t)\delta\psi,$$

where the second equality follows from (25). Thus,

$$\frac{\delta H}{\delta\psi} = \sqrt{1+|\nabla\eta|^2}\left[\nabla\phi\cdot\hat{\mathbf{n}}_\eta\right]_{z=\eta} = \frac{\partial\eta}{\partial t}, \tag{27}$$

as required, and so (16) is indeed the correct Hamiltonian for water waves.

3 Summary

In this lecture we have studied the foundations of Hamiltonian mechanics, and we have shown that the problem of inviscid, irrotational surface waves is Hamiltonian. It is desirable to formulate a system as a Hamiltonian because this immediately confers a number of useful properties. The fact that the "flow" in phase space is volume-preserving means that asymptotic stability of the system is impossible – only neutral stability is possible. It also means that we can not have attractors or repellers ("sources" and "sinks") in phase space. These properties make the system suitable for symplectic integration, a form of numerical integration that makes use of the preservation of volume in phase space.

An important property of a Hamiltonian formulation is that it allows us, in some cases, to identify systems of equations that are completely integrable, thus guaranteeing that the system can be solved explicitly. The following procedure is due to Liouville, and is described in more detail in [1]. We begin by defining the Poisson bracket $\{A,B\}$. For a Hamiltonian system of $2N$ ordinary differential equations (ODEs) with conjugate variables $(p_1,\ldots,p_N,q_1,\ldots,q_N)$, the Poisson bracket for any pair of functions on the phase space is

$$\{A,B\} = \sum_{j=1}^{N} \frac{\partial A}{\partial p_j}\frac{\partial B}{\partial q_j} - \frac{\partial A}{\partial q_j}\frac{\partial B}{\partial p_j}.$$

The Hamiltonian system of these 2N coupled ODEs is said to be *completely integrable* if (i) the Hamiltonian does not depend explicitly on time, and (ii) the equations admit N constants of the motion P_j (for $j=1,\ldots,N$) that are functionally independent of each other and that are "in involution", i.e.

$$\{P_j,P_k\} = 0 \ \text{ for any pair } j,k = 1,\ldots,N.$$

Suitable combinations of these constants of motion are called "action variables". Conjugate on them are N "angle variables", Q_j (for $j=1,\ldots,N$). The action and angle variables provide an alternative complete set of $2N$ coordinates on the phase space.

Note that while the Hamiltonian H might depend on all of the original variables $(p_1, \ldots, p_N, q_1, \ldots, q_N)$, it must now be independent of the angle variables since the action variables are constants of motion. Indeed,

$$\frac{\mathrm{d}P_j}{\mathrm{d}t} = -\frac{\partial H}{\partial Q_j} = 0, \text{ for } j = 1, \ldots, N,$$

so that

$$H = H(P_1, \ldots, P_N)$$

only, and $P_j(t)$ is constant. The other half of Hamilton's equations,

$$\frac{\mathrm{d}Q_j}{\mathrm{d}t} = \frac{\partial H}{\partial P_j}, \text{ for } j = 1, \ldots, N,$$

can be integrated trivially because $\omega_j = \partial H/\partial P_j$ is necessarily constant, so

$$Q_j(t) = \omega_j t + \phi_j, \text{ for } j = 1, \ldots, N.$$

The set of functions $P_j(t)$ and $Q_j(t)$ for $j = 1, \ldots, N$, as derived above, form the complete solution of the system for all time.

The geometry on the phase space is as follows. The N action variables define an N-dimensional manifold within the 2N-dimensional phase space. The trajectory of any solution must remain on this (time-independent) manifold for all time. If the manifold is compact, then one can show that it must be an N-dimensional torus, and that the solution must be either periodic or quasiperiodic in time. The N angle variables provide coordinates on the manifold (whether that manifold is a torus or not), and the solution consists of uniform translation along a straight line on the manifold. Thus, in terms of the action-angle variables, the solution of a completely integrable problem is very simple. However, note that the canonical transformation between the action-angle variables and the original variables $(p_1, \ldots, p_N, q_1, \ldots, q_N)$ can be complicated, so this inherently simple motion can appear quite complex when viewed in terms of the original variables.

How does this discussion of completely integrable Hamiltonian systems relate to the equations of surface water waves? As discussed above, these equations are Hamiltonian, but no one has shown that they are completely integrable (at least, not yet). However, in Lecture 5 we show that in some particular limit, the water wave problem is approximated by the Korteweg-de Vries (KdV) equation. Zakharov and Faddeev [3] took an important step in the development of soliton theory when they showed that the KdV equation can be viewed as a completely integrable, nonlinear Hamiltonian system with an infinite-dimensional phase space. In lecture 7, we discuss the Kadomtsev-Petviashvili (KP) equation, which is a natural generalization of the KdV equation; the KP equation is also completely integrable. In Lecture 13, we discuss the "three-wave equations", which approximate the water wave equations in another limit; these coupled equations are also completely integrable. In Lecture 14, we discuss the nonlinear Schrödinger equation, which approximates the water wave equations in yet another limit, and, once again, completely integrable (in one spatial dimension).

Completely integrable Hamiltonian systems are apparently rare: most Hamiltonian systems are not integrable. Before the discovery of soliton theory in the 1960s, no nontrivial

examples were known of nonlinear Hamiltonian partial differential equations that are completely integrable. Now we know of infinitely many examples of such equations, but very few of them are relevant to problems of physical interest. So it is remarkable that the water wave problem has so many approximations that are completely integrable. But we live in a remarkable world.

References

[1] V. Arnold, *Mathematical methods of classical mechanics*, Springer, 1989.

[2] V. Zakharov, *Stability of periodic waves of finite amplitude on the surface of a deep fluid*, Journal of Applied Mechanics and Technical Physics, 9 (1968), pp. 190–194.

[3] V. Zakharov and L. Fadeev, *Kortweg-de-vries equation: a complete hamiltonian system*, Funct. Annal. Pril., 5 (1971), p. 18.

Lecture 5: Waves in shallow water, part I: the theory

Lecturer: Harvey Segur. Write-up: Nicolas Grisouard

June 16, 2009

1 Introduction

We saw in Lecture 2 a linearized model of water waves, in which all waves (necessarily) have very small amplitude, and the longest waves (with wavenumbers near $k = 0$) propagate with essentially no dispersion. The next step towards a more realistic description of water waves is to introduce weakly dispersive and weakly nonlinear effects. The derivation of the Korteweg – de Vries (KdV) equation, which follows this approach, is the topic of the first part of this lecture. The properties of the KdV equation are presented in a second part, followed by a third part which discusses the accuracy of this equation for water waves in real-life settings and its relevance as an approximation of the water wave equations.

Let us recall the notations used so far:

- Cartesian coordinate system of variables (x, y, z), the origin of z being the mean altitude of the free-surface,
- position of a particle at a given time: $\mathbf{x} = (x(t), y(t), z(t))$
- position of the free-surface: $z = \eta(x, y, t)$,
- position of the bottom: $z = -h(x, y)$,
- fluid velocity in the interior of the fluid: $\mathbf{u}(x, y, z, t) = (u, v, w)$,
- density of water: ρ,
- surface tension: σ.

Let us also recall the approximations made so far, which still apply in this study:

- water is incompressible ($\nabla \cdot \mathbf{u} = 0$),
- uniform density ($\rho(x, y, z, t) = \rho$),
- no rotation,
- constant gravity vector pointing downwards,
- no viscous effects,
- irrotational flow ($\nabla \times \mathbf{u} = 0$), allowing us to define a velocity potential ϕ as $\mathbf{u} = \nabla\phi$.

The four equations of motion for the waves under these assumptions are:

$$\partial_t \eta + \nabla\phi \cdot \nabla\eta = \partial_z \phi \qquad \text{on } z = \eta(x,y,t), \tag{1}$$

$$\partial_t \phi + \frac{1}{2}\left|\nabla\phi\right|^2 + g\,\eta = \frac{\sigma}{\rho}\nabla \cdot \left\{ \frac{\nabla\eta}{\sqrt{1+\left|\nabla\eta\right|^2}} \right\} \qquad \text{on } z = \eta(x,y,t), \tag{2}$$

$$\nabla^2 \phi = 0 \qquad \text{for } -h(x,y) < z < \eta(x,y,t), \tag{3}$$

$$\partial_z \phi + \nabla\phi \cdot \nabla h = 0 \qquad \text{on } z = -h(x,y). \tag{4}$$

2 Derivation of the KdV equation

The following derivation of the KdV equation uses a general method to derive approximate models of waters waves, such as the Kadomtsev – Petviashvili (KP) equation or the Nonlinear Schrödinger (NLS) equation. It consists of four steps starting from what is considered as the exact set of equations of the problem. These four steps reflect the outline of this section:

1. to identify an appropriate limit of the problem. As mentioned earlier, we consider here 1D, long, weakly nonlinear waves.
2. to scale the equations in order to make these limits explicit,
3. to expand these equations in order to solve them approximately, order by order,
4. to introduce multiple scales as needed and solve the equations correctly.

Useful references for this section are Ablowitz & Segur [1] and Dauxois & Peyrard [7].

2.1 Limit of interest for KdV (or KP)

In this problem, a is the amplitude scale of η and L_x the characteristic scale along x. Five assumptions are made:

1. the propagation is along x only, implying that the partial derivatives along y are equal to zero. If it is only approximately true, then one can obtain the KP equation instead of KdV (see Lecture 3).
2. flat topography (h = constant),
3. shallow water (or long wave or weak dispersion) approximation: the horizontal scale of the waves is large compared with the mean water height. We then define δ such that $h = \delta L_x$ with $\delta \ll 1$,
4. weak nonlinearity: a is small compared to h but finite. We define ε such that $a = \varepsilon h$ with $\varepsilon \ll 1$,

5. weak nonlinear and weak dispersive effects balance: in the KdV case, it implies $\varepsilon = O(\delta^2)$.

The final point requires some clarification. The procedure outlined above in points 1 through 4 introduces two small parameters, ε and δ. Both are dimensionless and assumed to be small compared with unity, but little is know a priori on how they compare in size with each other. An "honest" procedure from here on would be to carry out a expansion of the original equations in both parameters ε and δ and deduce the correct scaling between the two numbers for which weakly nonlinear and weakly dispersive effects are comparable in size at the first non-linear order in the expansion. This method is unfortunately mathematically cumbersome.

For the sake of clarity, in these lectures, we use previous results which yielded the correct scalings, namely $\varepsilon = O(\delta^2)$. This assumption is revisited and discussed in Section 2.4. Hence, in the rest of the lecture, we set $\varepsilon = \delta^2$ and only retain ε as the scaling parameter.

2.2 Variable scaling

As announced at the beginning of this section, we introduce new non-dimensional variables (to be recognized by *) in order to make the previous assumptions explicit. The vertical coordinates is scaled by the total water height:

$$z = h z^*.$$

Following Section 2.1, we now rescale the horizontal variable with $L_x = h/\delta$ and the amplitude of the perturbation η with $a = \varepsilon h$, and use the identity $\varepsilon = \delta^2$:

$$x = L_x\, x^* = \frac{h\, x^*}{\sqrt{\varepsilon}}, \qquad \eta = a\, \eta^* = \varepsilon\, h\, \eta^*.$$

The linear phase speed of infinitely long waves $\sqrt{gh}$ can be used to define a characteristic speed. As the motion is driven by the free-surface and as the free-surface motion is very gentle, the characteristic fluid velocity is also very gentle compared to the phase speed of the linear wave. We choose the ratio of the two typical velocities to be of $O(\varepsilon)$, and carry this scaling over to the velocity potential:

$$u = u^* \varepsilon \sqrt{gh} \quad \Rightarrow \quad \phi = h\sqrt{\varepsilon\, g\, h}\, \phi^*.$$

Finally, a characteristic timescale is naturally defined as the characteristic period of the linear waves:

$$T = \frac{L_x}{\sqrt{gh}} = \sqrt{\frac{h}{g\varepsilon}} \quad \Rightarrow \quad t^* = \sqrt{\frac{g\varepsilon}{h}}\, t.$$

In the new coordinate system, the solutions of the problem (η^*, u^* and ϕ^* for example) are now functions of the re-scaled variables x^* and t^*, and of the parameter ε, e.g.

$$\eta^* = \eta^*(x^*, t^*; \varepsilon),\ \phi^* = \phi^*(x^*, z^*, t^*; \varepsilon),\ u^* = u^*(x^*, t^*; \varepsilon), \ldots$$

2.3 Solve equations approximately, order by order

2.3.1 Approximate velocity field

Let us go back to the original function ϕ for a moment and write ϕ as a convergent Taylor series near the bottom boundary (i.e. a series in $z+h$) :

$$\phi(x,z,t) = \sum_{n=0}^{\infty} \frac{(h+z)^n}{n!} \phi^{(n)}(x,t) \tag{5}$$

where we use the notation $\phi^{(n)}(x,t) = \partial^n \phi / \partial z^n |_{x,t,z=-h}$. If we take the Laplacian of the last equation and use (3), we find that:

$$\nabla^2 \phi = \sum_{n=0}^{\infty} \frac{(h+z)^n}{n!} \left[\frac{\partial^2 \phi^{(n)}}{\partial x^2} + \phi^{(n+2)} \right] = 0. \tag{6}$$

Each term of the sum being independent, in order to satisfy this equation, one has to make sure that for each n, and at each point (x,t),

$$\frac{\partial^2 \phi^{(n)}}{\partial x^2} + \phi^{(n+2)} = 0. \tag{7}$$

Equation (4) tells us that $\phi^{(1)} \equiv 0$ since $\partial\phi/\partial z = 0$ at $z = -h$ for a flat bottom boundary. As a consequence, it follows from equation (7) that for each n being odd, $\phi^{(n)}(x,t) \equiv 0$. Let's define $\phi_0(x,t) \equiv \phi^{(0)}$. Equation (7) can be used recursively starting from $n = 0$ to find all of the functions $\phi^{(n)}(x,t)$ for n even:

$$\phi^{(2m)}(x,t) = (-1)^m \frac{\partial^{2m} \phi_0}{\partial x^{2m}}.$$

We can now rewrite ϕ as:

$$\phi(x,z,t) = \phi_0(x,t) + \sum_{m=1}^{\infty} \frac{(-1)^m}{(2m)!} (h+z)^{2m} \frac{\partial^{2m} \phi_0}{\partial x^{2m}}$$

$$\text{or as} \qquad \phi = h\sqrt{\varepsilon\, g\, h} \left[\phi_0^*(x^*,t^*;\varepsilon) + \sum_{m=1}^{\infty} \frac{(-\varepsilon)^m}{(2m)!} (1+z^*)^{2m} \frac{\partial^{2m} \phi_0^*}{\partial (x^*)^{2m}} \right] \tag{8}$$

in terms of the rescaled variables. We have not made any extra approximation so far, meaning that equation (8) provides an exact form of ϕ^* (provided the series converges).

If we differentiate equation (8) along the two directions and only retain the terms of the first two orders, we find:

$$u(x,z,t) = \varepsilon \sqrt{g\, h} \left[\hat{u}^*(x^*,t^*;\varepsilon) - \frac{\varepsilon}{2} (1+z^*)^2 \frac{\partial^2 \hat{u}^*}{\partial (x^*)^2} + O(\varepsilon^2) \right] \tag{9}$$

where we defined $\hat{u}^* = \partial \phi_0^* / \partial x^*$,

$$w(x,z,t) = \sqrt{\varepsilon\, g\, h}\, w^* = -\varepsilon \sqrt{\varepsilon\, g\, h} \left[(1+z^*) \frac{\partial \hat{u}^*}{\partial x^*} - \frac{\varepsilon}{6} (1+z^*)^3 \frac{\partial^3 \hat{u}^*}{\partial (x^*)^3} + O(\varepsilon^2) \right] \tag{10}$$

2.3.2 At the free surface

We now focus on equations (1) and (2). They both involve η and ϕ, or in other words $\eta^*(x^*,t^*;\varepsilon)$ and $\phi^*(x^*,z^*,t^*;\varepsilon)$. The other variable to be solved for is $\hat{u}^*(x^*,t^*;\varepsilon)$. We only have to solve these equations order-by-order and for that purpose, we expand η^* and $\hat{u}^*$ in power series in ε:

$$\eta^*(x^*,t^*;\varepsilon) = \eta_1(x^*,t^*;\varepsilon) + \varepsilon\,\eta_2(x^*,t^*;\varepsilon) + O(\varepsilon^2) \tag{11}$$

$$\text{and} \qquad \hat{u}^*(x^*,t^*;\varepsilon) = u_1(x^*,t^*;\varepsilon) + \varepsilon\,u_2(x^*,t^*;\varepsilon) + O(\varepsilon^2) \tag{12}$$

Equation (1), in non-dimensional form, is:

$$\varepsilon\frac{\partial\eta^*}{\partial t^*} + \varepsilon^2 u^*\frac{\partial\eta^*}{\partial x^*} = w^* \quad \text{on } z^* = \varepsilon\eta^*. \tag{13}$$

Using equations (10), (11) and (12) to expand (13) in a power series of ε until $O(\varepsilon^2)$, we find:

$$\varepsilon\frac{\partial\eta_1}{\partial t^*} + \varepsilon^2\frac{\partial\eta_2}{\partial t^*} + \varepsilon^2 u_1\frac{\partial\eta_1}{\partial x^*} = -\varepsilon\frac{\partial u_1}{\partial x^*} - \varepsilon^2\frac{\partial u_2}{\partial x^*} - \varepsilon^2\,\eta_1\frac{\partial u_1}{\partial x^*} + \frac{\varepsilon^2}{6}\frac{\partial^3 u_1}{\partial(x^*)^3}. \tag{14}$$

Note that we have kept the multiplicative ε factor everywhere to emphasize the order to which the equation is valid.

We also have to expand the dynamical equation of the free-surface (2). In non-dimensional variables, this equation reads:

$$\varepsilon\frac{\partial\phi^*}{\partial t^*} + \frac{\varepsilon}{2}\left(\varepsilon\,(u^*)^2 + (w^*)^2\right) + \varepsilon\,\eta^* = \varepsilon^2\frac{\sigma}{\rho g h^2}\frac{\partial}{\partial x^*}\left\{\frac{\frac{\partial\eta^*}{\partial x^*}}{\sqrt{1+\varepsilon^3\left|\frac{\partial\eta^*}{\partial x^*}\right|^2}}\right\} \quad \text{on } z^* = \varepsilon\eta^* \tag{15}$$

If we differentiate the latter along x^* to eliminate ϕ^*, we get

$$\varepsilon\frac{\partial u^*}{\partial t^*} + \frac{\varepsilon^2}{2}\frac{\partial\,(u^*)^2}{\partial x^*} + \frac{\varepsilon}{2}\frac{\partial\,(w^*)^2}{\partial x^*} + \varepsilon\,\frac{\partial\eta^*}{\partial x^*} = \varepsilon^2\frac{\sigma}{\rho g h^2}\frac{\partial^2}{\partial(x^*)^2}\left\{\frac{\frac{\partial\eta^*}{\partial x^*}}{\sqrt{1+\varepsilon^3\left|\frac{\partial\eta^*}{\partial x^*}\right|^2}}\right\} \quad \text{on } z^* = \varepsilon\eta^* \tag{16}$$

And finally, using the Taylor expansions (9) and (12), an approximation of this equation up to $O(\varepsilon^2)$ reads:

$$\varepsilon\frac{\partial u_1}{\partial t^*} - \frac{\varepsilon^2}{2}\frac{\partial^3 u_1}{\partial t^*\,\partial(x^*)^2} + \varepsilon^2\frac{\partial u_2}{\partial t^*} + \frac{\varepsilon^2}{2}\frac{\partial\left(u_1^2\right)}{\partial x^*} + \varepsilon\,\frac{\partial\eta_1}{\partial x^*} + \varepsilon^2\,\frac{\partial\eta_2}{\partial x^*} = \varepsilon^2\frac{\sigma}{\rho g h^2}\frac{\partial^3\eta_1}{\partial(x^*)^3}.\varepsilon\eta^*(x^*,t^*). \tag{17}$$

Note that in this last equation, because of the Taylor-expansion performed, u_i and η_i do not depend on z.

2.3.3 Free-surface equations, lowest order

If we reduce equations (14) and (17) to their $O(\varepsilon)$ terms, we get

$$\frac{\partial \eta_1}{\partial t^*} + \frac{\partial u_1}{\partial x^*} = 0 \qquad \text{and} \qquad \frac{\partial u_1}{\partial t^*} + \frac{\partial \eta_1}{\partial x^*} = 0. \tag{18}$$

At this order, these two equations can be written in the form of two d'Alembert wave equations and the solutions can be decomposed into leftward and rightward components, namely:

$$\eta_1 = f(x^* - t^*) + F(x^* + t^*), \tag{19}$$

$$u_1 = f(x^* - t^*) - F(x^* + t^*). \tag{20}$$

F and f are the same for both u_1 and η_1 as equations (18) are strongly coupled. This expression shows that, to $O(\varepsilon)$, the system recovers the well-known behavior of linear non-dispersive water waves, and there is no evidence at this order that either η_1 or u_1 depends on ε. In the following section, we assume η_1 and u_1 are indeed independent of ε, and see where this approximation takes us.

2.3.4 Free-surface equations, next order

Now balancing the terms in equations (14) and (17) that are $O(\varepsilon^2)$, we get a more complex behavior:

$$\frac{\partial \eta_2}{\partial t^*} + \frac{\partial u_2}{\partial x^*} = \frac{1}{6}\frac{\partial^3 u_1}{\partial (x^*)^3} - \frac{\partial (u_1 \eta_1)}{\partial x^*}, \tag{21}$$

$$\text{and} \qquad \frac{\partial u_2}{\partial t^*} + \frac{\partial \eta_2}{\partial x^*} = \left(\frac{\sigma}{\rho g h^2} - \frac{1}{2}\right)\frac{\partial^3 \eta_1}{\partial (x^*)^3} - \frac{1}{2}\frac{\partial u_1^2}{\partial x^*}. \tag{22}$$

where (18) was used to replace $\partial u_1/\partial t^*$ by $-\partial \eta_1/\partial x^*$.

The fact that u_1 and η_1 can be written as a combination of rightward- and leftward-traveling functions indicates that we may make the following change of variables:

$$\begin{aligned} (x^*, t^*) \rightarrow \quad & (r, s) = (x^* - t^*, x^* + t^*), \\ \frac{\partial}{\partial x^*} \rightarrow \quad & \frac{\partial}{\partial r} + \frac{\partial}{\partial s}, \\ \frac{\partial}{\partial t^*} \rightarrow \quad & -\frac{\partial}{\partial r} + \frac{\partial}{\partial s}. \end{aligned}$$

so that

$$\eta_1(x^*, t^*) = f(r) + F(s) \tag{23}$$

$$u_1(x^*, t^*) = f(r) - F(s). \tag{24}$$

If we rewrite equations (21) and (22) according to the new variables r and s, we get:

$$-\frac{\partial \eta_2}{\partial r} + \frac{\partial \eta_2}{\partial s} + \frac{\partial u_2}{\partial r} + \frac{\partial u_2}{\partial s} = \frac{1}{6}\left(\frac{\mathrm{d}^3 f}{\mathrm{d}r^3} - \frac{\mathrm{d}^3 F}{\mathrm{d}s^3}\right) - \frac{\mathrm{d}f^2}{\mathrm{d}r} + \frac{\mathrm{d}F^2}{\mathrm{d}s}, \tag{25}$$

$$-\frac{\partial u_2}{\partial r} + \frac{\partial u_2}{\partial s} + \frac{\partial \eta_2}{\partial r} + \frac{\partial \eta_2}{\partial s} = \left(\frac{\sigma}{\rho g h^2} - \frac{1}{2}\right)\left(\frac{\mathrm{d}^3 f}{\mathrm{d}r^3} + \frac{\mathrm{d}^3 F}{\mathrm{d}s^3}\right) - \frac{1}{2}\frac{\mathrm{d}f^2}{\mathrm{d}r} - \frac{1}{2}\frac{\mathrm{d}F^2}{\mathrm{d}s} + f\frac{\mathrm{d}F}{\mathrm{d}s} + F\frac{\mathrm{d}f}{\mathrm{d}r}. \tag{26}$$

Addition and subtraction of these two equations give:

$$2\frac{\partial(\eta_2+u_2)}{\partial s} = \left(\frac{\sigma}{\rho g h^2}-\frac{1}{3}\right)\underbrace{\frac{\mathrm{d}^3 f}{\mathrm{d}r^3}} + \left(\frac{\sigma}{\rho g h^2}-\frac{2}{3}\right)\frac{\mathrm{d}^3 F}{\mathrm{d}s^3} - \frac{3}{2}\underbrace{\frac{\mathrm{d}f^2}{\mathrm{d}r}} + \frac{1}{2}\frac{\mathrm{d}F^2}{\mathrm{d}s} + f\frac{\mathrm{d}F}{\mathrm{d}s} + F\frac{\mathrm{d}f}{\mathrm{d}r}, \tag{27}$$

$$2\frac{\partial(\eta_2-u_2)}{\partial r} = \left(\frac{\sigma}{\rho g h^2}-\frac{2}{3}\right)\frac{\mathrm{d}^3 f}{\mathrm{d}r^3} + \left(\frac{\sigma}{\rho g h^2}-\frac{1}{3}\right)\underbrace{\frac{\mathrm{d}^3 F}{\mathrm{d}s^3}} + \frac{1}{2}\frac{\mathrm{d}f^2}{\mathrm{d}r} - \frac{3}{2}\underbrace{\frac{\mathrm{d}F^2}{\mathrm{d}s}} + f\frac{\mathrm{d}F}{\mathrm{d}s} + F\frac{\mathrm{d}f}{\mathrm{d}r}. \tag{28}$$

When integrated along s and r respectively, the underbraced terms in equations (27) and (28) are multiplied by s and r respectively, which means that they diverge when $t^* \to \infty$. To be more specific, as we are talking about terms of $O(\varepsilon^2)$ compared to terms of $O(\varepsilon)$, the power series in ε no longer converges when $\varepsilon t^* \sim 1$. The other terms remain bounded as $t^* \to \infty$ and therefore do not cause any trouble.

One way to eliminate this problem would be to make sure that the underbraced terms cancel each other, leading to differential equations that do not include temporal derivative. The consequence would be that the condition should be satisfied also for the initial condition, which would be too restrictive.

2.4 Introducing the slow time-scale

In order to eliminate this problem, we need to go back to Section 2.3.3 and re-examine the assumption that u_1 and η_1 do not depend on ε. Note that if u_1 and η_1 depend on another, slowly-varying variable in addition to x^* and t^*, then derivatives with respect to that slow variable would not appear at $O(\varepsilon)$ in equation (17). We therefore introduce a slow time-scale τ, of the kind

$$\tau = \varepsilon^\alpha t^*.$$

This new variable effectively captures the possible ε dependence of η^* and $\hat{u}$ discussed in Section 2.2, and carries over to the coefficients in their respective Taylor expansions as

$$\eta_j(x^*,t^*;\varepsilon) \to \eta_j(x^*,t^*;\tau) \qquad \text{and} \qquad u_j(x^*,t^*;\varepsilon) \to u_j(x^*,t^*;\tau)$$

For example, the first-order functions u_1 and η_1 are now expressed as

$$\eta_1(r,s;\tau) = f(r,\tau) + F(s,\tau) \tag{29}$$

$$u_1(r,s;\tau) = f(r,\tau) - F(s,\tau). \tag{30}$$

Finally, any temporal derivative $\partial/\partial t^*$ then becomes

$$\frac{\partial}{\partial t^*} \to \frac{\partial}{\partial t^*} + \varepsilon^\alpha \frac{\partial}{\partial \tau}.$$

All that remains to be done is to select the relevant value for α. Going back to equations (14) and (17), note that each $\partial/\partial t^*$ term results in the addition of an extra "slow" temporal derivative that is ε^α-times smaller than the "fast" derivative. Taking the simplest possible case of $\alpha = 1$, the procedure adds one extra $O(\varepsilon^2)$ term in each equation. In particular, this results in the addition of respectively $-\partial_\tau \eta_1$ and $-\partial_\tau u_1$ to the right hand sides

of equations (21) and (22). Tracking these changes further down, equations (25) and (26) acquire extra $-\partial_\tau(f+F)$ and $-\partial_\tau(f-F)$ terms in their RHS respectively, which finally leads to this alternative set of equations for u_2 and η_2:

$$2\frac{\partial(\eta_2+u_2)}{\partial s} = \left(\frac{\sigma}{\rho g h^2}-\frac{1}{3}\right)\frac{\partial^3 f}{\partial r^3} - \frac{3}{2}\frac{\partial f^2}{\partial r} - 2\frac{\partial f}{\partial \tau} \quad + \mathcal{F}(r,s,\tau), \tag{31}$$

$$2\frac{\partial(\eta_2-u_2)}{\partial r} = \left(\frac{\sigma}{\rho g h^2}-\frac{1}{3}\right)\frac{\partial^3 F}{\partial s^3} - \frac{3}{2}\frac{\partial F^2}{\partial s} + 2\frac{\partial F}{\partial \tau} \quad + \mathcal{G}(r,s,\tau), \tag{32}$$

$\mathcal{F}$ and $\mathcal{G}$ being the sum of all terms in equations (27) and (28) respectively that do not cause any unphysical growth. It is now possible to eliminate all of the other terms in the RHS of each equations by requiring:

$$2\frac{\partial f}{\partial \tau} + 3\,f\frac{\partial f}{\partial r} + \left(\frac{1}{3}-\frac{\sigma}{\rho g h^2}\right)\frac{\partial^3 f}{\partial r^3} = 0, \tag{33}$$

$$\text{and} \qquad -2\frac{\partial F}{\partial \tau} + 3\,F\frac{\partial F}{\partial s} + \left(\frac{1}{3}-\frac{\sigma}{\rho g h^2}\right)\frac{\partial^3 F}{\partial s^3} = 0. \tag{34}$$

These equations are two KdV equations, each one being related to either a rightward or leftward propagating wave. It is interesting to notice that both are written in the reference frame that is moving at the linear long-wave velocity (and it has been the case since the definition of s and r).

The KdV equation can be physically interpreted in the following sense. If we ignore the dispersive term in the KdV equation (*i.e.* the third derivative term), it simply reduces to the inviscid Burgers equation, which is known to generate shock waves as the nonlinear term tends to steepen the wavefronts. In the full KdV, this effect is balanced by the dispersive term, which by contrast tends to spread the horizontal extent of the fronts. As we see from the equation, these two counteracting effects act on the slow timescale only, emphasizing how important the introduction of this extra timescale was. The balance between nonlinear and dispersive effects has been ensured in Section 2.1 by the fact that $\varepsilon = O(\delta^2)$, which might a priori seem arbitrary. It is in fact a very stable feature of the system dynamics, in the sense that this balance is often eventually reached. If it is not the case, the system simply does not evolve into solitons. As a matter of fact, this balance is a property of solitons much more general than the KdV model.

3 Properties of the KdV equation

3.1 What did Korteweg and de Vries know? First properties.

It is possible to rescale equation (33) to cast it in its canonical form:

$$\frac{\partial \nu}{\partial \tau} + 6\nu\frac{\partial \nu}{\partial \xi} + \frac{\partial^3 \nu}{\partial \xi^3} = 0. \tag{35}$$

This equation supports, among others, the following spatially localized type of solutions:

$$\nu(\xi,\tau) = 2p^2 \mathrm{sech}^2\left[p(\xi - 4p^2\tau + \xi_0)\right], \tag{36}$$

with the hyperbolic secant defined as sech $= 1/\cosh$ and $2p^2$ the amplitude of the wave. First note the wave has a positive propagation speed $4p^2$ *in the moving reference frame* and that this speed is proportional to the amplitude of the wave. Going back to the fixed reference frame of the laboratory, it means that such waves propagate faster than the linear long waves, especially when they are of high amplitude. Such waves are called *solitons*.

When looking for non-localized solutions, it is still possible to find solutions of permanent form by introducing the Jacobi elliptical functions or cnoidal functions, which give the following set of solutions :

$$\nu(\xi,\tau) = 2p^2 \mathrm{cn}^2\left[p(\xi - 4p^2\tau + \xi_0);\kappa\right] + \nu_0. \tag{37}$$

When $\kappa = 0$, cn $=$ cos and when $|\kappa| \to 1$, cn $\to$ sech. In other words, when κ varies from 0 to 1, the waves transform from regular sinusoidal waves to waves with flatter troughs and sharper crests. This is in fact what happens for example as waves approach the coast, when the nonlinearity increases, as can be seen in Figure 1.

Figure 1: Waves approaching the coast of Callao (Peru). It is possible to see that as the waves approach, their crests sharpen whereas their troughs flatten (Google Earth, around $12°01'$ S, $77°09'$ W).

De Vries's PhD thesis ([8]) was published in 1894 and Korteweg and de Vries's article in 1895 ([13]). After having introduced their equation, having determined its periodic solutions using elliptical functions and having proved that solitons can arise out of long waves, Korteweg and de Vries stopped their work on this subject.

Interest in these nonlinear waves rose again after 1953 when Fermi, Pasta and Ulam ran one of the first numerical experiments in history, trying to know how a crystal evolves towards thermal equilibrium. In numerical experiments, they injected energy into the longest Fourier mode that fit in their periodic domain, and expected that this energy would be redistributed among all modes so that their system would tend to thermal equilibrium. Instead, they found that the energy spread into the first few modes of the system, but then returned almost entirely into the initial mode. This happened over and over again, and is now known as "Fermi-Pasta-Ulam recurrence". This work was published in a Los Alamos technical report in 1955 ([9]) and it took 10 years until Zabusky & Kruskal (1965) ([19]) explained this phenomenon, introducing the concept (and the name) of *soliton* and reviving the KdV equation. Other names would further add their contribution to that (now) growing theory: Gardner, Miura, Greene... Let us bring up two of the most major properties associated with the KdV equation.

3.2 Infinite number of conservation laws

Here we only consider localized solutions of KdV, with ν infinitely differentiable and rapidly decreasing towards 0 as $|\xi| \to \infty$. A conservation law is a relation of the form

$$\frac{\partial(\text{density})}{\partial\tau} + \frac{\partial(\text{flux})}{\partial\xi} = 0.$$

Miura (1968) ([14, 15]) proved that KdV has an infinite number of conservation laws, such as:

$$\frac{\partial\nu}{\partial\tau} + \frac{\partial}{\partial\xi}\left(3\nu^2 + \frac{\partial^2\nu}{\partial\xi^2}\right) = 0,$$

$$\frac{\partial\nu^2}{\partial\tau} + \frac{\partial}{\partial\xi}\left(4\nu^3 + 2\nu\frac{\partial^2\nu}{\partial\xi^2} - \frac{1}{2}\left(\frac{\partial\nu}{\partial\xi}\right)^2\right) = 0,$$

$$\frac{\partial}{\partial\tau}\left(\nu^3 - \frac{1}{2}\left(\frac{\partial\nu}{\partial\xi}\right)^2\right) + \frac{\partial}{\partial\xi}\left(\frac{9}{2}\nu^4 + 3\nu^2\frac{\partial^2\nu}{\partial\xi^2} - 6\nu\left(\frac{\partial\nu}{\partial\xi}\right)^2 - \frac{\partial v}{\partial\xi}\frac{\partial^3 v}{\partial\xi^3} + \frac{1}{2}\left(\frac{\partial^2 v}{\partial\xi^2}\right)^2\right) = 0,$$

$$\vdots$$

This infinite number of conservation laws makes any solution of KdV *very* constrained.

3.3 Inverse scattering transform

Useful references for this section are again Ablowitz & Segur [1] for the derivation and Dauxois & Peyrard [7] for the similarities with quantum mechanics.

Let us now take an apparently unrelated detour, and introduce the time-invariant Schrödinger equation:

$$\partial_{\xi\xi}\psi + (\lambda + \nu(\xi,\tau))\psi = 0, \tag{38}$$

where the function $\nu(\xi,\tau)$ is assumed to be infinitely differentiable and rapidly decreasing towards 0 as $|\xi|^2 \to \infty$. Note that this equation is called "time-invariant" by contrast with

the standard Schrödinger equation discussed in Lecture 3, which contains a term of the kind $i\psi_t$. As such, ν can be viewed as a localized potential which depends on the parameter τ, and as we shall see, this parameter could also be viewed as a timescale much slower than t. There are two kinds of bounded solutions to this problem:

- localized (trapped) solutions associated with discrete negative eigenvalues λ_n
- oscillatory solutions associated with positive eigenvalues $\lambda = k^2$, in the continuous spectrum

To see this, simply note that if $|\nu|$ goes to 0 rapidly as $|\xi|$ goes to infinity, then (38) reduces to $\partial_{\xi\xi}\psi = -\lambda\psi$ as $|\xi| \to \infty$. The corresponding eigenmodes ψ then either tend to $e^{\pm ik\xi}$ with $k = \sqrt{\lambda}$ (solutions which exist for all values of $\lambda > 0$), or to the exponential solutions $e^{\pm\sqrt{-\lambda}\xi}$ (in which case only a discrete number of values of λ exist for which the solution is bounded).

One way of solving this equation of course, if $\nu(\xi, \tau)$ is known, is to integrate it in space for every value of τ. Generally speaking, the eigenmodes and eigenvalues depend on τ.

Let's then ask the two following questions: (1) What conditions does the potential ν have to satisfy for the eigenvalues to be independent of τ? Under these conditions of course, only the mode shape varies with τ which leads to the following question: (2) Can we derive a simpler evolution equation for a given mode as a function of τ, which does not rely on solving (38)?

The answer to both questions is fairly simple. Firstly, note that the only plausible evolution equation for a single mode to stay in its own phase space is

$$\partial_\tau \psi = A\psi + B\partial_\xi \psi. \tag{39}$$

(since (38) is a linear second-order equation), where A and B do not depend on ψ, but depend on ν of course. For this equation to hold as an equivalent solution of (38), a simple compatibility condition must be satisfied: $\partial_\tau \partial_\xi^2 \psi = \partial_\xi^2 \partial_\tau \psi$. A few manipulations reveal that this compatibility condition is equivalent to

$$\begin{aligned} \lambda_\tau = -\nu_\tau - A_{\xi\xi} + 2B_\xi(\lambda + \nu) + B\nu_\xi \text{ and} \\ B_{\xi\xi} = -2A_\xi \end{aligned} \tag{40}$$

where A and B are themselves functions of ν. For λ_τ to be zero, ν has to satisfy the PDE

$$\nu_\tau = \frac{B_{\xi\xi\xi}}{2} + 2B_\xi(\lambda + \nu) + B\nu_\xi. \tag{41}$$

In other words, any choice of the functional B yields a potential ν solution of (41) which has the desired property that the respective eigenvalues of the Schrödinger equation are independent of τ.

The relationship between the results presented above and the KdV equation becomes apparent as soon as one notes that (41) recovers KdV provided

$$B = 4\lambda - 2\nu. \text{ in which case } A = \alpha + \partial_\xi \nu \tag{42}$$

where α is an arbitrary constant. This fundamental point now provides a unique means to solve the KdV equation using Inverse Scattering Theory.

Indeed, note that as $|\xi| \to \infty$, equation (39) reduces to the very simple linear equation $\psi_\tau = \alpha\psi + 4\lambda\psi_\xi$. Hence, if $\lim_{\to\pm\infty} \psi(\xi,\tau)$ is known at time $\tau = 0$, simple linear PDE methods provide us with analytical expression for $\lim_{\to\pm\infty} \psi(\xi,\tau)$ at any time τ. This information, called the "scattering data", is *all that is needed* to reconstruct the "scattering potential" $\nu(\xi,\tau)$ at every time τ provided it is available for every eigenvalue of the system (in both the continuous and in the discrete part of the spectrum). This inversion process is called Inverse Scattering Theory.

We have therefore outlined the steps towards a novel method of computing solutions of the KdV:

1. at the initial time, given the initial condition (potential shape) $\nu(\xi, \tau = 0)$, solve equation (38) to calculate all the eigenmodes $\psi(\xi, 0)$ and associated eigenvalues λ. As it has been said previously, we ensured when choosing A and B in equation (39) that the λ_n's are invariant in time,

2. using (39) applied as $|\xi| \to \infty$, calculate the time evolution of the scattering data.

3. invert the scattering data at any time to reconstruct the shape of the potential ν.

It can be shown that each negative eigenvalue of the Schrödinger equation represents a soliton solution of the KdV, while the continuous spectrum represents oscillatory wave trains.

Any initial condition that is bounded and smooth for all ξ, and that decreases rapidly enough as $|\xi| \to \infty$, therefore generates a KdV solution consisting of a finite number of solitons (the negative discrete eigenvalues), plus an oscillatory wave train. It is easy to construct special initial conditions that generate only solitons, and other initial conditions that generate nothing but solitons (see Section 4.1).

Finally, note that this method is sometimes referred to as a "nonlinear Fourier analysis" as these two methods are somehow similar. In the Fourier analysis of a linear problem where a solution u of a PDE associated with a dispersion relationship $\omega = \omega(k)$ has to be found, the three equivalent steps are:

1. at the initial time, calculate the Fourier transform of u:

$$\tilde{u}(k) = \int_{-\infty}^{+\infty} u(x, t = 0)\, e^{-ikx} dx,$$

   which determines the importance of each k component. This distribution is "frozen" in time,

2. compute the evolution in time of each component. This is very easy to do as only plane waves have to be considered, the result being $\tilde{u}(k)\, e^{i(kx-\omega(k)t)}$,

3. reconstruct u at any time by summing all the k's / inverting the Fourier transform:

$$u(x,t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{u}(k)\, e^{i(kx-\omega(k)t)}\, dk.$$

4 Accuracy of the KdV equation

After deriving the KdV equation, the next natural step is to check whether it adequately models the behavior of inviscid water wave in the limit of long, weakly nonlinear waves.

4.1 KdV vs. experiment

Eight years after the rediscovery of the KdV equation by Zabusky & Kruskal (1965), Hammack (1973) [10] and Hammack & Segur (1974, 1978) [11, 12] published results comparing the solutions of the KdV equation with nonlinear wave profiles in an experimental tank. The apparatus, described in Figure 2, consists of a piston able to generate up– or downward displacement and moved by a motor that is able to generate a fairly broad range of modulations.

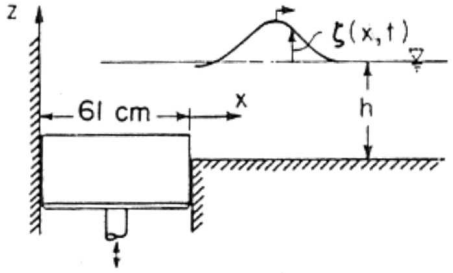


Figure 2: Description of Joe Hammack's apparatus for the generation of localized waves: a piston is able to move up and down to generate water displacements. From [11].

Three different experiments with positive mean displacement of the piston are presented in Figure 3.

In the first experiment, the piston moves monotonically upwards. The resulting free-surface displacement is therefore step-shaped on top of it and evolves into a train of four solitons, followed by a small oscillatory tail as the wave propagates down the tank.

In the second experiment, the constant upward motion of the piston is modulated by a sine wave, which has the same period as the time it takes for the piston to move up. Close to the piston, the free-surface displacement is quite different from the previous experiments with the presence of additional "noise", *i.e.* short, small waves superimposed on the previous step-shaped displacement. Interestingly, far away from the piston, we find again the same four solitary waves as in the first experiment.

In the third experiment, the motion of the piston is modulated with a sine wave of much higher frequency. The "noise" is initially much stronger but as the waves propagate, the "noise" propagates slower than the larger, nonlinear waves and disperses. Away from the piston, the solitary waves are recovered.

These experimental results can be understood from an analysis of the KdV equation using the IST. Solving equation (38), using the measured initial shape of the wave from the experiments (i.e., at $x/h = 0$), reveals that there are exactly four negative eigenvalues, and therefore exactly four solitons in the wave form that evolved from these initial data.

This statement holds for any of the three initial conditions shown in Figure 3, and the experimental results confirmed that four solitons indeed eventually emerge.

We saw in Lecture 3 (see also equation (36) that solitons with larger amplitudes travel faster, so it follows that the 4 solitons generated from the initial data will necessarily be ordered by amplitude for large τ, with the largest one in front. The experiments confirmed this prediction. Moreover, recall from Lecture 3 again that the KdV solutions predict a unique relationship between the amplitude and the shape of the solitons. Measurements of the amplitude of the solitons at large time, and comparison of the associated predicted profiles with the measured profiles again reveals good agreement.

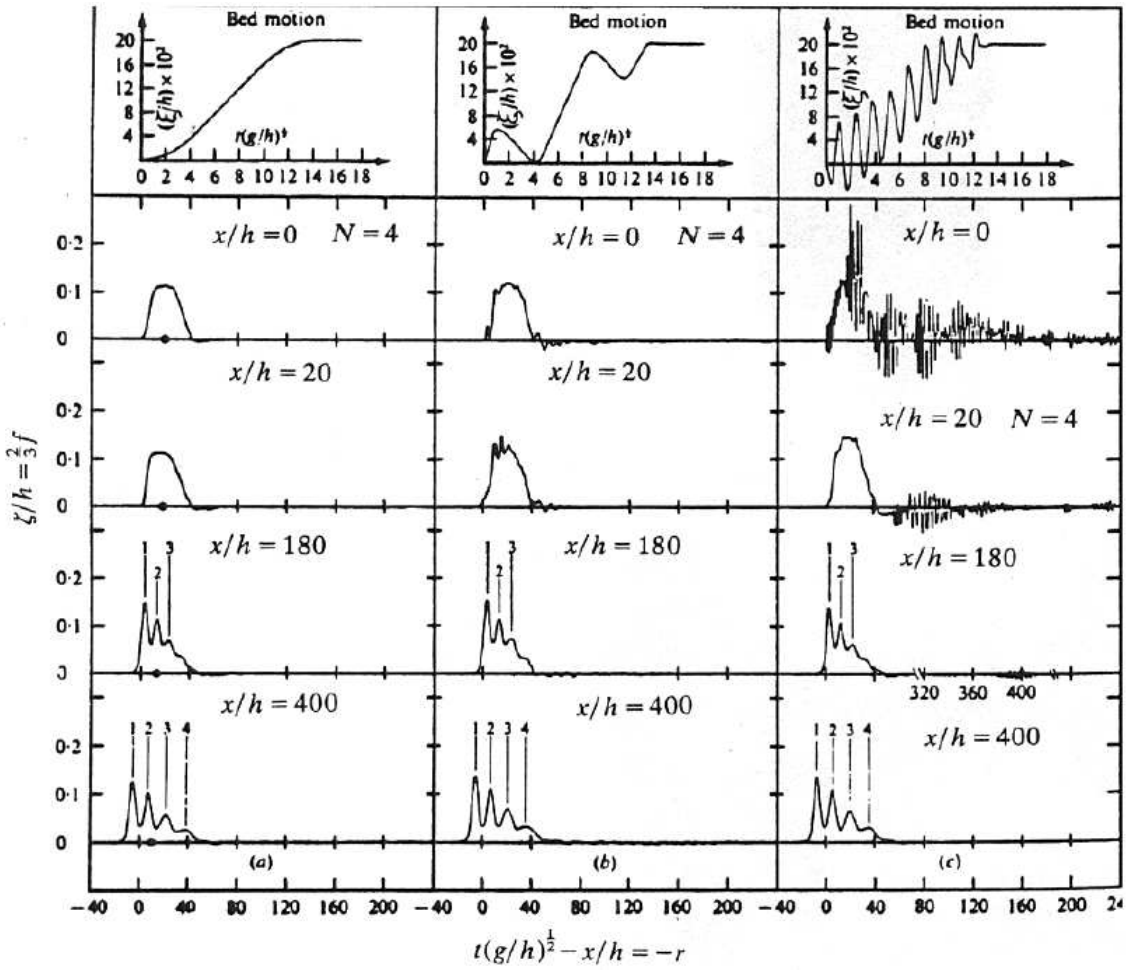


Figure 3: Various experiments resulting in the generation of KdV solitons (the general motion of the piston is positive). From top to bottom: – piston motion, – free-surface displacement on top of the piston, – free-surface displacement in three other locations further away from the piston. From left to right: three different types of piston motion. From [11].

When the piston is moved downwards ($\nu(\xi, 0) < 0$) it is possible to show that there are no discrete negative eigenvalues of (38), only modes with eigenvalues from the continuous

spectrum. As a result, we expect no generation of solitons, as can indeed be seen in Figure 4, only an oscillatory wave train. It can be seen that the amplitude is still quite large but with no implication on the appearance of solitons. This is a general result for the KdV equation: when the initial disturbance is a depression instead of an elevation, for surface waves long enough for the surface tension not to be significant, no solitary waves of depression can arise, only the continuous spectrum (cf. Section 3.3). Interestingly, if the waves are short enough for the surface tension to be significant, a surface wave of depression can correspond to an attractive potential that allows for depression solitons to exist. This corresponds to the inversion of the sign of $(\frac{1}{3} - \frac{\sigma}{\rho g h^2})$ in *e.g.* Equations (33) and (34), $\frac{\rho g h^2}{\sigma}$ being the Bond number of the problem.

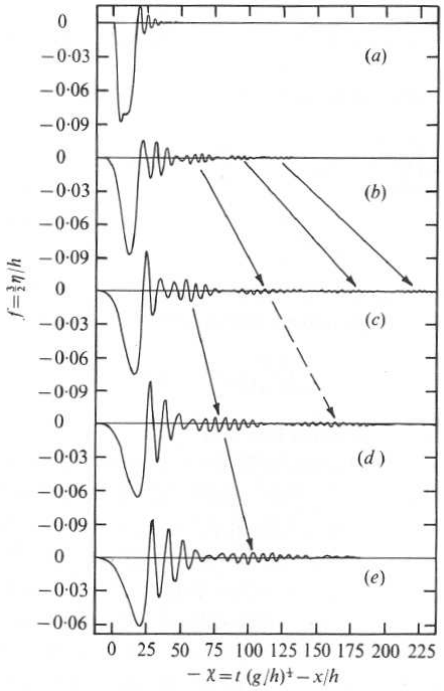


Figure 4: Experiment with a negative displacement of the piston: no solitons are generated, although the waves are large enough to produce nonlinear effects. From top to bottom: free-surface displacement in five locations with increasing distance from the piston. Arrows emphasize on the propagation of various linear wave packets. From [11].

4.2 KdV vs. full water wave equations

The question of how well the KdV equation approximates the water waves in the limit of long, nonlinear waves won't be discussed here. For more details, see [16], [20], [6, 5, 3, 4, 17, 2, 18].

5 Conclusion

This lecture was the occasion to derive the Korteweg – de Vries (KdV) equation using an asymptotic expansion of the inviscid, two-dimensional water waves in the long, weakly nonlinear limit. This expansion, restricted to the first nonlinear and dispersive terms, is derived assuming that these weak terms balance. This balance then provides a physical explanation of the fact that there exist solutions of the KdV that can propagate without any change in their shape: the solitons or solitary waves.

This equation has peculiar properties. It has an infinite number of conservation laws, making any of its solutions very constrained. It can be solved using the Inverse Scattering Transform (Ist), an exact method which involves solving the equation in three linear steps instead of the original nonlinear problem. This method sorts the solutions of KdV into two different categories: solitary waves that only arise in finite numbers and a continuous spectrum of linear, dispersive waves. This method is also useful to understand that no solitary waves can emerge out of an initial depression of the surface, unless the capillary effects are significant enough.

In a series of experiments by Hammack & Segur in the 70's, the accuracy of the KdV equation for two-dimensional, long and weakly nonlinear waves has been investigated, with good qualitative agreement with the predictions of the Ist and also a good quantitative agreement when dissipative correction is added.

References

[1] M. J. Ablowitz and H. Segur, *Solitons and the inverse scattering transform*, Society for Industrial and Applied Mathematics, 1981.

[2] J. T. Beale, *The approximation of weak solutions to the 2-D Euler equations by vortex elements*, Institute for Mathematics and Its Applications, 29 (1991), pp. 23–+.

[3] G. Bona, J. A. Chen, and J. P. Saut, *Boussinesq Equations and Other Systems for Small-Amplitude Long Waves in Nonlinear Dispersive Media. I: Derivation and Linear Theory*, Journal of NonLinear Science, 12 (2002), pp. 283–318.

[4] J. L. Bona, M. Chen, and J.-C. Saut, *Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media: II. The nonlinear theory*, Nonlinearity, 17 (2004), pp. 925–952.

[5] J. L. Bona, T. Colin, and D. Lannes, *Long Wave Approximations for Water Waves*, Arch. Rational Mech. Anal., 178 (2005), pp. 373–410.

[6] W. Craig, *An existence theory for water waves and the Boussinesq and Korteweg-de Vries scaling limits*, Communications in Partial Differential Equations, 10 (1985), pp. 787–1003.

[7] T. Dauxois and M. Peyrard, *Physics of solitons*, Cambridge University Press, 2006.

[8] G. de Vries, *Bidrage tot de Kebbis der lange golven*, PhD thesis, univ. Amsterdam, 1894.

[9] E. Fermi, J. Pasta, and S. Ulam, *Studies of nonlinear Problems I*, Tech. Rep. LA-1940, Los Alamos, 1955.

[10] J. L. Hammack, *A note on tsunamis: their generation and propagation in an ocean of uniform depth*, Journal of Fluid Mechanics, 60 (1973), pp. 769–799.

[11] J. L. Hammack and H. Segur, *The Korteweg-de Vries equation and water waves. Part 2. Comparison with experiments*, Journal of Fluid Mechanics, 65 (1974), pp. 289–314.

[12] ——, *The Korteweg de Vries equation and water waves. Part 3. Oscillatory waves*, Journal of Fluid Mechanics, 84 (1978), pp. 337–358.

[13] D. J. Korteweg and G. de Vries, *On the change of form of long waves advancing in a rectangular canal and on a new type of long stationnary waves*, Philosophical Magazine 5th Series, 36 (1895), pp. 422–443.

[14] R. M. Miura, *Korteweg-de Vries equation and generalizations. I. A remarkable explicit nonlinear transformation*, Journal of Mathematical Physics, 9 (1968), p. 1202.

[15] ——, *Korteweg-de Vries equation and generalizations. II. Existence of conservation laws and constants of motion*, Journal of Mathematical Physics, 9 (1968), p. 1204.

[16] G. Schneider and C. E. Wayne, *Kawahara dynamics in dispersive media*, Physica D Nonlinear Phenomena, 152 (2001), pp. 384–394.

[17] ——, *The Rigorous Approximation of Long-Wavelength Capillary-Gravity Waves*, Archive for Rational Mechanics and Analysis, 162 (2002), pp. 247–285.

[18] J.-M. Vanden-Broeck, *Elevation solitary waves with surface tension*, Physics of Fluids, 3 (1991), pp. 2659–2663.

[19] N. J. Zabusky and M. D. Kruskal, *Interaction of "Solitons" in a Collisionless Plasma and the Recurrence of Initial States*, Physical Review Letters, 15 (1965), pp. 240–243.

[20] D. Wright, *Corrections to the KdV approximation for water waves*. SIAM Journal on Mathematical Analysis, 37 (2005), pp. 1161–1206

Lecture 6: Derivation of the KdV equation for surface and internal waves

Lecturer: Roger Grimshaw. Write-up: Erinna Chen, Hélène Scolan, Adrienne Traxler

June 17, 2009

1 Introduction

We sketch here a different derivation of the Korteweg–de Vries equation, applicable to a wider range of problems, including internal waves and waves in sheared flows in addition to surface water waves. After beginning with the Euler equations and discarding small terms, the problem is reduced to a superposition of various modes — but through dispersion, after sufficient time has passed each mode becomes isolated, so we may consider solitary wave (soliton) solutions. The problem is then treated with an asymptotic expansion in small parameters characterizing the amplitude and dispersion. To first order we recover the linear long wave solutions, and to second order we find that the amplitude evolves according to the KdV equation. The extended KdV (eKdV) equation is discussed for critical cases where the quadratic nonlinear term is small, and the lecture ends with a selection of other possible extensions.

2 Derivation for surface and internal waves: Basic Setup

In the basic state, the motion is assumed to be two-dimensional and the fluid has a density $\rho_0(z)$, a corresponding pressure $p_0(z)$ such that the background state is hydrostatic ($p_{0z}(z) = -g\rho_0$), and a horizontal shear flow $u_0(z)$ in the x-direction. In this lecture we consider only the case where the bottom is flat (h is constant). Extensions considering variable depth are possible and lead to a variable-coefficient KdV equation (see Lecture 9).

The equations of motion relative to this basic state are the Euler equations:

$$\rho_0(u_t + u_0u_x + wu_{0z}) + p_x = -\rho_0(uu_x + wu_z) - \rho(u_t + u_0u_x + wu_{0z} + uu_x + wu_z) \quad (1)$$

$$p_z + g\rho = -(\rho_0 + \rho)(w_t + u_0w_x + uw_x + ww_z) \quad (2)$$

$$g(\rho_t + u_0\rho_x) - \rho_0N^2w = -g(u\rho_x + w\rho_z) \quad (3)$$

$$u_x + w_z = 0 \quad (4)$$

Here $(u_0 + u, w)$ are the velocity components in the (x, z) directions, $\rho_0 + \rho$ is the density, $p_0 + p$ is the pressure and t is time. The equations are arranged such that for linear long

waves all terms on the right hand side can be neglected (this statement is proved in Section 4). Here $N(z)$ is the buoyancy frequency, defined by

$$\rho_0 N^2 = -g\rho_{0z}. \tag{5}$$

The boundary conditions are

$$w = 0 \text{ at } z = -h, \tag{6}$$

$$p_0 + p = 0 \text{ at } z = \eta, \tag{7}$$

$$w = \eta_t + u_0\eta_x + u\eta_x \text{ at } z = \eta. \tag{8}$$

It is useful to use the vertical particle displacement ζ as the primary dependent variable. ζ is readily measured in the ocean and is related to the buoyancy frequency. The vertical particle displacement is defined by

$$\frac{D\zeta}{Dt} = \zeta_t + u_0\zeta_x + u\zeta_x + w\zeta_z = w. \tag{9}$$

The perturbed density field is given by the difference between the density at the vertical position a fluid particle originates from, and that of its current location: $\rho = \rho_0(z-\zeta) - \rho_0(z) \approx g\rho_0 N^2\zeta$ as $\zeta \to 0$, where we have assumed that as $x \to -\infty$, the density field relaxes to $\rho_0(z)$. Isopycnal surfaces (i.e. $\rho_0 + \rho$ =constant) are then given by $z = z_0 + \zeta$ where z_0 is the far-field level. The physical meaning of ζ is clearest on the free surface ($z = \eta$). In terms of ζ, the kinematic boundary condition (8) becomes $\zeta = \eta$ at $z = \eta$.

3 Linear long waves

To describe internal solitary waves, we seek solutions whose horizontal length scales are much greater than h (shallow water, long waves) and whose time scales are much greater than N^{-1} (linear behavior). We now solve the Euler equations by omitting the right-hand side of equations (1-4) and utilizing the linearized free-surface boundary conditions of (7,8). Solutions are sought in the form

$$\zeta = A(x - ct)\phi(z) \tag{10}$$

Plugging this solution for ζ into the linear Euler equations gives values for the remaining dependent variables:

$$u = A\left((c-u_0)\phi\right)_z, \; p = \rho_0(c-u_0)^2 A\phi_z, \; g\rho = \rho_0 N^2 A\phi. \tag{11}$$

Here c is the linear long wave speed, and the modal functions $\phi(z)$ are defined by the boundary-value problem

$$\{\rho_0(c-u_0)^2\phi_z\}_z + \rho_0 N^2\phi = 0 \text{ for } h < z < 0 \tag{12}$$

$$\phi = 0 \text{ at } z = -h, \; (c-u_0)^2\phi_z = g\phi \text{ at } z = 0. \tag{13}$$

Equation (12) is the well-known Taylor-Goldstein equation for perturbations in stratified shear flows, here expressed in the long-wave limit.

Typically, the boundary value problem (12,13) has an infinite sequence of linear long-wave modes solutions, $\phi_n^{\pm}(z)$, $n = 0, 1, 2, \ldots$, with corresponding wave speeds, $c_n^{\pm}(z)$. Here, the superscript "$\pm$" indicates waves with $c_n^+ > u_M = \max u_0(z)$ and $c_n^- < u_m = \min u_0(z)$. We shall confine our attention to these regular modes, and consider only stable shear flows. Mathematically, this implies that we do not consider modes with complex eigenvalues c, nor modes with $c \in [u_m, u_M]$. Analogous theory can be developed for singular modes with $u_m < c < u_M$.

In general, this boundary value problem has to be solved numerically. Typically, the $n = 0$ mode denotes the surface gravity waves for which c scales with $\sqrt{gh}$, while the $n = 1, 2, 3, \ldots$ modes denote internal gravity waves for which c scales with Nh. The surface mode ϕ_0 has no extrema in the interior of the fluid and takes its maximum value at $z = 0$. The internal modes $\phi_n^{\pm}(z)$, $n = 1, 2, 3, \ldots$, have $n - 1$ extremal points in the interior of the fluid and ϕ nearly vanishes near $z = 0$ (because $c^2 \ll gh$ for internal waves, and using equation (13)).

The solution of the linearized long wave equations is given asymptotically[1] by

$$\zeta \sim \sum_{n=0}^{\infty} A_n^{\pm}(x - c_n^{\pm} t)\phi_n^{\pm}(z) \text{ as } t \to \infty. \tag{14}$$

Here the amplitudes $A_n^{\pm}$ are determined from initial conditions. Assuming that the speeds $c_n^{\pm}$ of each mode are sufficiently distinct, the modes will separate spatially for large times, so we can consider a single mode in isolation. Henceforth, we shall omit the index n and assume that the single mode has speed c, amplitude A and modal function $\phi(z)$.

4 Asymptotic expansion

Having waited sufficiently long for the modes to separate also implies that hitherto neglected nonlinear terms may begin to have an effect. The nonlinear effects are balanced by dispersion (also neglected in linear long wave theory); this balance emerges as time increases and results in the Korteweg-deVries equation for the wave amplitude.

The formal derivation of the evolution equation requires the introduction of two small parameters, ε and δ, characterizing the wave amplitude and inverse wave-length, respectively. As seen in Lecture 5, for nonlinearity to be balanced by dispersion (KdV balance), it is required that $\varepsilon = \delta^2$ (note that the leading nonlinear term is of order ϵ^2 while the leading dispersive term is of order $\epsilon\delta^2$). It was also shown that the nonlinear dynamics take place on a slow timescale T.

As in Lecture 5, we first introduce the scaled variables

$$T = \delta\varepsilon t, \;\; X = \delta(x - ct). \tag{15}$$

We then assume solutions of the form

$$\zeta = \varepsilon A(X, T)\phi(z) + \varepsilon^2 \zeta_2 + \ldots \tag{16}$$

[1]The reason why this is only valid asymptotically and not for all times is a technical issue: the modes defined do not necessarily form an orthogonal set so additional transient terms must be added to match the initial conditions in some cases.

with similar expressions analogous to 11 for the other dependent variables. Plugging these solutions into the full Euler equations results in the linear long wave solution for the modal function, $\phi(z)$ and the speed, c, at leading order. Since the modal equations (12,13) are homogeneous, we are free to impose a normalization condition on $\phi(z)$. A commonly used condition is that $\phi(z_m) = 1$ where $|\phi(z)|$ achieves a maximum value at $z = z_m$. In this case, the amplitude εA is uniquely defined as the amplitude of ζ to $\mathrm{O}(\varepsilon)$ at the depth z_m.

Continuing the expansion to the next order in ε, it can be shown that this leads to the following equation for ζ_2:

$$\{\rho_0(c-u_0)^2\zeta_{2Xz}\}_z + \rho_0 N^2\zeta_{2X} = M_2, \text{ for } -h < z < 0 \tag{17}$$

and the corresponding boundary condition

$$\zeta_{2X} = 0 \text{ at } z = -h, \ \rho_0(c-u_0)^2\zeta_{2Xz} - \rho_0 g\zeta_{2X} = N_2 \text{ at } z = 0. \tag{18}$$

The inhomogeneous terms M_2, N_2 are due to nonlinearity and dispersion, and are known in terms of the first-order functions $A(X,T)$ and $\phi(z)$. They are given by

$$M_2 = 2\{\rho_0(c-u_0)\phi_z\}_z A_T + 3(\rho_0(c-u_0)^2\phi_z^2)_z AA_X - \rho_0(c-u_0)^2\phi A_{XXX}, \tag{19}$$

$$N_2 = 2\{\rho_0(c-u_0)\phi_z\}A_T + 3(\rho_0(c-u_0)^2\phi_z^2)AA_X \tag{20}$$

Equations (17,18) are identical to the equations defining the modal function (12,13), with an additional forcing term on the right-hand side in (17). There will be a solution for the forced equation (17) that satisfies the boundary conditions (18) only if a certain compatibility condition is satisfied. We can obtain this compatibility condition for example by a direct construction of ζ_2.

Let us first define the linear operator $\mathcal{L}$ such that

$$\mathcal{L}(\phi) = \{\rho_0(c-u_0)^2\phi_z\}_z + \rho_0 N^2\phi. \tag{21}$$

Any pair of functions ψ and ϕ satisfying the lower boundary condition of the problem ($\psi(-h) = 0$ and $\phi(-h) = 0$) also satisfies

$$\int_{-h}^{z}\{\psi\mathcal{L}(\phi) - \phi\mathcal{L}(\psi)\}\mathrm{d}z = \rho_0(c-u_0)^2(\psi\phi_z - \phi\psi_z) = W(\psi,\phi;z) \tag{22}$$

where we have defined the Wronskian functional of ψ and ϕ as

$$W(\psi,\phi;z) \equiv \rho_0(c-u_0)^2(\psi\phi_z - \phi\psi_z). \tag{23}$$

If ψ and ϕ are solutions to the modal equation (12), then $\mathcal{L}\phi = \mathcal{L}\psi = 0$ so that

$$\frac{\mathrm{d}W}{\mathrm{d}z} = \psi\mathcal{L}\phi - \phi\mathcal{L}\psi = 0 \tag{24}$$

Hence $W(\psi,\phi;z)$ is actually independent of z.

Now, for ζ_{2X} solution of the forced equation (17), we have

$$\phi\mathcal{L}(\zeta_{2X}) - \zeta_{2X}\mathcal{L}(\phi) = \phi M_2 = \frac{\mathrm{d}W(\phi,\zeta_{2X};z)}{\mathrm{d}z}, \text{ and} \tag{25}$$

and similarly for ψ. Integration of (25) and elimination of ξ_{2Xz} results in

$$\phi \int_{-h}^{z} \psi M_2 dz - \psi \int_{-h}^{z} \phi M_2 dz = \zeta_{2X} W(\psi, \phi; z), \tag{26}$$

The general solution of ζ_{2X} is then the sum of a general solution for the unforced problem plus the particular solution just identified,

$$\zeta_{2X} = A_{2X}\phi + \phi \int_{-h}^{z} \frac{M_2 \psi}{W(\psi,\phi)} dz - \psi \int_{-h}^{z} \frac{M_2 \phi}{W(\psi,\phi)} dz. \tag{27}$$

where we recall that $W(\psi, \phi)$ is constant. This general solution for ζ_{2X} was so far obtained by applying only the boundary condition on the bottom ($z = -h$). In the process, we have introduced another modal function ψ and the Wronskian, W. By applying the free surface boundary condition, and requiring that ψ be linearly independent of ϕ such that $\psi(0)$ does not satisfy the upper boundary condition of (17), we can now obtain a compatibility condition that is independent of ψ and W. Plugging the solution of ζ_{2X} into the free surface boundary condition in (18), we obtain

$$\rho(c-u_0)^2 \left\{ \phi_z \int_{-h}^{0} \frac{M_2\psi}{W} \mathrm{d}z - \psi_z \int_{-h}^{0} \frac{M_2\phi}{W} \mathrm{d}z \right\} - \rho_0 g \left\{ \phi \int_{-h}^{0} \frac{M_2\psi}{W} dz - \psi \int_{-h}^{0} \frac{M_2\phi}{W} \mathrm{d}z \right\} = N_2. \tag{28}$$

Recalling that $\rho(c-u_0)^2\phi_z = \rho_0 g \phi$ at $z = 0$ from (13), it then follows that the compatibility condition is

$$\int_{-h}^{0} M_2 \phi \mathrm{d}z = [N_2\phi]_{z=0}. \tag{29}$$

Here we have obtained the compatibility condition through the direct construction of ζ_2. However, the compatibility condition can be obtained more easily without knowledge of ζ_2 (or higher-order ζ_n), and without the need to introduce an additional function ψ.

The compatibility condition is that the inhomogeneous terms in (17,18) should be orthogonal to the solution of the adjoint of the modal equations (12,13). This construction is fairly straightforward. We first begin by combining equations (22) and (25) into

$$\phi M_2 = \{\rho_0(c-u_0)^2(\phi\zeta_{2Xz} - \zeta_{2X}\phi_z)\}_z. \tag{30}$$

Integrating (30) and applying the free-surface boundary condition for the modal function ϕ (13) first, then the free-surface boundary condition (18) results in the compatibility condition (29) found earlier. This last method can easily be applied at any order in the expansion.

Note that the amplitude A_2 is left undetermined at this stage; further expansion into higher orders will result in an evolution equation for A_2. In general, solutions for ζ_{n+1} will result in a compatibility condition and thus, an evolution equation for A_n.

5 Korteweg-deVries (KdV) equation

Substituting the expressions for M_2 and N_2 (19, 20) into the compatibility condition (29), we obtain the evolution equation for $A(X, T)$, namely the KdV equation

$$A_T + \mu A A_X + \lambda A_{XXX} = 0, \tag{31}$$

where the coefficients μ (nonlinearity) and λ (dispersion) depend on the modal function ϕ:

$$I\mu = 3\int_{-h}^{0} \rho_0 (c-u_0)^2 \phi_z^3 \mathrm{d}z, \tag{32}$$

$$I\lambda = \int_{-h}^{0} \rho_0 (c-u_0)^2 \phi^2 \mathrm{d}z, \tag{33}$$

$$\text{where } I = 2\int_{-h}^{0} \rho_0 (c-u_0) \phi_z^2 \mathrm{d}z, \tag{34}$$

The KdV equation (31) is solved with the initial condition $A(X, T=0) = A_0(X)$ where $A_0(X)$ is determined from linear long wave theory, and is in essence the projection of the original initial conditions onto the appropriate linear long wave mode. Localized initial conditions lead to (at sufficiently large time) the generation of a finite number of solitary waves, or internal solitons.

For waves moving to the right, where $c > u_M = \max u_0(z)$, I and λ are always positive. For the surface mode, $\phi(z) > 0$ and $\phi(0) = 1$ (no extrema in the interior) we see that $\mu > 0$. In general, μ can take either sign, and may be zero in some special situations. Explicit evaluation of the coefficients μ and λ requires knowledge of the modal function, and hence they are usually evaluated numerically. The modal function is known in several instances, and we illustrate the process with two simple examples.

5.1 Example 1: Surface water waves with no surface tension

For water waves, we set $\rho =$ constant so that $N^2 = 0$. We also assume that there is no background shear. The modal solution to equation (12) satisfying the boundary conditions (13) is then

$$\phi = \frac{z+h}{h} \text{ for } -h < z < 0,\ c = (gh)^{1/2}. \tag{35}$$

Note that there are no other modes in this system. Plugging in the modal function into equations (32-34), the coefficients I, μ and λ are

$$I = \frac{2\rho_0 c}{h} \text{ and } \mu = \frac{3c}{2h} \text{ and } \lambda = \frac{ch^2}{6}. \tag{36}$$

Thus the KdV equation for water waves is, in the original variables,

$$\zeta_t + c\zeta_x + \frac{3c}{2h}\zeta\zeta_x + \frac{ch^2}{6}\zeta_{xxx} = 0. \tag{37}$$

Note that here $z_m = 0$ so we identified A with $\zeta(x, 0, t)$, the free surface displacement, to leading order. For zero surface tension, this is the equation derived by Korteweg and de Vries in 1895 (and first by Boussinesq in the 1870's).

5.2 Example 2: Interfacial waves

For a two-layer fluid, waves may occur at the interface. Let the density be constant with value ρ_1 in an upper layer of height h_1 and ρ_2 in the lower layer of height $h_2 = h - h_1$. We assume the fluid is stably stratified such that $\rho_2 > \rho_1$. The density in the fluid is $\rho_0(z) = \rho_1 H(z + h_1) + \rho_2 H(-z - h_1)$ and the buoyancy frequency is $\rho_0 N^2 = g(\rho_2 - \rho_1)\delta(z + h_1)$. Here $H(z)$ is the Heaviside function and $\delta(z)$ is the Dirac δ-function. For simplicity, we assume that $\rho_1 \approx \rho_2$, the usual situation in the ocean. As mentioned earlier, the upper boundary condition for $\phi(z)$ then is approximately $\phi(0) \approx 0$. The modal function is then

$$\phi = \frac{z+h}{h_2} \text{ for } -h < z < -h_1, \ \phi = -\frac{z}{h_1} \text{ for } -h_1 < z < 0. \tag{38}$$

The corresponding coefficients are

$$\mu = \frac{3c(h_1 - h_2)}{h_1 h_2}, \ \lambda = \frac{c h_1 h_2}{6}, \ c^2 = \frac{g(\rho_2 - \rho_1)}{\rho_2} \frac{h_1 h_2}{h_1 + h_2}. \tag{39}$$

When the interface is closer to the free surface than to the bottom ($h_1 < h_2$), the nonlinear coefficient μ for these interfacial waves is negative. For single-layer water waves μ always remains positive. In the case when $h_1 \approx h_2$, μ nearly vanishes and it is necessary to introduce higher- order nonlinearity in order to balance the dispersion.

6 Extended Korteweg-deVries equation

As seen in the example of interfacial waves, the quadratic nonlinearity may vanish, and in this instance, it is necessary to use an extended KdV equation which contains higher-order nonlinearities and additional terms.

6.1 Higher-order expansions

Proceeding to the next highest order in the asymptotic expansion yields a set of equation analogous to (17,18) for ζ_3, whose compatibility condition then determines an evolution equation for the second-order amplitude A_2. Using the transformation $A + \varepsilon A_2 \to A$, and then combining the KdV equation (31) with the evolution equation for A_2 leads to a higher-order KdV equation

$$A_T + \mu A A_X + \lambda A_{XXX} + \varepsilon(\lambda_1 A_{XXXXX} + \sigma A^2 A_X + \mu_1 A A_{XXX} + \mu_2 A_X A_{XX}) = 0. \tag{40}$$

Explicit expressions for the coefficients are known analogs of (32-34). This higher-order KdV equation is Hamiltonian provided $\mu_2 = 2\mu_1$.

It is important to note that equation (40) is not unique: the near-identity transformation $A \to A + \varepsilon(aA^2 + bA_{XX})$ reproduces the same equation (40) to the same order in ε provided the coefficients are also changed to

$$(\lambda_1, \sigma, \mu_1, \mu_2) \to (\lambda_1, \sigma - a\mu, \mu_1, \mu_2 - 6a\lambda + 2b\mu).$$

Note that (40) at O(1) needs to be used to transform terms of the kind $\varepsilon A A_T$ and εA_{XXT}.

Furthermore, when $\mu \neq 0$, $\lambda \neq 0$, the enhanced transformation

$$A \to A + \varepsilon \left(aA^2 + bA_{XX} + a'A_X \int^X A dX + b'XA_T \right) \tag{41}$$

reduces (40) to the KdV equation. From a mathematical point of view, this implies that KdV is a normal form of the system, or in other words, the lowest order and simplest form characterizing the dynamics given the long-wave, small amplitude approximation made. Physically, this implies that for small perturbations with $\varepsilon \ll 1$, no additional dynamics are introduced by the higher-order terms.

6.2 Extended KdV equation

A particularly important case arises when the nonlinear coefficient μ is close to zero. In this case, the near-identity transformation (41) cannot cancel out the the cubic nonlinear term in the higher-order KdV equation (40) at this order. This identifies this particular higher-order term as being the most important one in balancing dispersion if $\mu \to 0$. The KdV equation (31) is then replaced by the extended KdV (or Gardner) equation,

$$A_T + \mu A A_X + \varepsilon\sigma A^2 A_X + \lambda A_{XXX} = 0. \tag{42}$$

For $\mu \approx 0$, a rescaling is needed, and the optimal choice is to assume μ is $\mathrm{O}(\delta)$, and then replace A with A/δ. The amplitude parameter is δ instead of δ^2. The resulting equation in the canonical form is

$$A_T + 6AA_X + 6\beta A^2 A_X + A_{XXX} = 0. \tag{43}$$

Like the KdV equation, the Gardner equation is integrable and can be solved using the Inverse Scattering method. The coefficient β can be either positive or negative, and the structure of the solutions depends on which sign is appropriate.

6.3 Solitary wave solutions

The solitary wave solutions for the extended KdV equation are given by

$$A = \frac{a}{b + (1-b)\cosh^2 \gamma(x - Vt)}, \tag{44}$$

$$\text{where } V = a(2 + \beta a) = 4\gamma^2,\ b = \frac{-\beta a}{(2 + \beta a)}. \tag{45}$$

There are two cases to consider. If $\beta < 0$, then there is a single family of solutions such that $0 < b < 1$ and $a > 0$. As b increases from 0 to 1, the amplitude a increases from 0 to a maximum of $-1/\beta$, while the speed V also increases from 0 to a maximum of $-1/\beta$. In the limiting case when $b \to 1$, the solution (44) describes the so-called "thick" or "table-top" solitary wave, which has a flat crest of amplitude $a_m = -1/\beta$ (see Figure 1). If $\beta > 0$, then the family of solutions allows both waves of depression and of elevation. In particular, there is a region of depression where solutions are not permitted, as indicated by the blue curve in Figure 1. As the amplitude is reduced, the solution becomes a "breather", a solitary wave with periodically-varying amplitude.

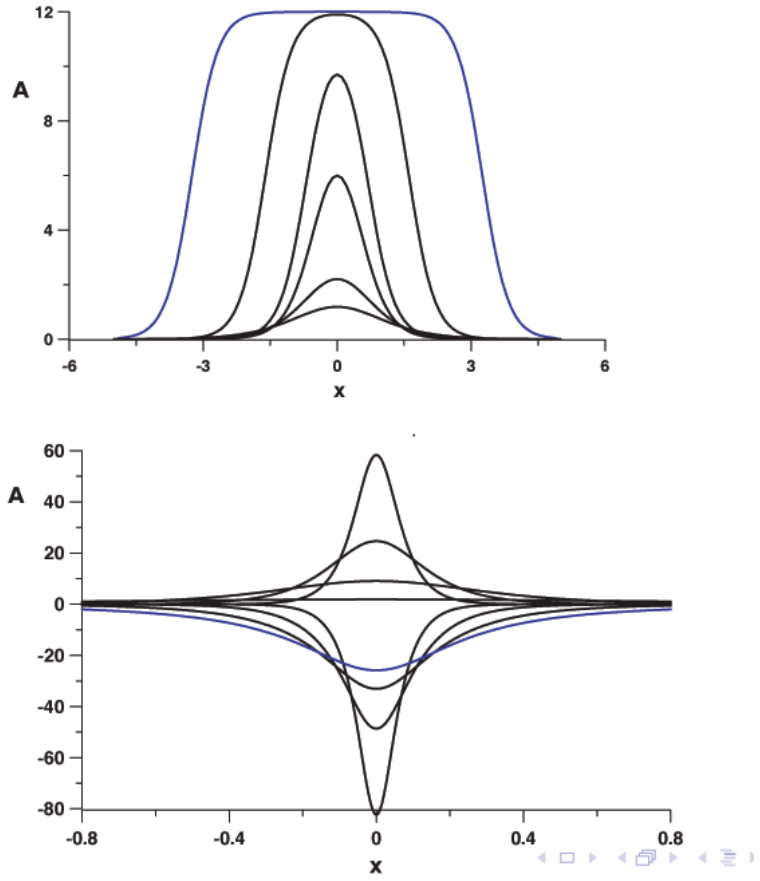


Figure 1: (top) Solitary wave solutions for $\beta < 0$. Note that there is a finite amplitude for the waves. Once this amplitude is achieved, the wave broadens and exhibits a "table top" behavior, indicated by the blue curve. (bottom) Solitary wave solutions for $\beta > 0$. Solutions can be both waves of elevation and depression. There is a minimum amplitude for waves of depression as indicated by the blue curve. Because of this, a breather solution is supported.

7 Other long-wave models

So far we have considered simple single-layer water wave examples. For a more realistic model of the ocean with a stratified near-surface layer lying above a deep ocean with constant density, a different scaling from the KdV equation is needed. In the surface layer, the long-wave scaling still holds, but this needs to be matched to a different scaling in the deep lower layer, where the vertical scale matches the horizontal scale. We therefore introduce a rescaled layer depth H, with $h = H/\delta$. In this scenario, the KdV equation is replaced by the intermediate long-wave (ILW) equation

$$A_\tau + \mu A A_X + \delta \mathcal{L}(A_X) = 0, \tag{46}$$

$$\text{where } \mathcal{L}(A) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} k \coth kH \exp(ikX)\mathcal{F}(A)dk \tag{47}$$

$$\text{and } \mathcal{F}(A) = \int_{-\infty}^{\infty} A \exp(-ikX)dX. \tag{48}$$

Here the nonlinear coefficient μ is again given by (32) with $-h$ now replaced by $-\infty$, while the dispersive coefficient δ is defined by $I\delta = \{\rho_0 c^2 \phi^2\}_{z\to\infty}$.In the limit $H \to \infty$, $k \coth kH \to |k|$, the ILW equation (46) becomes the Benjamin-Ono (BO) equation. In the opposite limit, where $H \to 0$, then (46) reduces to a KdV equation. Both equations are integrable.

References

[1] *Magic lectures on nonlinear waves.* http://www.maths.dept.shef.ac.uk/magic/course.php?id=21. (Mathematics Access Grid Instruction and Collaboration).

[2] *Nonlinear waves in fluids: Recent advances and modern applications*, in CISM Courses and Lectures No.483, R. Grimshaw, ed., Springer, Wien New York, 2005, p. 196.

[3] *Solitary waves in fluids*, in Advances in Fluid Mechanics, R. Grimshaw, ed., vol. 47, WIT Press, UK, 2007, p. 208.

[4] G. Whitham, *Linear and Nonlinear Waves*, Wiley, New York, 1974.

Lecture 7: Oceanographic Applications.

Lecturer: Harvey Segur. Write-up: Daisuke Takagi

June 18, 2009

1 Introduction

Nonlinear waves can be studied by a number of models, which include the Korteweg–de Vries (KdV) and the Kadomtsev-Petviashvili (KP) equations. These equations are reviewed briefly and applied here to investigate the dynamics of surface waves in the ocean. The resulting tsunami of the Indian Ocean earthquake in 2004 are considered as a specific example. The tsunami dynamics, as well as wave patterns observed near shore, are explained by ideas developed in the previous lectures.

2 Review of waves in shallow water

A theory of nonlinear surface waves in shallow water was presented in Lectures 5 and 6. Relevant aspects of this theory are reviewed first for later discussion of observed waves in the ocean. Note that the theory, which includes effects due to dispersion, is different from the hyperbolic partial differential equations called the shallow-water equations, which describe non-dispersive waves and are presented in Lecture 8.

Consider the water in the ocean as an incompressible, irrotational fluid with velocity potential ϕ. A Cartesian coordinate system is adopted with the x and y axes in the horizontal plane and the z axis pointing upwards from the mean level of the fluid. The fluid lies in the domain bounded below by a prescribed topography, $z = -h(x, y)$, and above by a free surface to be determined, $z = \eta(x, y, t)$. In this theoretical framework, the governing equations at any time t are given by

$$\nabla^2\phi = 0 \qquad -h < z < \eta, \tag{1}$$

$$\frac{\partial \eta}{\partial t} + \nabla\phi \cdot \nabla\eta = \frac{\partial \phi}{\partial z} \qquad z = \eta, \tag{2}$$

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}|\nabla\phi|^2 + g\eta = 0 \qquad z = \eta, \tag{3}$$

$$\mathbf{n} \cdot \nabla\phi = 0 \qquad z = -h. \tag{4}$$

The velocity potential satisfies the incompressibility condition (1), subject to the boundary conditions at the top and bottom of the domain. Equations (2) and (3) are respectively the kinematic and dynamic conditions on the free surface. Gravity g is dominant and surface tension is neglected in the dynamic condition (3), implicitly restricting the analysis to waves with horizontal lengthscales much larger than the capillary length scale. Equation

(4), where $\mathbf{n}$ is the unit vector normal to the topography, is the impermeability condition at the rigid bottom of the ocean.

A theory of nonlinear waves in shallow water is developed by introducing the following approximations. First, the characteristic variation a in η is small compared to the entire depth of the water h, under the small-amplitude approximation. Second, waves propagate in the x direction with a typical wavelength L_x, much longer than the depth of the water. This is the shallow water or the long wave approximation. Third, the motion is nearly one-dimensional, provided the horizontal length scale L_y in the transverse direction of the propagating wave is much longer than L_x. It is assumed that all small effects balance by order of magnitude such that $a/h = O(\epsilon)$, $(h/L_x)^2 = O(\epsilon)$ and $(L_x/L_y)^2 = O(\epsilon)$ or smaller, where $\epsilon \ll 1$ is the small parameter in the problem. See Lectures 3, 5 and 6 for the derivation and discussion of these scalings.

At leading order in ϵ, η satisfies the one-dimensional wave equation given by

$$\frac{\partial^2 \eta}{\partial t^2} = c^2 \frac{\partial^2 \eta}{\partial x^2}, \tag{5}$$

for constant topography h, where $c = \sqrt{gh}$ is the phase speed of the wave. The solution is a linear combination of traveling waves given by

$$\eta = \epsilon h[f(x - ct; \epsilon y, \epsilon t) + F(x + ct; \epsilon y, \epsilon t)] + O(\epsilon^2), \tag{6}$$

where f and F are amplitudes of the waves that propagate in the positive and negative x directions respectively. At next order in ϵ, either the KdV or the KP equation is obtained, depending on whether $(L_x/L_y)^2 \ll O(\epsilon)$ or $(L_x/L_y)^2 = O(\epsilon)$ respectively. Waves propagating along the x axis with $(L_x/L_y)^2 \ll O(\epsilon)$, are one-dimensional and described by the KdV equation [8]

$$\frac{\partial f}{\partial \tau} + f \frac{\partial f}{\partial \xi} + \frac{\partial^3 f}{\partial \xi^3} = 0, \tag{7}$$

where $\tau = \epsilon t$ is the slow time variable and $\xi = x - ct$ is the spatial coordinate in the reference frame of the moving waves. The KdV equation indicates that the wave amplitude evolves due to nonlinear and dispersive effects, corresponding to the second and third terms of (7) respectively. Two-dimensional waves, with $(L_x/L_y)^2 = O(\epsilon)$, disperse weakly in the transverse direction of propagation and are described by $f(\xi, \zeta, \tau)$ satisfying the KP equation [6] instead

$$\frac{\partial}{\partial \xi}\left(\frac{\partial f}{\partial \tau} + f \frac{\partial f}{\partial \xi} + \frac{\partial^3 f}{\partial \xi^3}\right) + \frac{\partial^2 f}{\partial \zeta^2} = 0, \tag{8}$$

where $\zeta = \epsilon y$ is the slowly-varying coordinate in the y direction. Both KdV and KP equations are integrable, meaning that they admit soliton solutions. A soliton is a special type of solitary wave, a localized wave that travels without any change in shape or size. The soliton has a permanent form in structure even after interacting with another oncoming soliton. Solitary waves occur frequently because they represent the long-time limit of waves that arise from a range of initial conditions. The possible appearance of solitary waves in the ocean is examined below.

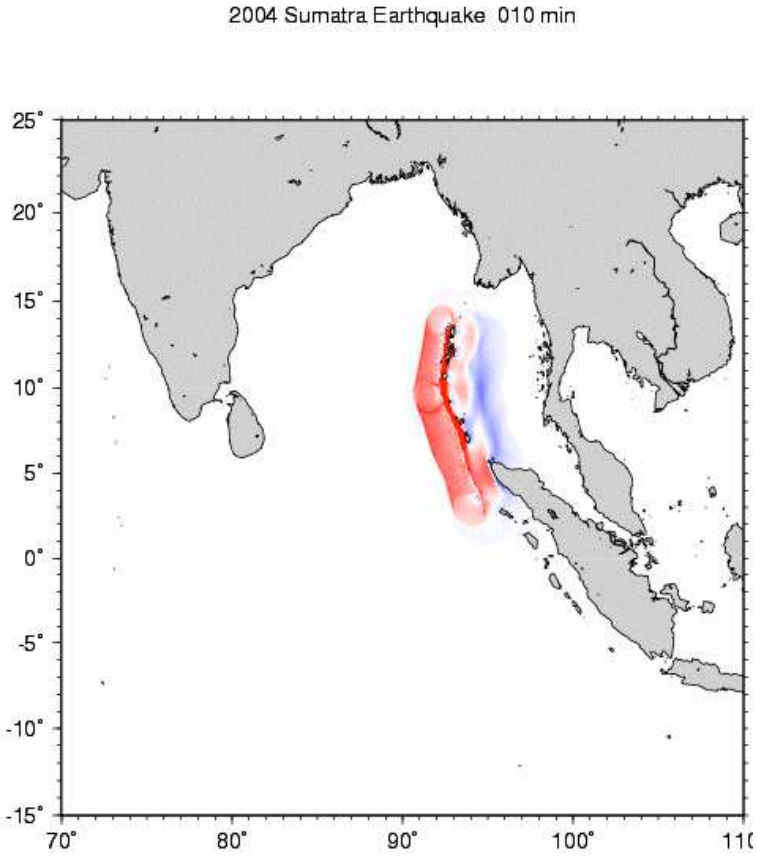


Figure 1: Numerical simulation of surface elevation (red) and depression (blue) of the Indian Ocean, soon after a series of undersea earthquakes occurred off the coast of Sumatra over an interval of 10 minutes on 26 December, 2004. An animation of the evolving surface of the ocean is available at http://staff.aist.go.jp/kenji.satake/.

Variable (units)	Magnitude
h (km)	3.5
a (km)	0.001
L_x (km)	100
L_y (km)	1000
c (km/h)	650
u (km/h)	0.2

Table 1: Typical magnitudes of the tsunami triggered on 26 December, 2004, in the Indian Ocean of depth h. Surface waves with amplitude a, wavelength L_x and width L_y traveled with phase speed c and fluid speed u.

3 Application to Tsunami waves

The sudden displacement of a large volume of water results in a series of surface waves in the ocean, called a tsunami. A famous example is the tsunami caused by a series of undersea earthquakes off the western coast of Sumatra on 26 December, 2004. These earthquakes occurred near-simultaneously along a 1000 km fault line. The surface elevation and depression of the ocean, soon after the earthquake, have been reproduced numerically as shown in Figure 1. The resultant tsunami waves devastatingly struck the coasts of the Indian Ocean and caused many casualties.

The magnitude of the tsunami in the Indian Ocean is estimated as presented in Table 1. Note that a/h, $(h/L_x)^2$ and $(L_x/L_y)^2$ are all small with a common order of magnitude ϵ in the narrow range from 10^{-2} to 10^{-3}, consistent with the approximations required to obtain solitary waves governed by the KP equation. However, we know from Lectures 5 and 6 that solitary waves only develop on the long time scale of order $1/\epsilon$, namely 10^2 to 10^3 hours in this case. Given that the tsunami traveled a distance of approximately 1500 km across the Bay of Bengal with phase speed 650 km/h, the initial displacement of water did not have sufficient time to evolve into a solitary wave. The tsunami would have needed to propagate a distance two or three orders of magnitude longer than across the Bay of Bengal to develop into a solitary wave, as governed by the KdV or the KP equation.

In contrast to the short distance traveled by the tsunami in the Indian Ocean, the tsunami triggered off the coast of Chile on 22 May, 1960, by the most powerful earthquake ever recorded, may have developed into solitary waves. The tsunami propagated tens of thousands of kilometers across the Pacific Ocean and reached the coast of Japan after 22 hours. It is possible that this tsunami developed into solitary waves, which propagated without change in their structural form.

In the deep and open ocean, the dynamics of tsunami waves are characterized by small amplitudes and long wavelengths, which are hardly detected by an observer on a boat on the surface. As the tsunami approaches near shore, the depth of the water decreases, resulting in a decelerating wave speed at the front while the back of the tsunami maintains speed. The waves consequently compress horizontally and grow vertically in a process called wave shoaling. The waves may grow to tens of meters in amplitude, causing much damage when they reach and strike coastlines.

The tsunami caused by the Sumatra earthquake in 2004 propagated eastwards to the

Figure 2: The tsunami of 26 December, 2004, approaching Hat Ray Leah beach on the Krabi coast, Thailand [1].

coast of Thailand with a wave of depression at the front. This reflects the downward displacement of water on the eastern side of the area where the tsunami originated, as shown in Figure 1. As shown in Lecture 5, laboratory experiments demonstrate that a downward displacement of water leads to a wave-train preceeded by a wave of depression[5]. When the front of the wave of depression arrived in Thailand, the water along the shoreline receded dramatically and exposed areas that are otherwise submerged, as shown in Figure 2. Soon after, successive waves of large amplitude struck the coast and destructed the area.

Risks posed to coastlines can be assessed by considering the tsunami in the deep ocean, from initiation to propagation. Tsunami is generated by a thrust fault, a normal fault or a landslide. A crucial quantity for estimating the size of the tsunami is the volume of water displaced by seismic events under the water. The time for the tsunami to propagate between two given positions, x_1 and x_2, is estimated by minimizing

$$\int_{x_1}^{x_2} \frac{ds}{c(s)} \tag{9}$$

over all possible paths from x_1 to x_2. Note that the shortest distance from x_1 to x_2 may not be the path that minimizes (9) because the wave speed $c(s) = \sqrt{gh(s)}$ may increase considerably with position s along another path.

It remains a challenge to predict the detailed dynamics of tsunami, particularly near the shore. As the tsunami approaches the coast, the dynamics are influenced by effects due to reflection, refraction and breaking of the waves. The near-shotre shape, size and speed of the tsunami are still poorly understood.

4 Oscillatory waves in shallow water

In this Section, further insight into waves in shallow water is provided by examining another special class of waves, which oscillate periodically. Indeed, in contrast to tsunami waves

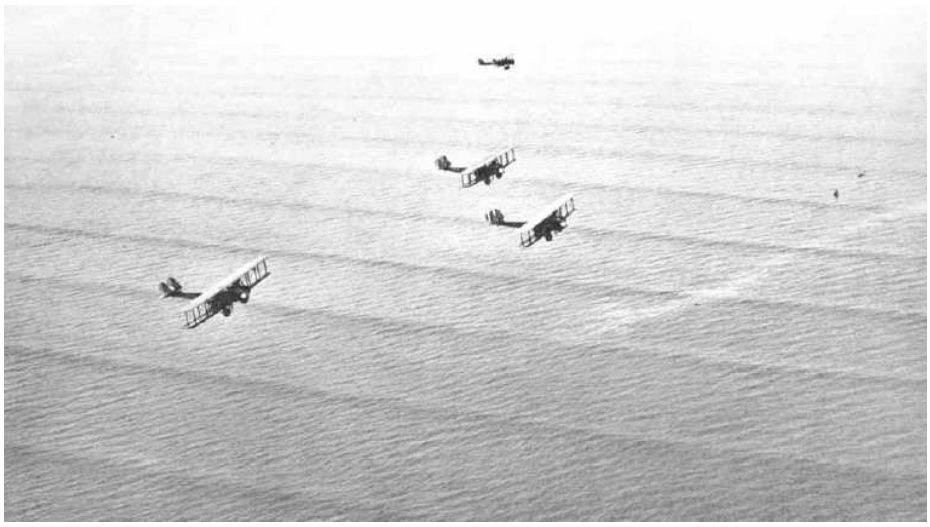


Figure 3: Cnoidal waves traveling in the ocean below flying aircrafts (National Geographic, 1933).

which are caused by earthquakes and landslides, most surface waves in the ocean are caused by storms and winds. These waves oscillate periodically and may develop patterns of permanent form, which are presented below.

The simplest model of long waves of small amplitude features linear, non-dispersive waves governed by (5), which all travel with speed $\sqrt{gh}$. For long waves of moderate amplitude, all traveling in approximately the same direction in water of uniform depth, a better approximation is the KP equation (8). It can be shown that the KP equation is completely integrable and admits solutions of the form[7]

$$f(\xi, \zeta, \tau) = 12 \frac{\partial^2}{\partial \xi^2} \log \Theta, \tag{10}$$

where Θ is a Riemann theta function of genus G. The genus is an integer corresponding to the number of independent phases in the solution. For example, solutions of genus 1 are the one-dimensional and periodic cnoidal wave solutions of the KdV equations discussed in Lectures 3, 5, and 6. Real cnoidal waves are shown in Figure 3; they propagate in the direction perpendicular to the wave crests with a coherent and permanent structure.

As discussed in Lecture 4, complete integrability guarantees the existence of quasi-periodic solutions obtained from (10), such as two-phase solutions of the KP equation. They are described by (10) with a Riemann theta function Θ of genus 2. The surface patterns are quasi-periodic in the sense that they cannot be characterized by a single period. The patterns are hexagonal in shallow water, as reproduced in Figure 4. The two-phase solutions can be interpreted as a combination of two cnoidal waves that meet at an angle, where a shift in phase occurs due to their interaction. The two-phase solutions agree remarkably well with surface patterns observed in laboratory experiments, demonstrating that the KP equation accurately describes real phenomena [3]. The excellent agreement between the theory and the experiments supports the existence of permanent patterns in the ocean, such as the approximately periodic patterns observes along the shoreline of Duck, North Carolina, as shown in Figure 5.

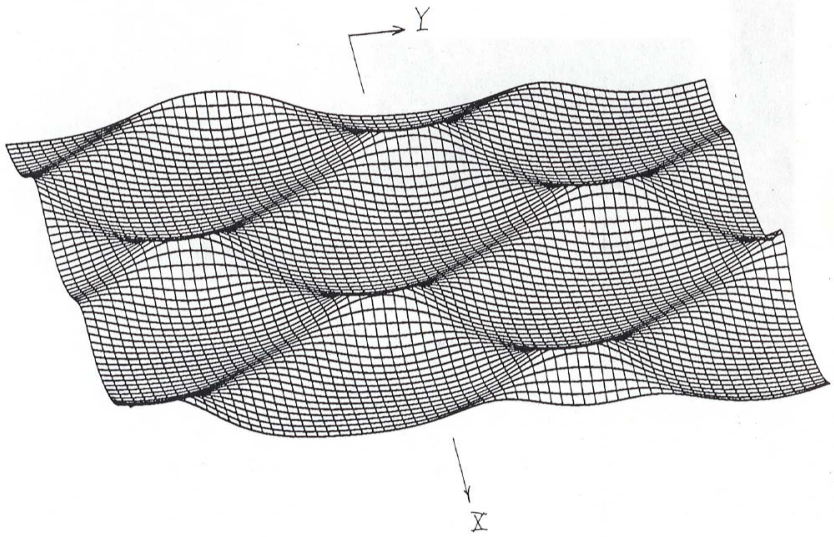


Figure 4: Two-phase solution of the KP equation propagating in the x direction with weak dispersion in the y direction [4].

Figure 5: Nearly periodic ocean waves with two-dimensional patterns along the shoreline of Duck, North Carolina.

Several open problems remain regarding these periodic wave patterns. Although it has been proved that they should exist in water of *any* depth [2], only the simplest patterns have ever been discovered. The KP equation admits other periodic solutions of permanent form with genus $G > 2$, which have not been explored yet. Another fundamental unsolved problem concerns the stability of these two-dimensional wave patterns of permanent form. A stability analysis of the patterns could determine whether one should expect to encounter them in nature or not, and how frequently. Another important problem is the effect of variable topography on the wave patterns, as is the case near shore. A better understanding of these problems could provide useful insight into the dynamics of surface waves in the ocean.

References

[1] A. Constantin and R. S. Johnson, *Propagation of very long water waves, with vorticity, over variable depth, with applications to tsunamis*, Fluid Dynamics Research, 40 (2008), pp. 175–211.

[2] W. Craig and D. P. Nicholls, *Traveling two and three dimensional capillary gravity water waves*, SIAM Journal on Mathematical Analysis, 32 (2000), pp. 323–359.

[3] J. Hammack, D. McCallister, N. Scheffner, and H. Segur, *Two-dimensional periodic waves in shallow water. Part 2. Asymmetric waves*, Journal of Fluid Mechanics, 285 (1995), pp. 95–122.

[4] J. Hammack, N. Scheffner, and H. Segur, *Two-dimensional periodic waves in shallow water*, Journal of Fluid Mechanics, 209 (1989), pp. 567–589.

[5] J. L. Hammack and H. Segur, *The Korteweg-de Vries equation and water waves. Part 2. Comparison with experiments*, Journal of Fluid Mechanics, 65 (1974), pp. 289–314.

[6] B. B. Kadomtsev and V. I. Petviashvili, *On the stability of solitary waves in weakly dispersing media*, in Sov. Phys. Dokl, vol. 15, 1970, pp. 539–541.

[7] I. Krichever, *Algebraic-geometric construction of the Zaharov-Sabat equations and their periodic solutions*, in Sov. Math. Dokl, vol. 17, 1976, pp. 394–397.

[8] G. B. Whitham, *Linear and nonlinear waves*, Wiley-Interscience, 1974.

Lecture 8: The Shallow-Water Equations

Lecturer: Harvey Segur. Write-up: Hiroki Yamamoto

June 18, 2009

1 Introduction

The shallow-water equations describe a thin layer of fluid of constant density in hydrostatic balance, bounded from below by the bottom topography and from above by a free surface. They exhibit a rich variety of features, because they have infinitely many conservation laws. The propagation of a tsunami can be described accurately by the shallow-water equations until the wave approaches the shore. Near shore, a more complicated model is required, as discussed in Lecture 21.

2 Derivation of shallow-water equations

To derive the shallow-water equations, we start with Euler's equations without surface tension,

$$\text{free surface condition :} \quad p = 0, \quad \frac{\mathrm{D}\eta}{\mathrm{D}t} = \frac{\partial \eta}{\partial t} + \boldsymbol{v} \cdot \nabla \eta = w, \quad \text{on } z = \eta(x, y, t) \tag{1}$$

$$\text{momentum equation :} \quad \frac{\mathrm{D}\boldsymbol{u}}{\mathrm{D}t} + \frac{1}{\rho}\nabla p + g\hat{z} = 0, \tag{2}$$

$$\text{continuity equation :} \quad \nabla \cdot \boldsymbol{u} = 0, \tag{3}$$

$$\text{bottom boundary condition :} \quad \boldsymbol{u} \cdot \nabla(z + h(x, y)) = 0, \quad \text{on } z = -h(x, y). \tag{4}$$

Here, p is the pressure, η the vertical displacement of free surface, $\boldsymbol{u} = (u, v, w)$ the three-dimensional velocity, ρ the density, g the acceleration due to gravity, and $h(x, y)$ the bottom topography (Fig. 1).

For the first step of the derivation of the shallow-water equations, we consider the global

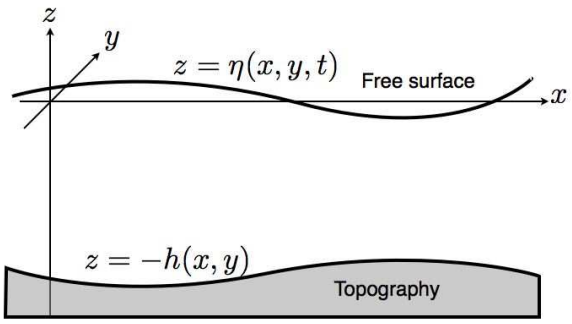


Figure 1: Schematic illustration of the Euler's system.

mass conservation. We integrate the continuity equation (3) vertically as follows,

$$
\begin{aligned}
0 &= \int_{-h}^{\eta} [\nabla \cdot \boldsymbol{u}] \mathrm{d}z, && (5)\\
&= \int_{-h}^{\eta} \left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right] \mathrm{d}z, && (6)\\
&= \frac{\partial}{\partial x} \int_{-h}^{\eta} u \mathrm{d}z - u|_{z=\eta} \frac{\partial \eta}{\partial x} + u|_{z=-h} \frac{\partial (-h)}{\partial x}, \\
&\quad + \frac{\partial}{\partial y} \int_{-h}^{\eta} v \mathrm{d}z - v|_{z=\eta} \frac{\partial \eta}{\partial y} + v|_{z=-h} \frac{\partial (-h)}{\partial y}, \\
&\quad + w|_{z=\eta} - w|_{z=-h}, && (7)\\
&= \frac{\partial}{\partial x} \int_{-h}^{\eta} u \mathrm{d}z - u|_{z=\eta} \frac{\partial \eta}{\partial x} + \frac{\partial}{\partial y} \int_{-h}^{\eta} v \mathrm{d}z - v|_{z=\eta} \frac{\partial \eta}{\partial y} + w|_{z=\eta}. && (8)
\end{aligned}
$$

where the bottom boundary condition (4) was used in the fourth row. With the surface condition (1), equation (8) becomes

$$
\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} \int_{-h}^{\eta} u \mathrm{d}z + \frac{\partial}{\partial y} \int_{-h}^{\eta} v \mathrm{d}z = 0. \tag{9}
$$

For the next step, we make the long-wave approximation, by assuming that the wave length is much longer than the depth of the fluid. However, we do not assume that perturbations have a small amplitudes, so that nonlinear terms are not neglected. Through the long-wave approximation, we can neglect the vertical acceleration term in (2), and deduce the hydrostatic pressure by integrating the vertical component of the momentum equation,

$$
\begin{aligned}
\int_{z}^{\eta} \frac{\partial p}{\partial z} \mathrm{d}z &= -\int_{z}^{\eta} \rho g \mathrm{d}z \\
p(x, y, \eta, t) - p(x, y, z, t) &= -\rho g (\eta(x, y, t)) - z) \\
p(x, y, z, t) &= \rho g (\eta(x, y, t) - z). && (10)
\end{aligned}
$$

where we used the surface condition $p(x, y, \eta, t) = 0$. Using this expression for the hydrostatic pressure (10) and further assuming that there are no vertical variations in (u, v), the

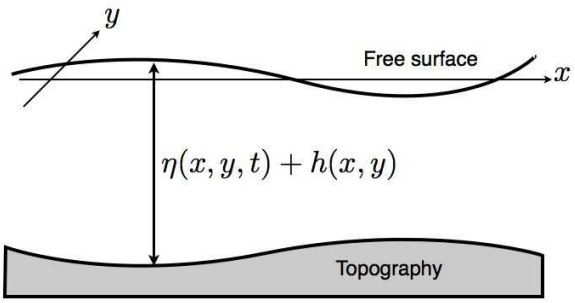


Figure 2: Schematic illustration of the shallow-water system.

horizontal momentum equations of the shallow-water system are obtained as follows,

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+g\frac{\partial \eta}{\partial x} = 0, \quad (11)$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+g\frac{\partial \eta}{\partial y} = 0. \quad (12)$$

The conservation of mass given by (9) becomes

$$\frac{\partial \eta}{\partial t}+\frac{\partial}{\partial x}[(\eta+h)u]+\frac{\partial}{\partial y}[(\eta+h)v]=0. \quad (13)$$

Then, equations (11), (12), and (13) are the shallow-water equations (Fig. 2).

These equations are similar to the equations for gas dynamics in 2-D. Indeed, the equations describing the dynamics of an inviscid, non-heat-conducting, isentropic (i.e. entropy is constant and $p \propto \rho^{\gamma}$) gas are (Kevorkian, 1990, chapter 3.3.4, [2])

$$\frac{\partial \rho}{\partial t}+\frac{\partial}{\partial x}(\rho u)+\frac{\partial}{\partial y}(\rho v) = 0, \quad (14)$$

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+\rho^{\gamma-2}\frac{\partial \rho}{\partial x} = 0, \quad (15)$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+\rho^{\gamma-2}\frac{\partial \rho}{\partial y} = 0, \quad (16)$$

where $\gamma \equiv C_p/C_v$ is the ratio of specific heats, C_p is the specific heat at constant pressure, and C_v is the specific heat at constant volume. These equations, (14), (15), and (16), are the exact analog of the shallow-water equations (13), (11), and (12) if we identify u, v in both cases, set $g(\eta+h)=\rho$, assume flat bottom topography (h = constant), and take $\gamma = 2$.

If we consider the shallow-water equations in a rotating frame (the rotation axis is perpendicular to x-y plane), the Coriolis term should be added to the momentum equation.

In that case (Vallis, 2006, chapter 3, [3]),

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}[(\eta + h)u] + \frac{\partial}{\partial y}[(\eta + h)v] = 0, \tag{17}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + g\frac{\partial \eta}{\partial x} = 0, \tag{18}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + g\frac{\partial \eta}{\partial y} = 0, \tag{19}$$

where f is the Coriolis parameter.

3 Mathematical structure

In this section, we explore the mathematical structure of the shallow-water equations.

3.1 Hyperbolic partial differential equations

The shallow-water equations can be written in the matrix form

$$\frac{\partial}{\partial t}\begin{pmatrix}\eta\\u\\v\end{pmatrix} + \begin{bmatrix}u & \eta+h & 0\\ g & u & 0\\ 0 & 0 & u\end{bmatrix}\frac{\partial}{\partial x}\begin{pmatrix}\eta\\u\\v\end{pmatrix} + \begin{bmatrix}v & 0 & \eta+h\\ 0 & v & 0\\ g & 0 & v\end{bmatrix}\frac{\partial}{\partial y}\begin{pmatrix}\eta\\u\\v\end{pmatrix} = -\begin{pmatrix}u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y}\\ 0\\ 0\end{pmatrix}. \tag{20}$$

The eigenvalues of the first coefficient matrix are

$$u, \quad u \pm \sqrt{g(\eta + h)}, \tag{21}$$

and those of the second coefficient matrix are

$$v, \quad v \pm \sqrt{g(\eta + h)}. \tag{22}$$

Since the eigenvalues (21) and (22) are real and distinct, the shallow-water equations are hyperbolic partial differential equations (PDEs).

The equations admit discontinuous (weak) solutions (see Kevorkian, 1990, chapter 5.3 for details). Such a discontinuity is called a "bore" and approximates a breaking wave. However, how do waves really break? This question is addressed in more detail in Lecture 21.

3.2 Method of characteristics

Because the shallow-water equations are hyperbolic PDEs, the method of characteristics can be applied to reduce them to a family of ordinary differential equations.

If we assume $v \equiv 0$ and $\frac{\partial}{\partial y} \equiv 0$ (i.e. 2-D in x-z plane), the shallow-water equations become,

$$\frac{\partial}{\partial t}(\eta + h) + u\frac{\partial}{\partial x}(\eta + h) + (\eta + h)\frac{\partial u}{\partial x} = 0, \tag{23}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + g\frac{\partial}{\partial x}(\eta + h) = g\frac{\partial h}{\partial x}. \tag{24}$$

If we define $c^2(x,y,t) = g(\eta + h)$, (23)$\times g$ can be written as

$$\frac{\partial c^2}{\partial t} + u\frac{\partial c^2}{\partial x} + c^2\frac{\partial u}{\partial x} = 0,$$

so

$$c\left[\frac{\partial (2c)}{\partial t} + u\frac{\partial (2c)}{\partial x} + c\frac{\partial u}{\partial x}\right] = 0. \tag{25}$$

Since $c(x,y,t) \neq 0$, (25) becomes

$$\frac{\partial (2c)}{\partial t} + u\frac{\partial (2c)}{\partial x} + c\frac{\partial u}{\partial x} = 0. \tag{26}$$

Likewise, (24) can be written as

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + c\frac{\partial (2c)}{\partial x} = g\frac{\partial h}{\partial x}. \tag{27}$$

Then (27) + (26) and (27) − (26) respectively give

$$\frac{\partial}{\partial t}(u+2c) + u\frac{\partial}{\partial x}(u+2c) + c\frac{\partial}{\partial x}(u+2c) = g\frac{\partial h}{\partial x}, \tag{28}$$

$$\frac{\partial}{\partial t}(u-2c) + u\frac{\partial}{\partial x}(u-2c) - c\frac{\partial}{\partial x}(u-2c) = g\frac{\partial h}{\partial x}. \tag{29}$$

Equation (28) states that along the curves in the (x,t) plane defined by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = u + c, \tag{30}$$

the quantity $u + 2c$ evolves according to

$$\frac{\partial}{\partial t}(u+2c) + \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial}{\partial x}(u+2c) = \frac{\mathrm{d}}{\mathrm{d}t}(u+2c) = g\frac{\partial h}{\partial x}. \tag{31}$$

And also along curves defined by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = u - c, \tag{32}$$

equation (29) can be written as

$$\frac{\partial}{\partial t}(u-2c) + \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial}{\partial x}(u-2c) = \frac{\mathrm{d}}{\mathrm{d}t}(u-2c) = g\frac{\partial h}{\partial x}. \tag{33}$$

Therefore, if h is constant, $(u + 2c)$ and $(u - 2c)$ are Riemann invariants (i.e. functions remain constant along the curves). If h is $h = mx + b$, then $\partial h/\partial x$ is constant and the characteristic equations can again be easily integrated, with this time $(u + 2c - gmt)$ and $(u - 2c - gmt)$ being the new Riemann invariants. For more complicated bottom topography more care must be taken. Finally, note that this method does not generalize to 3-D (x-y-t) problems.

3.3 Linearization of the shallow-water equations

To linearize the shallow-water equations, we consider small disturbances about a fluid at rest. That is,

$$\eta = 0 + \eta', \quad u = 0 + u', \quad v = 0 + v'. \tag{34}$$

By substituting (34) in the shallow-water equations (11), (12), (13) and neglecting the second-order terms, we obtain the linearized shallow-water equations as follows, (primes are omitted)

$$\begin{aligned} \frac{\partial \eta}{\partial t} + \frac{\partial (uh)}{\partial x} + \frac{\partial (vh)}{\partial y} &= 0, & (35)\\ \frac{\partial u}{\partial t} + g\frac{\partial \eta}{\partial x} &= 0, & (36)\\ \frac{\partial v}{\partial t} + g\frac{\partial \eta}{\partial y} &= 0. & (37) \end{aligned}$$

Multiplying (35) by $\sqrt{g}$, and both (36) and (37) by $\sqrt{h}$, we obtain,

$$\begin{aligned} \frac{\partial}{\partial t}(\eta\sqrt{g}) + \frac{\partial}{\partial x}(u\sqrt{h}\cdot\sqrt{gh}) + \frac{\partial}{\partial y}(v\sqrt{h}\cdot\sqrt{gh}) &= 0, & (38)\\ \frac{\partial}{\partial t}(u\sqrt{h}) + \sqrt{gh}\frac{\partial}{\partial x}(\eta\sqrt{g}) &= 0, & (39)\\ \frac{\partial}{\partial t}(v\sqrt{h}) + \sqrt{gh}\frac{\partial}{\partial y}(\eta\sqrt{g}) &= 0. & (40) \end{aligned}$$

If we eliminate $u\sqrt{h}$ and $v\sqrt{h}$ from the above equations, we obtain the linear 2-D wave equations,

$$\frac{\partial^2}{\partial t^2}(\eta\sqrt{g}) = \nabla\cdot[gh\cdot\nabla(\eta\sqrt{g})]. \tag{41}$$

The phase speed is, as expected from the long-wavelength limit discussed in Lecture 2, $c(x,y) = \sqrt{gh}$, although this time is explicitly derived for varying bottom topography (i.e. h is a function of both x and y).

The vorticity equation is derived by defining the vorticity as $\omega(x,y,t) = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$ and calculating $\frac{\partial}{\partial x}(40) - \frac{\partial}{\partial y}(39)$,

$$\frac{\partial \omega}{\partial t} = 0. \tag{42}$$

This means that the vorticity remains constant in time. In general, the velocity field (u, v) can be decomposed into rotational and irrotational parts, so u and v can be written as

$$\begin{aligned} u &= -\frac{\partial \psi}{\partial y} + \frac{\partial \phi}{\partial x}, & (43)\\ v &= \frac{\partial \psi}{\partial x} + \frac{\partial \phi}{\partial y}, & (44) \end{aligned}$$

where ψ and ϕ are called streamfunction and velocity potential, respectively. Using (43) and (44), equation (42) can be written as

$$\frac{\partial \omega}{\partial t} = \frac{\partial}{\partial t}(\nabla^2\psi) = 0. \tag{45}$$

Calculating $\frac{\partial}{\partial x}(39) + \frac{\partial}{\partial y}(40)$, we obtain

$$\frac{\partial}{\partial t}(\nabla^2\phi) + \sqrt{g}\nabla^2(\eta\sqrt{g}) = 0. \tag{46}$$

The rotational part of the velocity field is therefore obtained by integrating the time-independent vorticity field (45) (i.e. that of the initial conditions), while the irrotational part is obtained from the solution of (41) and (46). Nevertheless, ψ and ϕ are not entirely independent: they "interact" when applying the lateral boundary conditions on the total velocity field (43,44).

3.4 Tracking Vorticity in 2D

From the nonlinear shallow-water equations (11), (12), and (13), the nonlinear vorticity equation is derived as

$$\frac{\partial\omega}{\partial t} + \frac{\partial}{\partial x}(u\omega) + \frac{\partial}{\partial y}(v\omega) = 0. \tag{47}$$

This time, although $\partial\omega/\partial t \neq 0$, the total (integrated) vorticity

$$\iint \omega(x,y)\mathrm{d}x\mathrm{d}y$$

is conserved. Note that the vorticity equation can be written as

$$\frac{\partial\omega}{\partial t} + u\frac{\partial\omega}{\partial x} + v\frac{\partial\omega}{\partial v} = -\omega\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \tag{48}$$

and the equation (13) can be written as

$$\frac{\omega}{\eta+h}\left[\frac{\partial}{\partial t}(\eta+h) + u\frac{\partial}{\partial x}(\eta+h) + v\frac{\partial}{\partial y}(\eta+h)\right] = -\omega\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right). \tag{49}$$

From (48) and (49), we obtain

$$\frac{\partial\omega}{\partial t} + u\frac{\partial\omega}{\partial x} + v\frac{\partial\omega}{\partial v} = \frac{\omega}{\eta+h}\left[\frac{\partial}{\partial t}(\eta+h) + u\frac{\partial}{\partial x}(\eta+h) + v\frac{\partial}{\partial y}(\eta+h)\right], \tag{50}$$

which simplifies to

$$\frac{\partial}{\partial t}\left(\frac{\omega}{\eta+h}\right) + u\frac{\partial}{\partial x}\left(\frac{\omega}{\eta+h}\right) + v\frac{\partial}{\partial y}\left(\frac{\omega}{\eta+h}\right) = 0. \tag{51}$$

This illustrate that the quantity

$$q \equiv \frac{\omega}{\eta+h}, \tag{52}$$

is conserved by each fluid particle (water columns). It is called "potential vorticity" and plays a fundamental role in fluid dynamics. The equation (51) is called Ertel's theorem (1942) [1].

Moreover, let $G(\zeta)$ be any differentiable function. Given that,

$$\frac{\partial G(\zeta)}{\partial t} = \frac{\mathrm{d}G(\zeta)}{\mathrm{d}\zeta}\frac{\partial \zeta}{\partial t}, \tag{53}$$

we obtain

$$\begin{aligned}\frac{\partial G(q)}{\partial t} + u\frac{\partial G(q)}{\partial x} + v\frac{\partial G(q)}{\partial y} &= \frac{\mathrm{d}G(q)}{\mathrm{d}q}\frac{\partial q}{\partial t} + u\frac{\mathrm{d}G(q)}{\mathrm{d}q}\frac{\partial q}{\partial x} + v\frac{\mathrm{d}G(q)}{\mathrm{d}q}\frac{\partial q}{\partial y},\\ &= \frac{\mathrm{d}G(q)}{\mathrm{d}q}\left[\frac{\partial q}{\partial t} + u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y}\right],\\ &= 0. \end{aligned} \tag{54}$$

Therefore any smooth function of q is also conserved by each water column.

From (47) and (54), we obtain

$$\frac{\partial}{\partial t}[\omega \cdot G(q)] + \frac{\partial}{\partial x}[u\omega \cdot G(q)] + \frac{\partial}{\partial y}[v\omega \cdot G(q)] = 0, \tag{55}$$

which means that there are infinitely many conservation laws, because G can be chosen from arbitrary differential functions.

4 Applications: Tsunami

The linearized shallow-water equation derived in the previous section

$$\frac{\partial^2}{\partial t^2}(\eta\sqrt{g}) = \nabla[gh \cdot \nabla(\eta\sqrt{g})], \tag{56}$$

is a good model for the propagation of tsunami across the open ocean, away from shore. Indeed, as seen in Lecture 7, the typical depth of the ocean is about 4 km while the wave length of tsunami is about 100 km. This equation shows that the local speed of propagation of the Tsunami, in any direction, is

$$c_p = c_g = \sqrt{gh(x,y)}, \tag{57}$$

with no dispersion.

Closer to the shore, the wave compresses horizontally and grows vertically, so the linear approximation is no longer valid. The nonlinear shallow-water equations being hyperbolic, they allow for wave breaking. However, the breaking mechanism is presumably much more complex than that captured by shallow-water equations. This topic is addressed in more detail in Lecture 21.

5 Summary so far

Let us complete this lecture by summarizing the various equations and respective assumptions made so far. In Lecture 1, we presented the full nonlinear surface water-wave equations. In Lecture 2, we considered the linear approximation to these equations in the case of flat

bottom topography, but allowed for waves of any horizontal wavelength. In the following lectures (4-7), we dropped the linear approximation, and considered the weakly non-linear case. However, in order to do this we had to focus on long-wavelength dynamics only, with flat-bottom topography. In this last lecture, we changed approach to consider the effect of varying bottom topography, and derived the fully nonlinear shallow-water equations (and its linearized counterpart). These equations also assume long-wavelength perturbations. However, an additional approximation had to be made in this case, in assuming that the vertical structure of the flow is entirely uniform ($\partial u/\partial z = 0 = \partial v/\partial z$).

References

[1] H. Ertel, *Ein neuer hydrodynamischer wirbelsatz (a new hydrodynamic eddy theorem)*, Meteorolol. Z., 59 (1942), pp. 277–281.

[2] J. Kevorkian, *Partial Differential Equations: Analytical Solutions Techniques*, Chapman & Hall, New York, 1990.

[3] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, Cambridge University Press, Cambridge, U.K., 2006.

Lecture 9 - Nonlinear waves in a variable medium

Lecturer: Roger Grimshaw. Write-up: Hélène Scolan

June 17, 2009

1 Introduction

The usual Korteweg-De-Vries equation, which assumes a uniform background state, is not sufficient to describe internal solitary waves in the coastal ocean. Indeed, the topography can vary horizontally, and the waves produced are not clean wave trains. This can be seen for example in the measurements of currents in the Australian Northwest Shelf reproduced in Figure 1.

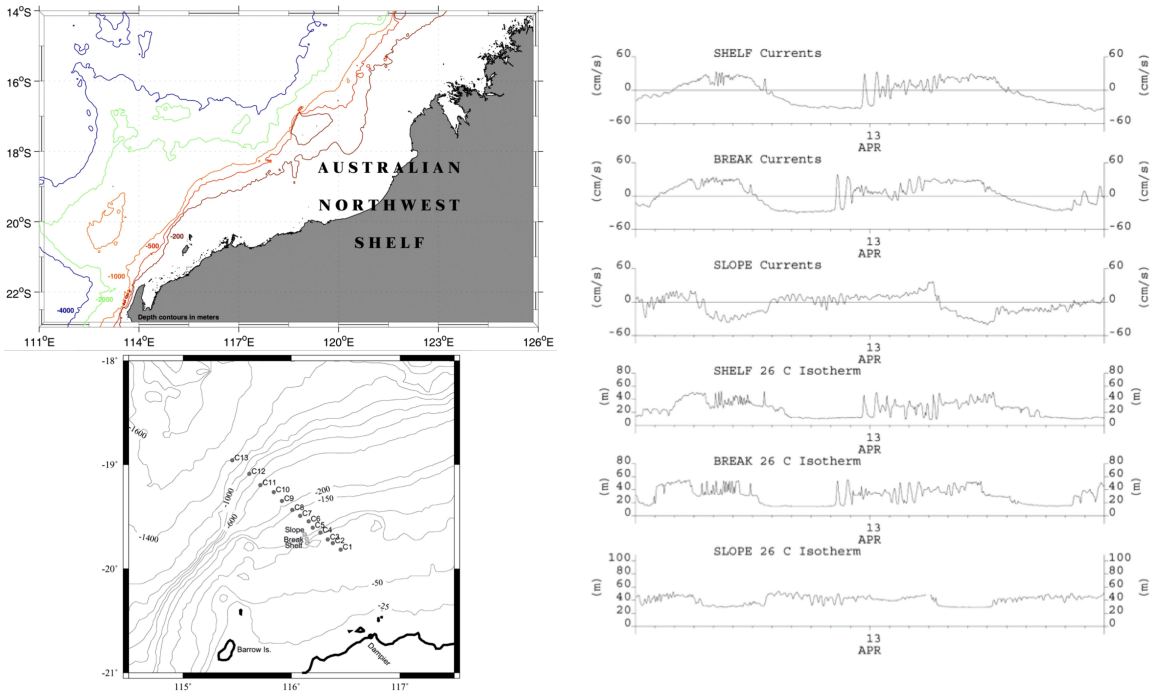


Figure 1: Time series of isotherm displacements and onshore currents are shown from 3 moorings, (Slope, Break and Shelf), located in 78 to 109 m water depths, and a few kilometers apart at the outer edge of the Australian Northwest continental shelf. The plots show a variety of nonlinear wave forms including bores on both the leading and trailing faces of the long internal tide, as well as short period (approximately 10 minutes, close to the buoyancy period) internal solitary waves. [After Holloway and Pelinovsky, 2001]. *An Atlas of Oceanic Internal Solitary Waves (February 2004) by Global Ocean Associates Prepared for Office of Naval Research Code 322 PO*

By incorporating a variable medium in the model, one can build the variable coefficient Korteweg-de Vries equation and find asymptotic and numerical solutions of the problem.

2 Waves in inhomogeneous medium

2.1 Linear waves and wave-action conservation law

First, we recall the properties of linear waves propagating through an inhomogeneous medium. Because of the presence of the variable background, the usual wave equation is modified. In many examples, the wave equation becomes

$$u_{tt} - (c^2(x)u_x)_x = 0 \tag{1}$$

where the wave speed $c(x)$ varies with position (for example $c(x) = \sqrt{gh(x)}$ for waves supported by the shallow-water equations, see Lecture 8).

We assume that the medium is "slowly varying" which means that the lengthscale L over which the medium changes is greater than the typical wavelength λ ($\lambda \ll L$) and so the coefficient $c(x)$ can be considered almost constant on the wave scale ($c = c(\epsilon x)$ with $\epsilon \ll 1$). The WKB approximation consists in looking for a solution close to the solution for a homogeneous medium $a\exp(-i\omega(t - x/c))$. Using an ansatz of the form $a(x,t)\exp(-i\omega(t - \tau(x)))$ with $\tau(x) = \int \frac{\mathrm{d}x}{c(x)}$, and developing an asymptotic expansion in the powers of ϵ, it can be shown that $a(x,t) \propto \frac{1}{\sqrt{c(x)}}$ (proof in [5] for example).

More generally (i.e. for wave equations not necessarily in the form of (1)), the WKB asymptotic solution can be written: $u \approx a(x,t)f(t - \tau(x)))$ where $\tau(x) = \int \frac{\mathrm{d}x}{c(x)}$ and where the phase $t - \tau(x)$ is assumed to vary rapidly compared with the amplitude function $a(x,t)$ and the speed $c(x)$. Then it can be shown that $a(x,t)$ verifies:

$$(a^2)_t + (ca^2)_x = 0 \tag{2}$$

This equation is called the wave action conservation law for wavetrains in slowly spatially varying medium.

The most general form of the wave action conservation law, for waves propagating in a non-uniform, time-dependent medium which may also sustain a mean flow $\mathbf{U}$ is

$$\left(\frac{E}{\hat{\omega}}\right)_t + \left(\frac{E}{\hat{\omega}}c_g\right)_x = 0, \tag{3}$$

where E is the wave energy density (which is usually related, but no necessarily equal to the square of the wave amplitude), $\hat{\omega}$ is the intrinsic frequency defined by

$$\hat{\omega} = \omega - \mathbf{k}\cdot\mathbf{U},$$

(i.e. the frequency of the wave seen by an observer moving with a mean flow if there is one) and c_g the group velocity. For more information about this conservation law see [2]. The quantity $E/\hat{\omega}$ is called the wave action density, so that the wave action flux is $\frac{E}{\hat{\omega}}c_g$ ($c_g \approx c$ if the medium is weakly dispersive).

This final equation can be interpreted in the following way: in the limit where $\hat{\omega}$ is constant (i.e the medium does not vary with time or there is no mean flow) then the wave action conservation law reduces to an energy conservation law. If $\hat{\omega}$ is not constant, then energy density is not conserved, but the wave action density $E/\hat{\omega}$ is.

2.2 Non linear waves: variable-coefficient KDV equation

To describe nonlinear internal waves in a variable medium, we begin with the basic non linear KdV equation:

$$u_t + cu_x + \mu u u_x + \beta u_{xxx} = 0 \tag{4}$$

in which we introduce the possibility of a variable background. Thus the linear phase speed and the coefficients μ and β have a spatial dependency. Furthermore, another term is needed if we want the variable-coefficient KdV equation to verify the general wave-action conservation law (3) in the limit where non-linear terms are negligible:

$$u_t + c(x)u_x + \frac{cQ_x}{2Q}u + \mu(x)uu_x + \beta(x)u_{xxx} = 0. \tag{5}$$

This additional term is written so that in the linear wave theory $u_t + c(x)u_x + \frac{cQ_x}{2Q}u = 0$ is transformed into $(\frac{Qu^2}{c})_t + (Qu^2)_x = 0$. Then, Qu^2 can be interpreted as the wave action flux (and so the wave-action density is $\frac{Qu^2}{c}$ since $c \approx c_g$ in the long wave/weak dispersion limit). The exact expression for Q depends on the original physical problem considered.

To maintain the balance between terms in the new equation including the effect of variable medium (5), we need the dispersion term, the non-linearity and the weak inhomogeneity term to be of the same order of magnitude. If $\frac{\partial}{\partial x} \sim \epsilon \ll 1$ and we suppose $u \sim \epsilon^a$ and $\frac{Q_x}{Q} \sim \epsilon^b$, the terms will be of the same order of magnitude if $a+b = 2a+1 = a+3$ which gives $a = 2$ and $b = 3$. So $\frac{Q_x}{Q}$ scales as ϵ^3. This implies that the variable-medium KdV is only valid in the limit where the medium varies very slowly compared with the horizontal scale of the wave.

As in the homogeneous KdV, it is useful to recast the governing equations in a moving coordinate system which follows the propagation of the wave, i.e. perform a change of variable in which $\xi \sim x/c - t$ where c is the phase speed. Here, the procedure is slightly more complex since c may vary with position. By analogy with the WKB approximation technique, we introduce

$$\tau = \int_0^x \frac{dx'}{c(x')},$$

This change of coordinate can be viewed as a mapping of the original spatial coordinate into a time-like coordinate, since the new variable τ is simply the travel time between the original position of the wave and its present location. If we also use the change of variable $X = \tau - t$, we then have

$$u_t = -u_X \tag{6}$$

$$u_x = \frac{u_X}{c} + \frac{u_\tau}{c}. \tag{7}$$

When the background varies very slowly compared with the size of the wave, it can be demonstrated that $\partial/\partial\tau \ll \partial/\partial X$, so that

$$u_x \simeq \frac{u_X}{c}. \tag{8}$$

Then, within the balance seen before the equation can be written:

$$u_\tau + \frac{Q_\tau}{2Q}u + \frac{\mu}{c}uu_X + \frac{\beta}{c^3}u_{XXX} = 0. \tag{9}$$

The two equations (5) and (9) are asymptotically equivalent and differ just by terms of $O(\epsilon^7)$. It is interesting and important to note that the coefficients μ, β and c now vary with the time-like variable τ. Physically this simply models the fact that as the wave propagates through the inhomogeneous medium, it "sees" a slowly time-dependent, but nearly homogeneous background around itself.

The more commonly used form of the KdV equation is obtained by putting $A = Q^{1/2}u$ which gives the variable-coefficient KdV equation:

$$A_\tau + \alpha AA_X + \lambda A_{XXX} = 0 \qquad \text{(vKDV)}$$

where $\alpha = \dfrac{\mu}{c\sqrt{Q}}$ and $\lambda = \dfrac{\beta}{c^3}$.

It can be verified that the variable-coefficient KdV has two conservation laws:

$$\frac{d}{d\tau}\int_{-\infty}^{\infty} A\mathrm{d}X = 0, \tag{10}$$

$$\frac{d}{d\tau}\int_{-\infty}^{\infty} A^2\mathrm{d}X = 0. \tag{11}$$

They are often referred as conservation of "mass" and momentum even if these are not the physical ones. The latter equation is in fact the conservation of the wave action flux. The former is asymptotically that for the physical mass.

3 Slowly varying periodic waves

3.1 Asymptotic expansion

As we suppose that the medium is slowly varying, we write $\alpha = \alpha(T)$ and $\lambda = \lambda(T)$ with $T = \sigma\tau$ with $\sigma \ll 1$. We can develop a multiscale expansion in powers of the small parameter σ for a modulated periodic wave by looking for solutions of the kind:

$$A = A_0(\theta, T) + \sigma A_1(\theta, T) + ...$$

where A is periodic in the phase $\theta = k(X - \frac{1}{\sigma}\int^T V(T)dT)$ with a fixed period of 2π, where k is a fixed constant and V remains to be determined.

As $\frac{\partial A_i}{\partial \tau} = \frac{\partial A_i}{\partial \theta}\frac{\partial \theta}{\partial \tau} + \frac{\partial A_i}{\partial T}\frac{\partial T}{\partial \tau} = A_{i\theta}\cdot(-\frac{k}{\sigma}V(T)) + A_{iT}\cdot\sigma$ and $\frac{\partial A_i}{\partial X} = kA_{i\theta}$, the expansion introduced in the vKDV gives:

$$O(\sigma^0): \qquad -VA_{0\theta} + \alpha A_0 A_{0\theta} + \lambda k^2 A_{0\theta\theta\theta} = 0 \tag{12}$$

$$O(\sigma): \qquad -VA_{1\theta} + \alpha(A_0A_1)_\theta + \lambda k^2 A_{1\theta\theta\theta} = -\frac{1}{k}A_{0T} \tag{13}$$

These are ordinary differential equation in θ with T as a parameter. A solution of (12) can be written:

$$A_0 = a\{b(m) + \mathrm{cn}^2(\gamma\theta; m)\} + d,$$

where

$$b = \frac{1-m}{m} - \frac{E(m)}{mK(m)}, \quad \alpha a = 12m\lambda\gamma^2k^2, \quad V = \alpha d + \frac{\alpha a}{3}\left\{\frac{2-m}{m} - \frac{3E(m)}{mK(m)}\right\}.$$

This solution is a typical *cnoidal wave* $\mathrm{cn}(x; m)$, which is a Jacobian elliptic function of modulus m $(0 < m < 1)$. The functions $K(m)$ and $E(m)$ are the elliptic integrals of the first and second kind, a is the amplitude and d is the mean value of A over one period $\gamma = K(m)/\pi$, and the spatial period is $2\pi/k$.

This solution contains three free parameters which depend on T: for example the amplitude a, the mean level d and the modulus m. We can consider the two limit cases:

- $m \to 1$: This is the solitary wave case.

  Indeed, $b \to 0$ and $\mathrm{cn}^2(x) \to \mathrm{sech}^2(x)$. $\gamma \to \infty$ and $k \to 0$ with $\gamma k = K$ held fixed.

- $m \to 0$: This gives sinusoidal waves of small amplitude $a \sim m$ and wavenumber k.

3.2 Modulation equations

To completely describe the solution we must now find how a, d and m depend on the slow variable T. There are two methods for this: the Whitham avering method or the asymptotic expansion continued at higher level.

3.2.1 Whitham averaging method

The method is developed as follows:

- Step 1: Determine the three conservation laws for the vKDV equation.
- Step 2: Insert the periodic cnoidal wave into the conservation laws.
- Step 3: Average over the phase θ.

Conservation laws: We already have the mass and momentum conservation laws (10) and (11):

$$\frac{\partial}{\partial T}\int_0^{2\pi} A\mathrm{d}\theta = 0 \quad \text{and} \quad \frac{\partial}{\partial T}\int_0^{2\pi} A^2\mathrm{d}\theta = 0. \tag{14}$$

Since we are dealing with slowly varying waves, an additional conservation law is derived from the law of "conservation of waves" (or "conservation of crests"). Indeed for slowly varying waves we have the definitions $k = \frac{\partial\theta}{\partial X}$ and $\omega = -\frac{\partial\theta}{\partial t}$ so that:

$$k_T + \omega_X = 0.$$

Since ω does not depend on X, k is constant.

Substitution of cnoidal wave into conservation laws.

- The mass equation implies that the mean level d is constant.
- After averaging over θ, the momentum equation produces a relationship between a and m:

$$a^2\left\{\frac{1}{2\pi}\int_0^{2\pi}\mathrm{cn}^4(\gamma\theta;m)\mathrm{d}\theta - b(m)^2\right\} = \text{constant}$$

$$\Rightarrow\ \frac{a^2}{m^2}\left\{(2-3m)(1-m) + \frac{(4m-2)E(m)}{K(m)} - 3m^2b(m)^2\right\} = \text{constant}$$

which uniquely determines the evolution of the modulus m:

$$F(m) \equiv K(m)^2\{(4-2m)E(m)K(m) - 3E(m)^2 - (1-m)K(m)^2\} = \text{constant}\frac{\alpha^2}{\lambda^2}$$

since α and λ vary with T in a known way.

The function $F(m)$, as seen in Figure 2 for example, is usually a monotically increasing function of m so if α/λ increases, m increases too. This implies that if the dispersive coefficient λ tends to zero then m tends to 1 and the waves become more like solitary waves.

For example for water waves, $c = \sqrt{gh}$, $Q = c$, $\mu = 3c/2h$ and $\beta = ch^2/6$ which leads to $\alpha/\lambda \propto h^{-9/4}$ and $F(m) \propto h^{-9/2}$. As the wave approaches the beach $h \to 0$ and $m \to 1$, which means that the wave gradually transforms into a solitary wave. It's amplitude goes as $h^{-3/4}$ so that the surface elevation varies as h^{-1}.

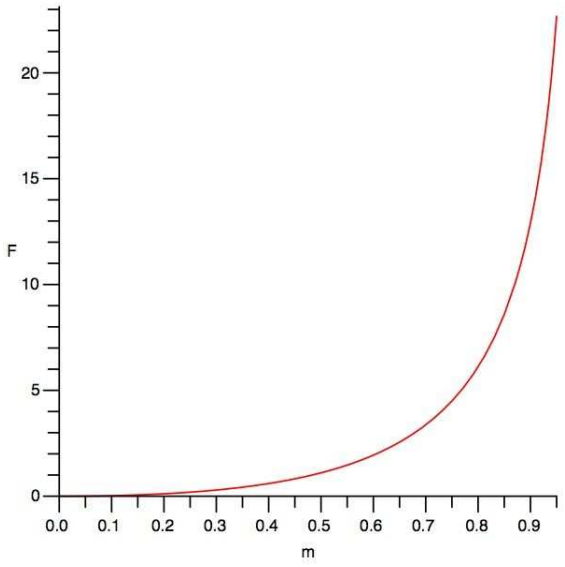


Figure 2: $F(m)$ in the case of water waves.

3.2.2 Asymptotic expansion continued

To find the conservation laws, we can also alternatively continue the method of asymptotic expansion to the next order. As A_1 must be periodic, we must force the right-hand side of (13) to be orthogonal to the periodic solutions of the adjoint to the homogeneous operator on the left-hand side (see the discussion for this point in Lecture 6).

Indeed let us define the operator L as

$$L = -V\frac{\partial}{\partial\theta} + \alpha(A_0.\ )_\theta + \lambda k^2\frac{\partial}{\partial\theta^3}.$$

Then (13) can be written: $L(A_1) = F$ where $F = -\frac{A_{0T}}{k}$.

By definition, the adjoint L^* verifies for any periodic function B and A_1:

$$< BL(A_1) >=< A_1L^*(B) >$$

where $<\quad>= \int_0^{2\pi} d\theta$. Thus if B is a solution of $L^*(B) = 0$, $< BF >=< BL(A_1) >=< A_1L^*(B) >= 0$.

The adjoint equation $L^*(B) = 0$ is :

$$-VB_\theta + \alpha A_0B_\theta + \lambda k^2B_{\theta\theta\theta} = 0$$

$B = 1$ and $B = A_0$ are two periodic solutions of this equation. A third solution can be found but is not periodic. So we have two conditions $< 1(-\frac{A_{0T}}{k}) >= 0$ and $< A_0(-\frac{A_{0T}}{k}) >= 0$ which coincide to the statement that d is constant and to the momentum conservation law equation (14).

4 Slowly varying solitary waves

It is important to note that the results for a slowly-varying periodic wave cannot directly be extrapolated to the solitary-wave case: indeed, the limit $m \to 1$ requires $k \to 0$ and so the period becomes infinite. Accordingly, the condition that the local period $(1/kV)$ should be much smaller than the scale of the variable medium $(1/\sigma)$ is no longer satisfied.

Thus, we must refine the definition of "slowly-varying" for the case of solitary waves. The solitary wave will be considered slowly-varying if the half-width is much less than the scale of the variable medium $(1/\sigma)$. An asymptotic expansion can then be developed in the same way as before but with a new expression for the phase:

$$\phi = X - \frac{1}{\sigma}\int^{T} V(T)dT.$$

A is not required to be periodic in ϕ, and is defined in $-\infty < \phi < \infty$ and bounded in $\phi \to \pm\infty$. Without changing the problem we can choose $\lambda > 0$ so that small-amplitude waves propagate in the negative x-direction (a transposition A, x with $-A$, $-x$ gives the other side). We can also assume $A \to 0$ as $\phi \to \infty$ without imposing anything in the other boundary condition as $\phi \to -\infty$.

The resulting ODEs of the asymptotic expansion are:

$$-VA_{0\phi} + \alpha A_0 A_{0\phi} + \lambda A_{0\phi\phi\phi} = 0 \tag{15}$$

$$-VA_{1\phi} + \alpha(A_0 A_1)_\phi + \lambda A_{1\phi\phi\phi} = -A_{0T}. \tag{16}$$

The solution A_0 is now a solitary wave: $A = a\mathrm{sech}^2(K\phi)$ with $V = \frac{\alpha a}{3} = 4\lambda K^2$ and only has one free parameter (a for instance). A background d can be added, but is constant and can be removed by a Galilean transformation.

At the next order, we require that $A_1 \to 0$ as $\phi \to \infty$. This imposes a new compatibility equation $< B(-A_{0T}) >= 0$ where $<>= \int_{-\infty}^{\infty} \mathrm{d}\phi$ and with the adjoint equation $L^*(B) = 0$:

$$-VB_\phi + \alpha A_0 B_\phi + \lambda B_{\phi\phi\phi} = 0.$$

Among the two possible bounded solutions $B = 1$ and $B = A_0$, only the latter satisfies the condition $A_1 \to 0$ as $\phi \to \infty$. So there is only one orthogonality condition which can be imposed which corresponds to the right-hand side of (16) being orthogonal to A_0 ie $< A_0.(-A_{0T}) >= 0$ ie :

$$\frac{\partial}{\partial T}\int_{-\infty}^{\infty} A_0^2 \mathrm{d}\phi = 0. \tag{17}$$

Even though there is only one equation this time, it is enough to determine the evolution of the free-parameter a. Substituting the sech^2 solution into the condition (17) yields

$$a^3 = \text{constant}\frac{\alpha}{\lambda}$$

which agrees with the limit $m \to 1$ of the periodic wave case.

4.1 Trailing shelf

A problem nevertheless occurs with the preceeding derivation since the vKDV equation has two conservation laws (momentum and mass) whereas only one condition can be imposed (17), which happens to coincide with the momentum equation (11). This means that for solitary waves we can not simultaneously require conservation of the total mass and momentum. This can also be seen by examining the solution of (16) for A_1: indeed if we integrate in ϕ, with the boundary condition $A_1 \to 0$ as $\phi \to \infty$ and $A_1 \to H_1$ as $\phi \to -\infty$ and using the properties of A_0 we get:

$$VH_1 = -\frac{\partial M_0}{\partial T} \quad \text{where} \quad M_0 = \int_{-\infty}^{\infty} A_0 \, \mathrm{d}\phi \quad \text{and} \quad H_1 = \frac{6}{\alpha K}\frac{a_T}{a} \tag{18}$$

which illustrates how the "total mass" changes as the solitary wave propagates.

The solution to this problem consists in constructing a "trailing shelf" A_s such that $A = A_0 + A_s$. A_s is of small amplitude $O(\sigma)$ but with a long length-scale $O(1/\sigma)$ which has $O(1)$ mass but $O(\sigma)$ for the momentum. It is located behind the solitary wave and to leading order has a value independent of T so that $A_s = \epsilon A_s(X)$ with $X = \sigma x$ for $X < \phi(T) = \int^T V(T)dT$.

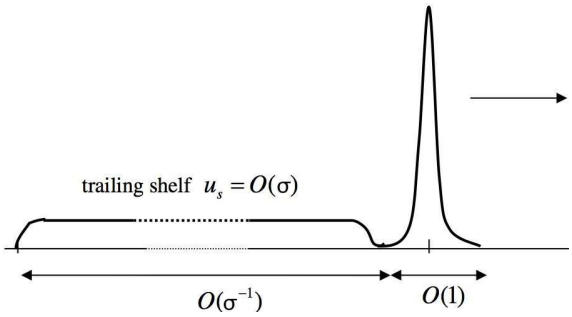


Figure 3: Trailing shelf residing behind the solitary wave.

The trailing shelf is determined by its value at the location $X = \phi(T)$ of the solitary wave, in particular $A_s(\phi(T)) = H_1(T)$. It can have a negative or positive polarity depending on the sign of λa_T and so on the growth or decay of the wave amplitude. It may be verified that the slowly-varying solitary wave and the trailing shelf together satisfy conservation of mass. Continuing the expansion to higher orders in σ reveals how the shelf itself evolves and generates secondary solitary waves.

4.2 Critical case

If we reconsider the expression for the free parameter a:

$$a^3 = \text{constant}\frac{\alpha}{\lambda}$$

we see that there is a critical point when $\alpha = 0$ where we may expect a dramatic change in the wave structure. Indeed, the wave amplitude goes to 0 if $\alpha \to 0$, and decreases as $|\alpha|^{1/3}$

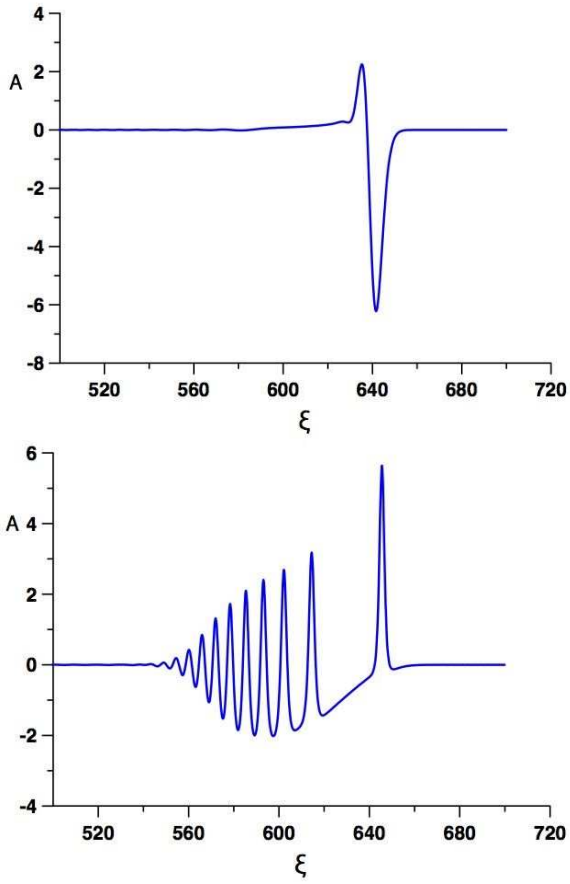


Figure 4: $\lambda = 1$ and α varies from -1 to 1. Upper panel: $\alpha = 0$. Lower panel: $\alpha = 1$

while the mass M_0 of the solitary wave only grows as $|\alpha|^{-1/3}$. Meanwhile, the amplitude A_s of the trailing shelf grows as $|\alpha|^{-8/3}$ with the opposite polarity of the wave.

Essentially the trailing shelf passes through the critical point as a disturbance of the opposite polarity to that of the original solitary wave, which then being in an environment with the opposite sign of α, can generate a train of solitary waves of the opposite polarity, riding on a pedestal of the same polarity as the original wave. Figure 4 shows for instance the possibility of conversion of a depression wave (with a positive shelf) into a train of elevation waves riding on a negative pedestal. The mean level of the new wave-train is negative corresponding to the initial negative mass of the depression wave.

References

[1] *Magic lectures on nonlinear waves.* http://www.maths.dept.shef.ac.uk/magic/course.php?id=21. (Mathematics Access Grid Instruction and Collaboration).

[2] F. Bretherton and C. Garrett, *Wavetrains in inhomogeneous moving media.*, Proc. Roy. Soc., 302 (1968), pp. 529–54.

[3] R. Grimshaw, *Internal solitary waves in a variable medium.*, Gesellschaft fur Angewandte Mathematik, 30, pp. 96–109.

[4] R. Grimshaw, E. Pelinovsky, and TalipovaT., *Modeling internal solitary waves in the coastal ocean..*, Surveys in Geophysics, 28 (2007), pp. 273–298.

[5] J. Simmonds and M. J.E., *A first look at perturbation theory*, Dover Publications, Inc. Mineola, New York, 1997.

Lecture 10: Whitham Modulation Theory

Lecturer: Roger Grimshaw. Write-up: Andong He

June 19, 2009

1 Introduction

The Whitham modulation theory provides an asymptotic method for studying *slowly varying periodic waves*, and is essentially a nonlinear WKB theory. Equations are derived which describe the slow evolution of the governing parameters for these nonlinear periodic waves (such as the amplitude, wavelength, frequency, etc.), and are called the *modulation (or Whitham) equations*. The Whitham equations have a remarkably rich mathematical structure, and are at the same time a powerful analytic tool for the description of nonlinear waves in a wide variety of physical contexts. One of the most important aspects of the Whitham theory is the analytic description of the formation and evolution of dispersive shock waves, or undular bores. These are coherent nonlinear wave-structures which resolve a wave-breaking singularity when it is dominated by dispersion rather than by dissipation. There are also a number of important connections between the Whitham theory, the inverse scattering transform (IST), and the general theory of integrable hydrodynamic systems.

Figure 1: Satellite view of a Morning Glory wave

2 KdV Cnoidal waves

Let us consider the Korteweg de Vries (KdV) equation as an example, namely,

$$u_t + 6uu_x + u_{xxx} = 0\,. \tag{1}$$

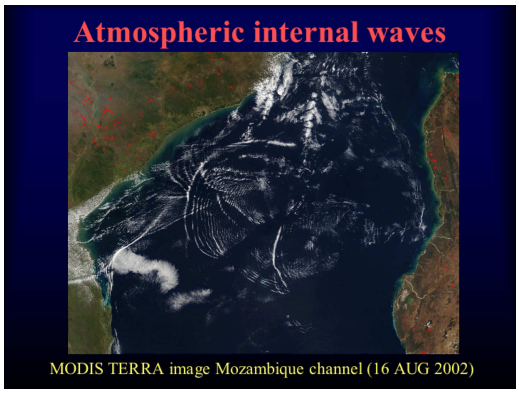


Figure 2: Atmospheric solitary waves near Mozambique

Figure 3: Undular Bore on the Dordogne river

As seen in many of the previous lectures, this equation has a one-phase periodic traveling wave solution, which is also called the *cnoidal wave* of the KdV equation (1):

$$u(x,t) = r_2 - r_1 - r_3 + 2(r_3 - r_2)\mathrm{cn}^2(\sqrt{r_3 - r_1}\,\theta; m)\,, \tag{2}$$

where $\mathrm{cn}(y;m)$ is the Jacobi elliptic cosine function defined as

$$\mathrm{cn}(y;m) = \cos\phi,$$

with ϕ satisfying

$$y = \int_0^{\phi} \frac{dt}{\sqrt{1 - m^2\sin^2 t}}.$$

This form of the wave solution is slightly different from the one introduced in previous lectures. Here the three parameters are $r_1 \le r_2 \le r_3$, and the phase variable θ and the modulus $m \in (0,1)$ are related to these parameters in the following way:

$$\theta = x - Vt\,, \quad V = -2(r_1 + r_2 + r_3)\,, \tag{3}$$

$$m = \frac{r_3 - r_2}{r_3 - r_1}\,, \quad L = \oint d\theta = \frac{2K(m)}{\sqrt{r_3 - r_1}}, \tag{4}$$

where $K(m) = \int_0^{\frac{\pi}{2}} \frac{dt}{\sqrt{1-m^2\sin^2 t}}$ is the complete elliptic integral of the first kind, and L is the "wavelength" along the x-axis. When $m \to 1$, $\mathrm{cn}(y;m) \to \mathrm{sech}(y)$ and (2) becomes a solitary wave; when $m \to 0$, $\mathrm{cn}(y;m) \to \cos(y)$ reducing to a sinusoidal wave (see [6]).

It is sometimes advantageous to use the parameters r_1, r_2 and r_3 instead of more physical parameters (such as the amplitude, speed, wavelength etc.), as they arise directly from the basic ordinary differential equation for the KdV traveling wave solution (2). Indeed, if we substitute (2) into the KdV equation (1), and integrate it once we get

$$u_\theta^2 = -2u^3 + Vu^2 + Cu + D\,, \tag{5}$$

where C, D are constants. This can be transformed to

$$w_\theta^2 = -4\mathbb{P}(w), \tag{6}$$

where

$$w = \frac{u}{2} - \frac{V}{4}\,, \quad \text{and} \quad \mathbb{P}(w) = \prod_{i=1}^{3}(w - r_i)\,.$$

This means that the cnoidal wave (2) is parameterized by the zeros r_1, r_2, r_3 of the cubic polynomial $\mathbb{P}(w)$. In a modulated periodic wave, the parameters r_1, r_2, r_3 are slowly varying functions of x and t, described by the Whitham modulation equations. These can be obtained either by a multi-scale asymptotic expansion, or more conveniently by averaging conservation laws, as described below.

3 Averaged conservation laws

Let us introduce an average over the period of the cnoidal wave (2) as

$$\langle \mathfrak{F} \rangle = \frac{1}{L}\oint \mathfrak{F} d\theta = \frac{1}{L}\int_{r_2}^{r_3} \frac{\mathfrak{F} d\mu}{\sqrt{-\mathbb{P}(\mu)}}\,, \tag{7}$$

for any function $\mathfrak{F}(x,t)$ such that (7) is finite. It can then be shown that

$$\langle u \rangle = 2(r_3 - r_1)\frac{E(m)}{K(m)} + r_1 - r_2 - r_3\,, \tag{8}$$

$$\langle u^2 \rangle = \frac{2}{3}[V(r_3 - r_1)\frac{E(m)}{K(m)} + 2Vr_1 + 2(r_1^2 - r_2 r_3)] + \frac{V^2}{4}\,, \tag{9}$$

where $E(m) = \int_0^{\pi/2} \sqrt{1 - m^2\sin^2\theta}\, d\theta$ is the complete elliptic integral of the second kind, and $K(m)$ was described earlier.

Next, recall that the KdV equation has a set of conservation laws (see Lecture 4):

$$(P_j)_t + (Q_j)_x = 0\,, \quad j = 1, 2, 3\,, \tag{10}$$

The averaging (7) procedure, when applied to these laws, yields equations of the kind:

$$\langle P_j \rangle_t + \langle Q_j \rangle_x = 0, \quad j = 1, 2, 3\,. \tag{11}$$

When combined with (8) and (9), the system (11) then describes the slow evolution of the parameters r_j for the cnoidal wave (2).

The first two conservation laws of the KdV equation (1) are

$$u_t + (3u^2 + u_{xx})_x = 0\,, \tag{12}$$

$$(u^2)_t + (4u^3 + 2uu_{xx} - u_x^2)_x = 0\,. \tag{13}$$

These respectively describe "mass" and "momentum" conservation. The next conservation law would be that of "energy", although here only two are needed, since after averaging the third equation can be replaced by the law for the conservation of waves:

$$k_t + \omega_x = 0\,, \quad \text{where} \quad k = \frac{2\pi}{L}\,, \quad \omega = kV\,. \tag{14}$$

This must be consistent with the modulation system (11), and can be introduced instead of any of three averaged conservation laws (11). In fact, any three independent conservation laws can be used, and will lead to equivalent modulation systems.

4 Whitham modulation equations

In general the Whitham modulation equations have the structure

$$\mathbf{b}_t + \mathbf{A}(\mathbf{b})\mathbf{b}_x = 0\,. \tag{15}$$

Here $\mathbf{b} = (r_1, r_2, r_3)^t$, and the coefficient matrix $\mathbf{A}(\mathbf{b}) = \mathbf{P}^{-1}\mathbf{Q}$ where the matrices $\mathbf{P}, \mathbf{Q}$ have the entries $P_{ij} = \langle P_i \rangle_{r_j}$ and $Q_{ij} = \langle Q_i \rangle_{r_j}$ for $i, j = 1, 2, 3$. The eigenvalues of the coefficient matrix $\mathbf{A}$ are called the characteristic velocities. If all the eigenvalues $v_j(\mathbf{b})$ of $\mathbf{A}(\mathbf{b})$ are real-valued, then the system is nonlinear hyperbolic and the underlying traveling wave is modulationally stable. Otherwise the traveling wave is modulationally unstable.

For this KdV case all the eigenvalues are real so the cnoidal wave is modulationally stable. It can be shown that

$$v_j = -2\sum r_j + \frac{2L}{\partial L/\partial r_j}\,, \quad j = 1, 2, 3. \tag{16}$$

The parameters r_j have been chosen because they are the Riemann invariants of the system (15) for the present case of the KdV equation. Thus this system has the diagonal form

$$(r_j)_t + v_j(r_j)_x = 0\,, \quad j = 1, 2, 3, \tag{17}$$

where we recall that $v_j(r_1, r_2, r_3)$ are the characteristic velocities (16)

$$\begin{aligned} v_1 &= -2\sum r_j + 4(r_3 - r_1)(1 - m)K/E\,, \\ v_2 &= -2\sum r_j - 4(r_3 - r_2)(1 - m)K/(E - (1 - m)K)\,, \\ v_3 &= -2\sum r_j + 4(r_3 - r_2)K/(E - K)\,. \end{aligned}$$

5 Limiting cases of Whitham modulation equations

In the sinusoidal wave limit ($m \to 0$), a solution of (17) is $r_2 = r_3, m = 0$, $v_1 = -6r_1, v_2 = v_3 = 6r_1 - 12r_3$ so that the system collapses to

$$\begin{gathered} r_{1t} - 6r_1 r_{1x} = 0\,, \quad r_{3t} + (6r_1 - 12r_3) r_{3x} = 0\,, \\ \text{or} \quad d_t + 6d d_x = 0\,, \quad k_t + \omega_x = 0\,. \end{gathered} \tag{18}$$

Here $-r_1 = d$ is the mean level, $r_3 - r_1 = k^2/4$ is the wavenumber, and the dispersion relation is $\omega = 6dk - k^3$. An expansion for small m is needed to recover the wave action equation.

In the solitary wave limit ($m \to 1$), a solution of (17) is $r_1 = r_2, m = 1$, $v_1 = v_2 = -4r_1 - 2r_3, v_3 = -6r_3$, so that the system collapses to

$$\begin{gathered} r_{1t} + (-4r_1 - 2r_3) r_{1x} = 0\,, \quad r_{3t} - 6r_3 r_{3x} = 0\,. \\ \text{or} \quad d_t + 6d d_x = 0\,, \quad a_t + V a_x = 0\,. \end{gathered} \tag{19}$$

Now $-r_3 = d$ is the background level, $2(r_3 - r_1) = a$ is the solitary wave amplitude, and $-4r_1 - 2r_3 = 6d + 2a = V$ is its speed.

6 Shocks and undular bores

Let's now consider the similarity solution of the modulation system (17) which describes an undular bore developing from an initial discontinuity:

$$u(x,0) = \Delta \quad \text{for} \quad x < 0\,, \quad \text{and} \quad u(x,0) \quad \text{for} \quad x > 0\,, \tag{20}$$

where $\Delta > 0$ is a constant.

When the dispersive term in the KdV equation (1) is omitted, the KdV becomes the Hopf equation:

$$u_t + 6uu_x = 0\,. \tag{21}$$

This is readily solved by the method of characteristics and the solution is multivalued, $u = \Delta$ for $-\infty < x < 6\Delta t$ and $u = 0$ for $0 < x < \infty$. A shock is needed with speed 3Δ. See Fig. 6 for a solution of this equation with $\Delta = 1$, at $t = 3$.

When the dispersive term is retained, the shock is replaced by a modulated wave train (dispersive shock wave or *undular bore*). Behind the undular bore $u = \Delta$ (or in terms of the Riemann invariants $r_1 = -\Delta, r_2 = r_3$), while ahead of it $u = 0$ ($r_1 = r_2, r_3 = 0$). Because of the absence of a length scale in this problem, the corresponding solution of the Whitham modulation system must depend on the self-similar variable $\tau = x/t$ alone, which reduces the system (17) to

$$(v_j - \tau)\frac{dr_j}{d\tau} = 0\,, \quad i = 1,2,3\,. \tag{22}$$

Hence two Riemann invariants must be constant, namely $r_1 = -\Delta$, $r_3 = 0$ while r_2 varies in the range $-\Delta < r_2 < 0$, given by $v_2 = \tau$.

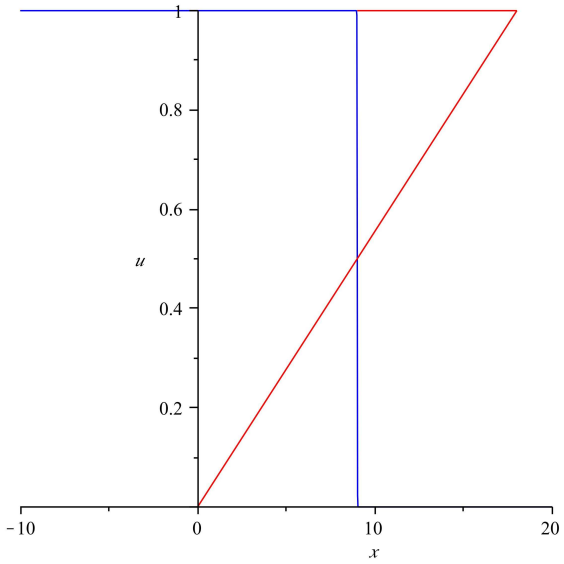


Figure 4: A shock solution to Hopf equation (21) equation, for $\Delta = 1$ (blue) and $\Delta = 3$ (red).

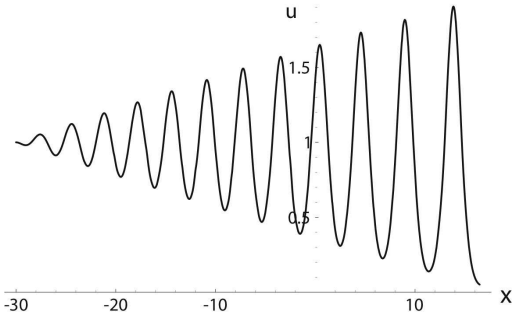


Figure 5: An undular bore

Finally, using the expressions (2, 3, 4) for the cnoidal wave, we get the solution for the undular bore expressed in terms of the modulus m,

$$\frac{x}{t\Delta} = 2(1+m) - \frac{4m(1-m)K(m)}{E(m)-(1-m)K(m)}\,, \tag{23}$$

$$\frac{u}{\Delta} = 1 - m + 2m\,\mathrm{cn}^2(\Delta^{1/2}(x-Vt);m)\,, \quad \frac{V}{\Delta} = -2(1+m)\,. \tag{24}$$

The leading and trailing edges of the undular bore are determined from (23) by putting $m = 1$ and $m = 0$: the undular bore is thus found to exist in the zone

$$-6 < \frac{x}{\Delta t} < 4\,. \tag{25}$$

Note that this solution is an unsteady undular bore which spreads out with time – a steady undular bore would require some friction. The leading solitary wave amplitude is 2Δ, exactly twice the height of the initial jump. Also the wavenumber is constant. For each wave in the wave train, $m \to 1$ as $t \to \infty$, so each wave tends to a solitary wave.

7 Rarefraction wave

When $\Delta < 0$, the initial discontinuity (20) creates a rarefraction wave

$$\begin{aligned} &u = 0\,, \quad \text{for} \quad x > 0\,, \\ &u = \frac{x}{6t}\,, \quad \text{for} \quad 6\Delta t < x < 0\,, \\ &u = \Delta\,, \quad \text{for} \quad x < 6\Delta t\,. \end{aligned} \tag{26}$$

This is a solution of the full KdV equation (1), but needs smoothing at the corners with a weak modulated periodic wave (see Fig. 6).

8 Further developments

1. The Whitham theory can be applied to any nonlinear wave equation which has a (known) periodic travelling wave solution. These include the NLS equations, Boussinesq equations, Su-Gardner equations.

2. For a broad class of integrable nonlinear wave equations, a simple universal method has been developed by Kamchatnov (2000), enabling the construction of periodic solutions and the Whitham modulation equations directly in terms of Riemann invariants.

3. The "undular bore" solution can be extended to the long-time evolution of a system from arbitrary localized initial conditions, described by Gurevich and Pitaevskii (1974) (and many subsequent works), and by Lax and Levermore (1983).

4. There have been applications in many physical areas, including surface and internal undular bores, collisionless shocks in rarefied plasmas (e. g. Earth's magnetosphere bow shock), nonlinear diffraction patterns in laser optics, and in Bose-Einstein condensates.

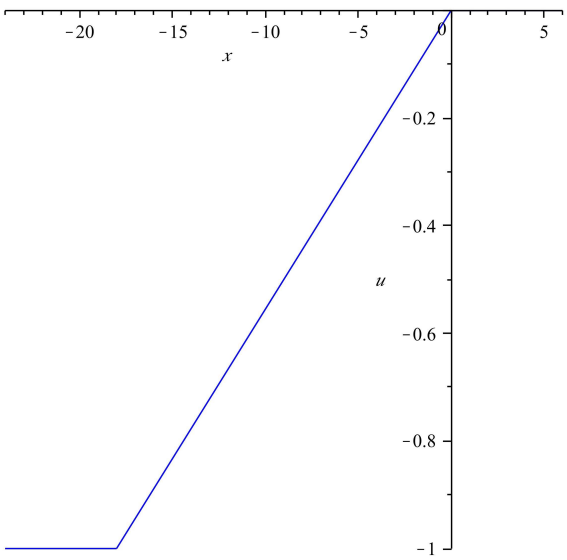


Figure 6: A rarefaction wave, with $\Delta = -1$, at instant $t = 3$.

References

[1] R. Grimshaw, *Solitary Waves in Fluids*, Advances in Fluid Mechanics, Vol 47, WIT press, UK, 208pp, (2007)

[2] A. V. Gurevich & L. P. Pitaevskii, *Nonstationary structure of a collisionless shock wave*, Sov. Phys. JETP **38**, 291-297, (1974)

[3] A. M. Kamchatnov, *Nonlinear Periodic Waves and Their Modulations—An Introductory Course*, World Scientific, Singapore, (2000)

[4] P. D. Lax & C. D. Levermore, *The small dispersion limit of the Korteweg de Vries equation I, II, III.*, Comm. Pure Appl. Math., **36**, 253-290, 571-593, 809-829, (1983)

[5] G. B. Whitham, *Linear and Nonlinear Waves*, Wiley, New York, (1974)

[6] E. T. Whittaker & G. N. Watson, *A Course in Modern Analysis*, 4th ed. Cambridge, England: Cambridge University Press, (1990)

Lecture 11: Internal solitary waves in the ocean

Lecturer: Roger Grimshaw. Write-up: Yiping Ma.

June 19, 2009

1 Introduction

In Lecture 6, we sketched a derivation of the KdV equation applicable to internal waves, and discussed the extended KdV equation for critical cases where the quadratic nonlinear term is small.horizontal direction. In this lecture, we describe internal solitary waves in the ocean, where the bottom topography may vary from the deep ocean to the shallow seas of the coastal oceans, and the background hydrography can also vary along the path of the wave. Hence the asymptotic models must incorporate a variable background state. On the assumption that this is slowly varying relative to the waves, the outcome is a KdV-type equation, but with variable coefficients, namely the variable-coefficient extended KdV (veKdV) equation. When references to Lecture 6 are made, Eq. (n) in Lecture 6 will be referred to as $(6.n)$ here. The presentation of the properties and results for the veKdV equation closely follows that of the vKdV equation discussed in Lecture 9.

2 Variable-coefficient extended KdV equation

The ocean has variable depth, as well as variations in the basic state hydrology and background currents. As seen in the previous lecture, these effects can be formally incorporated into the theory by supposing that the basic state is a function of the slow spatial variable $\chi = \epsilon^3 x$. Thus here we assume a depth $h(\chi)$, a horizontal shear flow $u_0(\chi, z)$ with a corresponding vertical velocity field $\epsilon^3 w_0(\chi, z)$, a density field $\rho_0(\chi, z)$, a corresponding pressure field $p_0(\chi, z)$ and a free surface displacement $\eta_0(\chi)$. With this scaling, the slow background variability enters the asymptotic analysis at the same order as the weakly nonlinear and weakly dispersive effects, and an asymptotic analysis produces a variable coefficient extended KdV equation. The modal system is again defined by (6.12-13) (N is the buoyancy frequency)

$$\{\rho_0(c-u_0)^2\phi_z\}_z + \rho_0 N^2\phi = 0\,, \quad \text{for} \quad -h < z < 0\,, \tag{1}$$

$$\phi = 0 \quad \text{at} \quad z = -h\,, \quad (c-u_0)^2\phi_z = g\phi \quad \text{at} \quad z = 0\,, \tag{2}$$

but now the linear long wave speed $c = c(\chi)$ and the modal functions $\phi = \phi(\chi, z)$, where the χ-dependence is parametric.

With all small parameters removed, the governing equation is

$$A_\tau + \alpha A A_\xi + \alpha_1 A^2 A_\xi + \lambda A_{\xi\xi\xi} = 0\,. \tag{3}$$

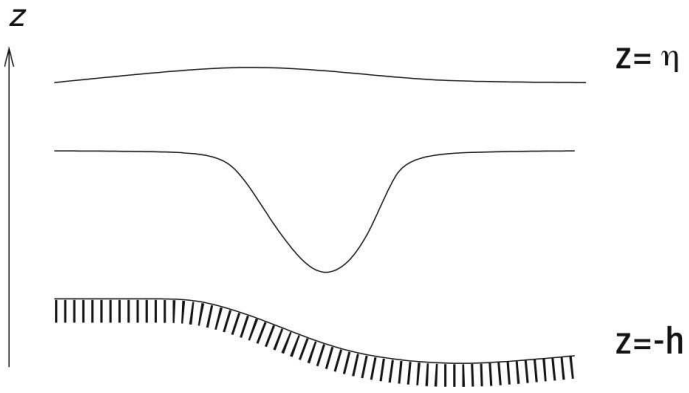


Figure 1: The basic coordinate system.

where we have moved to the new coordinate system, based on the travel time of the wave as introduced in Lecture 9, namely

$$\tau = \int^x \frac{\mathrm{d}x}{c}\,, \quad \xi = t - \tau\,, \tag{4}$$

and where the original amplitude A (defined by $\zeta = A(x-ct)\phi(z)$ where ζ is the vertical particle displacement) has been replaced by $\sqrt{Q}\,A$. Here Q is the linear magnification factor, defined so that QA^2 is the wave action flux. The linear long-wave speed c, and the coefficients $\alpha, \alpha_1, \lambda$ depend on x, and hence on the evolution variable τ. The coefficients $\alpha(\tau), \alpha_1(\tau), \lambda(\tau)$ and $Q(\tau)$ are given by

$$\alpha = \frac{\mu}{cQ^{1/2}}\,,\ \alpha_1 = \frac{\mu_1}{cQ}\,,\ \lambda = \frac{\delta}{c^3}\,,\ Q = c^2 I\,, \tag{5}$$

in terms of the coefficients in the extended KdV equation (6.42) (in different notations)

$$A_T + \mu A A_X + \mu_1 A^2 A_X + \delta A_{XXX} = 0\,, \tag{6}$$

and the definition (6.34)

$$I = 2\int_{-h}^{0} \rho_0 (c-u_0)\phi_z^2\,\mathrm{d}z\,. \tag{7}$$

Unlike the KdV or extended KdV equations, this variable coefficient equation (3) is not integrable in general, so we must seek a combination of asymptotic and numerical solutions.

2.1 Slowly-varying solitary waves

The veKdV equation (3) possesses two relevant conservation laws,

$$\int_{-\infty}^{\infty} A\,\mathrm{d}x \;=\; \text{constant}\,, \tag{8}$$

$$\int_{-\infty}^{\infty} A^2\,\mathrm{d}x \;=\; \text{constant}\,, \tag{9}$$

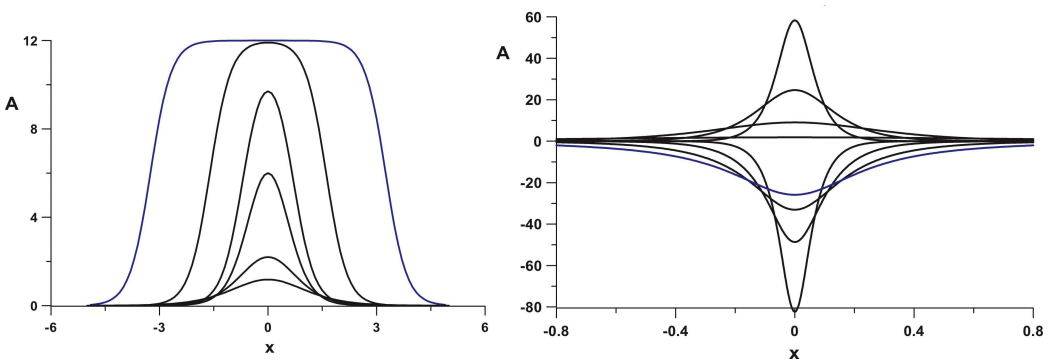


Figure 2: The family of solitary waves for (a) $\delta\mu_1 < 0$; (b) $\delta\mu_1 > 0$.

representing conservation of "mass" and "momentum" respectively (more strictly speaking, an approximate representation of the physical mass and wave action flux).

From Lecture 6, the solitary wave family of the eKdV equation (6) is given by

$$A = \frac{H}{1 + B\cosh K(X - VT)}\,, \tag{10}$$

$$\text{where } V = \frac{\mu H}{6} = \delta K^2\,,\ B^2 = 1 + \frac{6\delta\mu_1 K^2}{\mu^2}\,, \tag{11}$$

with a single parameter B. For $\delta\mu_1 < 0$ (Figure 2.1), $0 < B < 1$, and the family ranges from small-amplitude waves of KdV-type ("sech^2"-profile as $B \to 1$) to a limiting flat-topped wave of amplitude $-\mu/\mu_1$ ("table-top" wave as $B \to 0$). For $\delta\mu_1 > 0$ (Figure 2.1), there are two branches. One has $1 < B < \infty$ and ranges from small-amplitude KdV-type waves ($B \to 1$), to large waves with a "sech"-profile ($B \to \infty$). The other branch, $-\infty < B < 1$, has the opposite polarity and ranges from large waves with a "sech"-profile to a limiting algebraic wave of amplitude $-2\mu/\mu_1$. Waves with smaller amplitudes do not exist, and are replaced by breathers.

In the veKdV equation, we have a family of solitary waves as before, but its parameter $B(\tau)$ now varies slowly in a manner determined by conservation of momentum (9), which requires

$$\begin{aligned} G(B) &= \text{constant}|\frac{\alpha_1^3}{\lambda\alpha^2}|^{1/2}\,, \\ \text{where} \qquad G(B) &= |B^2 - 1|^{3/2}\int_{-\infty}^{\infty}\frac{\mathrm{d}u}{(1 + B\cosh u)^2}\,. \end{aligned} \tag{12}$$

The integral term in $G(B)$ can be explicitly evaluated, and so these expressions provide explicit formulas for the variation of $B(\tau)$ as the environmental parameters vary. However, since the conservation of momentum completely defines the slowly-varying solitary wave, total mass (8) is only conserved provided one adds a "trailing shelf" (linear long wave) whose amplitude A_{shelf} at the rear of the solitary wave is

$$VA_{\text{shelf}} = -\frac{\partial M_{\text{sol}}}{\partial\tau}\,, \qquad M_{\text{sol}} = \int_{-\infty}^{\infty} A_{\text{sol}}\,\mathrm{d}\xi\,, \tag{13}$$

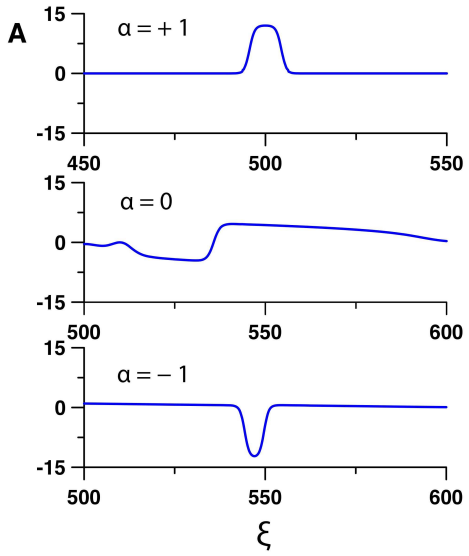


Figure 3: Critical point $\alpha = 0, \alpha_1 < 0$: eKdV case. Here $\lambda = 1, \alpha_1 = -0.083$ and α varies from 1 to -1 (that is, the variable coefficient eKdV equation, with a negative cubic nonlinear coefficient). This shows the conversion of an elevation "table-top" wave into a depression "table-top" wave, riding on a positive pedestal.

where $A_{\rm sol}$ is the solitary wave solution.

The adiabatic expressions (12, 13) show that the critical points where $\alpha = 0$ (or where $\alpha_1 = 0$) are sites where we may expect a dramatic change in the wave structure. There are two qualitatively different cases to consider.

2.2 Critical point $\alpha = 0, \alpha_1 < 0$

First, as α passes through zero, assume that $\alpha_1 < 0, 0 < B < 1$ at the critical point $\tau = 0$ where $\alpha = 0$. Then as $\alpha \to 0$, it follows from (12) that $B \to 0$ and the wave profile approaches the limiting "table-top" wave. But in this limit, $K \sim |\alpha|$, and so the amplitude approaches the limiting value $-\alpha/\alpha_1$. Thus the wave amplitude decreases to zero, the mass M_0 of the solitary wave grows as $|\alpha|^{-1}$ and the amplitude A_1 of the trailing shelf grows as $1/|\alpha|^4$. Essentially the trailing shelf passes through the critical point as a disturbance of the opposite polarity to that of the original solitary wave, which then being in an environment with the opposite sign of α, can generate a train of solitary waves of the opposite polarity, riding on a pedestal of the same polarity as the original wave (see Figure 3).

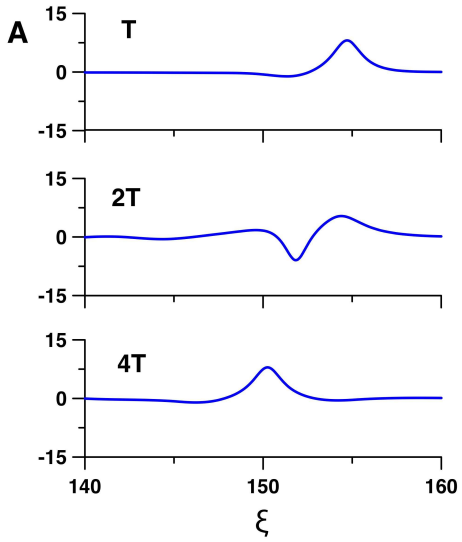


Figure 4: Critical point $\alpha = 0, \alpha_1 > 0$: eKdV case. Here $\lambda = 1, \alpha_1 = 0.3$ and α varies from 1 to -1 for $-T < \tau < T$ (that is, the variable coefficient eKdV equation, with a positive cubic nonlinear coefficient). This shows the adiabatic evolution of an elevation wave from $\tau = -T$ to $\tau = T$, where its amplitude is too small, and so the wave becomes a breather.

2.3 Critical point $\alpha = 0, \alpha_1 > 0$

Next, let us suppose that at the critical point where $\alpha = 0$, $\alpha_1 > 0$. In this case, $1 < |B| < \infty$ and there are the two sub-cases to consider, $B > 1$ or $B < -1$, when the the solitary wave has the same or opposite polarity to α. Then, as $\alpha \to 0, |B| \to \infty$ as $|B| \sim 1/|\alpha|$. It follows from (11) that then $K \sim 1$, $H \sim 1/|\alpha|, a \sim 1, M_0 \sim 1$. Therefore, the wave adopts the "sech"-profile, but has *finite* amplitude, and so can pass through the critical point $\alpha = 0$ without destruction. But the wave changes branches from $B > 1$ to $B < -1$ as $|B| \to \infty$, or *vice versa*. An interesting situation then arises when the wave belongs to the branch with $-\infty < B < -1$ and the amplitude is reducing. If the limiting amplitude of $-2\alpha/\alpha_1$ is reached, then there can be no further reduction in amplitude for a solitary wave, and instead a breather will form (Figure 4).

3 Wave propagation, deformation and disintegration

For real oceanic shelves, there can be wave paths along which the parameters in the veKdV equation may not vary sufficiently slowly, and they may also contain several critical points. As a result, an internal solitary wave loses its identity as a soliton within a finite lifetime. In this section, we describe direct numerical simulations performed using the veKdV equation

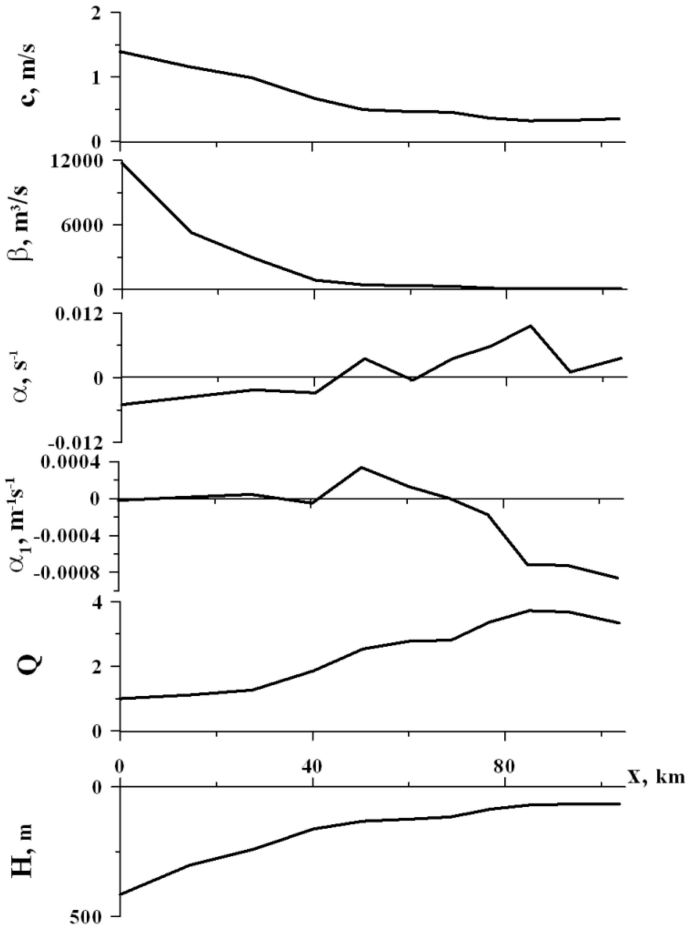


Figure 5: The coefficient of veKdV (3) for the passage of an initial solitary wave of depression across the North West Shelf of Australia.

(3) with the coefficients calculated for three oceanic shelves: the NWS of Australia, the MSE (west of Scotland), and the Arctic shelf (in the Laptev Sea). The initial condition is a typical solitary wave for each shelf. In all simulations, we only use the density stratification of the coastal zone, and ignore any background current. The notation for the quantities plotted in the figures is consistent with (3) with λ in (3) denoted as β in the figures.

3.1 North West shelf of Australia

The measured coefficients of the veKdV equation are presented in Figure 5. The hydrology can indeed be considered as slowly varying because the characteristic horizontal variation scale of the oceanic parameters (more than 10km) exceeds the characteristic soliton wavelength (about 1-2km). This shelf is distinguished by the property that the coefficient α of the quadratic nonlinear term has several sign changes.

A simulation for an initial wave amplitude of 15m is shown in Figure 6. The initial

solitary wave has negative polarity, and transforms at a distance after 68 km into a nonlinear dispersive tail and a group of secondary solitons.

3.2 Malin shelf edge

The measured coefficients of the veKdV equation are presented in Figure 7. The coefficient of the quadratic nonlinear term is everywhere negative. There is only one critical point at a distance of about 8km associated with the sign change of the coefficient of the cubic nonlinear term.

A simulation for an initial wave amplitude of 21m is shown in Figure 8. The soliton-like shape is maintained for a distance of about 20km. Subsequently, the significant decrease of the dispersion parameter leads to the formation of a shock wave. For distances more than 25km, the borelike disturbance transforms into solitary waves.

3.3 Arctic shelf

The measured coefficients of the veKdV equation are presented in Figure 9. The nonlinear coefficients, as well as the dispersion parameter, vary significantly but quite slowly. The coefficient of the cubic nonlinear term is everywhere negative, and increases by three times. The critical point (a zero value of the coefficient of the quadratic nonlinear term) appears at the end of the wave path at a distance of 155km.

A simulation for an initial wave amplitude of 13m is shown in Figure 10. Before reaching the critical point, the solitary wave maintains its soliton-like shape for very long distances (140 km) as the background varies sufficiently slowly.

4 World map of eKdV coefficients

The previous simulations show the key role played by the coefficients of the veKdV equation. Thus in Figure 11, we display world maps of these coefficients (same notations as (6)). As expected, the linear phase speed and the linear dispersive coefficient scale respectively with $h^{1/2}$ and $h^{5/2}$. Hence, as is well known, the largest amplitude internal solitary waves will generally be found in the shallow seas of the coastal zones. However, the quadratic and cubic coefficients show considerable variability, with many sign changes, thus emphasizing again the importance of critical points.

References

[1] Grimshaw, R., *Internal solitary waves*, in *Environmental Stratified Flows*, (2001), ed. R. Grimshaw, Kluwer, Boston.

[2] Grimshaw, R., Pelinovsky, E., Talipova, T. and Kurkin, A. , *Simulation of the transformation of internal solitary waves on oceanic shelves*, J. Phys. Ocean., 34 (2004), pp. 2774-2779.

[3] Grimshaw, R., Pelinovsky, E. and Talipova, T. *Modeling internal solitary waves in the coastal ocean*, Surveys in Geophysics, 28 (2007), pp. 273-298.

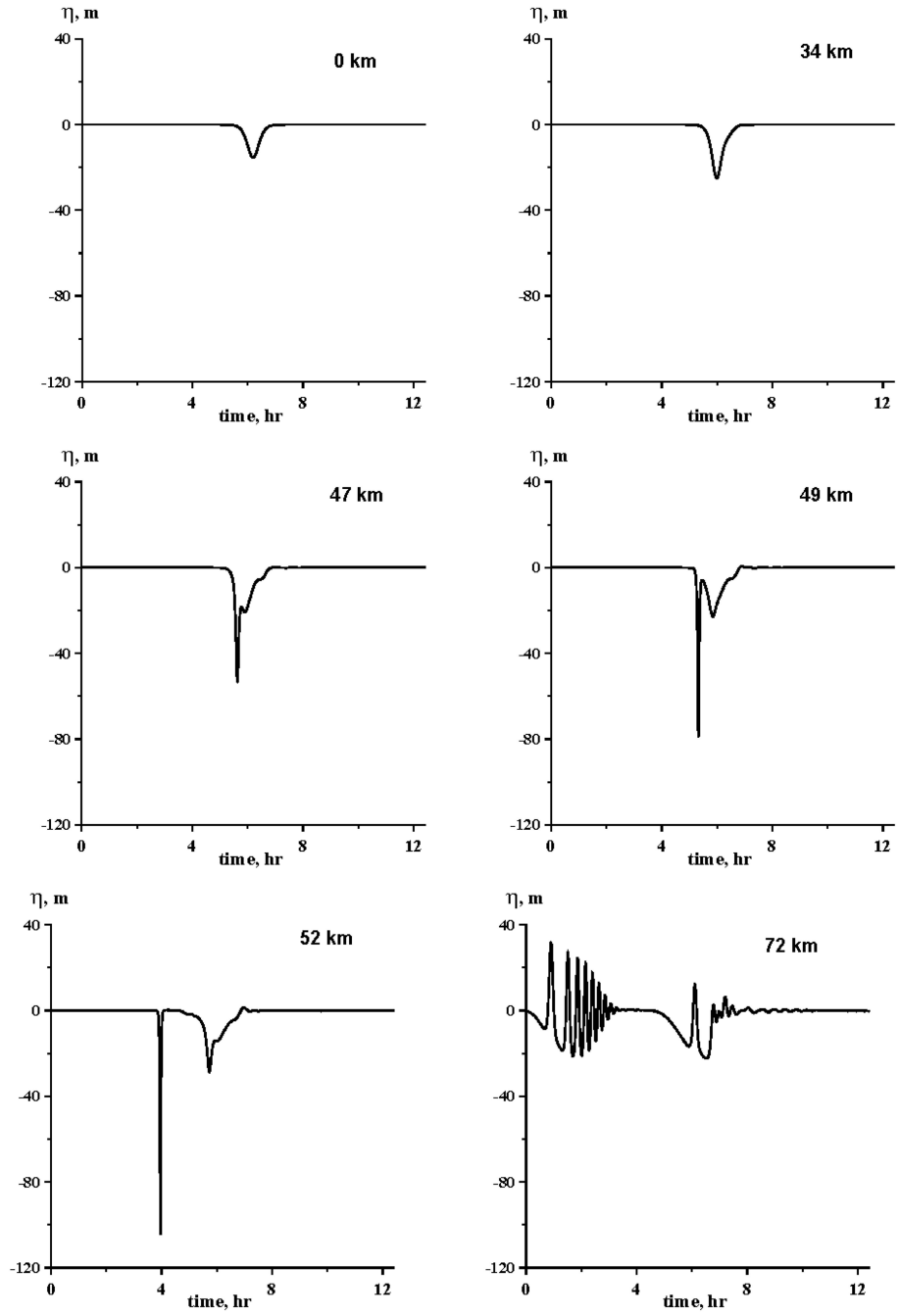


Figure 6: NWS, initial depression wave of 15m amplitude.

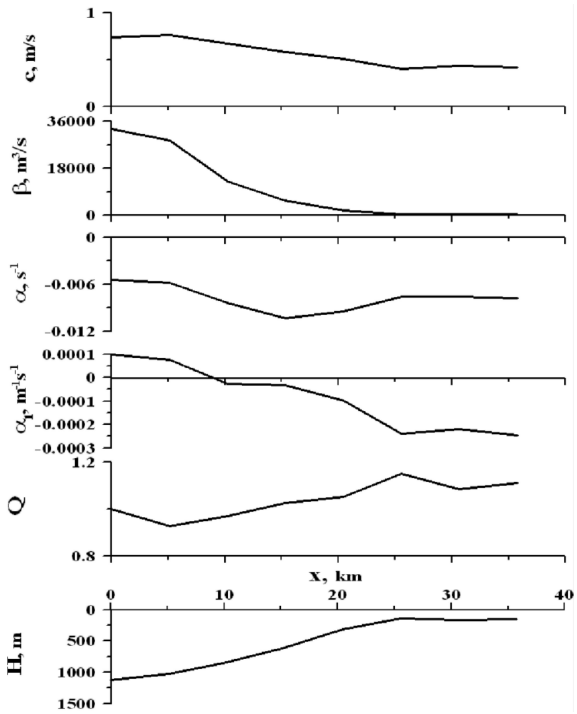


Figure 7: The coefficient of veKdV (3) for the passage of an initial solitary wave of depression across the Malin shelf off west coast of Scotland.

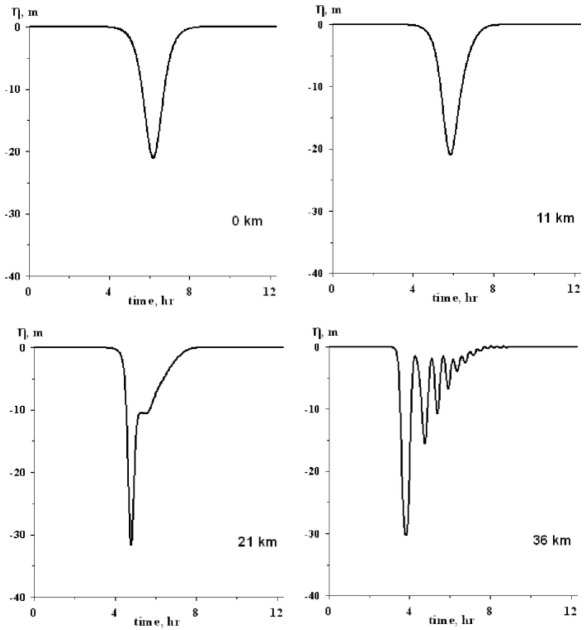


Figure 8: Malin Shelf, fission, initial amplitude of 21m.

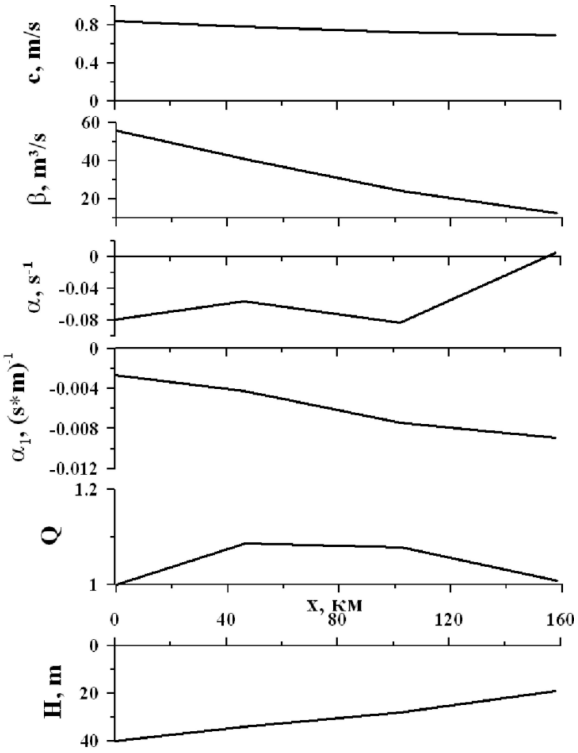


Figure 9: The coefficient of veKdV (3) for the passage of an initial solitary wave of depression across the Arctic shelf off north coast of Russia.

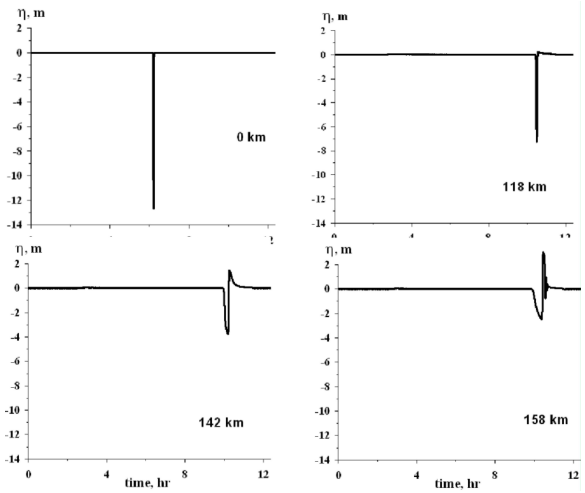
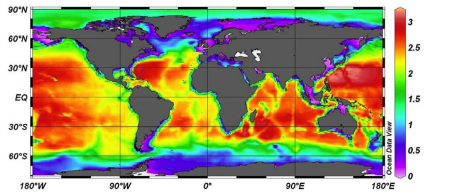
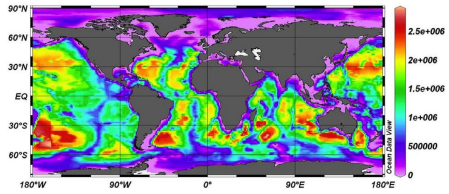


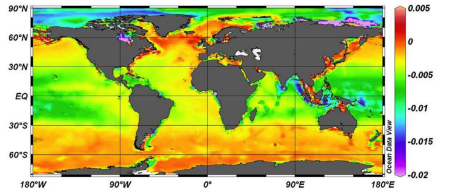
Figure 10: Arctic Shelf, adiabatic, initial amplitude of 13m.



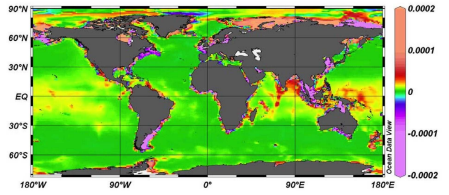
a) speed of propagation, c (m/s)



b) dispersion coefficient, δ (m^3/s)



c) coefficient of quadratic nonlinearity, μ (s^{-1})



d) coefficient of cubic nonlinearity, μ_1 ($m^{-1}s^{-1}$)

Figure 11: World map of eKdVcoefficients.

Lecture 12: Transcritical flow over an obstacle

Lecturer: Roger Grimshaw. Write-up: Erinna Chen

June 22, 2009

1 Introduction

The flow of a fluid over an obstacle is a classical and fundamental problem in fluid mechanics. When the flow velocity is near the local wave speed, obstacles in the path of this resonant flow tend to generate interesting wave fields around them. This lecture explores transcritical flow over a step primarily in the framework of the forced Korteweg-deVries equation using both asymptotic analysis and numerical simulations. This lecture follows from Lecture 11 in that resonant flow over a step may be resolved with undular bores.

2 Linearized shallow-water theory

Consider one-dimensional shallow-water flow past topography. The flow variables we are interested in are the total local depth $H = \zeta + h - F(x)$ and the depth-averaged horizontal velocity V. Here ζ is the surface elevation above the undisturbed depth h and the bottom is located at $z = -h + F(x)$ where $F(x)$ is the obstacle. The fully nonlinear shallow water equations for conservation of mass and momentum are

$$\zeta_t + (HV)_x = 0, \tag{1}$$

$$V_t + VV_x + g\zeta_x = -\frac{(H^2D^2H)_x}{3H} - \frac{(H^2D^2F)_x}{2H} - \frac{F_xD^2(\zeta + F)}{2}, \tag{2}$$

$$\text{where } D = \frac{\partial}{\partial t} + V\frac{\partial}{\partial x}.$$

Equation (1) is exact, but equation (2) is a long-wave approximation; the terms on the right-hand side are the leading-order effects of wave dispersion. They form the Su-Gardner equations (also known as the Green-Naghdi equations).

If the Su-Gardner equations are linearized about the constant state U, h, where $V = U + u$, $|u| << V$, $|\zeta| << h$, they reduce to the forced linear wave equation

$$D_I^2\zeta - c^2\zeta_{xx} - U^2F_{xx},\ D_I = \frac{\partial}{\partial t} + U\frac{\partial}{\partial x},\ c = \sqrt{gh}. \tag{3}$$

Here c is the linear long-wave speed, and a key parameter is the Froude number, $Fr = U/c$. Provided that background flow velocity is not critical ($Fr \neq 1$), as $t \to \infty$, there is a steady solution

$$\zeta = \frac{U^2}{U^2 - c^2}F(x). \tag{4}$$

The solution is a stationary depression over the obstacle for subcritical flow ($Fr < 1$), and a stationary elevation for supercritical flow ($Fr > 1$).

3 Forced Korteweg-deVries equation

The linear solution clearly fails as the flow nears criticality $Fr \approx 1$. The wave energy cannot propagate away from the obstacle. In this case, it is necessary to invoke weak nonlinearity and weak dispersion. This results in the forced Korteweg-deVries (fKdV) equation.

For water waves, the fKdV equation is, in non-dimensional form with scaling of characteristic length h and velocity c,

$$-\zeta_t - \Delta\zeta_x + \frac{3}{2}\zeta\zeta_x + \frac{1}{6}\zeta_{xxx} + \frac{1}{2}F_x = 0. \tag{5}$$

Here $\Delta = Fr - 1$ measures the degree of criticality, subcritical for $\Delta < 0$ or supercritical for $\Delta > 0$. The equation describes the usual KdV balance between nonlinearity, dispersion and time evolution, supplemented here by forcing and criticality. The asymptotic regime where equation (5) holds is characterized by a small parameter $\epsilon << 1$, where a balance between all terms requires the scaling $F \sim \epsilon^4$, $A \sim \epsilon^2$, $\partial/\partial x \sim \epsilon^4$, $\partial/\partial t \sim \epsilon^3$, and $\Delta \sim \epsilon^2$. Note that ζ scales with $\sqrt{F}$ and with the detuning Δ, a feature typical of forced resonant systems.

The canonical form of the fKdV equation,

$$-A_t - \Delta A_x + 6AA_x + A_{xxx} + F_x(x) = 0, \tag{6}$$

is obtained by putting $\zeta = 2A/3$, $\Delta = \tilde{\Delta}/6$, $t = 6\tilde{t}$, $F = 2\tilde{F}/9$, and then omitting the "tilde." Equation (6) is typically solved with the initial condition that $A(x, 0) = 0$, which corresponds to the introduction of the topographic obstacle $F(x)$ at $t = 0$.

4 Transcritical flow over a localized obstacle

Figures (1-3) are a series of solutions with varying criticality (Δ) for the canonical form of the fKdV equation where the forcing function is localized at $x = 0$ with a maximum height $F_M > 0$ and is "Gaussian" in shape. In these figures, F_M is specifically $F_M = 1$.

In Figures 1-3 it can be seen that solutions of the fKdV equation typically consist of upstream and downstream nonlinear wavetrains connected by a locally steady solution over the obstacle. The origin and behavior of the nonlinear wavetrains is due to the structure of the locally steady solution over the obstacle. Assuming that the dispersionless or "hydraulic" limit is applicable in the obstacle region, the linear dispersive term in equation (6) can be neglected. A result of this is that for all localized $F(x)$ with a maximum height $F_M > 0$,

$$6A_{\pm} = \Delta \mp (12F_M)^{1/2}. \tag{7}$$

The transcritical regime is thus defined as

$$|\Delta| < (12F_M)^{1/2}. \tag{8}$$

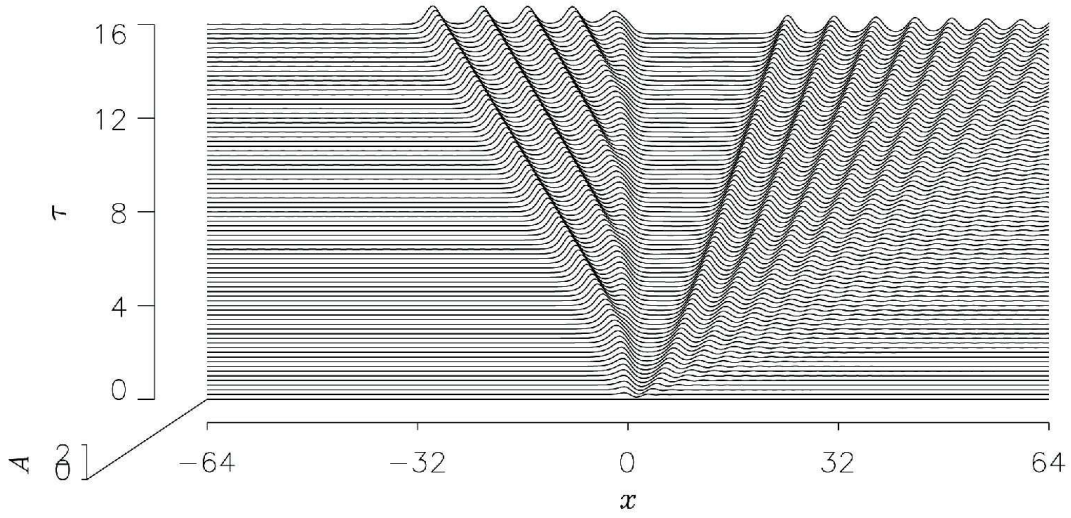


Figure 1: Solution to the fKdV equation (6) at exact criticality, $\Delta = 0$. The forcing (not shown in the plot) is located at $x = 0$ and has a maximum height of $F_M = 1$

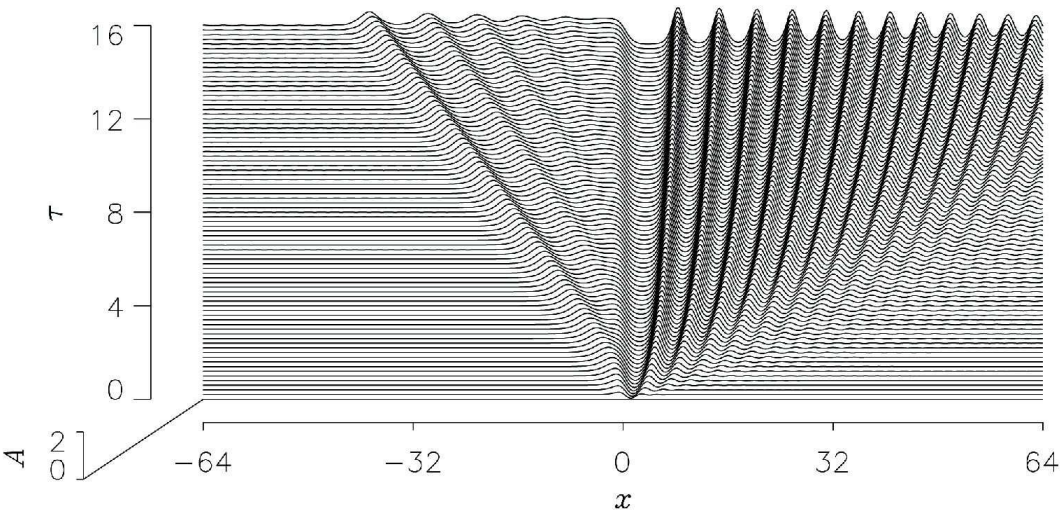


Figure 2: Solution to the fKdV equation (6) for a subcritical case, $\Delta = -1.5$.

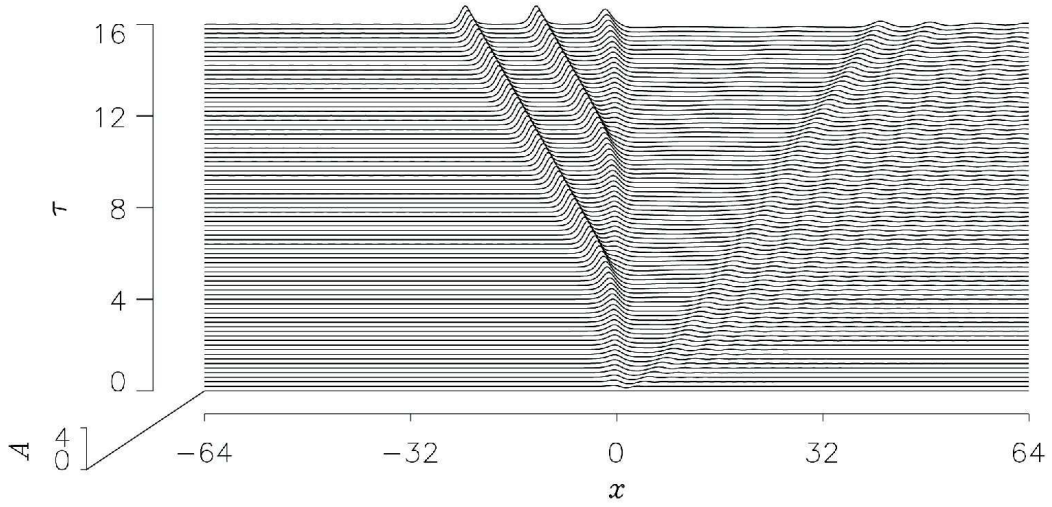


Figure 3: Solution to the fKdV equation (6) for a supercritical case, $\Delta = 1.5$.

In the transcritical regime, the local steady solution is characterized by a transition from a constant elevation $A_- > 0$ upstream ($x < 0$) of the obstacle to a constant depression $A_+ < 0$ downstream ($x > 0$) of the obstacle, where $\Delta = 3(A_+ + A_-)$, independent of the details of the localized forcing term $F(x)$. The explicit values of A_+ and A_- are determined by the forcing term $F(x)$, as shown in Sections 5.1 and 5.2. The upstream and downstream solutions are solved with undular bores.

4.1 Undular bores

A simple representation of an undular bore can be obtained from the solution of the KdV equation

$$A_t + 6AA_x + A_{xxx} = 0 \tag{9}$$

with the initial condition of a step, $A = A_0 H(-x)$ with $A_0 > 0$ and $H(x)$ the Heaviside function. An asymptotic solution can be found using Whitham's modulation theory. This formalism was presented in Lecture 10. The relevant asymptotic solution corresponding to the "step" initial condition is constructed in terms of a similarity variable x/t. The undular bore wavetrain is located in the region

$$-6 < \frac{x}{A_0 t} < 4. \tag{10}$$

Also recall that if the initial condition A_0 corresponds to a negative step $A_0 < 0$ then the undular bore solution does not exist. Instead the asymptotic solution is a rarefaction wave,

$$A = 0 \text{ for } x > 0,$$
$$A = \frac{x}{6t} \text{ for } A_0 < \frac{x}{6t} < 0,$$

$$A = A_0, \text{ for } \frac{x}{6t} < A_0 < 0. \tag{11}$$

Small oscillatory wave trains are needed to smooth out the discontinuities in A_x at $x = 0$ and $\frac{x}{t} = -6A_0$.

4.2 Asymptotic analysis for localized forcing

We now return to the asymptotic solution of the fKdV equation, and resolve the upstream and downstream transitions by the "undular bore" solutions. Making the appropriate transformations for the upstream wavetrain in (10) (see [2]), the upstream wavetrain ($x < 0$) occupies the region

$$\Delta - 4A < \frac{x}{t} < \min\{0, \Delta + 6A\}. \tag{12}$$

The upstream wavetrain cannot move beyond the obstacle (at $x = 0$), and thus is only fully realized if $\Delta < -6A_-$. Combining this criterion with (7) and (8) defines the regime

$$-(12F_M)^{1/2} < \Delta < -\frac{1}{2}(12F_M)^{1/2} \tag{13}$$

where a fully developed undular bore solution can develop upstream. On the other hand, when $\Delta > -6A_-$ or

$$-\frac{1}{2}(12F_M)^{1/2} < \Delta < (12F_M)^{1/2} \tag{14}$$

the undular bore cannot develop beyond the obstacle. A partially formed undular bore develops upstream, and this is attached to the obstacle.

The downstream wavetrain similarly is constrained to lie in $x > 0$. The undular bore solution occupies the zone

$$\max\{0, \Delta - 2A_+\} < \frac{x}{t} < \Delta - 12A_+ \tag{15}$$

The downstream wavetrain is only fully realized if $\Delta > 2A_+$. Combining this criterion with (7) and (8), the region where a fully developed downstream undular bore coincides with (14), the region where the upstream wavetrain is partially formed and attached to the obstacle. Similarly, when the downstream undular bore ($\Delta < 2A_+$) is only partially developed and attached to the obstacle, the upstream undular bore is fully developed and detached from the obstacle.

Figures (1-3) show this behavior in the regime defined by (14) where the upstream wavetrain is partially formed and attached to the obstacle while the downstream wavetrain is fully developed. In the subcritical case ($\Delta < 0$), the upstream wavetrain weakens and for sufficiently large $|\Delta|$ separates from the obstacle (i.e. makes the transition from the regime defined by (14) to the regime defined by (13)). When this transition is reached, the downstream wavetrain intensifies and forms a stationary lee wave field. For supercritical flow ($\Delta > 0$), the behavior is always governed by (14). As $|\Delta|$ increases the upstream wavetrain develops into well-separated solitary waves attached at the obstacle while the downstream wavetrain weakens and moves downstream.

For the case where the obstacle has negative polarity ($F_M < 0$), the upstream and downstream solutions are qualitatively similar (i.e. can be described in terms of an undular

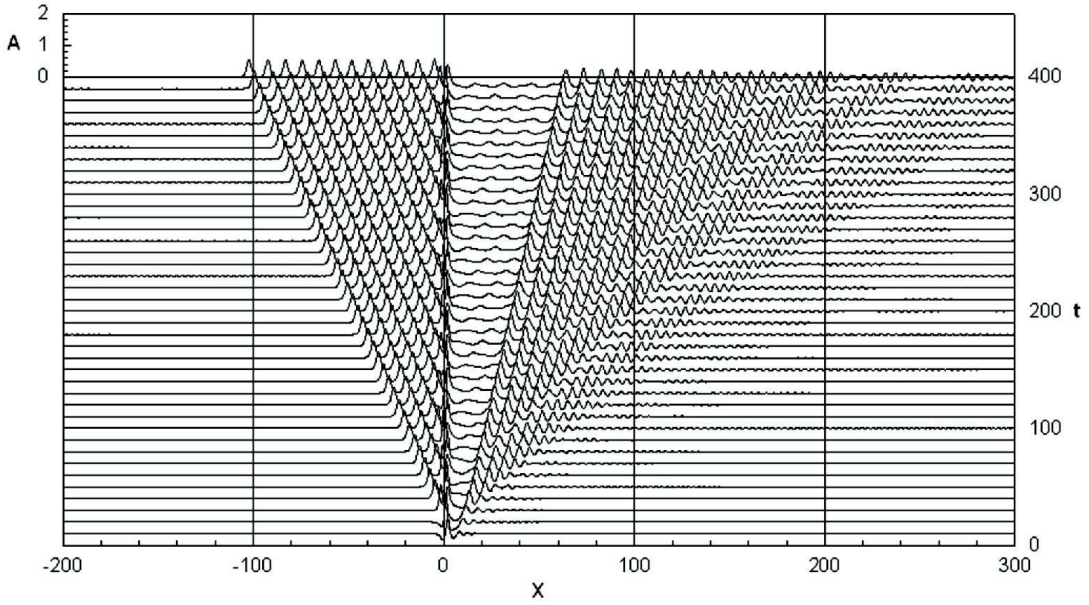


Figure 4: Solution to the fKdV equation (6) for exact criticality, $\Delta = 0$. The forcing (not shown in the plot) is located at $x = 0$ and has localized negative forcing $F_M = -0.1$.

bore solution). However, the local solution around the obstacle is transient, and this causes a modulation of the undular bore solutions seen previously. A typical solution of the fKdV equation with a negative polarity obstacle is shown in Figure (4). Note the perturbations in the solution around the obstacle.

5 Transcritical flow over a step

5.1 Asymptotic analysis for positive step forcing

First, consider a broad positive step, where

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ F_M & \text{for } x > W \end{cases} \tag{16}$$

and $F(x)$ varies monotonically in $0 < x < W$, and $F_M > 0$. Strictly $F(x)$ should return to zero for some $L >> W$. For this analysis, we ignore this and assume that $L \to \infty$. We revisit the step of finite length in the numerical simulations of Section 5.3. We now modify the asymptotic solution found for localized forcing in Section 4 for an infinitely long step. The first step is to construct the local steady-state solution, using the hydraulic limit. In the forcing region ($0 < x < W$), $A = A(x)$, while otherwise

$$A = A_- \text{ for } x \to -\infty, \tag{17}$$

$$A = A_+ \text{ for } x \to \infty. \tag{18}$$

The steady-state fKdV equation (6) ignoring the dispersive term is

$$-\Delta A_x + 6AA_x + F_x = 0. \tag{19}$$

Integrating (19) with respect to x results in the equation

$$-\Delta A + 3A^2 + F = C, \tag{20}$$

whose solutions are

$$6A = \Delta \pm (\Delta^2 + 12C - 12F)^{1/2}. \tag{21}$$

Thus, there are two branches to the solution. Applying the far-field limits (17, 18) to equation (20) yields

$$C = -\Delta A_- + 3A_-^2 = -\Delta A_+ + 3A_+^2 + F_M. \tag{22}$$

The constant C is obtained by taking the long-time limit of the unsteady hydraulic solution. The fKdV equation (6) with no dispersion is a nonlinear hyperbolic equation that can be solved by the the method of characteristics. This equation written in characteristic form is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \Delta - 6A, \ \frac{\mathrm{d}A}{\mathrm{d}t} = F_x(x) \tag{23}$$

with the initial condition that $A = 0$ at $t = 0$. For a positive step, F_x is zero for $x < 0$ and $x > W$ and $F_x > 0$ for $0 < x < W$. From (23), all characteristics begin with an initial slope Δ and then decrease. The key issue is whether the characteristics ever reach a turning point ($\mathrm{d}x/\mathrm{d}t = 0$, $\Delta = 6A$).

For $\Delta \leq 0$, all characteristics have a negative slope and there are no turning points. In this case, $A_+ = 0$ and from (22), $C = F_M$ and the upper branch must be chosen in (21). For $\Delta > (12F_M)^{1/2}$, there are also no turning points, and all characteristics have positive slope. In this case, $A_- = 0$, $C = 0$ and the lower branch is chosen. For $0 < \Delta < (12F_m)^{1/2}$ there is a turning point. Characteristics emerging from the step ($0 < x < W$) with $F = F_0$, $0 < 12F_0 < 12F_M - \Delta^2$ have a turning point and then go upstream into $x < 0$. From this, $12C = 12F_M - \Delta^2$ and from (22), $6A_+ = \Delta$, while A_- is then obtained from the upper branch of (21). In summary, the upstream and downstream solutions for a positive step forcing are

$$\Delta \leq 0: \ 6A_- = \Delta + (\Delta^2 + 12F_M)^{1/2}, 6A_+ = 0, \tag{24}$$

$$0 < \Delta < (12F_M)^{1/2}: \ 6A_- = \Delta + (12F_M)^{1/2}, 6A_+ = \Delta, \tag{25}$$

$$\Delta > (12F_M)^{1/2}: \ 6A_- = 0, 6A_+ = \Delta - (\Delta^2 - 12F_M)^{1/2}. \tag{26}$$

In all cases, the upstream solution A_- is a "jump" in the hydraulic limit, and thus needs to be resolved by an undular bore. The upstream undular bore is located in the region

$$\Delta - 4A < \frac{x}{t} < \min\{0, \Delta + 6A\}, \tag{27}$$

the region similar to that for the upstream jump for a localized forcing (12). For a fully detached undular bore, $\Delta + 6A_- < 0$, and combining this criterion with (24, 25, 26), the defining regime is

$$\Delta < -2(F_M)^{1/2} < 0. \tag{28}$$

In the regime where $\Delta + 6A_- > 0$ but $\Delta - 4A_- < 0$, or

$$-2(F_M)^{1/2} < \Delta < (12F_M)^{1/2}, \tag{29}$$

the upstream undular bore is only partially formed and is attached to the positive step. Unlike the localized obstacle, the downstream profile A_+ is not a jump ($A_+ > 0$). The solution downstream is terminated instead by a rarefaction wave.

5.2 Asymptotic analysis for negative step forcing

A similar analysis to Section 5.1 for a positive step can be carried out for a negative step ($F_M < 0$). The results are

$$\Delta \geq 0: \ 6A_- = 0, 6A_+ = \Delta - (\Delta^2 - 12F_M)^{1/2} \text{ with } C = 0, \tag{30}$$

$$-(|12F_M|)^{1/2} < \Delta < 0: \ 6A_- = \Delta, 6A_+ = \Delta - (|12F_M|)^{1/2} \text{ with } C = -\Delta^2/12, \tag{31}$$

$$\Delta < -(|12F_M|)^{1/2}: \ 6A_- = \Delta - (\Delta^2 - 12F_M)^{1/2}, 6A_+ = 0 \text{ with } C = F_M. \tag{32}$$

In all cases, the downstream solution A_+ in (30-32) is negative and contains a jump. This jump needs to be resolved by an undular bore, occupying the zone similar to (13)

$$\max\{0, \Delta - 2A_+\} < \frac{x - W}{t} < \Delta - 12A_+. \tag{33}$$

For a fully detached undular bore ($\Delta - 2A_+ > 0$) is obtained in the regime

$$\Delta > -(|3F_M|)^{1/2}. \tag{34}$$

The downstream undular bore is partially attached when $\Delta - 2A_+ < 0$ corresponding to a regime where

$$-(|12F_M|)^{1/2} < \Delta < -(|3F_M|)^{12} < 0. \tag{35}$$

A stationary lee-wave train forms downstream when $\Delta < -(|12F_M|)^{1/2}$. The upstream solution A_- is less than zero, and thus terminated by a rarefaction wave.

5.3 Numerical solutions for the fKdV equation with a step forcing

Combining the results from Sections 5.1 and 5.2, we can now understand the behavior of flow over a finite-length positive step forcing; that is, a step up, followed by a region of constant elevation and terminated by a step down. Figure 5 shows numerical simulations of the fKdV equation for water waves (5) with varying Δ for a step forcing of the form

$$F(x) = \frac{F_M}{2}(\tanh \gamma x - \tanh \gamma(x - L)), \tag{36}$$

where $F_M = 0.1$, $\gamma = 0.25$ and $L = 50 >> 1$. Note these are parameter values for the unscaled parameters in the water wave formulation.

When the flow is critical ($\Delta = 0$), the theory predicts an upstream undular bore attached to the obstacle (regime given by (29)) and a fully detached downstream undular bore that

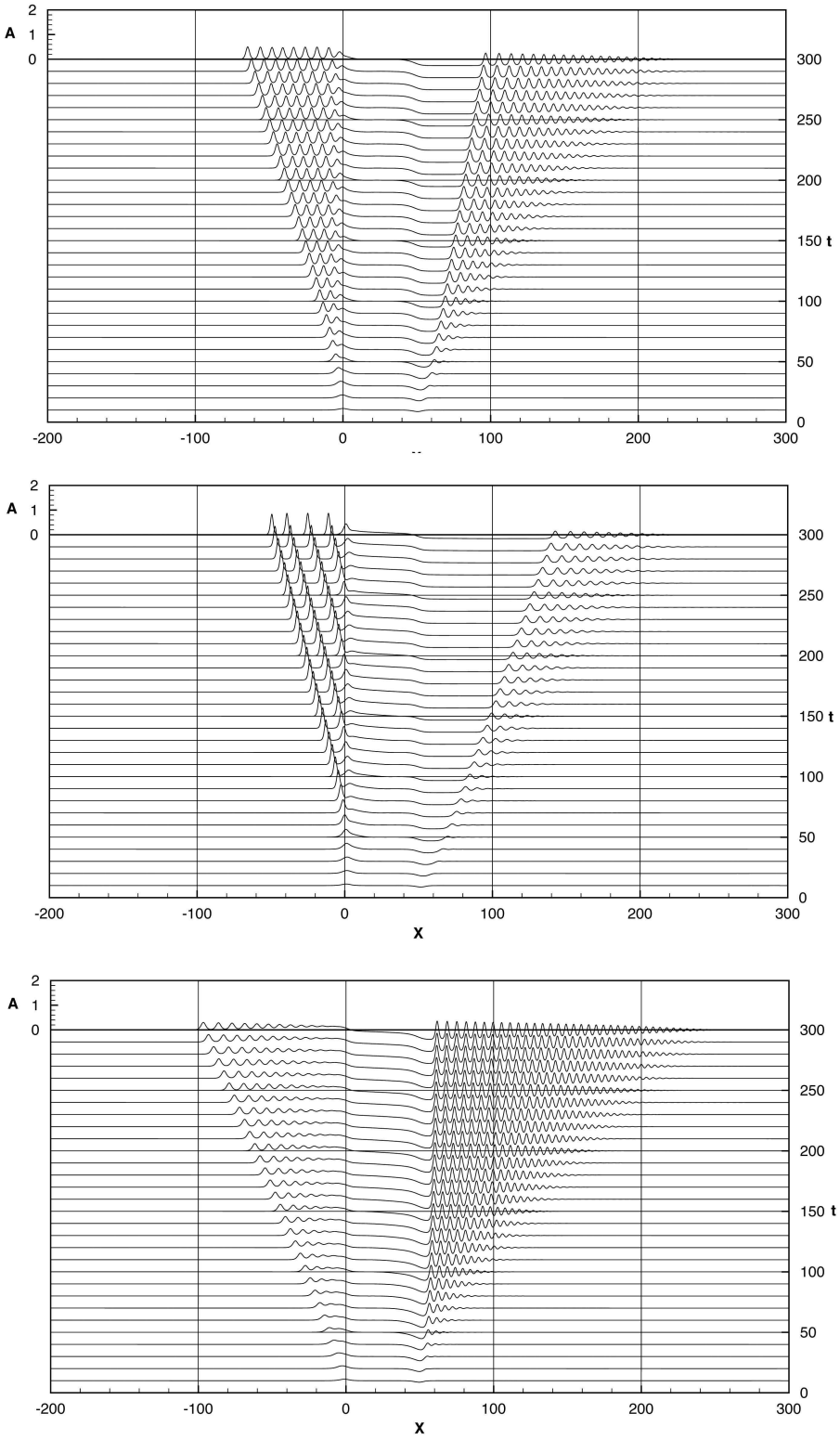


Figure 5: Numerical simulations of the fKdV equation (5) for the forcing (36), where $F_M = 0.1$, $\gamma = 0.25$, $L = 50 \gg 1$ and (a)$\Delta = 0.$, (b)$\Delta = 0.2$, (c)$\Delta = -0.2$. Note, to obtain the correct boundary values given by (28, 29) and (34, 35), the parameters must be scaled to those for the canonical fKdV (6).

propagates downstream (34). This is seen in Figure 5a, and is similar to the behavior of exactly resonant flow over a localized obstacle (Figure 1).

The behavior of a supercritical flow is shown in Figure 5b. Once again, the upstream undular bore is only partially formed and the downstream undular bore is fully developed and propagates downstream. The upstream solution however has transitioned from the solution defined by (24) to that defined by (25); the downstream solution still remains in the regime given by (34). A rarefaction wave thus propagates downstream from the positive step at $x = 0$. The rarefaction wave persists until it reaches the end of the step at $x = 50$, and the solution adjusts to a solution predicted for simply a localized forcing (one upstream undular bore and one downstream undular bore).

Figure 2c shows a numerical simulation for subcritical flow where $\Delta = -0.2$. In this case, there is a detached undular bore propagating upstream and a detached downstream bore that intensifies and propagates slowly. The parameter values in this simulation are close to the boundaries predicted by the asymptotic theory, and while the upstream bore falls in the regime (29) given for an attached undular bore, the behavior in the simulation suggests that the upstream solution actually follows regime (28). Similarly, the downstream bore should fall into regime (35), but actually follows regime (34). This is attributed to errors in the estimates of $A_{\mp}$ and the resulting boundaries derived from those quantities. Note that the downstream solution is in the regime defined by 31 and thus a rarefaction wave propagating upstream is generated at the negative step ($x = 50$).

5.4 Comparison of the fKdV solutions to solutions of the full Euler equations

Figure 6 shows numerical simulations of the full Euler equations for the same positive step forcing as before. In these simulations, there seems to be good qualitative agreement with the behavior of the solutions of the fKdV equation. Table 1 shows the quantitative differences between the simulations. The amplitudes of the leading upstream and downstream waves are consistently larger in the fKdV simulations likely due to nonlinearity. The variation of all the predicted amplitudes and elevations as Δ is varied follows the same trend for both the fKdV and Euler equations. Furthermore, in all of the Euler equation simulations, there were no other wavetrains generated than those seen in the simulations of the fKdV equation. Therefore, for small-amplitude steps, the fKdV equation likely provides a good guide for transcritical flow over a step.

5.5 Solutions to a negative step forcing

Similar numerical simulations can be carried out for an obstacle where the step forcing in (36) is negative, $F_M < 0$. The flow first encounters a negative step, followed by a constant depression section, and is terminated by a positive step. The asymptotic analyses from Sections 5.1 and 5.2 suggest that the resulting undular bore propagates downstream of the negative step and another undular bore similarly propagates upstream of the positive step. These undular bores interact with each other over the step itself.

In the depression region, the interactions between the two undular bores can be quite complex as seen in Figures 7-8. In Figure 7, the undular bores pass through each other

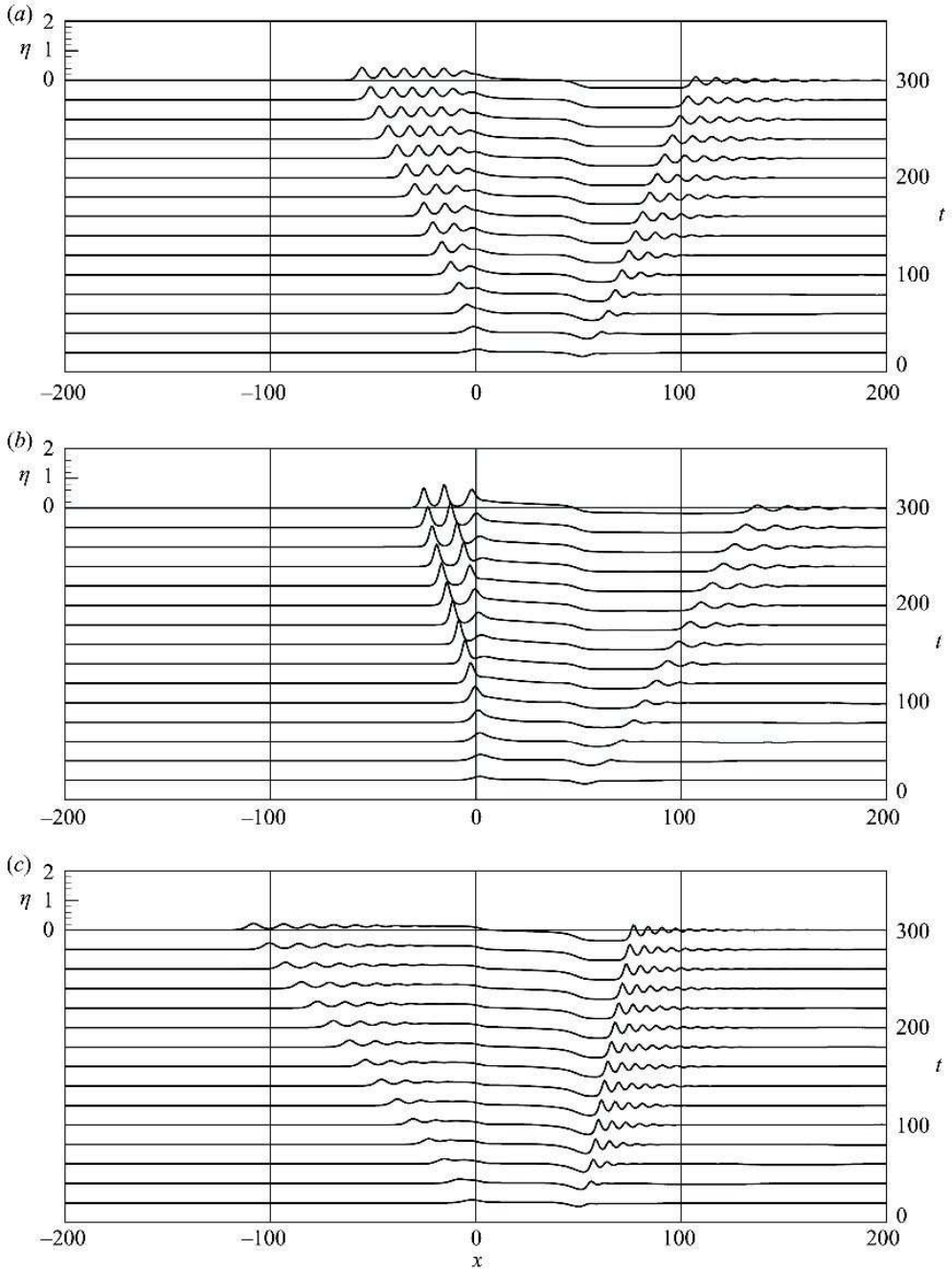


Figure 6: Numerical simulations of the full Euler equations for the forcing (36), where $F_M = 0.1$, $\gamma = 0.25$, $L = 50 \gg 1$ and (a)$\Delta = 0.$, (b)$\Delta = 0.2$, (c)$\Delta = -0.2$.

Δ	fKdV				Euler			
	A_{W-}	A_-	A_{W+}	A_+	A_{W-}	A_-	A_{W+}	A_+
0.2	0.83	0.44	0.31	-0.16	0.75	0.40	0.28	-0.18
0.1	0.66	0.38	0.39	-0.20	0.57	0.36	0.32	-0.21
0.0	0.50	0.30	0.51	-0.26	0.44	0.33	0.37	-0.25
-0.1	0.39	0.22	0.64	-0.33	0.32	0.20	0.43	-0.30
-0.2	0.30	0.16	0.84	-0.40	0.23	0.13	0.53	-0.36
-0.3	0.24	0.13	0.64	-0.38	0.16	0.08	0.57	-0.38
-0.4	0.19	0.10	0.00	0.00	0.10	0.01	0.00	0.00

Table 1: Quantitative comparison of the results from the fKdV equation (5) and the Euler equations. $A_-(A_+)$ is the elevation just upstream (downstream) of the positive (negative) step at $x = 0(50)$ respectively, and $A_{W-}(A_{W+})$ is the amplitude of the leading wave in the corresponding undular bore.

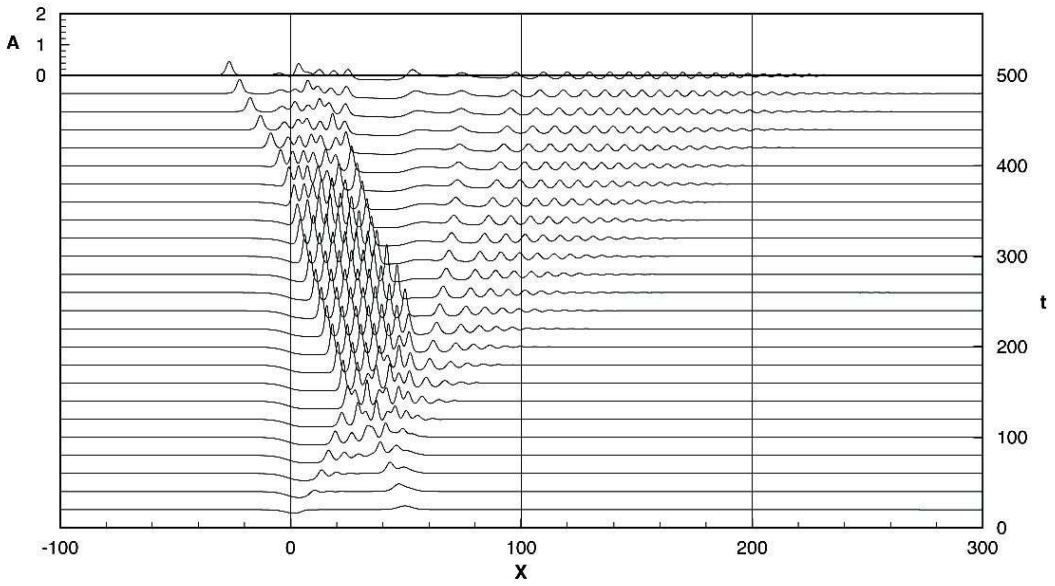


Figure 7: Numerical simulation of the fKdV equation (6) for the step forcing (36) where $F_M = -0.1$, $\gamma = 0.25$, $L = 50 \gg 1$ and $\Delta = 0.0$.

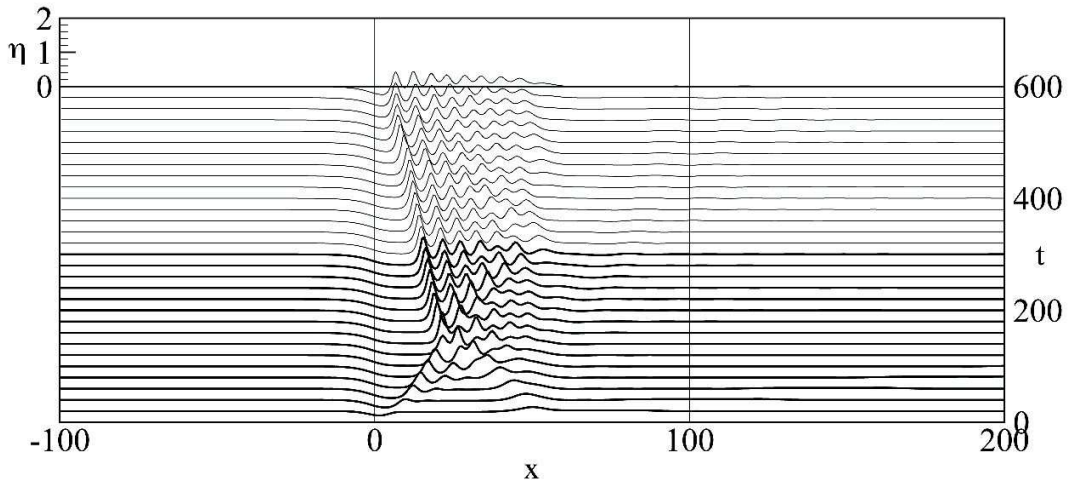


Figure 8: Numerical simulation of the full Euler equations for the step forcing (36) where $F_M = -0.1$, $\gamma = 0.25$, $L = 50 \gg 1$ and $\Delta = -0.2$.

with interference in the depression region. For moderate time, both the upstream and downstream solutions are fairly unperturbed by each other.

Figure 7 shows a solution of the fKdV equation to a negative step forcing, while Figure 8 shows a solution to the full Euler equations for the same forcing. For the positive step, it seemed that there were no qualitative differences between the fKdV solutions and the solutions to the full Euler equations. However, the differences in Figures 7 and 8 are significant. In Figure 8 the wave propagation is largely restricted to the depression region. The waves in the full Euler solutions are likely confined to the step region due to the interaction of two effects not accounted for in the fKdV equation. First, the steps have a finite amplitude that the wave must pass over in order to propagate. Small amplitude waves may reflect at the steps, instead of propagating forward. In addition, the depression also causes a depression in the local wave speed in the step region, and waves may have lower amplitude due to this effect.

References

[1] R. Grimshaw and N. Smyth, *Resonant flow of a stratified fluid over topography*, J. Fluid Mech., 169 (1986), pp. 429–464.

[2] R. Grimshaw, D. Zhang, and K. Chow, *Generation of solitary waves by transcritical flow over a step*, J. Fluid Mech., 587 (2007), pp. 235–254.

[3] R. Grimshaw, D. Zhang, and K. Chow, *Transcritical flow over a hole*, Stud. Appl. Math., 122 (2009), pp. 235–248.

[4] D. Zhang and A. Chwang, *Generation of solitary waves by forward- and backward-step bottom forcing*, J. Fluid Mech., 432 (2001), pp. 341–350.

Lecture 13: Triad (or 3-wave) resonances.

Lecturer: Harvey Segur. Write-up: Nicolas Grisouard

June 22, 2009

1 Introduction

The previous lectures have been dedicated to the study of specific nonlinear problems, most of them involving solitons. It is now time to discuss the problem of triad resonances, as they are actually the most basic nonlinear phenomenon that can occur between two waves. As their name indicates, triad resonances involve the weak interaction between two waves combining to form a third wave.

We construct a set of equations describing single-triad interactions in Section 2, and study their mathematical structure in Section 3. Insights into the behavior of a real system, subject to multiple triad resonances, are presented in Sections 4 and 5.

2 Derivation of the 3-wave equations

2.1 Weakly nonlinear interactions

For dispersive waves of small amplitude, resonant triad interactions are the first nonlinear interactions to appear (if they are possible). Let's start with a physical system of evolutionary partial differential equations with no dissipation, described by a set of equations

$$N(u) = 0$$

with $u = 0$ being the state of rest.

To study the nonlinear interaction of three waves of weak amplitude, the procedure is, as usual, to linearize around $u = 0$. From previous lectures, we expect the linear solution to take the form:

$$u(\boldsymbol{x}, t, \varepsilon) = \varepsilon \left[\sum_k A(\boldsymbol{k}) e^{i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega(\boldsymbol{k})t)} + c.c. \right] + O(\varepsilon^2)$$

This linear problem is associated with a dispersion relation:

$$\omega = \omega(\boldsymbol{k}).$$

Once this dispersion relation is found, the next step is to build a weakly nonlinear model. In order to do so, one has to see if the linear problem admits three pairs $\{\boldsymbol{k}_i, \omega(\boldsymbol{k}_i)\}$, with $\boldsymbol{k}_i \neq 0$ for $i = 1, 2, 3$, satisfying a triad relationship:

$$\boldsymbol{k_1} \pm \boldsymbol{k_2} \pm \boldsymbol{k_3} = 0, \tag{1}$$
$$\omega(\boldsymbol{k_1}) \pm \omega(\boldsymbol{k_2}) \pm \omega(\boldsymbol{k_3}) = 0. \tag{2}$$

A graphical procedure to investigate this possibility has been discovered several times (*e.g.* Ziman (1960) [18], Ball (1964) [1]) and is illustrated in Figure 1. In Figure 1(a), the solid line represents the linearized dispersion relation of pure gravity waves in deep water:

$$\omega = \pm\sqrt{g|k|},$$

while Figure 1(b) shows the linearized dispersion relation of capillary-gravity waves in deep water:

$$\omega = \pm\sqrt{g|k| + \frac{\sigma}{\rho}|k|^3}.$$

The graphical procedure to detect resonant triads is the following. First, pick any point (k_1, ω_1) on one of the branches (solid lines) of the dispersion curve. Then reproduce all branches of the dispersion relation with the origin translated to (k_1, ω_1), here drawn as the dashed lines. Let P be a point where the two curves intersect. Its coordinates in the original coordinate system are identified as (k_3, ω_3), and as (k_2, ω_2) in the translated coordinate system. Then by construction,

$$k_1 + k_2 = k_3,$$
$$\omega_1 + \omega_2 = \omega_3.$$

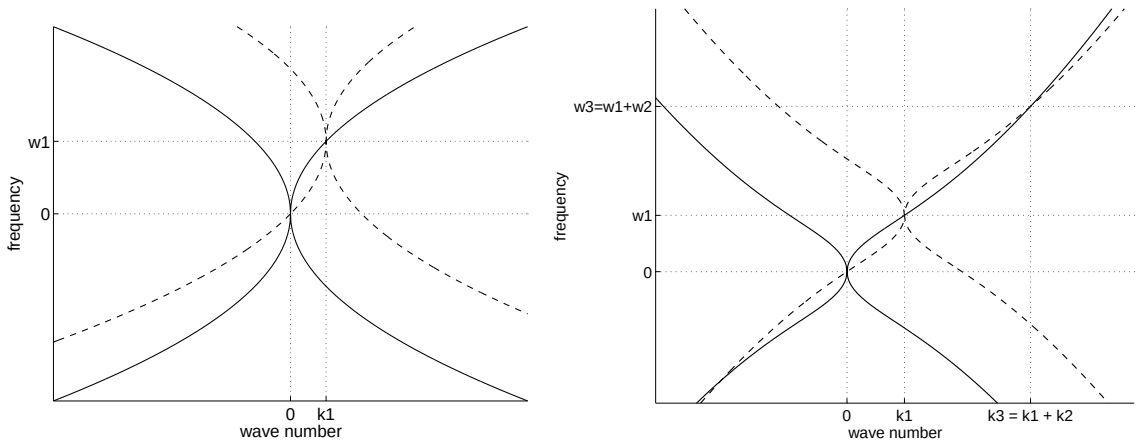
In Figure 1(a), the only point of intersection is at the origin on the solid curve ($k_3 = 0$) contrary to our requirements. Hence, there is no possibility of forming any resonant triad with pure gravity waves. In Figure 1(b) on the other hand one can notice that there are at least two possibilities of forming a resonant triad in top right and in the bottom left hand corner of the figure. In fact, for most choices of k_1 there will be at least two intersection points, leading to another two new triads. Hence, for capillary-gravity waves in 1D, there are infinitely many possible triads.

We have just described the 1-D case in which k_1, k_2 and k_3 are collinear, so they are effectively scalars. In 2D, $\boldsymbol{k_1}$, $\boldsymbol{k_2}$ and $\boldsymbol{k_3}$ are two-component vectors. A similar procedure can be applied, except that both solid and dashed curves now become 2-D surfaces. The two surfaces typically intersect on 1-D curves (or not at all), showing that in 2D there are an infinite number of possible triads associated with a given pair $(\boldsymbol{k_1}, \omega(\boldsymbol{k_1}))$.

To conclude, there are two simple alternatives: if triplets of waves satisfying equations (1) and (2) exist, then triad resonances are expected. If there are no such triplets, one has to look for resonant quartets (4-waves interactions). The latter alternative is discussed in Lectures 14, 15 and 20.

2.2 Single triad

Suppose that equations (1) and (2) are satisfied for exactly one triad. Let us write u as a superposition of the three interacting waves plus weak interaction terms:



(a) Gravity waves, deep water

(b) Capillary-gravity waves, deep water

Figure 1: Linearized dispersion relations of: – (a) pure gravity waves in deep water, – (b) capillary-gravity waves in deep water. In (a), no resonant triad can be formed and in (b), (k_1, ω_1), (k_2, ω_2) and (k_3, ω_3) form a resonant triad.

$$u(\boldsymbol{x}, t; \varepsilon) = \varepsilon \left[\sum_{m=1}^{3} A_m e^{i(\boldsymbol{k_m} \cdot \boldsymbol{x} - \omega_m t)} + c.c. \right] + \varepsilon^2 \left[\sum_{m=1}^{3} \sum_{n=-m}^{m} B_{mn}(t) e^{i((\boldsymbol{k_m} + \boldsymbol{k_n}) \cdot \boldsymbol{x} - (\omega_n + \omega_m) t)} + c.c. \right] + O(\varepsilon^3)$$

If we substitute this ansatz in the original set of equations $N(u) = 0$, we expect to recover the linearized equations to $O(\varepsilon)$. At $O(\varepsilon^2)$, we obtain a set of equations for the coefficients $B_{nm}(t)$, which, in a manner similar to the cases described in Lectures 5 and 6, have a LHS equal to the LHS of the homogeneous linearized equations, and a RHS which depends on the first order amplitudes A_m times a phase function $e^{i(\boldsymbol{k_m} \cdot \boldsymbol{x} - \omega_m t)}$. By construction of the resonant triad, some of these phases are in resonance with the LHS, leading to solutions $B_{mn}(t)$ which grow linearly with time. If that is the case, after a duration $t \sim 1/\varepsilon$, the $O(\varepsilon^2)$ terms are as important as the $O(\varepsilon)$ terms and the asymptotic expansion is no longer valid.

To solve the problem, we introduce as in Lectures 5 and 6 a slower timescale[1] in the wave amplitudes, perform a multiple-scale analysis, and find a compatibility condition to prevent the amplitudes from blowing up. Hence, let u now be:

$$u(\boldsymbol{x}, t; \varepsilon) = \varepsilon \left[\sum_{m=1}^{3} A_m(\varepsilon \boldsymbol{x}, \varepsilon t) e^{i(\boldsymbol{k_m} \cdot \boldsymbol{x} - \omega_m t)} + c.c. \right] + O(\varepsilon^2).$$

In the general case, as found by Benney & Newell (1967) ([2]) among others, the amplitudes have to satisfy this set of compatibility conditions in order to prevent any unphysical growth in the equations:

[1] and/or a longer lengthscale...

$$\begin{aligned}
\partial_\tau A_1 + \boldsymbol{c_1}\cdot\nabla A_1 &= i\delta_1 A_2^* A_3^*,\\
\partial_\tau A_2 + \boldsymbol{c_2}\cdot\nabla A_2 &= i\delta_2 A_1^* A_3^*,\\
\partial_\tau A_3 + \boldsymbol{c_3}\cdot\nabla A_3 &= i\delta_3 A_1^* A_2^*,
\end{aligned} \tag{3}$$

where $\tau = \varepsilon t$ is the slow timescale, $\boldsymbol{c_n} = \mathrm{d}\omega/\mathrm{d}k|_{k=k_n}$ is the group velocity of the n^{th} mode, and δ_n is a real-valued constant which depends on the nonlinear terms of the original set of equations $N(u) = 0$ and can be deduced from an analysis of the original Hamiltonian. These equations can be applied in various contexts, *e.g.* capillary-gravity waves ([13]) or χ_2 materials in optics ([3]). Equations (3) are the so-called PDE version of the equations of triad resonances.

3 Mathematical structure of a single triad of ODEs

A special case occurs when the complex wave amplitude depends on only one independent variable. This can happen for example for systems in which the spatial planform is fixed but its amplitude varies with time. In this case, equations (3) reduce to:

$$\begin{aligned}
A_1' &= i\delta_1 A_2^* A_3^*,\\
A_2' &= i\delta_2 A_1^* A_3^*,\\
A_3' &= i\delta_3 A_1^* A_2^*,
\end{aligned} \tag{4}$$

where A_n' is either $\partial_\tau A_n$ or $c_n\partial_x A_n$. These equations are the so-called ODE version of the equations of triad resonances.

3.1 Hamiltonian structure and integrability

If the original system is hamiltonian, so is this reduced system. The conjugate variables for the single-triad ODEs are

$$\begin{aligned}
q_n(\tau) &= \mathrm{sign}(\delta_n)\frac{A_n(\tau)}{\sqrt{|\delta_n|}},\\
p_n(\tau) &= \frac{A_n^*(\tau)}{\sqrt{|\delta_n|}} \text{ for } n = 1,2,3
\end{aligned}$$

and the Hamiltonian of the system is

$$\begin{aligned}
H &= i(A_1A_2A_3 + A_1^*A_2^*A_3^*)\\
&= i\sqrt{|\delta_1\delta_2\delta_3|}(\mathrm{sign}(\delta_1\delta_2\delta_3)q_1q_2q_3 + p_1p_2p_3).
\end{aligned}$$

It is easy to verify that H and all the $\{p_n, q_n\}$ verify

$$q_n' = \frac{\partial H}{\partial p_n} \quad \text{and} \quad p_n' = -\frac{\partial H}{\partial q_n} \quad \text{for } n = 1,2,3.$$

The following quantities are constants of the motion:

$$-iH = A_1 A_2 A_3 + A_1^* A_2^* A_3^*, \tag{5}$$

$$J_1 = \frac{A_1 A_1^*}{\delta_1} - \frac{A_3 A_3^*}{\delta_3}, \tag{6}$$

$$J_2 = \frac{A_2 A_2^*}{\delta_2} - \frac{A_3 A_3^*}{\delta_3}. \tag{7}$$

(and so is the equivalently defined J_3, although J_3 is not independent of J_1 and J_2). Equations (6) and (7) are known as the Manley-Rowe equations.

We can now use some of the properties of Hamiltonian systems discussed in Lecture 4 to study the integrability of the problem. Recall that the notion of integrability first requires the definition of the Poisson bracket $\{F, G\}$ for any two real-valued functions $F(p_n, q_n, t)$ and $G(p_n, q_n, t)$:

$$\begin{aligned}\{F, G\} &= \sum_{m=1}^{3} \left(\frac{\partial F}{\partial p_m}\frac{\partial G}{\partial q_m} - \frac{\partial F}{\partial q_m}\frac{\partial G}{\partial p_m} \right) \\ &= \sum_{m=1}^{3} \delta_m \left(\frac{\partial F}{\partial A_m^*}\frac{\partial G}{\partial A_m} - \frac{\partial F}{\partial A_m}\frac{\partial G}{\partial A_m^*} \right).\end{aligned}$$

We have already identified 3 constants of motion for our 6-dimensional phase space. We next have to verify that these constants of motion are *in involution*, i.e. that any two pairs have a null Poisson Bracket. This can indeed be verified, as for example:

$$\begin{aligned}\{-iH, J_1\} &= \sum_{m=1}^{3} \delta_m \left(\frac{\partial(-iH)}{\partial A_m^*}\frac{\partial J_1}{\partial A_m} - \frac{\partial(-iH)}{\partial A_m}\frac{\partial J_1}{\partial A_m^*} \right) \\ &= (A_1^* A_2^* A_3^* - A_1 A_2 A_3) + 0 + (-A_1^* A_2^* A_3^* + A_1 A_2 A_3) \\ &= 0.\end{aligned}$$

It can be easily shown in a similar way that $\{-iH, J_2\} = \{J_1, J_2\} = 0$. Moreover, any 3 independent linear combination of these constants of motion will then also be *in involution*. All that remains to be done is to construct the 3 action variables P_n $(n = 1, 2, 3)$ as linear combinations of $-iH$, J_1 and J_2, together with the 3 conjugate angle variables, such that the $\{P_n, Q_n\}$ pairs satisfy

$$\frac{\partial H}{\partial Q_n} = -\frac{\mathrm{d}P_n}{\mathrm{d}t} = 0$$

$$\frac{\partial H}{\partial P_n} = \frac{\mathrm{d}Q_n}{\mathrm{d}t}$$

The resultant system is completely integrable. As we saw in Lecture 4, the 3 action variables define a 3-D surface in the 6-D phase space. Every solution of the ODEs consists of straight-line motion on this surface.

The topology of the 3-D surface depends uniquely on the signs of δ_1, δ_2 and δ_3. For the Hamiltonian system defined above, it can be shown that the general solution can be written

in terms of elliptic functions. In this case, the surface defined by $\{-iH, J_1, J_2\}$ is compact if and only if $\{\delta_1, \delta_2, \delta_3\}$ do not all have the same sign. To see this, first note that if the signs of δ_1 and δ_3 are different for example, then $J_1 =$ constant is the equation for an ellipse in the $(|A_1|, |A_3|)$ plane. It follows that provided at least one of the δ_n has a different sign, it is always possible to select a pair among the three possible J_n for which the 2 equations J_i=const and J_j=const combined describe a torus.

In the unusual situation however where δ_1, δ_2 and δ_3 all have the same sign, Coppi, Rosenbluth & Sudan (1969) [7] showed that A_1, A_2 and A_3 can all blow up together, in finite time: indeed, as J_1 is a constant of the motion, if $|A_1|$ increases, $|A_3|$ has to increase also according to equation (6) and as J_2 is also a constant of the motion, $|A_2|$ has to increase as well... This is the explosive instability, which is discussed in Lecture 21. It has applications in plasma physics ([9]), density-stratified shear flows ([6, 8]) and for vorticity waves ([14]).

3.2 Properties of single-triad systems

Consider a single triad of ODEs, without dissipation. In this example we consider a system which has coefficients such that $\delta_1 > 0$, $\delta_2 > 0$ and $\delta_3 < 0$.

One property of such a configuration is that only A_3 can share energy with the other modes. This can be readily seen from the definition of J_1 and J_2 in equations (6) and (7), which can be rewritten here as:

$$J_1 = \frac{|A_1|^2}{|\delta_1|} + \frac{|A_3|^2}{|\delta_3|} \quad \text{and} \quad J_2 = \frac{|A_2|^2}{|\delta_2|} + \frac{|A_3|^2}{|\delta_3|}.$$

Therefore, if $|A_1|$ (respectively $|A_2|$) initially has almost all the energy, then J_2 (respectively J_1) is initially small. Since J_2 (J_1) is a constant of motion, it necessarily remains small thereby limiting the amplitude of the other modes. By contrast, if $|A_3|$ starts with almost all the energy, it can distribute it to the whole system.

Moreover, Hasselmann (1967) [10] proved that in a single triad of ODEs in which $\{\delta_1, \delta_2, \delta_3\}$ do not all have the same sign, the interaction coefficient δ_n with the opposite sign from the other two (in our example, that would be $|A_3|$) is always the wave mode with the highest frequency in the triad.

This section has shown that the case of a single triad of ODEs is well understood. If the complex wave amplitude do *not* depend on only one independent variable, we have to move on to the case of a single triad of PDEs.

4 Triads of PDEs

Zakharov & Manakov (1976) [17] showed that the system of equations (3) of PDEs is completely integrable. Kaup (1978) [11] partly solved the initial-value problem in 1-D on $-\infty < x < \infty$, with restrictions. Kaup, Reiman & Bers (1980) solved the initial-value problem in 3-D in all space, under some restrictions. Besides that, few physical applications of this theory have been developed so far.

Zakharov (1968) [16] went further than a single resonant triad interaction of PDEs and considered all possible interactions, including the non-resonant ones. This led to the Zakharov's integral equation for the amplitudes:

$$\partial_t A(\boldsymbol{k}) + i\omega(\boldsymbol{k})A(\boldsymbol{k}) = -i\iint [V(\boldsymbol{k},\boldsymbol{k_1},\boldsymbol{k_2})\delta(\boldsymbol{k}+\boldsymbol{k_1}+\boldsymbol{k_2})A^*(\boldsymbol{k_1})A^*(\boldsymbol{k_2}) + perm.]\, d\boldsymbol{k_1}d\boldsymbol{k_2}$$
$$-i\iiint [W(\boldsymbol{k},\boldsymbol{k_1},\boldsymbol{k_2},\boldsymbol{k_3})\delta(\boldsymbol{k}+\boldsymbol{k_1}+\boldsymbol{k_2}+\boldsymbol{k_3})A^*(\boldsymbol{k_1})A^*(\boldsymbol{k_2})A^*(\boldsymbol{k_3}) + perm.]\, d\boldsymbol{k_1}d\boldsymbol{k_2}d\boldsymbol{k_3}. \tag{8}$$

V and W are generalized interaction coefficients and *perm.* stands for the permutations in the combinations of $\boldsymbol{k}$'s, *e.g.* $\delta(\boldsymbol{k} \pm \boldsymbol{k_1} \pm \boldsymbol{k_2})\ldots$ This equation acts on the fast timescale whereas equations (3) acted on the slow timescale τ. As the slow timescales are still present in equation (8), everything has to be resolved and a numerical integration is therefore slow and expensive.

In general, single triads of ODEs are insufficient (i) when wave envelopes have spatial variability, in which case PDEs are required, (ii) when there are multiple triad interactions, in which case more ODEs (or PDEs) are required and (iii) when dissipation occurs, in which case one needs to study non-Hamiltonian ODEs (the case of a single dissipative triad is studied in Lecture 20). Having discussed a few properties of single triads of PDEs, let us now discuss problems associated with multiple triad interactions under experimental conditions.

5 Application to capillary-gravity waves

In Section 2.1, we saw that for capillary-gravity waves (and most realistic systems) there exists a continuum of triads. So one can rightfully wonder whether any of the results described in Section 3 for single-triad resonance remain applicable. If not, under what conditions is it possible to model the problem with just a small number of triads? Under what conditions are we instead forced to consider all possible triads, as in Zakharov's integral formulation seen in Section 4? This section discusses these questions in the light of experimental results.

Simmons (1969) [15] conjectured that the magnitudes of the interaction coefficients δ_n do not vary much across the continuum of possible capillary-gravity waves resonant triads. Should this be correct, then any energy input into a single wave mode will eventually be transferred to all the other modes, triad by triad, thus generating a broad-banded response of the system to the applied forcing.

Perlin, Henderson and Hammack(1990) [12] attempted to test this conjecture experimentally. In their work, a tank of typical size $10\,\mathrm{cm} \times 1\,\mathrm{m}$ filled with water is forced with a paddle oscillating at a frequency $f_0 = 25\,\mathrm{Hz}$, therefore exciting capillary-gravity waves that are subject to triad interactions. As we have seen in Section 2.1, in 2D and unbounded configurations, the spectrum of possible interactions is a continuum. Although this geometry is bounded, it is large enough to have a significant number of possible modes and one can therefore expect to have a broad response in the frequency space.

A first series of experiments was performed and contrary to expectations did not display the expected broad-band response, as can be seen in Figure 2. All figures show that, no matter what the forcing amplitude is, the response of the fluid is localized in a few modes with frequency mainly around f_0. Figures 2(a) and 2(b) show results of the experiment

with two different forcing amplitudes. The components that are to be seen in this picture are:

- a strong signal at f_0,
- a second harmonic at $f_0 + f_0 = 50\,\mathrm{Hz}$,
- a third harmonic at $2f_0 + f_0 = 75\,\mathrm{Hz}$,
- a very weak component at $f_c = 60\,\mathrm{Hz}$ that can only be seen in Figure 2(b). It is an internal frequency of the computer that controls the forcing mechanism and that is transmitted to the tank as mechanical vibration.
- a frequency $f_\alpha = 35\,\mathrm{Hz} = f_0 - f_c$,
- two frequencies $f_\beta = 10\,\mathrm{Hz}$ and $f_\gamma = 15\,\mathrm{Hz}$, characterized by $f_\beta + f_\gamma = f_0$.

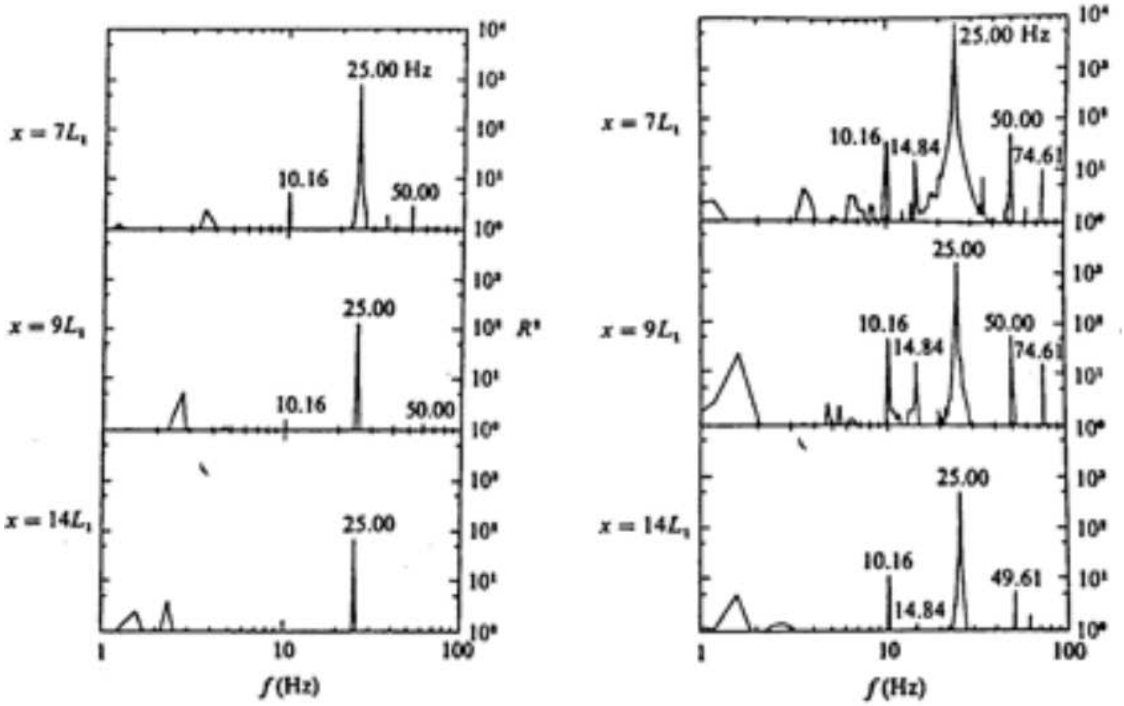
If we ignore the second and third harmonics of the forcing, the specific transmission chain is then the following: f_c and f_0, the forcing frequencies, interact to form f_α, then f_0 and f_α interact to form f_β which then interacts with f_0 to form f_γ.

A second series of experiments was later performed with a newer computer which did not perturb the system with the additional frequency at $f_c = 60\,\mathrm{Hz}$. In these new experiments, a broad frequency spectrum was observed in response to the forcing. It is therefore quite remarkable to note how the whole system dynamics change depending on the presence or absence of the additional forcing at f_c, even when the forcing is so weak as to be barely detected in Figure 2. The difference between the two sets of experiments suggests the presence of a selection mechanism for individual triads through the additional forcing.

Now let us focus on the experiment related to Figure 2. In the triad involving $f_\alpha = f_c - f_0$ and following the notations of Section 3.2, A_1, A_2 and A_3 correspond to f_0, f_α and f_c respectively. A_3 is the highest frequency but A_1 has almost all the energy initially, so A_2 and A_3 should remain small. Although they indeed remain small compared to A_1, A_2 gets significantly bigger than A_3, in contradiction with Section 3.2. In the triad involving $f_\alpha - f_0 = f_\beta$, (with A_1, A_2 and A_3 corresponding to f_0, f_β and f_α this time), again we have $A_3 \ll A_1$, but A_2 gains some energy in the process. Only in the last triad involving $f_0 - f_\beta = f_\gamma$ is the highest mode also the most energetic one. Hasselmann's results presented in Section 3.2, which were valid for a single triad, are clearly not relevant here.

These experiments have revealed two surprising features: the selection of a small number of triads among the possible continuum by a weak periodic signal, and the invalidation of single-triad theory (see Section 3.2) as soon as more than one triad is involved. In an actual physical problem involving 3-wave interactions, these two conclusions show how difficult it may be, in practice, to make even qualitative predictions about a system's response to forcing. In general, it is indeed very hard to predict whether a single, a few, or all triads must be taken into account. Any process has to be investigated in detail before answering this question, following Prof. Einstein[2]:

[2] **Albert Einstein** (14 March 1879, Ulm, Germany – 18 April 1955, Princeton, USA) German-born scientist who was somehow active as a theoretical physicist but who unfortunately died four years too soon to benefit from any Geophysical Fluid Dynamics summer course. If it had been the case, which could have been possible given the quality of his resume, there is no doubt that he could have achieved much more.



(a) Lower amplitude

(b) Higher amplitude

Figure 2: Periodograms resulting from the excitation of capillary-gravity waves. The forcing frequency is $f_0 = 25\,\text{Hz}$. The figures from top to bottom show the field at various locations, further away from the forcing apparatus.

A good mathematical model of a physical problem should be as simple as possible, and no simpler.

6 Summary and conclusion

Among all the nonlinear phenomena studied in this summer course, triad interactions are, at first glance, the simplest, most natural and most intuitive ones. This lecture introduced the mathematical theory of triad interactions, beginning with the simplest possible example, namely the single triad of ODEs. It is interesting to note that despite its "simplicity", the problem of single triads of ODEs was only solved approximately at the same time as the first results on solitons were published. The cases of a single triad of PDEs, and/or multiple triad interactions, were solved later and are very briefly described here.

In the last part of these notes, theoretical results are confronted to an experimental situation which sheds light on the difficulty in isolating single-triad interactions in practice. These results should cast some doubt on the belief that these problems are simple, if such a belief ever existed.

Triad interactions have a lot in common with other topics discussed in this lecture course: for example, the KdV model seen in Lecture 5 could be seen as a triad interaction between spatial frequencies 0, 0 and 0: $0 + 0 = 0$. Another more explicit example is the NLS model (see Lectures 3 and 13), which can also be seen as the interaction between a wave of spatial frequency k and a wave of spatial frequency $k + \delta k$ close to k.

This problem, even in its most simple configurations such as single triads or simple triad clusters, is still the center of active mathematical attention (see *e.g.* Bustamante & Kartashova (2009) [4, 5]).

References

[1] F. K. Ball, *Energy transfer between external and internal gravity waves*, Journal of Fluid Mechanics, 19 (1964), pp. 465–478.

[2] D. J. Benney and A. C. Newell, *The propagation of nonlinear wave envelopes*, Journal of Mathematical Physics, 46 (1967), pp. 133–139.

[3] R. W. Boyd, *Nonlinear Optics*, Academic Press, London, 2nd ed., 2003.

[4] M. D. Bustamante and E. Kartashova, *Dynamics of nonlinear resonances in Hamiltonian systems*, Europhysics Letters, 85 (2009), pp. 14004–+.

[5] ——, *Effect of the dynamical phases on the nonlinear amplitudes' evolution*, Europhysics Letters, 85 (2009), pp. 34002–+.

[6] R. A. Cairns, *The role of negative energy waves in some instabilities of parallel flows*, Journal of Fluid Mechanics Digital Archive, 92 (1979), pp. 1–14.

[7] B. Coppi, M. N. Rosenbluth, and R. N. Sudan, *Nonlinear interactions of positive and negative energy modes in rarefied plasmas. I, II (Energy modes nonlinear*

interactions in rarefied plasmas, presenting wave energy and momentum definitions and kinetic equations for mode density in wave number space), Annals of Physics, 55 (1969), pp. 207–270.

[8] A. D. D. Craik, *Wave interactions and fluid flows*, Cambridge Univrsity Press, 1985, ch. 5.

[9] R. C. Davidson, *Methods in nonlinear plasma theory.*, Academic Press, 1972.

[10] K. Hasselmann, *A criterion for nonlinear wave stability*, Journal of Fluid Mechanics, 30 (1967), pp. 273–281.

[11] D. J. Kaup, *Simple harmonic generation: an exact method of solutions*, Studies in Applied Mathematics, 59 (1978), pp. 25–35.

[12] M. Perlin, D. M. Henderson, and J. Hammack, *Experiments on ripple instabilities. Part 2. Selective amplification of resonant triads*, Journal of Fluid Mechanics, 219 (1990), pp. 51–80.

[13] O. M. Phillips, *The dynamics of the upper ocean*, Cambridge Press, 2nd ed., 1977.

[14] V. I. Shrira, V. V. Voronovich, and N. G. Kozhelupova, *Explosive Instability of Vorticity Waves*, Journal of Physical Oceanography, 27 (1997), pp. 542–554.

[15] W. F. Simmons, *A variational method for weak resonant wave interactions*, Proceedings of the Royal Society, London, 309 (1969), pp. 551–575.

[16] V. E. Zakharov, *Stability of periodic waves of finite amplitude on the surface of a deep fluid*, Journal of Applied Mechanics and Technical Physics, 9 (1968), pp. 190–194.

[17] V. E. Zakharov and S. V. Manakov, *The theory of resonance interaction of wave packets in nonlinear media*, Sov. Phys. JETP, 42 (1976), pp. 842–850.

[18] J. M. Ziman, *Electrons and Phonons*, Oxford, 1960.

Lecture 14: Waves on deep water, I

Lecturer: Harvey Segur. Write-up: Adrienne Traxler

June 23, 2009

1 Introduction

In this lecture we address the question of whether there are stable wave patterns that propagate with permanent (or nearly permanent) form on deep water. The primary tool for this investigation is the nonlinear Schrödinger equation (NLS). Below we sketch the derivation of the NLS for deep water waves, and review earlier work on the existence and stability of 1D surface patterns for these waves. The next lecture continues to more recent work on 2D surface patterns and the effect of adding small damping.

2 Derivation of NLS for deep water waves

The nonlinear Schrödinger equation (NLS) describes the slow evolution of a train or packet of waves with the following assumptions:

- the system is conservative (no dissipation)
- then the waves are dispersive (wave speed depends on wavenumber)

Now examine the subset of these waves with

- only small or moderate amplitudes
- traveling in nearly the same direction
- with nearly the same frequency

The derivation sketch follows the by now normal procedure of beginning with the water wave equations, identifying the limit of interest, rescaling the equations to better show that limit, then solving order-by-order.

We begin by considering the case of only gravity waves (neglecting surface tension), in the deep water limit ($kh \to \infty$). Here h is the distance between the equilibrium surface height and the (flat) bottom; a is the wave amplitude; η is the displacement of the water surface from the equilibrium level; and ϕ is the velocity potential, $\mathbf{u} = \nabla\phi$. (See also Lecture 4.) In terms of our small parameter ϵ ($= ah$), we will consider waves that are nearly monochromatic:

$$\begin{aligned} \vec{k} &= (k_0, 0) + O(\epsilon) \\ \omega &= \sqrt{gk_0} + O(\epsilon) \end{aligned}$$

and small in amplitude:

$$k_0\|\eta\| = O(\epsilon)$$

We use a coordinate $\theta = k_0 x - \omega(k_0)t$ and a solution form

$$\eta(x,y,t;\epsilon) = \epsilon[A(\epsilon x, \epsilon y, \epsilon t, \epsilon^2 t)e^{i\theta} + A^* e^{-i\theta}] + \epsilon^2[\text{stuff}_2] + \epsilon^3[\text{stuff}_3] + O(\epsilon^4)$$

and then insert the formal expansions for $\eta(x,y,t;\epsilon)$ and $\phi(x,y,t;\epsilon)$ into the full equations.

Solving order by order, the algebra is nasty, but mostly avoidable via one of the usual suspects such as maple. What we find is the following:

- At $O(\epsilon)$, $\omega^2 = gk$, we recover the linearized dispersion relation for gravity-induced waves on deep water.
- At $O(\epsilon^2)$, the expansion blows up and becomes disordered unless

$$\frac{\partial A}{\partial(\epsilon t)} + c_g \frac{\partial A}{\partial(\epsilon x)} = 0$$

  where c_g is the group velocity (the propagation speed of the wave envelope).

In η above, now define a coordinate moving with the wave $\xi = (\epsilon x) - c_g(\epsilon t)$, a slowly varying transverse direction $\zeta = \epsilon y$, and a slowly varying time $\tau = \epsilon^2 t$. Now examine $O(\epsilon^3)$, where the expansion again becomes disordered unless $A(\xi,\zeta,\tau)$ satisfies

$$i\partial_\tau A + \alpha\partial_\xi^2 A + \beta\partial_\zeta^2 A + \gamma|A|^2 A = 0 \tag{1}$$

which is the 2D nonlinear Schrödinger equation. Here α, β, and γ are real numbers determined by the problem. For the case of deep water waves, the signs of the coefficients (independent of scaling choice) are $\alpha < 0$, $\beta > 0$, and $\gamma < 0$.

The NLS or some equivalent has been derived at many times and in many contexts, including: Zahkarov in 1968 for water waves [14], Ostrovsky in 1967 for optics [8], Benjamin and Feir in 1967 for water waves [3], a general formulation by Benney and Newell in 1967 [4], Whitham in 1965 [12], whose formulation was used by Lighthill in 1965 [6], and finally by Stokes in 1847 for water waves with no spatial dependence. A historical overview is provided by Zakharov and Ostrovsky [15].

3 Waves of permanent form on deep water

Now we come to the overarching question of interest in this and the next lecture: do stable waves of permanent form in deep water exist? Starting back at the beginning, Stokes [9] considered a spatially uniform train of plane waves,

$$\eta(x,y,t;\epsilon) = \epsilon[A(\epsilon^2 t)e^{i\theta} + A^* e^{-i\theta}] + O(\epsilon^2)$$

which after removing the spatial derivatives from (1) must satisfy

$$i\partial_\tau A + \gamma|A|^2 A = 0 \ , \ \gamma = -4k_0^2$$

where the expression for γ depends on the choice of scalings (see [1]). If we take the complex conjugate of this equation, multiply the original by A^* and the conjugate by A, and subtract them, we find:

$$iA^*\partial_\tau A + iA\partial_\tau A^* = 0$$

In other words, $\partial_\tau(A^*A) = 0$, so the square of the amplitude is a constant. With that, the equation is easy to solve, yielding

$$A(\tau) = (A_0 e^{i\phi})e^{i\gamma|A_0|^2\tau}$$

Putting this into the η expression above, and using θ from earlier, we have

$$\eta(x,t;\epsilon) = 2\epsilon|A_0|\cos\left[k_0 x - \omega(k_0)t - (2\epsilon k_0|A_0|)^2 t\right] + O(\epsilon^2)$$

where the final term inside the cosine is Stokes' nonlinear correction to the frequency.

Although Stokes found a nonlinear correction for water waves of permanent form with finite amplitude, he did not prove that such waves existed. Nekrassov (available in [7]) and Levi-Civita [5] accomplished this task in the 1920s, and Struik [10] extended their work from deep water to water of any constant depth. Finally, Amick and Toland [2] obtained optimal results about the existence of waves of permanent form in 2D. The solutions in question, a periodic train of plane waves, are approximated by experimental data such as in the top panel of Figure 1.

3.1 Why don't we see them?

Despite the above work to prove the existence of such waves, they are not commonly observed in nature, but perhaps they can be produced in more controlled conditions. Photos from Benjamin in 1967, in Figure 1, show a wave train disintegrating in 60 meters, as discussed in the stability analysis of [3].

Returning to the NLS in Zakharov 1968,

$$i\partial_\tau A + \alpha\partial_\xi^2 A + \beta\partial_\zeta^2 A + \gamma|A|^2 A = 0$$

for the case of gravity waves on deep water, we have $\alpha < 0$, $\beta > 0$, and $\gamma < 0$. The "Stokes wave" above is the solution for a spatially uniform, finite amplitude train of plane waves to third order expansion in the water wave equations. We can now check its stability by linearizing the NLS around a Stokes wave and looking for unstable modes.

$$A(\xi,\zeta,\tau) = |A_0|\, e^{i\gamma|A_0|^2\tau}$$

Assume additional small perturbations:

$$A(\xi,\zeta,\tau) = e^{i\gamma|A_0|^2\tau}[|A_0| + \mu\cdot u(\xi,\zeta,\tau) + i\mu\cdot v(\xi,\zeta,\tau)] + O(\mu^2)$$

Putting this into (1), above, and keeping only terms up to $O(\mu)$, we have:

$$\begin{aligned} e^{i\gamma|A_0|^2\tau}\big[&-\gamma|A_0|^2(|A_0| + \mu u + i\mu v) + i\mu(\partial_\tau u + i\partial_\tau v) \\ &+\alpha\mu(\partial_\xi^2 u + i\partial_\xi^2 v) + \beta\mu(\partial_\zeta^2 u + i\partial_\zeta^2 v) + \gamma|A_0|^2(|A_0| + \mu u + i\mu v) + 2\gamma|A_0|^2\mu u\big] = 0 \end{aligned}$$

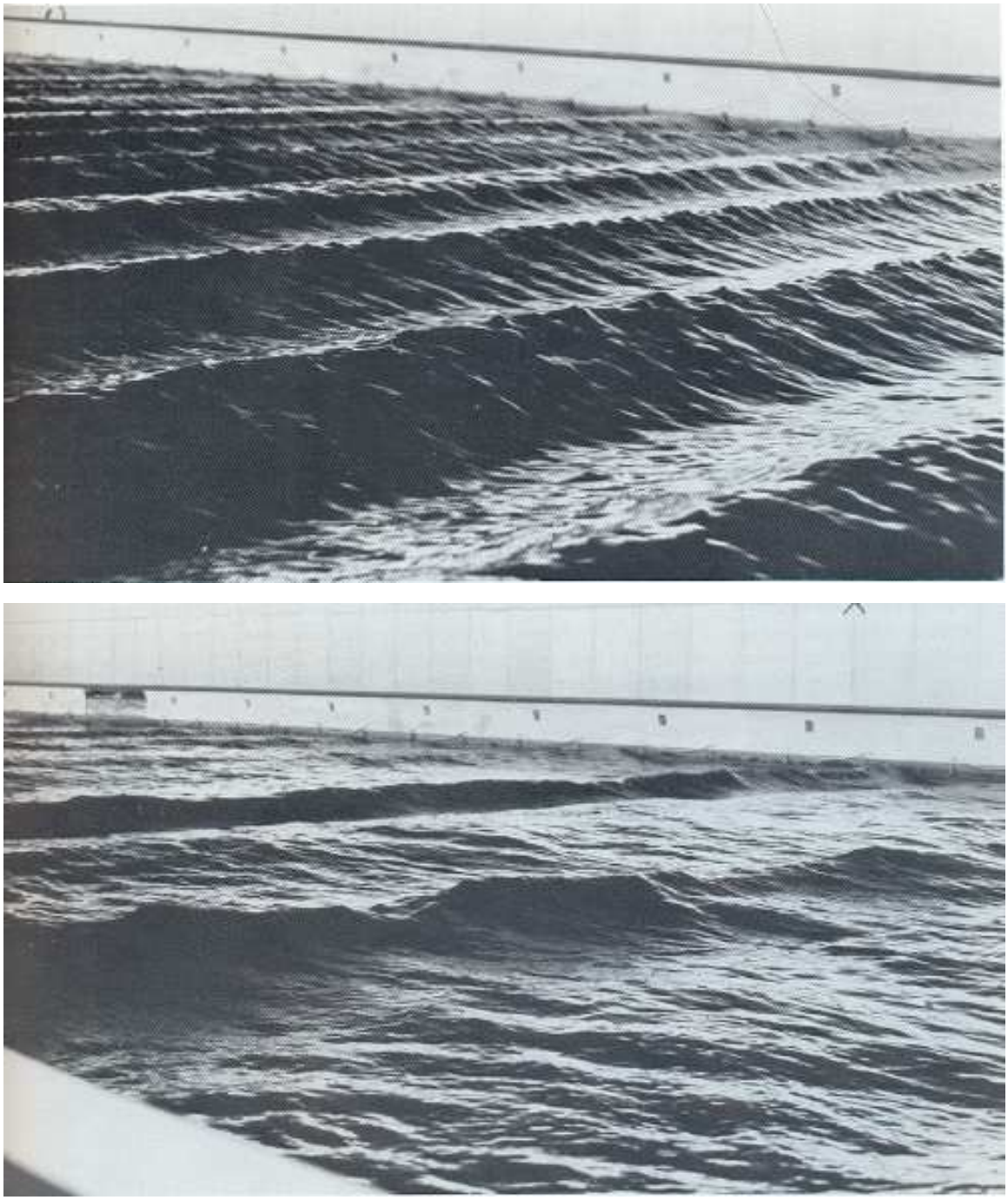


Figure 1: Wavetrain in deep water ($L = 2.3$ m, $h = 7.6$ m), with 60 meters between photos.

where the real and imaginary parts respectively break into a pair of linear PDEs with constant coefficients,

$$\begin{aligned} \partial_\tau v &= \alpha\partial_\xi^2 u + \beta\partial_\zeta^2 u + 2\gamma|A_0|^2 u \quad (2) \\ -\partial_\tau u &= \alpha\partial_\xi^2 v + \beta\partial_\zeta^2 v \quad (3) \end{aligned}$$

We seek a solution of the form

$$\begin{aligned} u &= U \cdot e^{im\xi + il\zeta + \Omega\tau} + (\text{c.c.}) \\ v &= V \cdot e^{im\xi + il\zeta + \Omega\tau} + (\text{c.c.}) \end{aligned}$$

Putting these into (2) and (3), above, and eliminating U and V, we are left with an algebraic equation for the linear stability:

$$\Omega^2 + (\alpha m^2 + \beta l^2)(\alpha m^2 + \beta l^2 - 2\gamma|A_0|^2) = 0 \quad (4)$$

where $Re(\Omega) > 0$, positive real growth rate, indicates linear instability. Marginal stability occurs at

$$\Omega = \pm\sqrt{-(\alpha m^2 + \beta l^2)(\alpha m^2 + \beta l^2 - 2\gamma|A_0|^2)} = 0$$

for which the first root (recalling that here $\alpha < 0$, $\beta > 0$) is

$$\begin{aligned} \alpha m^2 &= -\beta l^2 \\ l &= \pm\sqrt{\frac{|\alpha|}{\beta}}\, m \end{aligned}$$

which defines a pair of lines with opposite slope that cross at the origin. The second root is

$$\alpha m^2 + \beta l^2 - 2\gamma|A_0|^2 = 0$$

which defines a hyperbola. See Figure 2 for the regions described by this expression; instability occurs for points along the m-axis between the two sets of curves. (This can be seen by considering $l^2 = 0$ and m^2 small, for which $\alpha m^2 < 0$ and $\alpha m^2 - 2\gamma|A_0|^2 > 0$; similarly, a positive real root does not exist for $m^2 = 0$ and $l^2 \neq 0$.)

The maximum value of the growth rate can be easily determined by noting that l and m appear only in the combination $\alpha m^2 + \beta l^2$, so we can call that something easy like s and look for extrema of Ω^2:

$$\begin{aligned} \Omega^2 &= -s(s - 2\gamma|A_0|^2) \\ \frac{\partial\Omega^2}{\partial s} &= -(s - 2\gamma|A_0|^2) - s \\ 0 &= -2s_{max} + 2\gamma|A_0|^2 \\ s_{max} &= \gamma|A_0|^2 \end{aligned}$$

The growth rate evaluated at this point is

$$\begin{aligned} \Omega^2(s_{max}) &= -\gamma|A_0|^2(\gamma|A_0|^2 - 2\gamma|A_0|^2) \\ \Omega_{max} &= |\gamma|\,|A_0|^2 \end{aligned}$$

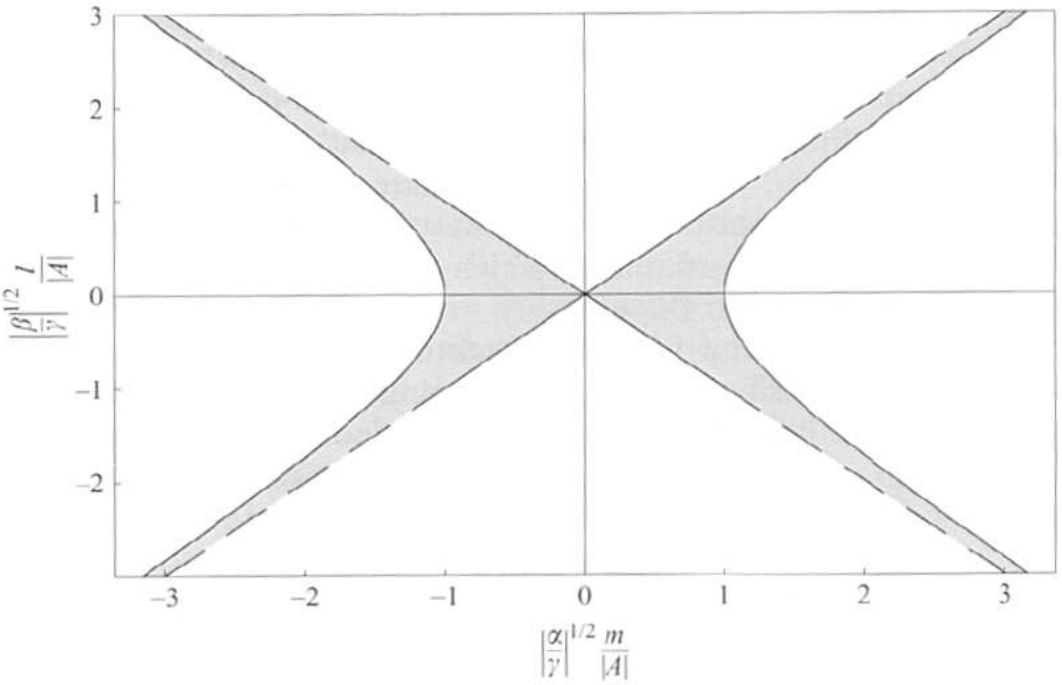


Figure 2: The regions of stability described by equation 4. Shaded regions are unstable.

Collecting the above, we have the result that a uniform train of finite amplitude plane waves is unstable in deep water, with the most unstable mode growing at $\Omega_{max} = |\gamma|\,|A_0|^2$. The instability is nonlinear in the sense that its growth rate depends on the amplitude $|A_0|$, so as $|A_0| \to 0$ then $\Omega_{max} \to 0$.

For applications other than deep-water gravity waves, we can repeat the above procedure for the stability diagram, where the respective signs of the coefficients will determine the existence of unstable regions. In general we have:

- $\alpha\beta < 0$, $\alpha\gamma > 0 \rightarrow$ unstable
- $\alpha\beta < 0$, $\alpha\gamma < 0 \rightarrow$ unstable
- $\alpha\beta > 0$, $\alpha\gamma > 0 \rightarrow$ unstable
- $\alpha\beta > 0$, $\alpha\gamma < 0 \rightarrow$ stable

One example of the instability of a uniform wavetrain was shown in Figure 1. For a second example in an electromagnetic context, see Figure 3 (which can be viewed as a cut across the center of the stability diagram, with the most unstable modes growing at either side).

3.2 A different wave pattern of permanent form

Having addressed the plane wave set of solutions above, and found them to be linearly unstable, another class of solutions is now our only hope for finding permanent stable wave

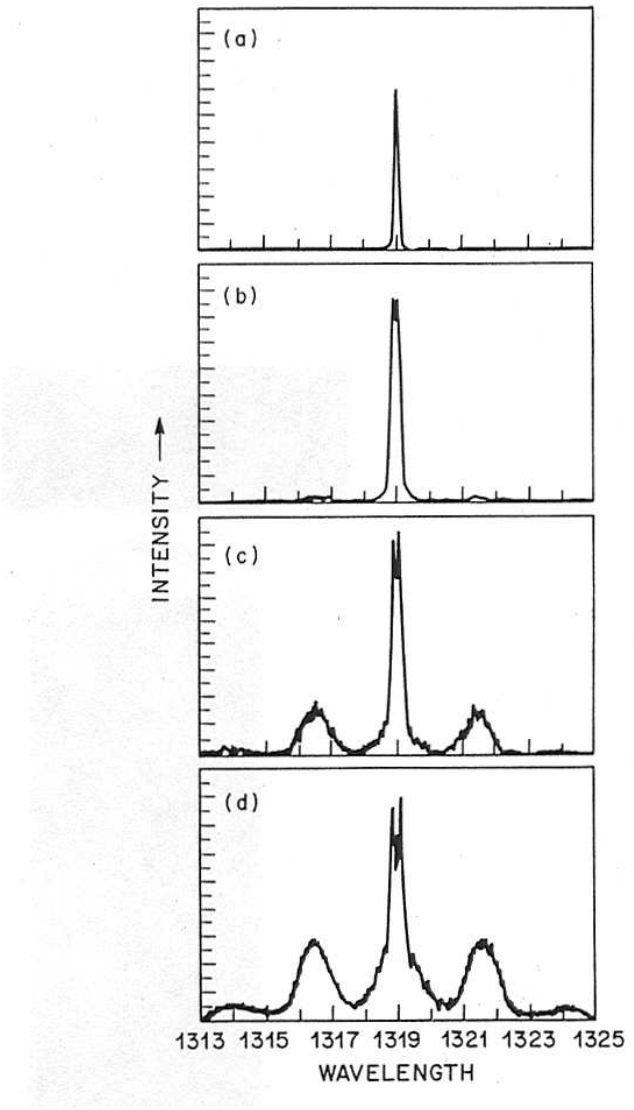


Figure 3: Experimental observation of modulational instability. The horizontal axis is in nanometers, so the length scale is $L = 1.3 \cdot 10^{-6}$ m, with timescale $T = 4 \cdot 10^{-15}$ s. Input power level low (a); 5.5 W (b); 6.1 W (c); 7.1 W (d). Figure adapted from [11].

patterns on deep water. This section will take a slightly different tack, using the Inverse Scattering Transform to solve the initial value problem (see Lecture 5) for the 1D NLS equation. This method enables us to show the existence of soliton solutions (which are by definition waves of permanent form) and deferring the question of their stability to transverse perturbations until the next lecture.

Continuing the search for another stable wave form, Zakharov and Shabat in 1972 [16] considered the NLS in 1D,

$$i\partial_\tau A = \partial_\xi^2 A + 2\sigma|A|^2 A \tag{5}$$

where $\sigma = -1$ is the "defocusing" case and $\sigma = +1$ is the "focusing" case appropriate for deep water. Setting $\sigma = 1$ and looking for traveling waves of "permanent form," there is a special case:

$$A(\xi,\tau) = 2a \cdot e^{-i(2a)^2\tau} \operatorname{sech} 2a(\xi + \xi_0)$$

The corresponding shape of the free surface, as seen in Section 2, takes the form $\eta(x,t;\epsilon) = \epsilon(Ae^{i\theta} + A^* e^{-i\theta}) + O(\epsilon^2)$ where $\theta = kx - \omega(k)t$ so that

$$\eta(x,t;\epsilon) = (2\epsilon a) \operatorname{sech}\left[(2\epsilon a)(x - c_g t)\right] \cos\left\{kx - [\omega(k) + (2\epsilon a)^2]t\right\} \tag{6}$$

which is a wave packet with special shape (see Figure 4) discussed further below. This solution is reminiscent of the sech^2 soliton solutions of the KdV. This remark naturally leads us to ask which other properties of the KdV, and associated solution techniques, apply to the 1D NLS.

Zakharov and Shabat (1972) [16] demonstrated that the 1D NLS (5) is also completely integrable for either sign of σ. As for the KdV, there are an infinite number of explicit, local conservation laws, of which the first three are

$$\begin{aligned} i\partial_\tau(|A|^2) &= \partial_\xi(A^*\partial_\xi A - A\partial_\xi A^*) \\ i\partial_\tau(A^*\partial_\xi A - A\partial_\xi A^*) &= \partial_\xi(\ldots) \\ i\partial_\tau(|\partial_\xi A|^2 + \sigma|A|^4) &= \partial_\xi(\ldots) \end{aligned}$$

The other major consequence of complete integrability (see notes for Lecture 5) is that we can treat the problem with the inverse scattering transform. However, this time the scattering problem is no longer defined by the time-invariant Schrödinger equation, but is found instead to be given by the set of equations

$$\begin{aligned} \partial_\xi v_1 &= -i\lambda v_1 + A v_2 \\ \partial_\xi v_2 &= -\sigma A^* v_1 + i\lambda v_2. \end{aligned}$$

The requirement that the eigenvalues be independent of time, together with the compatibility condition

$$\partial_\xi \partial_\tau \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \partial_\tau \partial_\xi \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

recover the 1D NLS provided

$$\begin{aligned} \partial_\tau v_1 &= (-2i\lambda^2 + i\sigma|A|^2)v_1 + (2A\lambda + i\partial_\xi A)v_2 \\ \partial_\tau v_2 &= (-2\sigma A^*\lambda + i\sigma\partial_\xi A^*)v_1 - (-2i\lambda^2 + i\sigma|A|^2)v_2 \end{aligned}$$

Indeed,

$$\begin{aligned}
\partial_\xi \partial_\tau v_1 &= (i\sigma A \partial_\xi A^* + i\sigma A^* \partial_\xi A) v_1 + (-2i\lambda^2 + i\sigma |A|^2)(-i\lambda v_1 + A v_2) \\
&\quad + (2\lambda \partial_\xi A + i\partial_\xi^2 A) v_2 + (2A\lambda + i\partial_x iA)(-\sigma A^* v_1 + i\lambda v_2) \\
&= (i\sigma A \partial_\xi A^* - 2\lambda^3 - \sigma\lambda |A|^2) v_1 + (i\sigma |A|^2 A + \lambda \partial_\xi A + i\partial_\xi^2 A) v_2 \\
\partial_\tau \partial_\xi v_1 &= -i\lambda[(-2i\lambda^2 + i\sigma |A|^2) v_1 + (2A\lambda + i\partial_\xi A) v_2] + (\partial_\tau A) v_2 \\
&\quad + A[(-2\sigma A^* \lambda + i\sigma \partial_\xi A^*) v_1 + (2i\lambda^2 - i\sigma |A|^2) v_2] \\
&= (-2\lambda^3 - \sigma\lambda |A|^2 + i\sigma A \partial_\xi A^*) v_1 + (\lambda \partial_\xi A + \partial_\tau A - i\sigma |A|^2 A) v_2
\end{aligned}$$

Equating the two, all v_1 terms cancel, but for the two v_2 sides to balance, we must have

$$\partial_\tau A = i\partial_\xi^2 A + 2i\sigma |A|^2 A$$

Repeating the process for the v_2 equations produces the complex conjugate of the above. From this we arrive back at the 1D NLS,

$$i\partial_\tau A = \partial_\xi^2 A + 2\sigma |A|^2 A$$

as should be the case.

As a consequence of constructing this Inverse Scattering Transform method for the 1D NLS, it can be shown that for $\sigma = 1$ (focusing NLS), any smooth initial data $A(\xi, 0)$ with $\int |A| d\xi < \infty$ evolves into N "envelope solitons" which persist forever, plus an oscillatory wavetrain that decays in amplitude as $\tau \to \infty$. Envelope solitons are stable for the focusing case in the 1D NLS.

Figure 4 (from unpublished work by Hammack) shows experimental evidence for the existence of these envelope solitons, and reveal that they match the sech solution (6) very well. Figure 5 shows data from three other experiments on envelope solitons: the first shows the evolution of two solitons in isolation; the second has some additional noise that eventually separates from the soliton; the third shows one of the wave pulses overtaking and passing through another, emerging on the other side in (mostly) unchanged form. In contrast to Figure 1, where the uniform train of plane waves disintegrates quickly, the experiments in Figures 4 and 5 show the relative persistence of soliton solutions, providing some confirmation of the stability predicted by this 1D theory.

4 Conclusions

Finally, we reach the point of drawing tentative conclusions. According to the 1D or 2D NLS, a uniform train of plane waves is unstable in deep water. However, according to focusing NLS in 1D with initial data in L_1 (i.e., meeting the integrability condition above), envelope solitons are stable in deep water. Experimental evidence seems to support both of these conclusions, but we have not yet addressed the stability of the 1D soliton solutions to transverse perturbations. For more on that topic, see the next lecture.

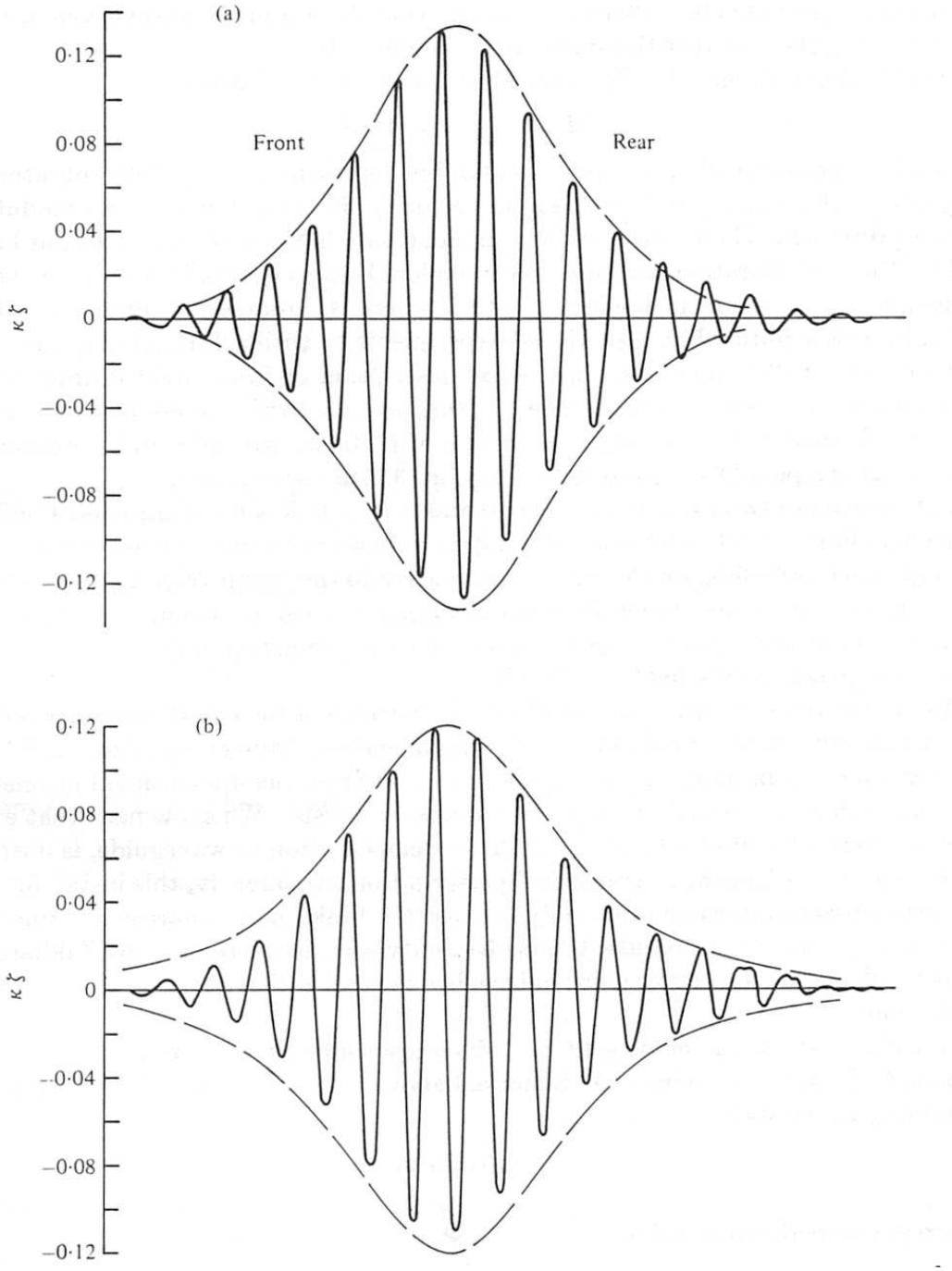


Figure 4: Measured surface displacement, showing evolution of envelope soliton at two downstream locations. Dimensions are $h = 1$ m, $kh = 4.0$, $\omega = 1$ Hz. The solid line is the measured history of surface displacement; the dashed line is the theoretical envelope shape. Data was taken at 6 m (top) and 30 m (bottom) downstream of the wavemaker.

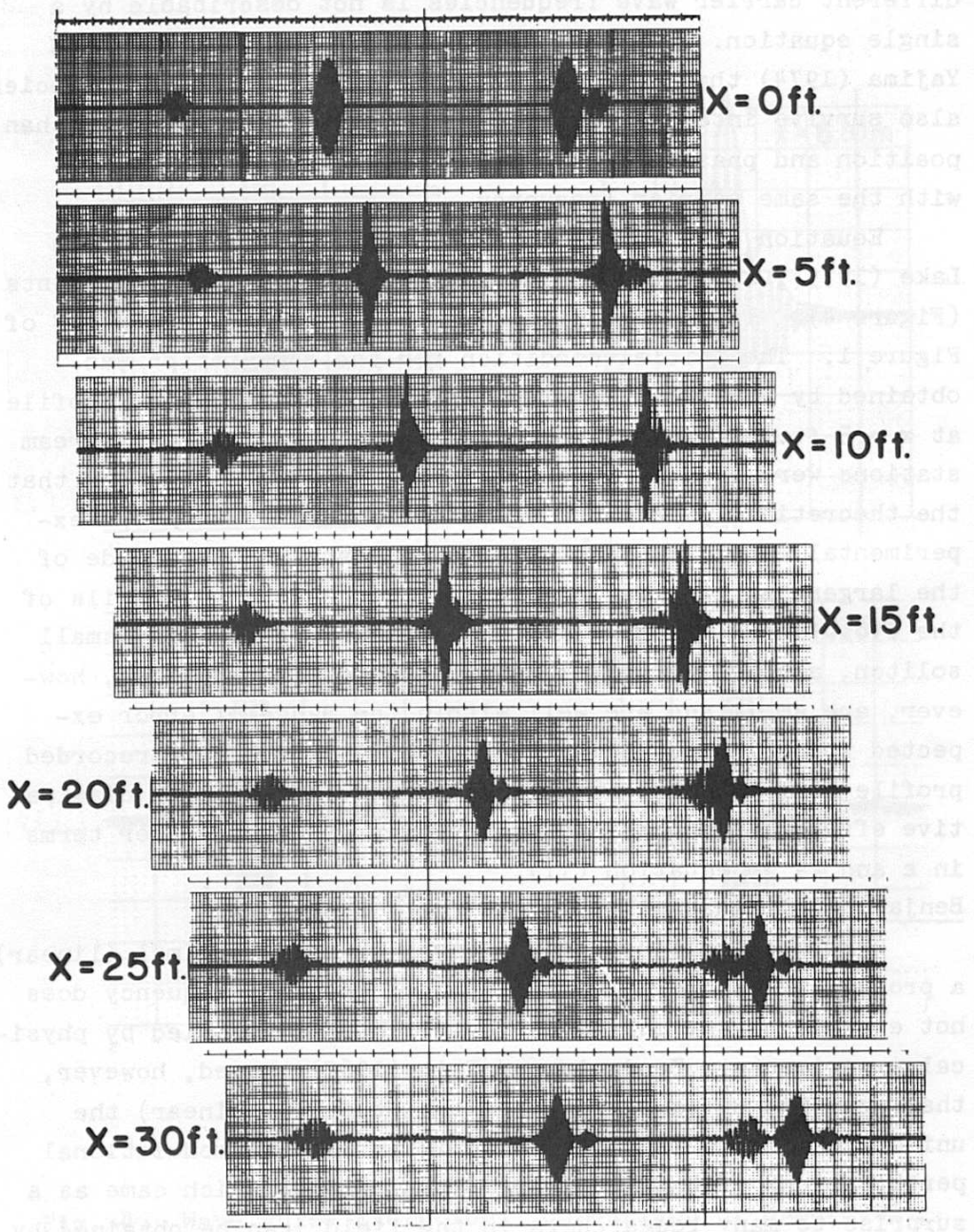


Figure 5: Wave pulse interaction: one wave pulse overtaking and passing through another wave pulse. Left-hand trace: first pulse alone, $\omega_0 = 1.5$ Hz, initial $(ka)_{max} \simeq 0.01$, six-cycle pulse. Center trace: second pulse alone, $\omega_0 = 3$ Hz, initial $(ka)_{max} \simeq 0.2$, 12-cycle pulse which disintegrates into two solitons. Right-hand traces: interaction of the two pulses. Adapted from [13].

References

[1] M. Ablowitz and H. Segur, *Solitons and the inverse scattering transform*, SIAM, Philadelphia, 1981.

[2] C. Amick and J. Toland, *On solitary water-waves of finite amplitude*, Archive for Rational Mechanics and Analysis, 76 (1981), pp. 9–95.

[3] T. Benjamin and J. Feir, *The disintegration of wave trains on deep water Part 1. Theory*, Journal of Fluid Mechanics, 27 (1967), pp. 417–430.

[4] D. Benney and A. Newell, *Propagation of nonlinear wave envelopes*, Journal of Mathematical Physics, 46 (1967), pp. 133–139.

[5] T. Levi-Civita, *Determination rigourese des ondes permanentes d'ampleur finie*, Math. Ann., 93 (1925), p. 264.

[6] M. Lighthill, *Contribution to the theory of waves in non-linear dispersive systems*, J. Inst. Math. Appl., 1 (1965), pp. 269–306.

[7] A. Nekrasov, *The exact theory of steady state waves on the surface of a heavy liquid*, University of Wisconsin, Madison, 1967.

[8] L. Ostrovskii, *Propagation of wave packets and space-time self-focusing in a nonlinear medium*, Sov. Phys. JETP, 24 (1967), pp. 797–800.

[9] G. Stokes, *On the theory of oscillatory waves*, Trans. Camb. Philos. Soc., 8 (1847), pp. 441–455.

[10] D. Struik, *Détermination rigoureuse des ondes irrotationelles périodiques dans un canal à profondeur finie*, Mathematische Annalen, 95 (1926), pp. 595–634.

[11] H. A. Tai, K. and A. Tomita, *Observation of modulational instability in optical fibers*, Physical Review Letters, 56 (1986), pp. 135–138.

[12] G. Whitham, *A general approach to linear and nonlinear dispersive waves using a Lagrangian*, Journal of Fluid Mechanics, 22 (1965), pp. 273–283.

[13] H. Yuen and B. Lake, *Nonlinear deep water waves: Theory and experiment*, Physics of Fluids, 18 (1975), pp. 956–960.

[14] V. Zakharov, *Stability of periodic waves of finite amplitude on the surface of a deep fluid*, J. Appl. Mech. and Tech. Phys., 9 (1968), pp. 190–194.

[15] V. Zakharov and L. Ostrovsky, *Modulation instability: The beginning*, Physica D, 238 (2009), pp. 540–548.

[16] V. Zakharov and A. Shabat, *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in non linear media*, Soviet Physics JETP, 34 (1972), pp. 62–69.

Lecture 15: Waves on deep water, II

Lecturer: Harvey Segur. Write-up: Andong He

June 23, 2009

1 Introduction.

In the previous lecture (Lecture 14) we sketched the derivation of the nonlinear Schrödinger equation (NLS)

$$i\partial_\tau A + \alpha\partial_\xi^2 A + \beta\partial_\zeta^2 A + \gamma|A|^2 A = 0, \tag{1}$$

where $\{\alpha, \beta, \gamma\}$ are real constants related to the original physical system from which the NLS was derived. Equation (1) is an important approximate model to describe deep water waves. In probing into the existence of stable wave patterns that propagate with (nearly) permanent form in deep water, we encountered the modulational (or Benjamin-Feir) instability. In most cases, 1D plane wave solutions of the NLS were found to be unstable to 1D perturbations (along the wave propagation axis) and 2D perturbations (transverse to the wave propagation axis).

We continue studying the existence and stability of waves with either 1-D or 2-D surface patterns in this lecture, focusing on more recent work. We discuss two topics: (1) what happens after the initial development of the Benjamin-Feir instability for "high"-amplitude nonlinear plane waves, namely Fermi-Pasta-Ulam (FPU) recurrence (with additional subtleties), and (2) the possible stabilization of "low"-amplitude nonlinear plane waves against the Benjamin-Feir instability by dissipation.

2 Near recurrence of initial states.

Benjamin & Feir [2] showed that a periodic wave train with initially uniform finite amplitude is unstable to infinitesimal perturbations. The instability takes the form of a growing "modulation" of the plane wave, or when viewed in terms of the Fourier spectrum, as exponentially growing sidebands to the plane wave frequency. But this instability is only the beginning of the story: the long-time behavior of such wave trains is arguably even more remarkable.

Lake, Yuen, Rungaldier & Ferguson [10] proposed that with periodic boundary conditions, the focusing 1-D NLS ($\sigma = 1$ in (3)) should exhibit near recurrence of initial states just as the Korteweg-de Vries equation does (see Lecture 5).

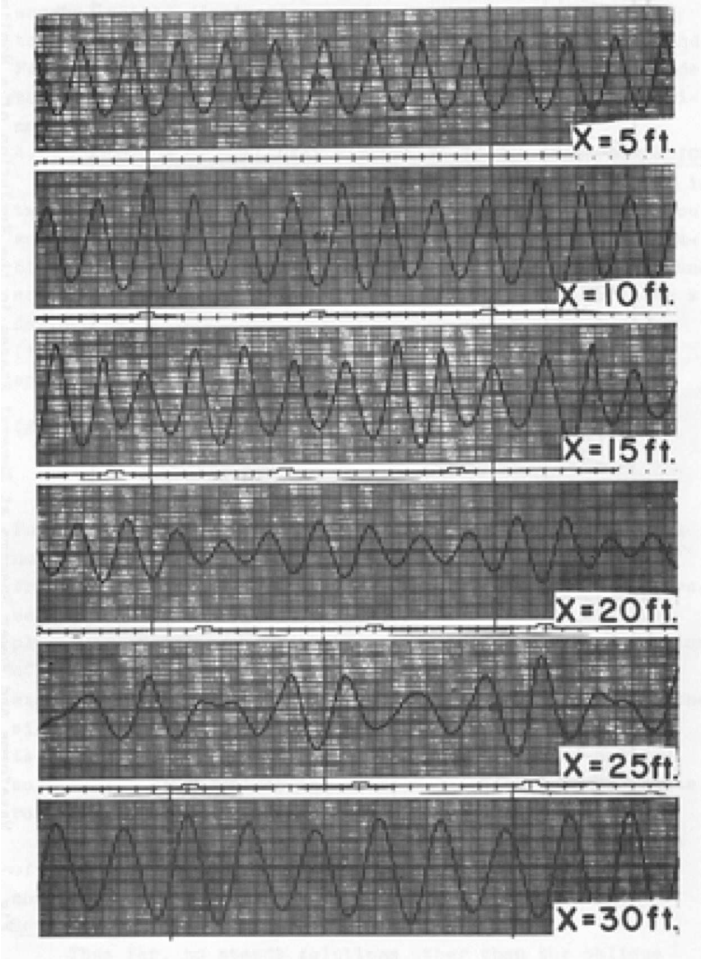


Figure 1: Example of the long-time evolution of an initially nonlinear wave train. Initial wave frequency is 3.6 Hz; oscillograph records shown on expanded time scale to display individual wave shapes; wave shapes are not exact repetitions each modulation period because modulation period does not contain integral number of waves.

To illustrate very qualitatively what "near recurrence of initial states" may mean, we take the linearized equations on deep water with periodic boundary conditions as an example, one solution of which will be

$$\eta(x,t) = \sum_{m=1}^{N} a_m \cos\{mx - \omega_m t + \phi_m\}, \quad \text{where } \omega_m^2 = gm. \tag{2}$$

Since frequencies are not rationally related (recall that $\omega_m = \omega_1\sqrt{m}$), η is not periodic in time. But for functions η that can be written as a sum over a *finite* number N of such terms, the solution returns *close* to its initial state.

The situation for solutions of the NLS is of course different, owing to the nonlinear nature of the equation, but is not entirely unrelated. Just as for the KdV, we saw that

the NLS can be solved using an inverse scattering transform. In the previous lecture, we discussed the case of an infinite domain, which yields, just as for the KdV, a finite number of soliton solutions, each related to a discrete eigenvalue and corresponding eigenmode of the scattering problem. A related scattering problem can be constructed in the periodic domain, and similarly yields a *finite* number of eigenvalues and eigenmodes. Hence, the dynamics of the solutions of the NLS in the periodic domain are limited to understanding the evolution of a finite number of periodic modes. By contrast to the linear case described above, the solution of the NLS is not a linear superposition of these modes, but nevertheless exhibits a similar recurrence phenomenon. This type of long-time behavior, generic to many nonlinear systems (KdV, NLS, ...) was first discovered by Fermi, Pasta & Ulam [5] in numerical experiments, and has become known as the of Fermi-Pasta-Ulam (or FPU) recurrence phenomenon.

Lake *et. al.* [10] investigated experimentally the long-time behavior of nonlinear wave trains generated by a wavemaker in a water tank. During the early stages of evolution, an initially unmodulated wave train develops an amplitude modulation as predicted by the analysis of Benjamin & Feir (see results for $x = 5, 10$ and 15 ft in Figure 1). As the wave train evolves further, the modulation increases in amplitude and the results of their linearized stability analysis no longer apply. However, as the system continues evolving, the wave train is observed to "demodulate" and the wave form returns to a relatively uniform state ($x = 30$ ft in Figure 1). Therefore the recurrence of wave patterns in deep water observed by Lake *et. at.* might be the first physical evidence of the FPU recurrence phenomenon.

3 2-D free surface.

Let us first briefly summarize the results from the last lecture. We saw that the 1D nonlinear Schrödinger equation,

$$i\partial_\tau A + \alpha\partial_\xi^2 A + \gamma|A|^2 A = 0 \tag{3}$$

has one-soliton solution,s as for example

$$A = a\left|\frac{2\alpha}{\gamma}\right|^{1/2} \operatorname{sech}\{a(\xi - 2b\tau)\}\exp\{ib\xi + i\alpha(a^2 - b^2)\tau\}, \tag{4}$$

in the case of $\alpha\gamma > 0$ (a and b are constants). These solutions are envelope solitons. There are also "dark solitons" when $\alpha\gamma < 0$. The dark solitons are a local reduction in the amplitude of a wave train.

The 1D NLS is often rewritten as

$$i\partial_\tau A + \partial_\xi^2 A + 2\sigma|A|^2 A = 0. \tag{5}$$

by a change of variables, with no loss of generality. In that case, envelope solitons occur for $\sigma = 1$ and dark solitons for $\sigma = -1$. The stability of solutions of the 1D NLS to transverse perturbations is studied using the 2D NLS:

$$i\partial_\tau A + \partial_\xi^2 A + \beta\partial_\zeta^2 A + 2\sigma|A|^2 A = 0. \tag{6}$$

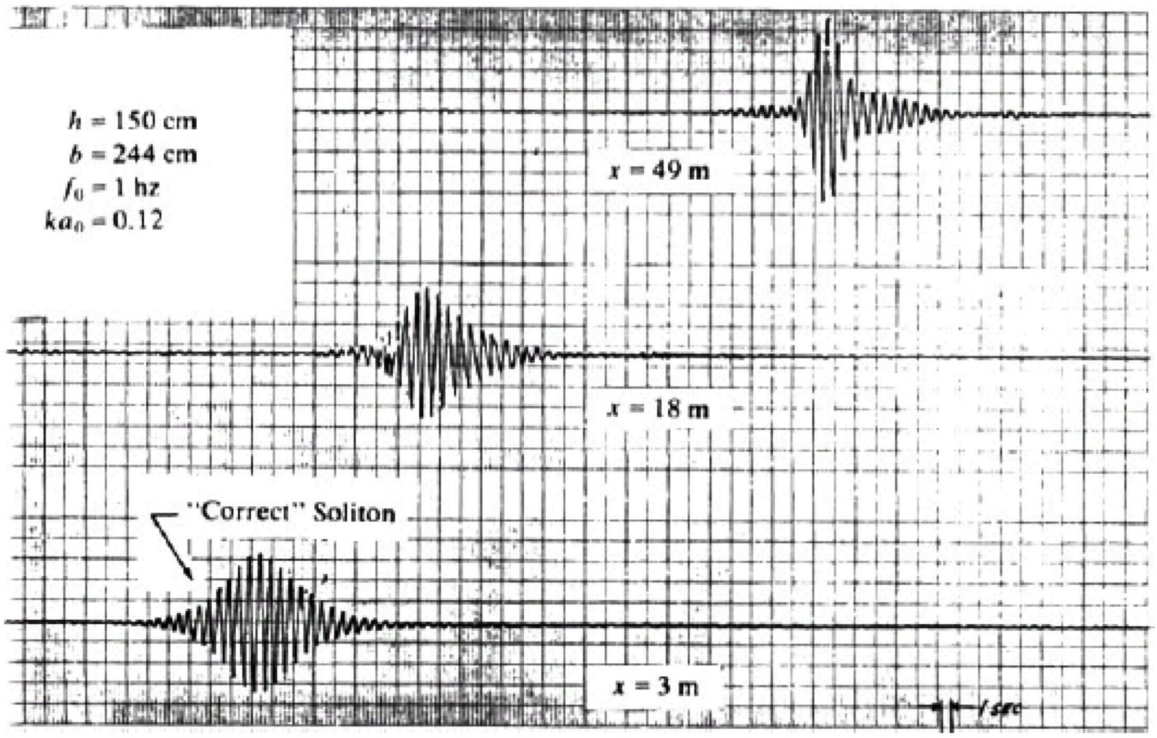


Figure 2: Evolution of water packet in a long tank, showing the transverse instability that was absent in the shorter tank of Figure 2. (Courtesy of J. Hammack)

Zakharov & Rubenchik [12] showed that in the case of $\sigma = 1$, for either sign of β, envelope solitons are unstable to 2-D perturbations; in the case of $\sigma = -1$, for either sign of β, dark solitons are unstable to 2-D perturbations. They also found that the unstable perturbations will have long transverse wavelengths.

As discussed in the previous lecture, Hammack had found "stable" 1D envelope solitons (where the only evolution of the soliton was caused by a weak damping), in apparent contradiction to the aforementioned results. However he later performed additional experiments using the same wavemaker and imposing nearly identical initial conditions, but in a longer tank. It turns out that the longer tank admitted the destabilizing transverse modes that were artificially excluded in the shorter tank, as seen in Figure 2.

Let us now discuss aspects of the 2D NLS which are specific to 2D solutions, first in theory, then in experiments.

Zakharov & Synakh [13] considered the elliptic focusing NLS in 2-D case by choosing $\beta = \sigma = 1$ in equation (6) to obtain

$$i\partial_\tau A + \partial_\xi^2 A + \partial_\zeta^2 A + 2|A|^2 A = 0. \tag{7}$$

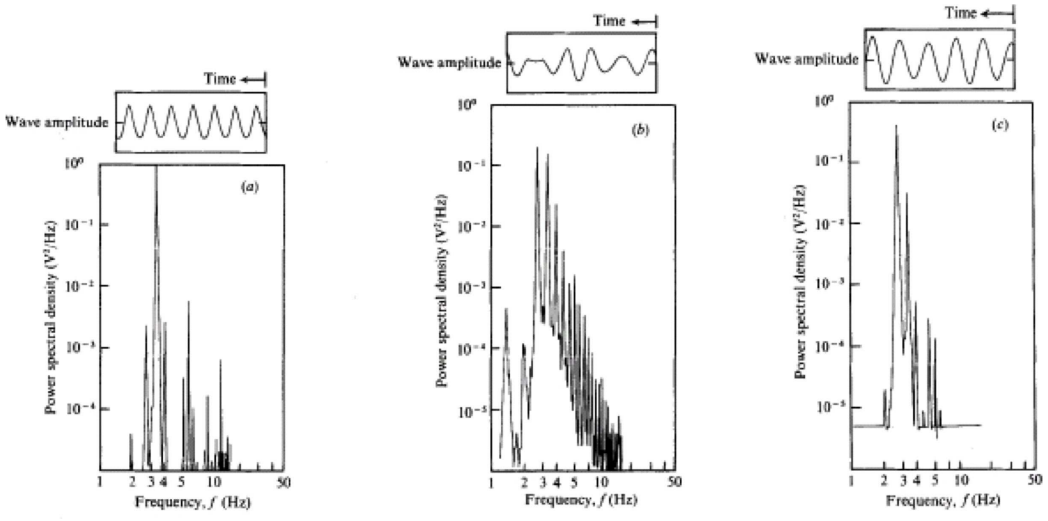


Figure 3: Evolution of a nonlinear finite amplitude wave train: wave forms and power spectral densities vs. propagation distance. (a) Initial stage of side-band growth, $z = 5$ ft, carrier wave with small amplitude modulation. (b) $z = 10$ ft, strong amplitude modulation, energy spread over many frequency components. (c) $z = 25$ ft, reduced amplitude modulation, return of energy to frequency components of original carrier wave, its side bands and harmonics. $f_0 = 3.25$Hz, $(ka)_0 = \delta = 0.23$, $(ka)_{5ft} = 0.29$.

They were able to find the following four conservation laws:

$$I_1 = \int\int (|A|^2) \mathrm{d}\xi \mathrm{d}\zeta, \tag{8}$$

$$I_2 = \int\int (A\partial_\xi A^* - A^*\partial_\xi A) \mathrm{d}\xi \mathrm{d}\zeta, \tag{9}$$

$$I_3 = \int\int (A\partial_\zeta A^* - A^*\partial_\zeta A) \mathrm{d}\xi \mathrm{d}\zeta, \tag{10}$$

$$H = I_4 = \int\int (|\nabla A|^2 - |A|^4) \mathrm{d}\xi \mathrm{d}\zeta. \tag{11}$$

If we consider the function $J(\tau) = \int\int ((\xi^2 + \zeta^2)\,|A|^2) d\xi \mathrm{d}\zeta$, and interpret $|A|^2(\xi, \zeta, \tau)$ as the "mass density", then I_1 is the "total mass" and $J(\tau) \geq 0$ is the "moment of inertia". It follows by direct calculations that

$$\frac{\mathrm{d}^2 J}{\mathrm{d}\tau^2} = 8H, \tag{12}$$

where H is defined in (11). If $H < 0$, $J(\tau)$ will become negative in finite time, and it may happen while quantities I_1, I_2, I_3 and H are kept conserved. This phenomenon, which is so called "wave collapse", has been important in nonlinear optic; nevertheless, since the governing equation (6) (when $\beta = \sigma = 1$) does not apply to the surface-wave case, it is not our major concern in present discussion.

Experiments of 2D NLS wave patterns have revealed a variety of puzzling features. For example, we discussed in Section 2 the notion of FPU recurrence, and its apparent observations in the experiments of Lake, Yuen, Rungaldier & Ferguson [10]. However, a closer

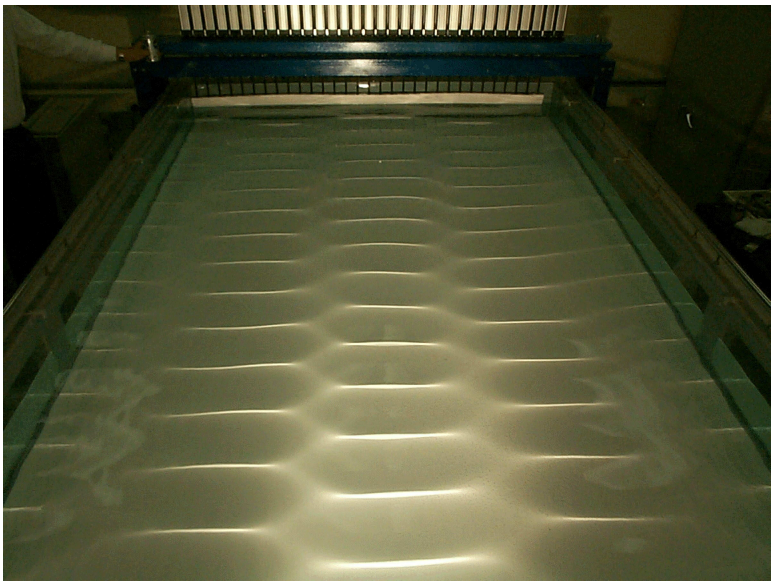


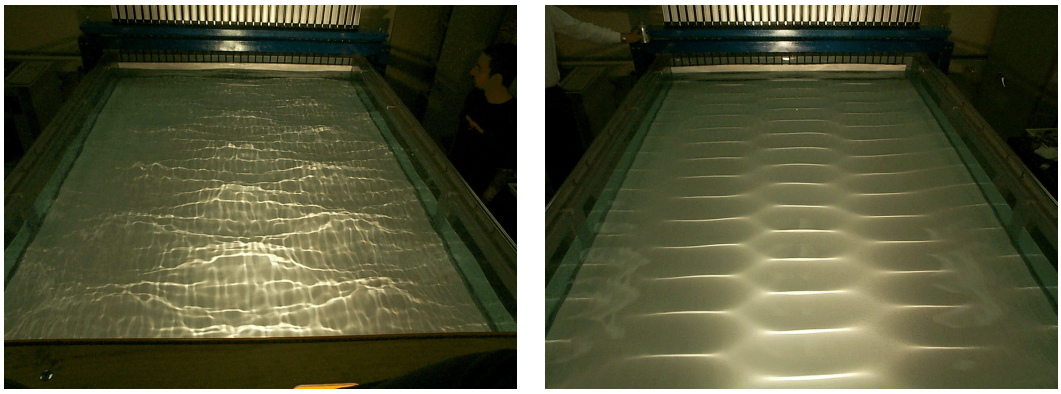
Figure 4: Experimentally stable wave patterns in deep water. Frequency=3Hz, wavelength=17.3cm.

inspection of the results presented in Figure 1 reveals that the period of the recurring pattern increases with distance in the tank (compare the first and last panel), a phenomenon called frequency downshifting. Downshifting has also been observed and studied in optics (see Figure 3, and [6], [7]). Interestingly, however, downshifting does not occur in simulations based on 1-D or 2-D NLS, neither in those based on Dysthe's generalization of NLS ([4]). It remains poorly understood.

In 1990s, Hammack built a new tank to study 2-D wave patterns on deep water. He found experimental evidence, in some situations, of apparently stable wave patterns in deep water (see Figure 4), despite the theoretical results described above [13]. How do we reconcile the experimental observations with Benjamin-Feir instability? There are a few possible explanations.

The first is that the modulational instability only appears in 1-D plane waves, but not in 2-D periodic patterns. Motivated by Hammack's experiments, many researchers began to investigate theoretically the existence of 2-D periodic surface patterns of permanent form on deep water. Craig & Nicholls [3] proved that such solutions are admitted in the full equations of inviscid water waves with gravity and surface tension. Recently, Iooss & Plotnikov [9] proved the existence of such patterns for pure gravity waves on deep water. But neither of these papers considers the stability of the solutions, and the question remains open as to whether they are stable or not.

Another possibility, explored in the next Section, has recently been proposed based on even more puzzling experimental results: Hammack found that conditions on the water surface may be essential for the instability. Figure 5(a) shows a 2D NLS experiment in a tank using recently poured water, which clearly exhibits strong instabilities. Meanwhile using the same volume of water after leaving it sit in the tank for a few days, in exactly



(a) (b)

Figure 5: Surface conditions may play an important role of determining the Benjamin-Feir instability. (a) $f = 2Hz$, new water; (b) $f = 3Hz$, old water.

the same experiment otherwise, reveals the presence of much more stable wave patterns, as shown in Figure 5(b). (For more pictures and movies, visit Pritchard Lab's website at www.math.psu.edu/dmh/FRG). Why should "old" water stabilize the wave pattern? A possible explanation is that impurities accumulating on the water surface provide an additional damping mechanism, to suppress the growth of the slowly-growing modulational instability.

4 Stabilization by damping

To study the possibility of stabilization by damping, let us consider the 1-D NLS with an added damping effect.

$$i(\partial_t A + c_g \partial_x A) + \epsilon(\alpha \partial_x^2 A + \gamma |A|^2 A + i\delta A) = 0, \tag{13}$$

where c_g is the group velocity, ϵ is a small parameter, and δA is the small damping term with $\delta \geq 0$. By introducing variables $\xi = t - \frac{x}{c_g}$ and $X = \epsilon \frac{x}{c_g}$, we obtain

$$i\partial_x A + \alpha \partial_\xi^2 A + \gamma |A|^2 A + i\delta A = 0. \tag{14}$$

Define $A(\xi, X) = e^{-\delta X} \mathcal{A}(\xi, X)$, then (14) becomes

$$i\partial_X \mathcal{A} + \alpha \partial_\xi^2 \mathcal{A} + \gamma e^{-2\delta X} |\mathcal{A}|^2 \mathcal{A} = 0. \tag{15}$$

Equation (15) is in fact a Hamiltonian equation, with a Hamiltonian

$$\hat{H}(X) = i \int \left(\alpha |\partial_\xi \mathcal{A}|^2 - \frac{\gamma}{2} e^{-2\delta X} |A|^4 \right) \mathrm{d}\xi. \tag{16}$$

It follows immediately that $\frac{d\hat{H}}{dX} \neq 0$.

There is a solution to (15) corresponding to a wave train that is uniform in ξ

$$\mathcal{A}_1 = \mathcal{A}_0 \exp\left\{ i\gamma |\mathcal{A}_0|^2 \left(\frac{1 - e^{-2\delta X}}{2\delta} \right) \right\} \tag{17}$$

If we perturb $\mathcal{A}$ around $\mathcal{A}_1$ by setting

$$\mathcal{A}(X,\xi) = \exp\left\{ i\gamma |\mathcal{A}_0|^2 \left(\frac{1 - e^{-2\delta X}}{2\delta} \right) \right\} [|\mathcal{A}_0| + \mu(u + iv)] + O(\mu^2), \tag{18}$$

and insert (4) into (15), then equating $O(\mu)$ terms to zero yields

$$\partial_X v = 2\gamma e^{-2\delta X} |A_0|^2 u + \alpha \partial_\xi^2 u, \tag{19}$$

$$\partial_X u = -\alpha \partial_\xi^2 v \qquad . \tag{20}$$

Without loss of generality, we seek solutions in the form

$$u(X,\xi) = \hat{u}(X) e^{im\xi} + \hat{u}^*(X) e^{-im\xi}, \quad \text{and} \quad v(X,\xi) = \hat{v}(X) e^{im\xi} + \hat{v}^*(X) e^{-im\xi}, \tag{21}$$

where $*$ stands for the complex conjugate. It follows from (19) and (20) that

$$\frac{d^2 \hat{u}}{dX^2} + \left[\alpha m^2 \left(\alpha m^2 - 2\gamma e^{-2\delta X} |\mathcal{A}_0|^2 \right) \right] \hat{u} = 0. \tag{22}$$

By Lyapunov's definition, a uniform wave train solution is said to be *linearly stable* if for every $\epsilon > 0$ there is a $\Delta(\epsilon) > 0$ such that if a perturbation (u, v) satisfies

$$\int [u^2(\xi, 0) + v^2(\xi, 0)] d\xi < \Delta(\epsilon) \quad \text{at } X = 0, \tag{23}$$

then necessarily

$$\int [u^2(\xi, X) + v^2(\xi, X)] d\xi < \epsilon \quad \text{for all } X > 0. \tag{24}$$

It follows that there is a universal bound B which the total growth of any Fourier mode can not exceed. To demonstrate the stability, one can choose $\Delta(\epsilon)$ so that

$$\Delta(\epsilon) < \frac{\epsilon}{B^2}. \tag{25}$$

Using Lyapunov's definition, we can see that there is a growing mode of equation (22) if

$$\alpha m^2 \left(\alpha m^2 - 2\gamma e^{-2\delta X} |\mathcal{A}_0|^2 \right) < 0. \tag{26}$$

It is not difficult to see that the growth stops eventually for any $\delta > 0$. Moreover, the total growth is bounded (see Figure 6.).

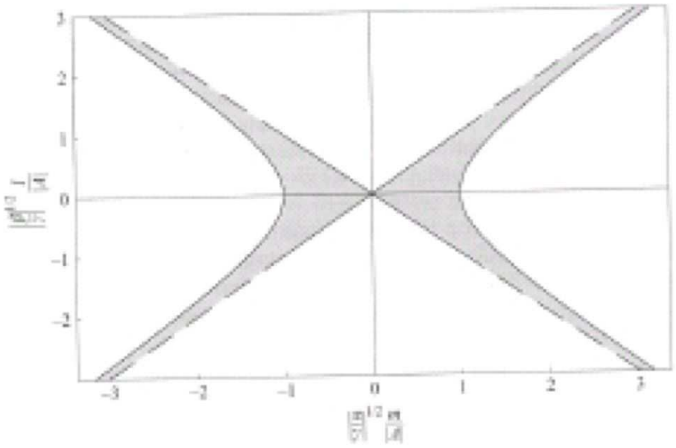


Figure 6: The shaded region shows the location of the growing modes in wavenumber space. The abscissa is $\left|\frac{\alpha}{\gamma}\right|^{\frac{1}{2}} \frac{m}{|A|}$, and the ordinate is $\left|\frac{\beta}{\gamma}\right|^{\frac{1}{2}} \frac{l}{|A|}$. Note that the experimental results presented are for the $\beta = 0$ case.

5 Experimental verification.

We compare the theory presented above with data from a series of experiments ([11]). Figure 7 shows measured water surface displacement in column 1 and modal amplitudes obtained from the corresponding Fourier transforms in column 2, measured at different distance from the wavemaker. At X_1, the Fourier spectrum shows that most of the power is in the carrier wave frequency (at 3.3Hz) and its first harmonic, as well as two small peaks very close to 3.3Hz. These provide a long-wave sinusoidal modulation of the carrier wave, as observed in the left-hand column. As the modulated wavetrain propagates downstream, we observe two changes: its overall amplitude decays, and the shape of the modulation changes. By inspection of the Fourier spectrum, we see that additional sidebands with frequencies near that of carrier wave 3.3Hz have grown between X_1 and X_8

Figure 8 compares the measured and predicted amplitudes of the set of initially seeded sidebands. Since the original Benjamin-Feir analysis did not include dissipation, direct comparison between their non-dissipative theory with the experiments shows very poor agreement (and is not shown in the Figure). A somewhat better agreement, as proposed by Benjamin (1968), can be obtained by calculating the growth rate of the side-bands and subtracting from this estimate their theoretical dissipation rate (which can be deduced from that of the carrier wave, since they nearly have the same frequency). As shown in the Figure 8, the amplitude predictions using this method agree with the experiments for short distances along the tank. However, for longer distances this simple estimate is no longer valid.

A much better agreement is obtained by using the full dissipative theory described above. This experiment thus clearly demonstrates that the stabilization of the side-bands results from the gradual decay of the amplitude of the *carrier* wave by dissipation, which

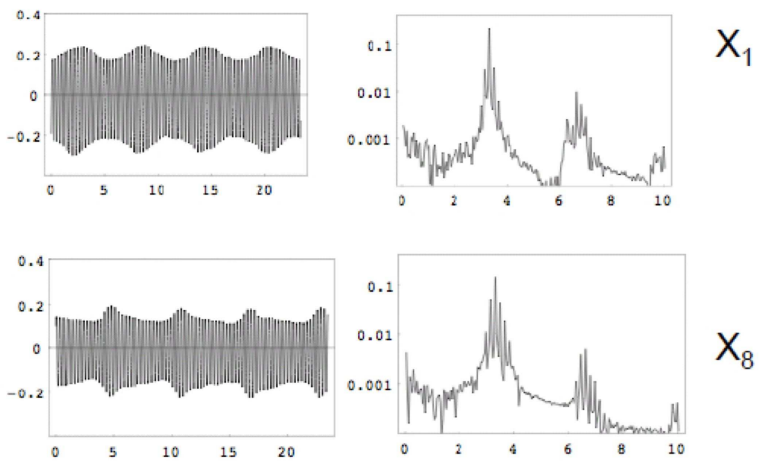


Figure 7: Experimental wave records. Column 1: water displacement (in cm) with time abscissa; column 2: corresponding Fourier coefficients (in cm) with frequency abscissa. Distance of the wave gauge from the wavemaker: 128 cm at X_1, and 478 cm at X_8. See also [11].

in turn stronly reduces the growth rate of the modulational instability.

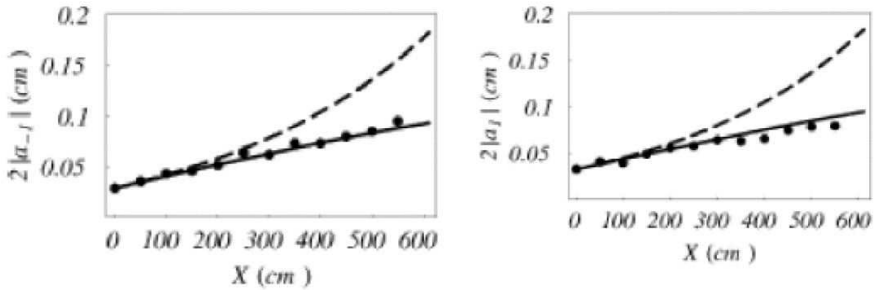


Figure 8: Prediction (solid curves) from the damped NLS theory (22) and measurements (dots) of the amplitudes of the two seeded sidebands $|a_{-1}|$ and $|a_1|$ as functions of distance from the wavemaker. $X = 0$ corresponds to 128 cm from the wavemaker. The dashed line corresponds to the Benjamin-Feir growth rate to which the side-band dissipation rate has been substracted.

The growth of the next two unseeded sidebands are shown in Figure 9. Since none of these were seeded, they started with smaller amplitudes than seeded the ones and remained smaller. Again, the damped NLS theory adequately predicts their evolution.

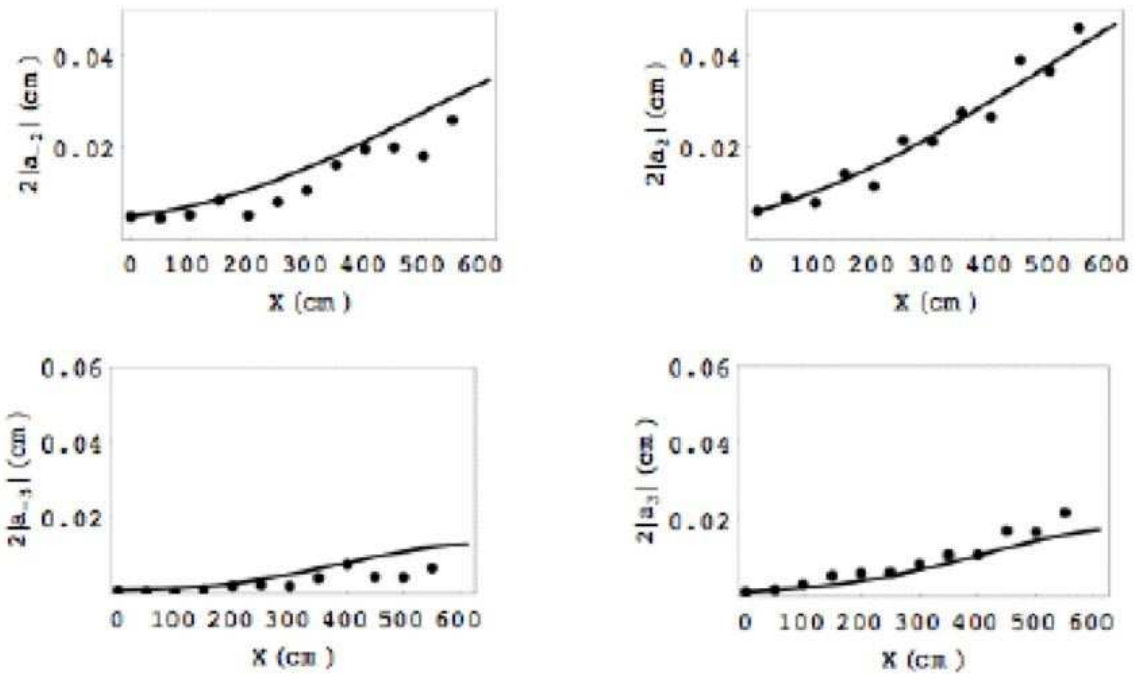


Figure 9: Prediction (solid curves) and measurements (dots) of the amplitudes of the two unseeded sidebands, (a) $|a_{-2}|$ and (b) $|a_2|$, and the amplitudes of the third set of sidebands, (c) $|a_{-3}$ and (d) $|a_3|$, as functions of distance from the wavemaker. $X = 0$ corresponds to 128 cm from the wavemaker, and the starting values of the amplitude were taken from data measured at the $n = 1$ location.

6 Summary.

The Benjamin-Feir instability implies that (1D) stable wave patterns propagating on deep water with nearly permanent form do no exist without presence of damping effects. But any amount of damping of the right kind can stabilize the instability. This dichotomy between with and without damping applies to both 1-D plane waves and to 2-D periodic surface patterns. However, this explanation is still somewhat controversial.

References

[1] M. Ablowitz & H. Segur, *Solitons and the inverse scattering transform*, Society for Industrial Mathematics, (2000)

[2] T. Benjamin & J. Feir, *The disintegration of wave trains on deep water Part 1. Theory*, J. Fluid Mech., **27**, 417, (1967)

[3] W. Craig & D. Nicholls, *Traveling two and three dimensional capillary gravity water waves*, SIAM journal on mathematical analysis, **32**, 2, 323-359, (2000)

[4] K. Dysthe, *Note on a Modification to the Nonlinear Schrodinger Equation for Application to Deep Water Waves*, Proc. R. Soc. Lond. A, **369**, 105-114, (1979)

[5] E. Fermi, J. Pasta & S. Ulam, *Studies of nonlinear problems*, 1940. In Collected Papers of Enrico Fermi, **2**, 978. University of Chicago Press, (1962)

[6] J. Gordon, *Theory of the soliton self-frequency shift*, Optics Letters, **11**, 10, 662-664, (1986)

[7] F. Mitschke & L. Mollenuar, *Discovery of the soliton self-frequency shift*, Optics Letters, **11**, 10, 659, (1986)

[8] A. Hasegawa & Y. Kodama, *Solitons in optical communications* , Clarendon press Oxford, (1995)

[9] G. Iooss & P. Plotnikov, *Small divisor problem in the theory of three-dimensional water gravity waves*, Memoirs of AMS. **200**, No 940, (2009)

[10] B. M. Lake, H. C. Yuen, H. Rungaldier & W. E. Ferguson, *Nonlinear deep-water waves: theory and experiment. Part 2. Evolution of a continuous wave train*, J. Fluid Mech., **83**, 1, 49-74, (1977)

[11] H. Segur, D. Henderson, J. Carter, J. Hammack, C. Li, D. Pheiff & K. Socha, *Stabilizing the Benjamin-Fair instability*, J. Fluid Mech., **539**, 229-271, (2005)

[12] V. Zakharov & A. Rubenchik, *Instability of waveguides and solitons in nonlinear media*, Sov. Phys. JETP, **38**, 494-500, (1974)

[13] V. Zakharov & V. Synakh, *The nature of self-focusing singularity*, Sov Phys JETP, **41**, 441-448, (1976)

Lecture 16: Solitary waves

Lecturer: Roger Grimshaw. Write-up: Daisuke Takagi, Yiping Ma and Andrew Stewart

June 24, 2009

1 Introduction

A solitary wave is a wave which propagates without any temporal evolution in shape or size when viewed in the reference frame moving with the group velocity of the wave. The envelope of the wave has one global peak and decays far away from the peak. Solitary waves arise in many contexts, including the elevation of the surface of water and the intensity of light in optical fibers. A soliton is a nonlinear solitary wave with the additional property that the wave retains its permanent structure, even after interacting with another soliton. For example, two solitons propagating in opposite directions effectively pass through each other without breaking.

Solitons form a special class of solutions of model equations, including the Korteweg de-Vries (KdV) and the Nonlinear Schrödinger (NLS) equations. These model equations are approximations, which hold under a restrictive set of conditions. The soliton solutions obtained from the model equations provide important insight into the dynamics of solitary waves. However, they are limited by the conditions under which the model equations hold. An alternative approach, which deals directly with the exact equations from which the model equations are derived, provides insight into a larger class of solitary waves than those obtained from the model equations.

In this lecture, we show that important properties of solitary waves can be determined directly from the exact equations governing a physical system by an asymptotic perturbation procedure. Information about the possible existence of certain types of solitary waves is obtained using a phase-plane formalism, a common technique in dynamical systems. In this framework, a solitary wave corresponds to a homoclinic orbit in a spatial dynamical system. Solitary waves exist in a number of different cases to be determined, given that they must decay far away from their global peak. A close examination of the tail regions, far away from the global peak, indicates that there are four possible cases of solitary waves. Three of the four possible cases are presented in detail. In the first case, the steady-state solution of the KdV equation is obtained. In the second case, a generalized solitary wave is obtained, which decays to non-zero oscillations of constant amplitude and wavenumber. In the third case, an envelope solitary wave is obtained, which satisfies the NLS equation.

2 Formulation of solitary waves as a dynamical system

Consider a solitary wave propagating with speed c in the positive direction of the x axis. The wave amplitude in the moving reference frame is the function $A(\xi)$ where $\xi \equiv x - ct$.

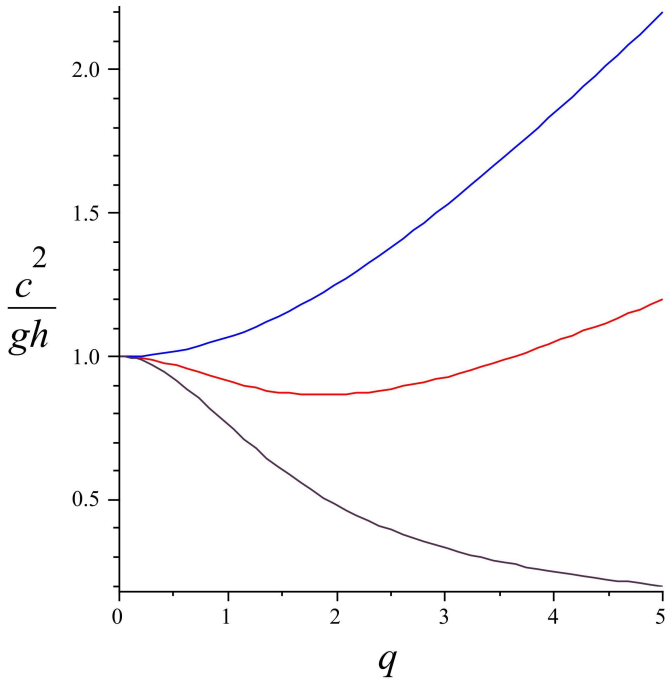


Figure 1: Plot of the dispersion relation for water waves given by (1) with Bond numbers $B = 0$ (violet), $B = 0.2$ (red) and $B = 0.4$ (blue).

Without loss of generality we can select ξ such that the peak of A is at $\xi = 0$, while A decays in the tail regions, in the limits as $\xi \to \pm\infty$. In the tail regions, where a linearized analysis is assumed to hold, it is fruitful to seek solutions proportional to the real part of $\exp(ik\xi)$, where k is a complex-valued wavenumber to be determined. The wavenumber must not be purely real, given that the wave amplitude must decay in the tail regions. This leads to the notion that solitary waves with a real phase speed c only exist within a limited range of wavenumbers in the spectrum, called gaps.

A relationship between the phase speed c and the wavenumber k is obtained by linearizing the governing equations of a physical system of interest, which yields a linearized dispersion relation. For example, the linearized dispersion relation of surface water waves is given by

$$c^2/gh = f(q) = [(1 + Bq^2)/q] \tanh q, \tag{1}$$

where g is the acceleration due to gravity, h the constant depth of the water when it is unperturbed and $q = kh$ the dimensionless wavenumber. The dimensionless number $B = \sigma/\rho g h^2$ is the Bond number, which measures the relative importance of surface tension and gravity, where σ is the coefficient of surface tension and ρ is the density of water. A plot of the function $f(q)$ given by equation (1) for three representative values of the Bond number B is shown in Figure 1.

Solitary waves can exist provided that no real value of k satisfies the dispersion relation, meaning that k has a non-zero imaginary part. In the case of $B = 0$ (no surface tension), solitary waves can exist only when $c^2 > gh$. When effects due to surface tension are

significant, such that $B > 1/3$, q is an increasing function of c with $q \geq 0$ for $c^2/gh \geq 1$. An example with $B = 0.4$ is shown in figure 1 (blue curve). Thus, solitary waves with $B > 1/3$ can only exist when $c^2/gh < 1$. Finally, when $0 < B < 1/3$, a real-valued q satisfies (1) for any speed $|c| > c_m$ for some positive c_m. Figure 1 shows a plot of (1) with $B = 0.2$, which attains a local minimum c_m^2/gh at some non-zero wavenumber k_m. This indicates that solitary waves with $0 < B < 1/3$ can exist only when $c^2 < c_m^2$.

Solitary waves are found by reformulating the problem in the framework of a dynamical system. Let's assume the system is conservative, which is the common and traditional scenario for solitary waves. The underlying physical system is Hamiltonian and reversible, meaning that the system conserves energy and is symmetric under the transformation $\xi \to -\xi$. This implies that if k is a solution then $-k$ is also a solution due to reversibility. If k has an imaginary part (as is required for a solitary wave), then k^* is a solution because the dispersion relation for real-valued phase speed c has real coefficients. The solutions generically form a quartet $(k, k^*, -k, -k^*)$ with an associated four-dimensional subspace for the corresponding wave mode. In the general case when the imaginary part of k is non-zero, there are typically two roots in the limit as $\xi \to \infty$ and two other roots as $\xi \to -\infty$. In the limit as $\xi \to \infty$, the imaginary part of k must be positive provided that the solitary wave decays in the tail regions. Likewise, the imaginary part of k must be negative as $\xi \to -\infty$.

Let us now study the amplitude $A(\xi)$ in the framework of a dynamical system. Consider the phase plane where the ξ-derivative of the wave amplitude is plotted against the amplitude. Trajectories in this phase planes are the curves $(A(\xi), A_\xi(\xi))$. This can either represent the whole system (for a 2D problem) or a projection of the whole system (for a higher-dimensional problem). Given the form of A at infinity discussed above, we deduce that the trajectory is a homoclinic orbit from the origin. Indeed, the origin corresponds to the tail regions as $\xi \to \pm\infty$, where both A and A_ξ tend to 0. Meanwhile, the peak of the wave ($\xi = 0$ point) lies somewhere on the $A_\xi = 0$ axis. If k is pure imaginary then the trajectories satisfy $A_\xi = -|k|A$ near the origin, and approach it directly. If k has a real part, then the trajectories spiral into the origin instead.

Consider now how the trajectory in the phase plane may change as parameters governing the original system are varied. Changes in the trajectory far from the origin of the phase plane mean that the overall shape of the wave $A(\xi)$ is qualitatively changed, but that its asymptotic behavior at $|\xi| \to \infty$ remains the same. However, much more dramatic changes in $A(\xi)$ occur if the behavior near the phase plane origin changes (for example when k goes from being real to complex to pure imaginary, or vice-versa).

This leads us to consider how the quartet structure $(k, k^*, -k, -k^*)$ evolves as some global system parameter is varied. Clearly bifurcations arise when two solutions for k coalesce, for which the necessary condition is that $\partial c/\partial k = 0$. This condition is equivalent to the condition that the phase speed c is equal to the group velocity c_g (i.e. the weak dispersion limit), when the bifurcation occurs at a real value of k. The equivalence follows immediately from the equation $c_g = c + k\partial c/\partial k$. For example, from Figure 1 we see that bifurcations arise in water waves when (k, c^2) takes the values $(0, gh)$ or (k_m, c_m^2).

Generically, there are two possible quartet structures at each of the bifurcation points:

- at $(k = 0, c^2 = gh)$: we either have the quartet $(0, 0, i\gamma, -i\gamma)$ or $(0, 0, \beta, -\beta)$, (two roots have coalesced at the origin, the other two are either real or pure imaginary)

- at $(k = k_m, c^2 = c_m^2)$: we either have $(\beta, \beta, -\beta, -\beta)$ (cc pairs coalesce on the real axis) or $(i\gamma, i\gamma, -i\gamma, -i\gamma)$ (pairs coalesce on the imaginary axis), where β and γ are real-valued.

The first three of the four possible cases will be examined in detail later. The fourth case, for the quartet $(i\gamma, i\gamma, -i\gamma, -i\gamma)$, has only rarely been studied and will not be considered here.

Consider a projection of the full system onto the appropriate four-dimensional subspace. The resulting bifurcation is analyzed within the framework of this subspace. A 4-vector $\mathbf{W}(\xi)$, representing the structure of the subspace, satisfies a first-order differential equation of the form

$$d\mathbf{W}/d\xi = L(\mathbf{W}; \epsilon) + N(\mathbf{W}), \tag{2}$$

where $L(\mathbf{W}; \epsilon)$ is a linear operator and $N(\mathbf{W})$ contains all nonlinear terms. The parameter ϵ represents the distance from the bifurcation point. Near the bifurcation (which takes place at $\epsilon = 0$) the linear operator L can be written as $L(\mathbf{W}; \epsilon) = L_0(\mathbf{W}) + \epsilon L_1(\mathbf{W}) + ...$, where the eigenvalues $\lambda = ik$ of the operator L_0 reproduce one of the four possible quartet structures described above. In each of the first three cases, the structural form of the small-amplitude solutions $\mathbf{W}$ is examined to identify and describe the nature of corresponding solitary waves.

3 The three cases

The three types of bifurcations may now be studied with the standard techniques of center manifold reduction and normal form analysis in bifurcation theory. For more details about these techniques, we refer the reader to the review article by Crawford [3].

3.1 Case (1)

Let us first consider case (1). At the bifurcation point ($\epsilon = 0$) the linearized system (2) has eigenvalues $(0, 0, \mp\gamma)$. Since the $\lambda = 0$ eigenvalue is degenerate, there will be a corresponding single eigenvector $\mathbf{V}_0$, and a single generalized eigenvector $\mathbf{V}_1$ such that $L_0\mathbf{V}_1 = \lambda\mathbf{V_1} + \mathbf{V}_0 = \mathbf{V}_0$. Small-amplitude solutions are then sought in the form

$$\mathbf{W} = A(\xi)\mathbf{V}_0 + B(\xi)\mathbf{V}_1 + \mathbf{W}^{(2)}\,. \tag{3}$$

Here A, B are real variables of $O(\alpha), \alpha \ll 1$, where α measures the wave amplitude. The leading terms form a two-dimensional subspace (A, B), while $\mathbf{W}^{(2)}$ is a small error term of $O(\alpha^2, \alpha\epsilon)$, where $\epsilon, \alpha \ll 1$ are both small parameters. Note that the two remaining eigenvalues $\mp\gamma$ play no role at the leading order here, since they correspond to strong exponential decay at infinity, and their effects are included in the small error term $\mathbf{W}^{(2)}$.

Substituting (3) into (2) yields

$$\begin{aligned} A_\xi\mathbf{V}_0 + B_\xi\mathbf{V}_1 + \frac{\mathrm{d}}{\mathrm{d}\xi}\mathbf{W}^{(2)} = \quad & L_0\left(A(\xi)\mathbf{V}_0 + B(\xi)\mathbf{V}_1 + \mathbf{W}^{(2)}\right) \\ & +\epsilon L_1\left(A(\xi)\mathbf{V}_0 + B(\xi)\mathbf{V}_1 + \mathbf{W}^{(2)}\right) \\ & +N\left(A(\xi)\mathbf{V}_0 + B(\xi)\mathbf{V}_1 + \mathbf{W}^{(2)}\right) \end{aligned} \tag{4}$$

Projection onto the two-dimensional subspace $(\mathbf{V}_0, \mathbf{V}_1)$ yields, to lowest order in ϵ and α,

$$A_\xi = B + O(\epsilon, \alpha^2, ...)$$
$$B_\xi = \mathbf{V}_1 \cdot \left[\epsilon A L_1(\mathbf{V_0}) + \epsilon B L_1(\mathbf{V_1}) + L_0(\mathbf{W}^{(2)}) + N(A\mathbf{V}_0 + B\mathbf{V_1})\right] + O(\alpha\epsilon^2, \epsilon\alpha^2, \epsilon^2, \alpha^3, ..)$$

The first two terms in the B equation are linear terms of $O(\epsilon\alpha)$, and their projection onto $\mathbf{V}_1$ yields $\epsilon(c_1 A + c_2 B)$, where c_1 and c_2 are the projection coefficients. Similarly, the linear and nonlinear terms of $O(\alpha^2)$ yield contributions of the kind A^2, AB or B^2, so that

$$A_\xi = B$$
$$B_\xi = \epsilon(c_1 A + c_2 B) + (d_1 A^2 + d_2 AB + d_3 B^2) + ...$$

where the omitted terms are $O(\alpha\epsilon^2, \alpha^2\epsilon, \alpha^3)$. Finally, a normal form analysis reveals the normal form of the system near the bifurcation :

$$A_\xi = B\,,$$
$$B_\xi = \epsilon A + \mu A^2 + ... \tag{5}$$

where μ is a real-valued coefficient, specific to the system being considered.

The eigenvalues $\lambda = ik$ of this system are $\pm\epsilon^{1/2}$ if $\epsilon > 0$, and $\pm i|\epsilon|^{1/2}$ if $\epsilon < 0$. As discussed earlier, only the former case, $\epsilon > 0$, yields the solitary wave solution, with $k = \pm i\epsilon^{1/2}$. Finally, comparison with the dispersion relation expressed near $k = 0$

$$c(k) = c(0) + \frac{k^2}{2}\left.\frac{\mathrm{d}^2 c}{\mathrm{d}k^2}\right|_{k=0} + ... \tag{6}$$

leads to the identification of ϵ as

$$\epsilon = -\frac{2(c - c(0))}{c_{kk}(0)}\,. \tag{7}$$

Hence, for solitary waves to exist ($\epsilon > 0$), we either need $c > c(0)$ if $c_{kk}(0) < 0$, or $c < c(0)$ if $c_{kk}(0) > 0$. These inequalities recover the graphical argument discussed in the previous section.

When the error terms in (5) are omitted, we can eliminate B, and the resulting ODE for A can be recognized as the steady-state KdV equation, which has the well-known "sech^2" solution. It is then a delicate and intricate task to establish that this solitary wave solution persists when the error terms are restored.

This dynamical systems approach to the problem has therefore established the steady-state KdV equation as the normal form for weakly nonlinear, weakly dispersive solitary waves (see also Lecture 6) whenever the dispersion relation satisfies $dc/dk \to 0$ when $k \to 0$, and has provided mathematical conditions on their propagation speed. The KdV is thus seen to be strictly correct for any solitary wave of the kind described above, far from the wave peak (i.e. in the low-amplitude tails). Whether this statement still holds nearer the peak $\xi = 0$ depends on the peak amplitude $A(0)$.

3.2 Case (2)

Next consider case (2), where the linearized system (2) at the bifurcation point ($\epsilon = 0$) has eigenvalues $(0, 0, \pm i\beta)$. Again, the degeneracy of $\lambda = 0$ implies that there is a single eigenvector $\mathbf{V}_0$, and a single generalized eigenvector $\mathbf{V}_1$. However, account must now be taken of the other two eigenvalues $\pm i\beta$, with their associated eigenvectors $\mathbf{V}_2, \mathbf{V}_2^*$, since they do not now lead to decaying solutions at infinity. Small-amplitude solutions are sought in the form

$$\mathbf{W} = A(\xi)\mathbf{V}_0 + B(\xi)\mathbf{V}_1 + C(\xi)\mathbf{V}_2 + C^*(\xi)\mathbf{V}_2^* + \mathbf{W}^{(2)}\,. \tag{8}$$

Here C is a complex-valued variable, and hence the leading terms form a four-dimensional subspace (A, B, C), while $\mathbf{W}^{(2)}$ is a small error term. Projection onto this four-dimensional subspace, followed by normal form analysis now reveals that (A, B, C) satisfy the system

$$\begin{aligned} A_\xi &= B\,, \\ B_\xi &= \epsilon A + \mu A^2 + \nu|C|^2 + \cdots\,, \\ C_\xi &= i\gamma(1 + \delta A)C + \cdots\,. \end{aligned} \tag{9}$$

Here μ, ν, δ are real-valued coefficients specific to the system being considered, and the omitted terms are small error terms as above.

When the error terms are omitted the system is integrable. In that limit, it is easy to verify that $|C|_\xi^2 = 0$ by constructing the quantity $C^*C_\xi + C_\xi^*C$. Hence, $|C| = C_0$ is a constant. By a change of origin from $A \to A + A_0$ with $\epsilon A_0 + \mu A_0^2 + \nu C_0^2 = 0$, the system reduces to the same form as (5) in case (1). Thus, for the case $\epsilon > 0$ (when case (1) is a KdV-type solitary wave), the solution is a one-parameter family of homoclinic-to-periodic solutions, with $|C| = C_0$ constant and $(A, B) \to (A_0, 0)$ as $\xi \to \pm\infty$. The solution is a generalized solitary wave which typically has a "sech^2" core, and decays at infinity to non-zero oscillations of constant amplitude C_0 and wavenumber γ (Figure 2). A delicate analysis of the full system (2) with the small error terms shows that at least two of these solutions persist; the minimal amplitude C_0 being exponentially small, that is $O(\exp(-K/|\epsilon|^{1/2}))$ where K is a positive real constant. Although such waves are permissible as solutions of the steady-state equations, they have infinite energy and their associated group velocity is inevitably inward at one end and outward at the other end. Hence, they cannot be realized in a physical system from any localized initial condition because energy is finite. Instead localized initial conditions will typically generate a one-sided generalized solitary wave, whose central core is accompanied by small-amplitude outgoing waves on one side only. Such waves cannot be steady, and instead will slowly decay with time as energy is radiated away by outgoing waves from the core.

3.3 Case (3)

Finally we consider case (3), when there is a double eigenvalue $\lambda = i\beta$ with generically a corresponding single eigenvector $\mathbf{V}_0$, and a single generalized eigenvector $\mathbf{V}_1$, while the complex conjugate double eigenvalue $\lambda = -i\beta$ has corresponding complex conjugate eigenvectors. Small-amplitude solutions are now sought in the form

$$\mathbf{W} = A(\xi)\mathbf{V}_0 + B(\xi)\mathbf{V}_1 + A^*(\xi)\mathbf{V}_0^* + B^*(\xi)\mathbf{V}_1^* + \mathbf{W}^{(2)}\,. \tag{10}$$

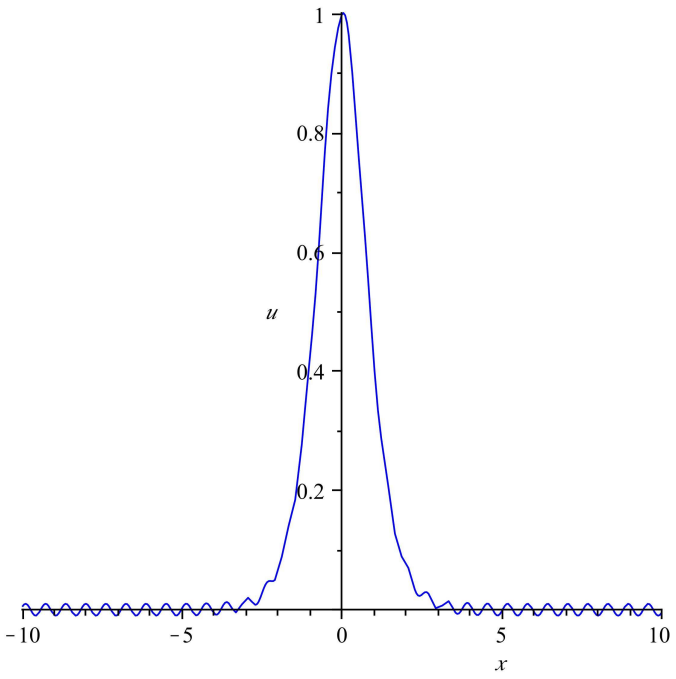


Figure 2: Generalized solitary wave.

Here A and B are complex-valued variables, forming a four-dimensional subspace while $\mathbf{W}^{(2)}$ is again a small error term. Projection onto this subspace and a normal form analysis reveal that

$$\begin{aligned} A_\xi &= i\beta A + B + iAP(\epsilon, |A|^2, K) + \cdots , \\ B_\xi &= i\beta B + iBP(\epsilon, |A|^2, K) + AQ(\epsilon, |A|^2, K) + \cdots , \end{aligned} \tag{11}$$

where $K = i(AB^* - A^*B)$ (note that K is real). Here P and Q are real-valued polynomials of degree 1, which take the form

$$\begin{aligned} P(\epsilon, |A|^2, K) &= \epsilon + \nu_1|A|^2 + \nu_2 K , \\ Q(\epsilon, |A|^2, K) &= 2\epsilon\beta + \mu_1|A|^2 + \mu_2 K , \end{aligned} \tag{12}$$

where all coefficients are real-valued. When the error terms in (11) are omitted, the resulting system is integrable. There are two constants of motion. The first one is K, which is verified by a direct construction of K_ξ. The second one is H, where

$$H = |B|^2 - (2\epsilon\beta|A|^2 + \frac{\mu_1}{2}|A|^4 + \mu_2 K|A|^2) . \tag{13}$$

To prove that H is constant, note that $(|A|^2)_\xi = A^*B + B^*A$ and that $(|B|^2)_\xi = Q(|A|^2)_\xi$. Hence, $(|B|^2)_\xi - Q(|A|^2)_\xi = 0$, so that the integral based on this quantity is constant. This integral is H.

For a solitary wave solution, conditions at infinity require $K = H = 0$. It then follows that

$$|B|^2 = 2\epsilon\beta|A|^2 + \frac{\mu_1}{2}|A|^4$$

$$(|A|_\xi)^2 = \left(\frac{(|A|^2)_\xi}{2|A|}\right)^2 = \left(\frac{A^*B}{|A|}\right)^2 = |B|^2 = 2\epsilon\beta|A|^2 + \frac{\mu_1}{2}|A|^4 \tag{14}$$

using the fact that $(|A|^2)_\xi = A^*B + B^*A$ and that $A^*B = B^*A$ is real. Thus solitary wave solutions exist provided that $\epsilon > 0$, and that the nonlinear coefficient $\mu_1 < 0$[1]. The condition $\epsilon > 0$ implies that the perturbed eigenvalues, $\lambda \approx i\beta \pm (2\epsilon\beta)^{1/2}$ have split off the imaginary axis, and so provide the conditions needed for exponential decay at infinity. The normal form coefficient μ_1 must be computed from the physical parameters through a reduction procedure [3], so the condition $\mu_1 < 0$ is problem-specific.

The solution of the truncated system governed by (14) is

$$A = a\exp\left(i[\beta+\epsilon]\xi\right)\mathrm{sech}(\gamma\xi)\,, \tag{15}$$

where $\gamma = (2\epsilon\beta)^{1/2}$, $|a|^2 = -4\epsilon\beta/\mu_1$. This solution describes an envelope solitary wave, with a carrier wavenumber $\beta + \epsilon$ and an envelope described by the "sech"-function. As we saw in Lecture 14, these solitary waves can also be obtained from the soliton solutions of the NLS equation, in the special case when the phase velocity equals the group velocity, $c = c_g$, or more precisely when $c + \Omega/K = c_g + V$, where V is the soliton speed, Ω the frequency and K the wavenumber correction. Note that the solution (15) contains an arbitrary phase in the complex amplitude a, meaning that the location of the crests of the carrier wave *vis-a-vis* the maximum of the envelope (here located at $\xi = 0$) is arbitrary. However, restoration of the error terms in (11) leads to the result that only two of these solutions persist, namely, those for which a carrier wave crest or trough is placed exactly at $\xi = 0$, so that the resulting solitary wave is either one of elevation or depression. This result requires very delicate analysis, but could be anticipated by noting that these are the only two solutions which persist under the symmetry transformation $\xi \to -\xi$.

4 Applications to water waves

The linearized dispersion relation holds the key to finding solitary waves. For water waves, for which the dispersion relation is (1), these two cases (1) and (2) imply that pure solitary waves of elevation exist for $B = 0$, and of depression for $B > 1/3$, while generalized solitary waves arise whenever $0 < B < 1/3$. For the case of generalized solitary waves, there is always the possibility that the amplitude of the oscillations is zero, and the solution then reduces to a pure solitary wave, called an "embedded" solitary wave. There are now many examples of such embedded solitary waves arising in various physical systems, notably for internal waves. This "dynamical-systems" approach to finding solitary waves has also been applied to interfacial waves, where again the linear dispersion relation holds the key to

[1]This may be understood by regarding $|A|$ as space, ξ as time, and $|A|_\xi$ as velocity in (14). Then (14) describes a velocity field on the positive real line. In order for an orbit starting from and returning to the origin as $\xi \to \mp\infty$ to exist, the velocity field must have a stagnation point, which occurs only when $\mu_1 < 0$.

determining where solitary waves can be found. However, various numerical and analytical studies suggest that embedded solitary waves do not arise in the context of water waves. Instead, case (3) implies that envelope solitary waves arise for capillary-gravity waves with $0 < B < 1/3$, where it can be shown that the coefficient μ_1 in (12) is negative as required.

Finally, we remark that the method of treating solitary waves as homoclinic orbits in a spatial dynamical system has many applications (cf. [2]) in general evolution equations, most of which are not integrable. This point of view, coupled with numerical continuation techniques, turns out to be fruitful in studying spatially localized states in both one (cf. [1]) and two (cf. [7]) dimensional pattern forming systems.

References

[1] Burke, J. and Knobloch, E., *Localized states in the generalized Swift-Hohenberg equation*, Physical Review E, 73 (2006), 056211.

[2] Champneys, A. R., *Homoclinic orbits in reversible systems and their applications in mechanics, fluids and optics*, Physica D, 112, (1998) pp. 158-186.

[3] Crawford, J. D., *Introduction to bifurcation theory*, Review of Modern Physics, 63, (1991) pp. 991-1037.

[4] Dias, F. and Iooss, G. , *Water waves as a spatial dynamical system*, *Handbook of Mathematical Fluid Dynamics*, ed. S. Friedlander and D. Serre, Elsevier, North Holland, Chapter 10, (2003), pp. 443-499.

[5] Grimshaw, R. and Iooss, G, *Solitary waves of a coupled Korteweg-de Vries system*, Mathematics and Computers in Simulation, 62, (2003) pp. 31-40.

[6] Solitary Waves in Fluids, ed. Grimshaw R. 2007, *Advances in Fluid Mechanics, Vol 47, WIT press, UK*, Chapter 1, 1-17 and Chapter 7, pp. 159-179.

[7] ,Lloyd, D. J. B., Sandstede B., Avitabile D. and Champneys A. R., *Localized hexagon patterns of the planar Swift-Hohenberg equation*, Siam J. Applied Dynamical Systems, 7, (2008), pp. 1049-1100.

Lecture17: Generalized Solitary Waves

Lecturer: Roger Grimshaw. Write-up: Andrew Stewart and Yiping Ma

June 24, 2009

We have seen that solitary waves, either with a "pulse"-like profile or as the envelope of a wave packet, play a key role in nonlinear wave dynamics. However, there are physical situations when such KdV-type waves may not be genuinely localized. Instead they are accompanied by co-propagating small oscillations which spread out to infinity without decay (see Figure 1). These are *generalized solitary waves*. As we saw in Lecture 16, they may occur for water waves with surface tension for Bond numbers less than 1/3. It can be shown that they can also occur for interfacial waves when there is a free surface, and for all internal waves with mode numbers $n \geq 2$. The underlying reason for their existence is the presence of a resonance between a long wave with wave number $k \approx 0$ and a short wave with a finite wave number. When the amplitude of the central core is small compared to its length, $O(\epsilon^2)$, the amplitude of the oscillations is exponentially small, typically $O(\exp(-C/\epsilon))$ where C is a positive constant. Hence generalized solitary waves cannot usually be found by conventional asymptotic expansions, and need *exponential asymptotics*.

Consider the dispersion relation for internal waves, shown in Figure 2. Normally any mode numbers higher than 1 will resonate with other modes, so typically these waves do not persist, and we see only the mode 1 (soliton) waves. However, sometimes the first mode resonates with the surface mode, in which case waves of mode 2 or higher become generalized solitary waves, and only the mode 1 wave is a pure solitary wave.

Steady generalized solitary waves are necessarily symmetric. However, this means they cannot be realized physically as then the group velocity of the small oscillations is the same in both tails, which implies that energy sources and sinks are needed at infinity. In practice, these waves are generated asymetrically, with a core and small oscillations only on one side, determined by the group velocity (see Figure 3). Consequently, they are unsteady and slowly decay due to this radiation. In Figure 4 we present an acoustic visualisation of the streamlines generated by stratified flow past a sill, as reported in the experiments of Farmer & Smith [3]. This phenomenon was subsequently explained in terms of generalized solitary waves by Akylas & Grimshaw [1].

1 The Coupled KdV Equations

The technique we use to find the tail oscillations is based on extending the usual asymptotic expansion into the complex plane, and using Borel summation. It is similar to the techniques used by [6] and [7].

We begin by using a model system of two coupled KdV equations, which can be shown to describe the interaction between two weakly nonlinear long internal waves whose linear

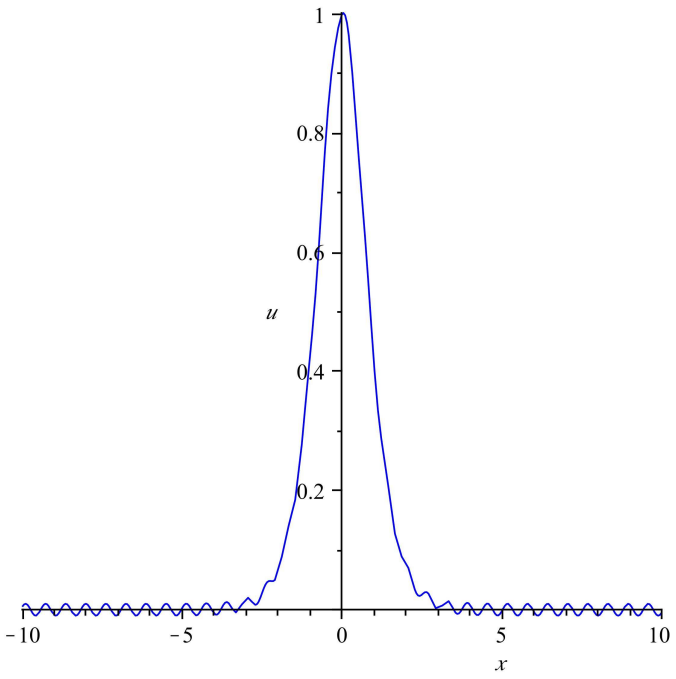


Figure 1: Schematic plot of a generalized solitary wave profile.

long wave speeds are nearly equal. The two coupled equations are

$$u_t + 6uu_x + u_{xxx} + \left(pv_{xx} + quv + \tfrac{1}{2}rv^2\right)_x = 0\,, \tag{1a}$$

$$v_t + \Delta v_x + 6vv_x + v_{xxx} + \lambda\left(pu_{xx} + ruv + \tfrac{1}{2}qu^2\right)_x = 0\,, \tag{1b}$$

where λ is the coupling parameter and Δ is the detuning parameter, proportional to the difference between the two linear long wave speeds, and p, q and r are real-valued constants. For stability we choose $\lambda > 0$, and we may also take $\Delta > 0$ without loss of generality. This system is Hamiltonian, and possesses conservation laws for the "mass" variables u and v, the "momentum" $\lambda u^2 + v^2$, and the Hamiltonian.

Let us first examine the linear spectrum for waves of wave number k and phase speed c for this system. Linearization of (1a) and (1b), followed by a search for solutions of the kind $e^{ik(x-ct)}$ yields

$$c = \tfrac{1}{2}\Delta - k^2 \pm \sqrt{\lambda p^2 k^4 + \tfrac{1}{4}\Delta^2}. \tag{2}$$

If we let the coupling parameter $\lambda \to 0$ these linear modes uncouple into a u-mode with spectrum $c = -k^2$ and a v-mode with spectrum $c = \Delta - k^2$. This situation persists for $\lambda > 0$, and there is a resonance between the long wave (u-mode) and a short wave (v-mode), with a resonant wavenumber $k_0 = \sqrt{\Delta/(1-\lambda p^2)}$ provided that $\lambda p^2 < 1$. A typical plot of these modes is presented in Figure 5.

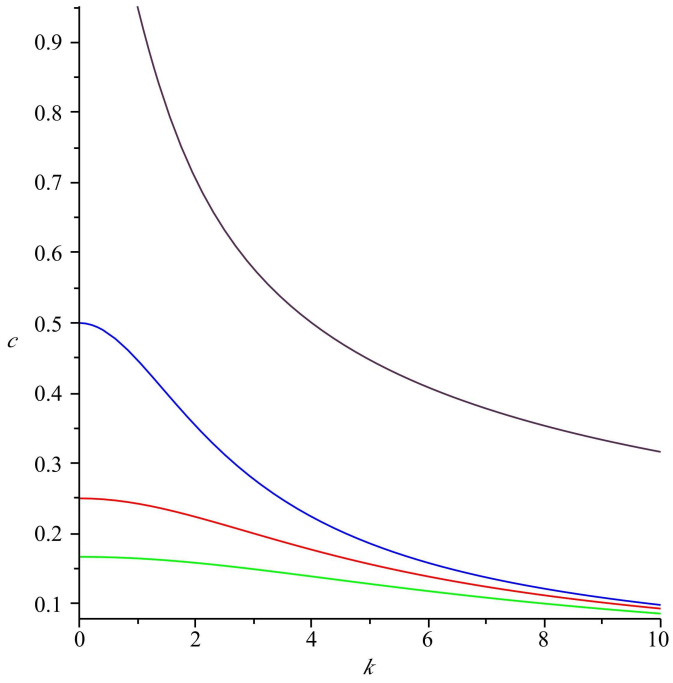


Figure 2: Plot of a schematic set of dispersion curves for internal waves: mode 1 (blue), mode 2 (red), mode 3 (green) and the surface mode (violet).

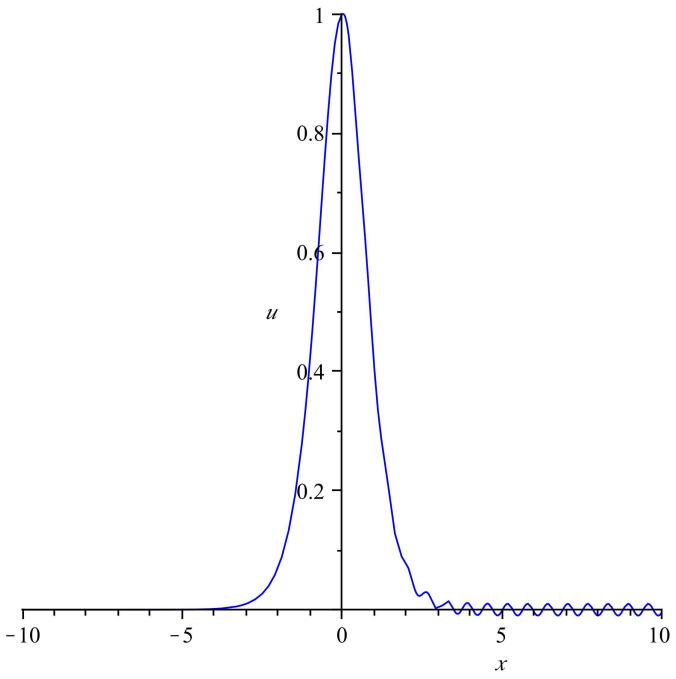


Figure 3: Schematic plot of an asymmetric generalized solitary wave profile.

We now seek nonlinear travelling wave solutions of the form

$$u = u(x - ct)\,, \quad v = v(x - ct)\,, \tag{3}$$

so that the coupled KdV system (1a, 1b) can be integrated once to become

$$-cu + 3u^2 + u_{xx} + pv_{xx} + quv + \tfrac{1}{2}rv^2 = 0\,, \tag{4a}$$

$$-cv + \Delta v + 3v^2 + v_{xx} + \lambda\left(pu_{xx} + ruv + \tfrac{1}{2}qu^2\right) = 0\,. \tag{4b}$$

Here the two constants of integration have been set to zero, which is achieved either by imposing solitary wave boundary conditions ($u, v \to 0$ as $|x| \to \infty$) or by translating u and v by constants. Equations (4a, 4b) form a fourth order ODE system. We shall show that they have symmetric generalized solitary wave solutions with co-propagating oscillatory tails of small amplitude. This amplitude will be found using either exponential asymptotics, or more directly by expanding in λ.

2 Exponential Asymptotics

A typical approach to these equations is to expand around $k = 0$ for the long (u-mode) wave. We introduce a small parameter $\epsilon \ll 1$, and seek an asymptotic expansion of the following form,

$$u_s(\epsilon x) = \sum_{n=1}^{\infty} \epsilon^{2n} u_n\,, \quad v_s(\epsilon x) = \sum_{n=1}^{\infty} \epsilon^{2n} v_n\,, \quad c = \sum_{n=1}^{\infty} \epsilon^{2n} c_n\,. \tag{5}$$

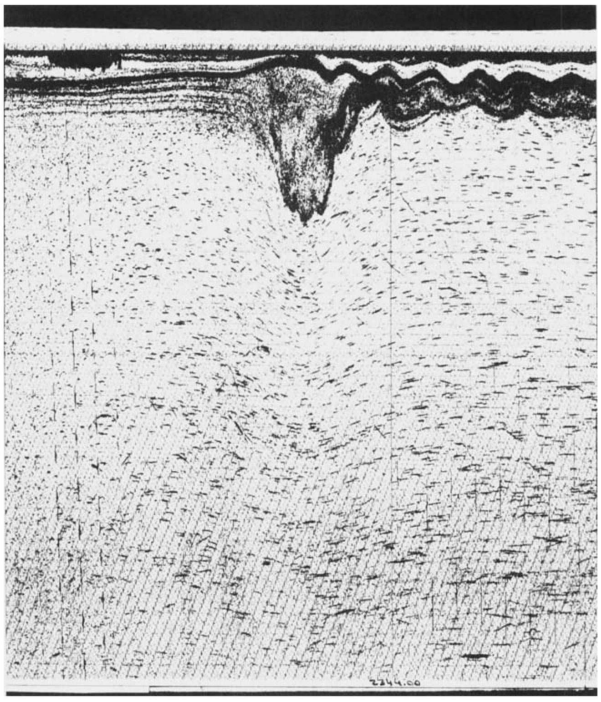


FIGURE 2. Acoustical image of internal-wave disturbances generated by stratified flow past a sill in the field experiments of Farmer & Smith (1980). The streamlines indicate that the main disturbance is a mode-2 solitary-like wave and is followed by a train of smaller-amplitude mode-1 short waves.

Recently, Turkington, Eydeland & Wang (1991), using a variational formulation of the governing equations, proposed a numerical technique for computing solitary-wave solutions in a stratified fluid, and presented several examples of mode-1 solitary waves; as expected, these waves are locally confined. In earlier related work, Tung, Chan & Kubota (1982) proved analytically and confirmed through numerical computations that large-amplitude locally confined mode-1 and mode-2 solitary waves are possible in a stratified fluid of finite depth, under the Boussinesq approximation. However, in discussing mode-2 solitary waves, they further assume that the density stratification is such that the Brunt–Väisälä frequency is symmetric about the fluid-layer centreline. This additional condition precludes the appearance of mode-1 oscillatory tails because waves of the first mode are symmetric while waves of the second mode are antisymmetric about the centreline. Nevertheless, mode-1 oscillations are still expected to develop at the tails of mode-3 solitary waves, which are also symmetric, but Tung *et al.* (1982) do not report calculations of solitary waves of mode-3 or higher.

Figure 4: An experimental observation from [1] of an asymmetric generalized solitary internal wave generated by stratified flow past a sill. The streamlines are visualised using acoustic imaging, and appear to show a mode 2 solitary wave followed by a train of smaller-amplitude mode-1 waves.

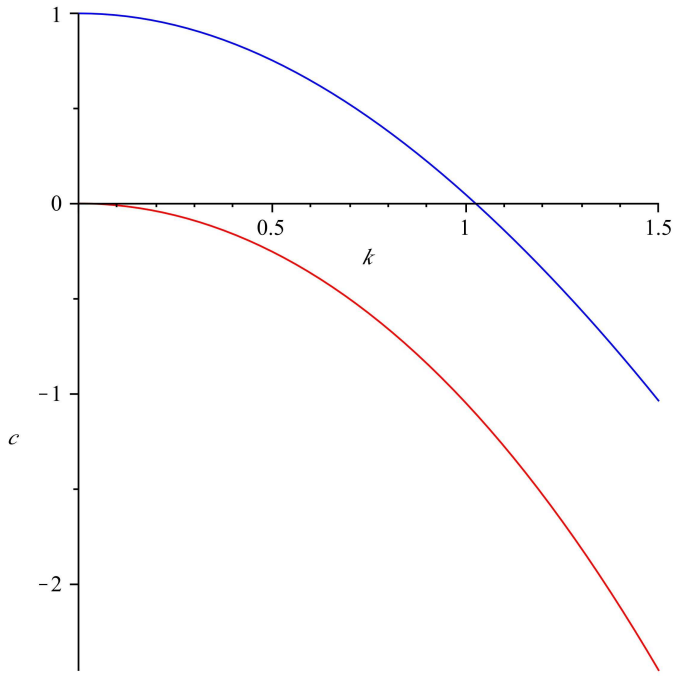


Figure 5: Plot of the linear phase speed c as a function of wave number k in the coupled KdV equations. Both the u-mode (red curve) and the v-mode (blue curve) are shown for the case $\Delta = 1$, $p = 0.5$, $\lambda = 0.2$.

Substituting this into (4a,4b) and solving order by order in ϵ yields

$$u_1 = 2\gamma^2 \mathrm{sech}^2(\epsilon\gamma x)\,, \quad v_1 = 0\,, \quad c_1 = 4\gamma^2\,, \tag{6a}$$

$$u_2 = \frac{\lambda}{\Delta}\left\{(20p^2+q^2-8pq)c_1u_1 - (q-6p)(q-10p)u_1^2\right\}\,, \tag{6b}$$

$$v_2 = -\frac{\lambda}{\Delta}\left\{pc_1u_1 + \tfrac{1}{2}(q-6p)u_1^2\right\}\,, \tag{6c}$$

$$c_2 = -\frac{\lambda}{\Delta}p^2c_1^2\,. \tag{6d}$$

The expansion can be continued to all orders in ϵ^2 without any oscillatory tail being detected. This is because the size of the tail depends exponentially on ϵ, and so it decays faster as $\epsilon \to 0$ than any power of ϵ.

To find the tail oscillations, we observe that u_n, v_n are singular in the complex plane at $x = (2m+1)i\pi/2\epsilon\gamma$, $m \in \mathbb{Z}$. This motivates a closer examination via the change of variable

$$x = \frac{i\pi}{2\epsilon\gamma} + z\,, \tag{7}$$

which allows us to consider the region of the complex plane close to the first singularity. Then as $\epsilon z \to 0$, $\mathrm{sech}^2(\epsilon\gamma x) \sim -1/\epsilon^2\gamma^2z^2$, and so, substituting back into our asymptotic expansion (6a)–(6d),

$$u_s \sim -\frac{2}{z^2} - \frac{\lambda}{2\Delta z^4}(q-6p)(q-10p) + \cdots + O(\epsilon^2)\,, \tag{8a}$$

$$v_s \sim -\frac{2\lambda}{\Delta z^4}(q-6p) + \cdots + O(\epsilon^2)\,. \tag{8b}$$

Next we consider the inner problem, in which we seek solutions of (4a, 4b) in the form $u = u(z), v = v(z)$, and for which the expressions (8a, 8b) form an outer boundary condition. The outcome is just the same system (4a, 4b) with x replaced by z,

$$-cu + 3u^2 + u_{zz} + pv_{zz} + quv + \tfrac{1}{2}rv^2 = 0\,, \tag{9a}$$

$$-cv + \Delta v + 3v^2 + v_{zz} + \lambda\left(pu_{zz} + ruv + \tfrac{1}{2}qu^2\right) = 0\,. \tag{9b}$$

Note that $c = O(\epsilon^2)$ from (6a) and can be omitted at the leading order. We proceed by applying a Laplace transform

$$[u,v] = \int_\Gamma e^{-zs}[U(s),V(s)]\,ds\,, \tag{10}$$

where the contour Γ runs from 0 to ∞ in the half-plane $\mathrm{Re}\{sz\} > 0$. We then seek a power series solution

$$[U(s),V(s)] = \sum_{n=1}^{\infty}[a_n,b_n]s^{2n-1}\,, \tag{11}$$

where $a_1 = -2$, $b_1 = 0$, $a_2 = -\lambda(q-6p)(q-10p)/12\Delta$, $b_2 = -\lambda(q-6p)/3\Delta$ from (8a, 8b). In general, substitution of (11) into the Laplace transform (10) generates the asymptotic series

$$[u,v] \sim \sum_{n=1}^{\infty}[\alpha_n,\beta_n]z^{-2n}\,, \quad [\alpha_n,\beta_n] = (2n-1)![a_n,b_n]\,. \tag{12}$$

This agrees with the asymptotic series (8a, 8b), and in effect the Laplace transform is a Borel summation of the asymptotic series.

Substitution of the Laplace transform (10) and the series (11) into the differential equation system (9a, 9b) yields a recurrence relation for $[a_n, b_n]$. Setting $\Delta[A_n, B_n] = (-k_0^2)^n[a_n, b_n]$, we find that

$$\frac{(n+1)(2n+5)}{(n-1)(2n-1)}A_{n-1} + \left(p - \frac{q}{(n-1)(2n-1)}\right)B_{n-1} = F_n\,, \tag{13}$$

$$\left(1-\lambda p^2\right)B_n - B_{n-1} - \lambda p A_{n-1} + \lambda\frac{rB_{n-1}+qA_{n-1}}{(n-1)(2n-1)} = G_n\,, \tag{14}$$

where F_n and G_n are quadratic convolution expressions in $A_2, \ldots, A_{n-2}$ and $B_2, \ldots, B_{n-2}$. As $n \to \infty$, these nonlinear terms can be neglected, and we find that

$$[A_n, B_n] \to [-p, 1]K \quad \text{as} \quad n \to \infty\,, \tag{15}$$

where K is a constant whose value depends on λ, p, q, r. It follows that the series (11) converges for $|s| < k_0$, $k_0^2 = \Delta/(1-\lambda p^2)$. The result (15) shows that as $|s| \to k_0$ there is a pole singularity given by

$$[U(s), V(s)] \approx \Delta\frac{[p, -1]K}{2(s - ik_0)}\,. \tag{16}$$

We have now established that the solution in the z-variable is given by (10) where $[U(s), V(s)]$ has a pole singularity at $s = ik_0$, at the complex conjugate point $s = -ik_0$, and at all of their harmonics $s = \pm imk_0$, $m \in \mathbb{N}$. Hence the contour Γ should be chosen to avoid the imaginary s-axis, and to be explicit we choose it to lie in $\mathrm{Re}\{s\} > 0$. However, we seek a symmetric solution, which in the z-variable requires that $\mathrm{Im}\{u, v\} = 0$ when $\mathrm{Re}\{z\} = 0$. The presence of the pole prevents (10) from satisfying this condition, and so we must correct it by adding a subdominant term

$$[u, v] = \int_\Gamma e^{-zs}[U(s), V(s)]\,ds + \frac{ib}{2}[p, -1]\exp(-ik_0 z + i\delta). \tag{17}$$

Here b, δ are real constants, and note that $|\exp(-ik_0 z)|$ is smaller than any power of $|z|^{-1}$ as $z \to \infty$ in $\mathrm{Re}\{z\} > 0\,, \mathrm{Im}\{z\} < 0$, recalling that $x = (i\pi/2\epsilon\gamma) + z$. The symmetry condition is now applied by bringing the contour Γ onto $\mathrm{Re}\{s\} = 0$ and deforming around the pole at $s = ik_0$. The outcome is

$$b\cos\delta = \pi K\,, \tag{18}$$

which we substitute into (17). The final step is to bring this solution back to the real axis, using $x = (i\pi/2\epsilon\gamma) + z$. Taking account of the corresponding singularity at $s = -ik_0$, we finally arrive at

$$[u, v] \sim [u_s, v_s] + b\Delta[-p, -1]\exp(-\pi k_0/2\epsilon\gamma)\sin(k_0|x| - \delta)\,. \tag{19}$$

This is a two-parameter family in γ and δ, where $0 < \delta < \pi/2$. The minimum tail amplitude occurs at $\delta = 0$, whilst the amplitude tends to infinity as $\delta \to \pi/2$. Note that the constant argument of the exponential is determined by the location of the singularity, but we require the exponential asymptotics to find the amplitude.

3 Embedded solitons

The constant K is determined by the recurrence relations (13, 14). It is a function of the system parameters λ, p, q, r and in general is found numerically. However, from our straightforward asymptotic solution (6a, 6b) we know that $K = 0$ for $q = 6p$, and in general we may find many parameter combinations where $K = 0$. In particular,

$$K \sim \frac{\lambda(6p - q)}{3\Delta} \quad \text{as} \quad \lambda \to 0\,. \tag{20}$$

These special values imply that the solitary wave decays to zero at infinity, even though its speed lies inside the linear spectrum, at least in this asymptotic limit. These are called embedded solitons. They are usually not stable, but are then *metastable*, or are said to exhibit *semi-stability*, in that they are unstable to small, but not infinitesimal, perturbations. Nevertheless, they are found to be useful in several applications, such as nonlinear optics and solid state physics. For water waves with surface tension, generalized solitary waves exist for Bond numbers $0 < B < 1/3$, although numerical simulations suggest that there are no embedded solitons.

4 One-sided generalized solitary waves

These symmetric solitary waves cannot be realized in practice, since they require an energy source and sink at infinity. Instead, they are replaced by solitary waves with radiating tails on one side only, determined by the group velocity. That is, in $x > 0$ for $c_g > c$, or in $x < 0$ for $c_g < c$, where c_g is the group velocity at the resonant wavenumber. For the present case, the linear dispersion relation is (2) and so for the relevant u-mode, $c_g = \Delta - 3k^2 < c = \Delta - k^2$. Hence there are no oscillations in $x > 0$, but they will appear in $x < 0$.

Thus in $x > 0$, or more generally in $\mathrm{Re}\{z\} > 0$, the solution is completely defined by the Laplace transform integral (10), with the contour Γ lying in $\mathrm{Re}\{s\} > 0$. Then for $x < 0$, or $\mathrm{Re}\{z\} < 0$, the contour Γ must be moved to $\mathrm{Re}\{z\} < 0$ across the axis $\mathrm{Re}\{s\} = 0$. In this process the solution collects a contribution from the pole at $s = ik_0$, which generates the tail oscillation. The final outcome is that (19) is replaced by

$$[u, v] \sim [u_s, v_s] - H(-x)2\pi K\Delta[-p, -1]\exp(-\pi k_0/2\epsilon\gamma)\sin(k_0 x) \tag{21}$$

where $H(\cdot)$ is the Heaviside function. In effect the phase shift $\delta = 0$, there are no oscillations in $x > 0$, and the amplitude in $x < 0$ is exactly twice the amplitude of the symmetric solution.

5 Weak Coupling

Let us now return to the coupled travelling wave equations (4a, 4b). Supposing that the coupling parameter is very small, $0 < \lambda \ll 1$, we may expand asymptotically in λ as follows,

$$[u, v] \sim \sum_{n=0}^{\infty} \lambda^n [u_n, v_n]\,, \quad c \sim \sum_{n=0}^{\infty} \lambda^n c_n\,. \tag{22}$$

Substituting this expansion into (4a 4b), we find that the leading order solution is

$$u_0 = 2\beta^2 \mathrm{sech}^2(\beta x)\,, \quad v_0 = 0\,, \quad c_0 = 4\beta^2\,. \tag{23}$$

This leading term is a u-mode solitary wave. A comparison with the previous expansion (5) suggests that $\beta = \epsilon\gamma$, but now the amplitude can be order unity. At the next order

$$-c_0 u_1 + 6u_0 u_1 + u_{1xx} + pv_{1xx} + qu_0 v_1 - c_1 u_0 = 0\,, \tag{24a}$$

$$(\Delta - c_0)v_1 + v_{1xx} + pu_{0xx} + \tfrac{1}{2}qu_0^2 = 0\,, \tag{24b}$$

We use the leading order solution for u (23) to rewrite (24b) as

$$(\Delta - c_0)v_1 + v_{1xx} = f(x) = -pc_0 u_0 + \tfrac{1}{2}(6p - q)u_0^2\,. \tag{25}$$

Note that in the limit $\lambda \to 0$, the resonant wavenumber is $k_0 \approx (\Delta - c_0)^{1/2}$, which takes account of the finite speed of the wave. We must now take $c_0 < \Delta$ to get tail oscillations, whilst for $c_0 > \Delta$ the expansion yields a genuine solitary wave. The general solution of (25) is

$$v_1 = A\sin k_0 x + B\cos k_0 x + \frac{1}{2k_0}\int_{-\infty}^{\infty} f(x')\sin\left(k_0|x - x'|\right)dx'\,. \tag{26}$$

To determine the constants A, B we impose a symmetry condition on v_1, so that $A = 0$, and then

$$v_1 \sim b_1 \sin\left(k_0|x| - \delta\right) \quad \text{as} \quad |x| \to \infty\,, \tag{27}$$

$$b_1 \cos\delta = L = \frac{1}{2k_0}\int_{-\infty}^{\infty} f(x)\cos\left(k_0 x\right)dx\,. \tag{28}$$

With v_1 known, we take the limit $|x| \to \infty$ in (24a) to find

$$u_1 \sim -p\frac{(\Delta - c_0)}{\Delta} b_1 \sin\left(k_0|x| - \delta\right), \quad \text{as} \quad |x| \to \infty\,, \tag{29}$$

Substituting (25) into (28), we find that

$$L = -\frac{\beta^2}{6k_0}\left\{k_0^2(q - 6p) + 4\beta^2 q\right\}\int_{-\infty}^{\infty} \mathrm{sech}^2(\beta x)\cos\left(k_0 x\right)dx\,. \tag{30}$$

Then, as $\beta = \epsilon\gamma \to 0$, this reduces to

$$L \sim \frac{\pi k_0^2}{3}(6p - q)\exp\left(-\pi k_0/2\epsilon\gamma\right), \tag{31}$$

which agrees with the previous result (20) from the exponential asymptotics, since $L = \pi K$. The one-sided solutions are obtained by setting $\delta = 0$, and replacing b_1 in (27, 29) by either 0 for $x > 0$ or $2b_1$ for $x < 0$.

References

[1] T. Akylas and R. Grimshaw, *Solitary internal waves with oscillatory tails*, J. Fluid Mech., 242 (1992), pp. 279–298.

[2] J. Boyd, *Weakly Nonlocal Solitary Waves and Beyond-All-Orders Asymptotics*, Kluwer, Amsterdam, 1998.

[3] D. Farmer and J. Smith, *Tidal interaction of stratified flow with a sill in Knight Inlet*, Deep-Sea Res, 27 (1980), pp. 239–254.

[4] R. Grimshaw and P. Cook, *Solitary waves with oscillatory tails*, in Proceedings of Second International Conference on Hydrodynamics, Hong Kong, 1996, "Hydrodynamics: Theory and Applications", A. Chwang, J. Lee, and D. Leung, eds., vol. 1, A.A. Balkema, Rotterdam, 1996, pp. 327–336.

[5] R. Grimshaw and N. Joshi, *Weakly non-local solitary waves in a singularly perturbed Korteweg-de Vries equation*, SIAM J. Appl. Math., 55 (1995), pp. 124–135.

[6] M. Kruskal and H. Segur, *Asymptotics beyond all orders in a model of crystal growth*, Stud. Appl. Math., 85 (1991), pp. 129–181.

[7] Y. Pomeau, A. Ramani, and B. Grammaticos, *Structural Stability of Korteweg de Vries solitons under a singular perturbation*, Physica D, 31 (1988), pp. 127–134.

Lecture 18: Wave-Mean Flow Interaction, Part I

Lecturer: Roger Grimshaw. Write-up: Hiroki Yamamoto

June 25, 2009

1 Introduction

Nonlinearity in water waves can lead to wave breaking. We can observe easily that waves break as they come to a beach. The waves on a current also may break. In this lecture we derive modulation equations of water waves using Whitham's averaged Lagrangian method. Then we consider the interaction of nonlinear water waves with currents and slopping bottom topography (e.g. waves on a beach).

2 Water waves in the linear approximation

In the linear approximation, the surface elevation ζ for sinusoidal unidirectional waves is

$$\zeta(x,t) = a\cos\theta, \quad \theta = kx - \omega t + \alpha, \tag{1}$$

for waves of amplitude a, wavenumber k (> 0), and frequency ω. Here α is an arbitrary constant ensemble parameter. When there is no mean current the dispersion relation is

$$\omega^2 = gk\tanh(kH), \tag{2}$$

where g is the acceleration due to gravity and H is the mean water depth.

If there is a constant horizontal mean current U, in the frame moving with the current ($x' = x - Ut$), the dispersion relation remains similar to (2). Then, θ can also be written as

$$\theta = kx - \omega t + \alpha = k(x' + Ut) - \omega t + \alpha = kx' - (\omega - kU)t + \alpha, \tag{3}$$

so that the dispersion relation of water waves on a horizontal mean current, in the *rest* frame is

$$\omega = Uk + \omega^*. \tag{4}$$

where ω^* is the intrinsic frequency (i.e. $\omega^* = \pm[gk\tanh(kH)]^{1/2}$), which has two branches. The total frequency ω is thus decomposed into the Doppler shift Uk and the intrinsic frequency ω^*.

3 Modulation equations of water waves

Now suppose that the amplitude, wave number, frequency, mean current and mean depth vary slowly relative to the wave field. Then (1) is replaced by the Fourier series expansions as

$$\zeta(x,t) \sim a(x,t)\cos\theta + a_2\cos 2\theta + O(a^3), \tag{5}$$

$$\theta = \phi(x,t) + \alpha, \quad k = \frac{\partial\phi}{\partial x}, \quad \omega = -\frac{\partial\phi}{\partial t}. \tag{6}$$

Here ϕ is the phase, and the ensemble parameter α is constant, and the coefficient $a_2 \sim O(a^2)$ depends on ω^*, k, U, H. It is convenient to introduce a velocity potential Ψ defined as

$$\Psi = Ux - Bt + \Phi(\theta, z), \tag{7}$$

where Φ is the wave component of Ψ, and B is related to the mean height of the waves. Now, Φ is expanded as Fourier series in the form

$$\Phi(\theta,z) = A_1\cosh(kz)\sin\theta + A_2\cosh(2kz)\sin 2\theta + O(a^3), \tag{8}$$

where $A_1 \sim O(a)$ and $A_2 \sim O(a^2)$. This is because of the kinematic boundary condition at free surface $\frac{\partial\Psi}{\partial z} = \frac{\partial\zeta}{\partial t} + \frac{\partial\Psi}{\partial x}\frac{\partial\zeta}{\partial x}$. From (6), the equation for conservation of waves is,

$$\frac{\partial k}{\partial t} + \frac{\partial\omega}{\partial x} = 0. \tag{9}$$

The issue is to determine how the amplitude, wavenumber, frequency, mean current and mean depth vary (slowly) in space and time. The mean current $U(x,t)$ and depth $H(x,t)$ can be decomposed into background components $u(x,t)$, $h(x)$ and a wave-induced $O(a^2)$ component.

The modulation equations for the wave amplitude, wavenumber, frequency, mean current and mean depth are found using Whitham's averaged Lagrangian method. The Lagrangian of the water wave system is (Whitham, 1974, chapter 13.2)

$$L = -\int_{-h}^{\zeta}\left\{\frac{\partial\Psi}{\partial t} + \frac{1}{2}\left(\frac{\partial\Psi}{\partial x}\right)^2 + \frac{1}{2}\left(\frac{\partial\Psi}{\partial z}\right)^2 + gz\right\}\mathrm{d}z. \tag{10}$$

Substituting (7) into (10), we obtain

$$\begin{aligned} L &= \int_{-h}^{\zeta}\left[B + \omega\frac{\partial\Phi}{\partial\theta} - \frac{1}{2}\left(U + k\frac{\partial\Phi}{\partial\theta}\right)^2 - \frac{1}{2}\left(\frac{\partial\Phi}{\partial z}\right)^2 - gz\right]\mathrm{d}z, \\ &= \left(B - \frac{1}{2}U^2\right)H - \frac{1}{2}gH^2 + gHh \\ &\quad +(\omega - Uk)\int_{-h}^{\zeta}\frac{\partial\Phi}{\partial\theta}\mathrm{d}z - \int_{-h}^{\zeta}\frac{1}{2}\left[k^2\left(\frac{\partial\Phi}{\partial\theta}\right)^2 + \left(\frac{\partial\Phi}{\partial z}\right)^2\right]\mathrm{d}z. \end{aligned} \tag{11}$$

Note that $\zeta + h = H$ and $\zeta^2 - h^2 = (\zeta+h)^2 - 2(\zeta+h)h$.

Averaging the Lagrangian (11) in α, we obtain

$$\overline{L} = \frac{1}{2\pi}\int_0^{2\pi} L \mathrm{d}\alpha = \overline{L}^{(\mathrm{m})}(U,B,H,h) + \overline{L}^{(\mathrm{w})}(E^*,\omega^*,k,H), \tag{12}$$

$$E^* = \frac{ga^2}{2}, \quad k = \frac{\partial\phi}{\partial x}, \quad \omega = -\frac{\partial\phi}{\partial t}, \quad U = \frac{\partial\Psi}{\partial x}, \quad B = -\frac{\partial\Psi}{\partial t}. \tag{13}$$

The functions $\overline{L}^{(\mathrm{m})}(U,B,H,h)$ and $\overline{L}^{(\mathrm{w})}(E^*,\omega^*,k,H)$ are mean and wave components of averaged Lagrangian respectively.

$$\text{Mean} \quad : \qquad \overline{L}^{(\mathrm{m})} = \left(B - \frac{U^2}{2}\right)H - \frac{gH^2}{2} + gHh, \tag{14}$$

$$\text{Wave} \quad : \qquad \overline{L}^{(\mathrm{w})} = \frac{DE^*}{2} + \frac{D_2 k^2 E^{*2}}{2g} + O(E^{*3}), \tag{15}$$

where

$$D = \frac{\omega^{*2}}{gkT} - 1, \quad D_2 = -\frac{9T^4 - 10T^2 + 9}{8T^4}, \quad T = \tanh(kH), \quad \omega^* = \omega - Uk. \tag{16}$$

These expressions are derived first by finding Ψ and thus L to get $\bar{L}$. The coefficients A_1, A_2, a_2 in (7) are obtained by solving the variational equations

$$\frac{\partial\overline{L}}{\partial A_1} = 0, \quad \frac{\partial\overline{L}}{\partial A_2} = 0, \quad \frac{\partial\overline{L}}{\partial a_2} = 0, \tag{17}$$

for A_1, A_2, a_2. The resulting Φ is then used to find Ψ (8), which is then re-substituted into L (11) and then $\overline{L}$. See Whitham (1974, chapter 16.6) and Whitham (1967) for details.

To obtain the modulation equations, the averaged variational principle

$$\delta\int\int \overline{L}dxdt = 0, \tag{18}$$

is used for variations in $\delta E^*, \delta\phi, \delta\psi, \delta H$, and we have

$$\delta E^* \quad : \quad \frac{\partial\overline{L}}{\partial E^*} = 0, \tag{19}$$

$$\delta\phi \quad : \quad \frac{\partial}{\partial t}\left(\frac{\partial\overline{L}}{\partial\omega}\right) - \frac{\partial}{\partial x}\left(\frac{\partial\overline{L}}{\partial k}\right) = 0, \quad \frac{\partial k}{\partial t} + \frac{\partial\omega}{\partial x} = 0, \tag{20}$$

$$\delta\Psi \quad : \quad \frac{\partial}{\partial t}\left(\frac{\partial\overline{L}}{\partial B}\right) - \frac{\partial}{\partial x}\left(\frac{\partial\overline{L}}{\partial U}\right) = 0, \quad \frac{\partial U}{\partial t} + \frac{\partial B}{\partial x} = 0, \tag{21}$$

$$\delta H \quad : \quad \frac{\partial\overline{L}}{\partial H} = 0. \tag{22}$$

From (19), the dispersion relation is obtained as

$$\frac{D}{2} + \frac{k^2 D_2 E^*}{g} + O(E^{*2}) = 0. \tag{23}$$

From (21) we obtain

$$\frac{\partial H}{\partial t}+\frac{\partial}{\partial x}\left[UH+k\left(E\frac{(\omega-Uk)}{gkT}\right)\right]=\frac{\partial H}{\partial t}+\frac{\partial}{\partial x}(HV)=0, \tag{24}$$

where

$$V=U+\frac{kA}{H},\quad A=\frac{\partial\overline{L}}{\partial\omega}=E\frac{(\omega-Uk)}{gkT}. \tag{25}$$

From (22) we obtain

$$B=\frac{1}{2}U^2+gH-gh+\frac{1}{2}\left(\frac{1-T^2}{T}\right)kE^*+O(E^{*2}). \tag{26}$$

Using (25) and (26), we obtain

$$\frac{\partial}{\partial t}(HV)+\frac{\partial}{\partial x}(HV^2)+\frac{\partial}{\partial x}\left(\frac{gH^2}{2}\right)+\frac{\partial S}{\partial x}=gH\frac{\partial h}{\partial x}, \tag{27}$$

where

$$S=k(F-VA)+\overline{L}^{(\mathrm{w})}-H\frac{\partial\overline{L}^{(\mathrm{w})}}{\partial H},\quad F=-\frac{\partial\overline{L}^{(\mathrm{w})}}{\partial k}. \tag{28}$$

Equation (20) can be written as

$$\frac{\partial A}{\partial t}+\frac{\partial F}{\partial x}=0. \tag{29}$$

Now, (19), (29), (24), and (27) are the dispersion relation, the wave action equation, the mean flow and mean momentum equations of the modulation equations respectively. To these we add equation (9) for conservation of waves.

$$\frac{\partial k}{\partial t}+\frac{\partial\omega}{\partial x}=0. \tag{30}$$

These equations are fully nonlinear. A is the wave action density and F is the wave action flux. In the linearized approximation ($H\approx h$, $U\approx u$) the dispersion relation (23) becomes

$$D(\omega^*,k,h)=0,\quad \omega=\omega^*+ku,\quad \omega^{*2}=gk\tanh(kh) \tag{31}$$

$$A=\frac{\partial D}{\partial\omega^*}\frac{E^*}{2}=\frac{E^*}{\omega^*},\quad F=c_gA=(c_g^*+u)A, \tag{32}$$

where $c_g^*=\frac{\partial\omega^*}{\partial k}$ is the intrinsic group velocity. S is the radiation stress, which in the linearized approximation reduces to

$$S=\left(kc_g^*+h\frac{\partial\omega^*}{\partial h}\right)A=\left(2kc_g^*-\frac{\omega^*}{2}\right)A, \tag{33}$$

since for water waves,

$$h\frac{\partial\omega^*}{\partial h}=kc_g^*-\frac{\omega^*}{2}. \tag{34}$$

The equation for conservation of waves (30) becomes

$$\frac{\partial k}{\partial t}+u\frac{\partial k}{\partial x}=-k\frac{\partial u}{\partial x}-\frac{\partial\omega^*}{\partial h}\frac{\partial h}{\partial x},\quad \frac{\partial\omega}{\partial t}+u\frac{\partial\omega}{\partial x}=k\frac{\partial u}{\partial t}. \tag{35}$$

Note that for steady backgrounds the frequency is conserved.

4 Waves on a current

Now we consider a unidirectional steady current $u = u(x)$, with constant depth h.

4.1 Linear approximation

In the linearized approximation ($H \approx h$, $U \approx u$), equation (30) becomes

$$\frac{\partial k}{\partial t} + \frac{\partial \omega}{\partial x} = 0, \quad \omega = uk + \omega^*, \quad \omega^{*2} = gk \tanh(kh). \tag{36}$$

The steady solution is $\omega = \omega_0$ (constant), with $k = k(x)$. The wave amplitude is obtained from the wave action equation (29), which reduces to

$$\frac{\partial A}{\partial t} + \frac{\partial}{\partial x}(c_g A) = 0, \quad c_g = u + c_g^*, \quad A = \frac{E^*}{\omega^*}. \tag{37}$$

The steady solution has constant wave action flux F_0,

$$2c_g A = \frac{c_g c^* a^2}{\tanh(kh)} = 2F_0, \quad c^* = \frac{\omega^*}{k}. \tag{38}$$

For simplicity, we now make the deep-water approximation $kh \to \infty$, so that $\omega^{*2} = gk$, $c_g^* = c^*/2$. Suppose that $u(x = 0) = 0$ and the intrinsic phase speed is $c^* = c_0 > 0$ at $x = 0$. Then the steady solution of (36) is obtained as follows. Since $\omega = \omega_0$,

$$u(x)k(x) + \sqrt{gk(x)} - \sqrt{gk(0)} = 0. \tag{39}$$

Dividing (39) by $k(x)$, we obtain

$$u(x) + c^*(x) - \frac{c^{*2}(x)}{c_0} = 0. \tag{40}$$

Here we used $c^*(x) = \sqrt{g/k(x)}$, $c_0 = c^*(0) = \sqrt{g/k(0)}$. Then the solution of (40) is

$$c^*(x) = \frac{c_0}{2} \pm \left\{ c_0 u(x) + \frac{c_0^2}{4} \right\}^{1/2}. \tag{41}$$

The condition at $x = 0$ means we choose the plus sign. Note that the group velocity is

$$c_g(x) = u(x) + \frac{c^*}{2} = u(x) + \frac{c_0}{4} \pm \frac{1}{2} \left\{ c_0 u(x) + \frac{c_0^2}{4} \right\}^{1/2}. \tag{42}$$

Thus, for an advancing current $u(x) > 0$, $x > 0$, we must choose only the plus sign, and so $c^*(x)$, $c_g(x)$ both increase as $u(x)$ increases, while then $k(x) = g/c^{*2}$ decreases. Since $c_g c^* a^2 = 2F_0$, the wave amplitude decreases.

For an opposing current $u(x) < 0$, $x > 0$, there is a stopping velocity at $x = x_c$, $u(x_c) = -c_0/4$, and the waves cannot penetrate past this point, since $c_g(x_c) = 0$. Instead the waves reflect, with the minus sign in (41, 42). Both $c^*(x)$, $c_g(x)$ decrease as $|u(x)|$ increases, while $k(x)$ increases. Since $c_g c^* a^2 = 2F_0 = c_0^2 a_0^2$, the wave amplitude increases

from the initial value a_0, and $a^2 \to \infty$ as $x \to x_c$. Of course, this result is outside the linear approximation, and in practice the waves will break at $x_b < x = x_c$. Here we use an experimental breaking criterion, $ak(x_b) = 0.44$ (Miche, 1944); note that x_b depends on a_0 and c_0. Using (41, 42) and $c_g c^* a^2 = 2F_0 = c_0^2 a_0^2$, the wave steepness ak can be calculated as follows,

$$ak = a_0 k_0 (G_1 G_2)^{-1/2} G_1^{-2}, \tag{43}$$

where

$$G_1 = \frac{1}{2} + \left(\frac{u}{c_0} + \frac{1}{4} \right)^{1/2}, \tag{44}$$

$$G_2 = \frac{u}{c_0} + \frac{1}{2} G_1. \tag{45}$$

The relation between u/c_0 and ak (i.e. equation (43)) is shown as Figure 1.

This rather simple theory has applications to the formation of giant (rogue, freak) waves in the ocean, for example on the Agulhas current. There are also applications to the modulation of water waves by an underlying internal solitary wave, whose surface current is $u(x) = u_0 \mathrm{sech}^2(Kx)$ say (Fig. 2). To explore these further, we take a wave packet solution of the wave action equation (37)

$$c_g A = c_g c^* a^2 = c_0^2 a_0^2 b^2 (t - \tau), \quad \tau = \int_0^x \frac{dx}{c_g}. \tag{46}$$

Here $a_0 b(t)$ is the wave amplitude at $x = 0$, and we assume that the shape function $b(t)$ is localized (e.g. Gaussian), varying from 0 to a maximum of 1 at $t = 0$. Then the waves break throughout the zone, $x_b < x < x_c$, over a time interval determined by the width of the packet.

4.2 Nonlinear effects

In deep water, the wave-induced components of U, H are negligible and so the Lagrangian (12) becomes just (15) given now by

$$\overline{L}^{(\mathrm{w})} = \left(\frac{\omega^*}{gk} - 1 \right) \frac{E^*}{2} - \frac{k^2 E^{*2}}{2g} + O(E^{*3}), \tag{47}$$

where now $\omega^* = \omega - ku(x)$. The nonlinear dispersion relation (23) becomes

$$\omega^{*2} = gk + 2k^3 E^* + O(E^{*2}). \tag{48}$$

Conservation of wave action (29) and conservation of waves (30) again yield for a steady solution ($\frac{\partial}{\partial t} = 0$)

$$F = -\frac{\partial \overline{L}^{(\mathrm{w})}}{\partial k} = F_0, \quad \omega_0 = \omega^* + u(x)k, \tag{49}$$

where F_0, ω_0 are constants. When combined with (48) these yield two coupled equations for k, E^* in terms of $u(x)$. Now the dispersion relation (48) depends on the amplitude,

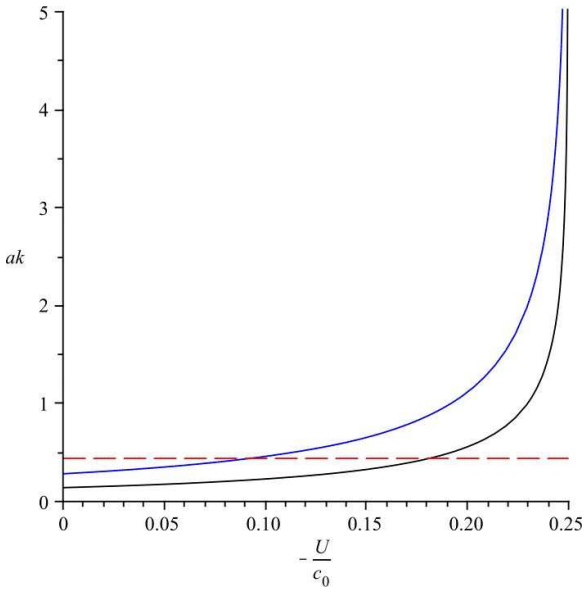


Figure 1: Wave steepness ak versus u/c_0; $a_0k_0 = 0.1, 0.2$ (black, blue), where a_0k_0 are wave steepness at $x = 0$. Wave breaking criterion $ak = 0.44$ (red dash), yields breaking for $|u|/c_0 > 0.18, 0.092$.

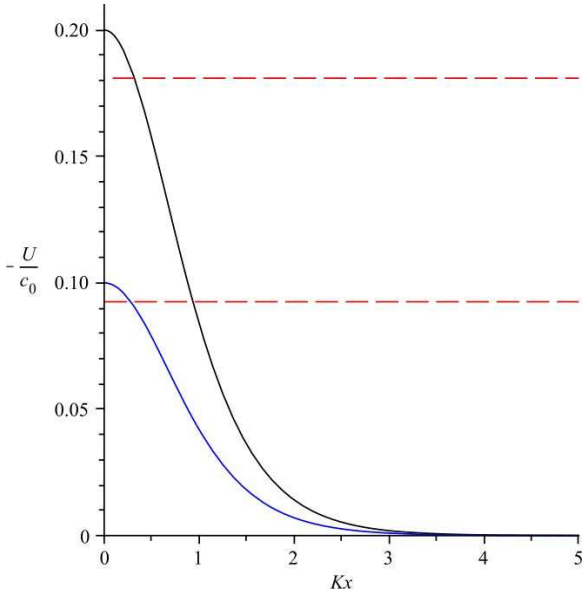


Figure 2: Breaking waves on the internal wave current $u = u_0\text{sech}^2(Kx)$, for $u_0/c_0 = -0.2, -0.1$ (black, blue), where the red lines give the breaking zones for $a_0k_0 = 0.1, 0.2$ (upper, lower). This shows that the waves of $a_0k_0 = 0.1$ (0.2) will break when they are on $|Kx| < 0.33$ (0.94) if the waves are on internal wave current indicated by black line.

$\omega^* = \omega^*(k, E^*)$ as well as the wavenumber. Conservation of wave action flux becomes

$$WA = F_0, \quad W = -\frac{\partial \overline{L}^{(\mathrm{w})}}{\partial k} \Bigg/ \frac{\partial \overline{L}^{(\mathrm{w})}}{\partial \omega} = u(x) + \frac{\omega^*}{2k} + k^2 A, \tag{50}$$

$$A = \frac{\partial \overline{L}^{(\mathrm{w})}}{\partial \omega} = \frac{E^*}{\omega^*}\left(1 + \frac{2k^2 E^*}{g}\right). \tag{51}$$

These are combined with (48) and (49),

$$\omega^{*2} = gk + 2k^3 \omega^* A, \quad \omega_0 = \omega^* + u(x)k, \tag{52}$$

to yield two equations for k, A in terms of $u(x)$. Note that for an opposing current $u(x) < 0$ ($x > 0$) there is now no stopping velocity, as $W \to 0$, $A \to \infty$ is not allowed.

The equation of wave steepness ak in terms of $u(x)/c_0$ is obtained by the same method used to derive (43), but is more complicated. For convenience, we define

$$s(ak) = \omega^* k^2 A, \quad s_0 = s(a_0 k_0), \tag{53}$$

and

$$\hat{s} = \frac{g + 2s_0}{g + 2s}. \tag{54}$$

Then the first equation of (52) is written as

$$\omega^* = k(g + 2s), \tag{55}$$

so A and W can be written as follows,

$$\begin{aligned} A &= \frac{E^*}{\omega^*}\left(1 + \frac{2k^2 E^*}{g}\right), \\ &= \frac{c^* a^2 g}{2(g + 2s)}(1 + a^2 k^2), \end{aligned} \tag{56}$$

$$\begin{aligned} W &= u + \frac{\omega^*}{2k} + k^2 A, \\ &= u + \frac{c^*}{2} + \frac{c^* a^2 k^2 g}{2(g + 2s)}(1 + a^2 k^2), \\ &= u + c^*\left[\frac{1}{2} + \frac{a^2 k^2 g}{2(g + 2s)}(1 + a^2 k^2)\right], \end{aligned} \tag{57}$$

where $c^* = \omega^*/k = \sqrt{g + 2s}/k$. Because $u(x = 0) = 0$, A_0 and W_0 are written as

$$A_0 = c_0 a_0^2 \frac{g}{2(g + 2s_0)}(1 + a_0^2 k_0^2), \tag{58}$$

$$W_0 = c_0\left[\frac{1}{2} + \frac{a_0^2 k_0^2 g}{2(g + 2s_0)}(1 + a_0^2 k_0^2)\right]. \tag{59}$$

From the second equation of (52), we obtain

$$\frac{k_0}{k} = \frac{c^*}{c_0} + \frac{u}{c_0}. \tag{60}$$

Using (54), k_0/k can be written as

$$\frac{k_0}{k} = \hat{s}\left(\frac{c^*}{c_0}\right)^2. \tag{61}$$

From (60) and (61), we obtain the quadratic equation for c^*/c_0 as

$$\hat{s}\left(\frac{c^*}{c_0}\right)^2 - \left(\frac{c^*}{c_0}\right) - \frac{u}{c_0} = 0, \tag{62}$$

and its solutions are

$$\frac{c^*}{c_0} = \frac{-1 \pm \sqrt{1 + 4\hat{s}(u_0/c_0)}}{2\hat{s}}. \tag{63}$$

Conservation of wave action flux can be written as

$$WA = W_0 A_0. \tag{64}$$

Multiplying (64) by $(60)^2$, we obtain

$$WA\left(\frac{c^*}{c_0} + \frac{u}{c_0}\right)^2 = W_0 A_0 \left(\frac{k_0}{k}\right)^2. \tag{65}$$

Substituting (56)-(59) to (65), we obtain

$$\begin{aligned}\left\{\frac{u}{c_0} + \frac{c^*}{c_0}\left[\frac{1}{2} + \frac{a^2k^2g}{2(g+2s)}(1+a^2k^2)\right]\right\}\left\{\frac{c^*}{c_0}a^2k^2\frac{g}{2(g+2s)}(1+a^2k^2)\right\}\left(\frac{c^*}{c_0}+\frac{u}{c_0}\right)^2 \\ = \left[\frac{1}{2} + \frac{a_0^2k_0^2g}{2(g+2s_0)}(1+a_0^2k_0^2)\right]a_0^2k_0^2\frac{g}{2(g+2s_0)}(1+a_0^2k_0^2),\end{aligned} \tag{66}$$

where c^*/c_0 can be calculated by (63). Because s and $\hat{s}$ are functions of ak, equation (66) describes the relation between ak and u/c_0, and it is shown as Figure 3.

5 Waves on a beach

In this section, we consider the waves on a beach. We recall that the full modulation equations are

$$\frac{\partial A}{\partial t} + \frac{\partial F}{\partial x} = 0, \quad A = \frac{\partial \overline{L}^{(\mathrm{w})}}{\partial \omega}, \quad F = -\frac{\partial \overline{L}^{(\mathrm{w})}}{\partial k}, \tag{67}$$

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(HV) = 0, \quad V = U + \frac{kA}{H}, \tag{68}$$

$$\frac{\partial V}{\partial t} + V\frac{\partial V}{\partial x} + g\frac{\partial \overline{\zeta}}{\partial x} + \frac{1}{H}\frac{\partial S}{\partial x} = 0, \tag{69}$$

$$S = k(F - VA) + \overline{L}^{(\mathrm{w})} - H\frac{\partial \overline{L}^{(\mathrm{w})}}{\partial H}, \tag{70}$$

$$\frac{\partial k}{\partial t} + \frac{\partial \omega}{\partial x} = 0, \tag{71}$$

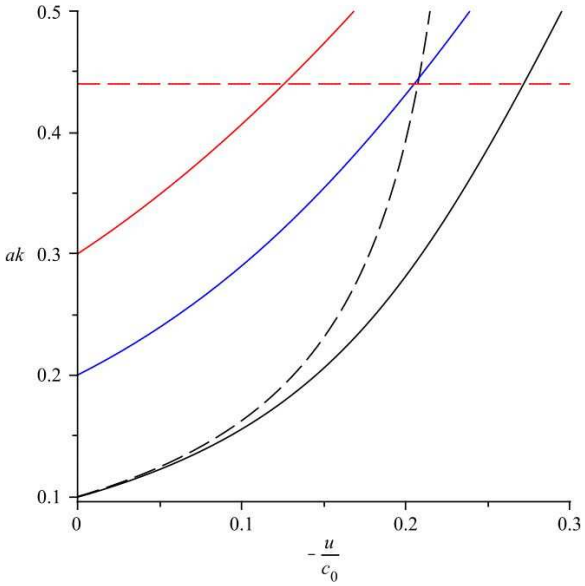


Figure 3: Wave steepness ak versus u/c_0; $a_0k_0 = 0.1, 0.2, 0.3$ (black, blue,red); wave breaking criterion $ak = 0.44$ (red dash) yields breaking for $|u|/c_0 > 0.27, 0.21, 0.13$. The dash line is the linear solution for $a_0k_0 = 0.1$.

where

$$\overline{L}^{(\mathrm{w})} = \frac{DE^*}{2} + \frac{D_2k^2E^{*2}}{2g} + O(E^{*3}), \tag{72}$$

and

$$D = \frac{\omega^{*2}}{gk\tanh(kH)} - 1, \quad \omega^* = \omega - Uk. \tag{73}$$

The mean momentum equation (69) has been rewritten.

Suppose that $h = h(x) \to 0$ as $x \to 0$, and that there is no background current. Then the steady solution ($\frac{\partial}{\partial t} = 0$) of these modulation equations yields the dispersion relation (71, 73), constant frequency $\omega = \omega_0$, and constant wave action flux and zero mass transport,

$$-\frac{\partial\overline{L}^{(\mathrm{w})}}{\partial k} = F_0, \quad V = U + \frac{kA}{H} = 0, \quad \omega^* = \omega_0 - Uk. \tag{74}$$

Thus there is a mean Eulerian flow $U = -kA/H$, opposing the Stokes drift due to the waves. The mean momentum equation (69) then yields the wave set-up $\overline{\zeta}$,

$$g\frac{\partial\overline{\zeta}}{\partial x} + \frac{1}{H}\frac{\partial S}{\partial x} = 0, \quad S = kF_0 + \overline{L}^{(\mathrm{w})} - H\frac{\partial\overline{L}^{(\mathrm{w})}}{\partial H}. \tag{75}$$

From (74), S as known in terms of H, and so

$$g\overline{\zeta} = -\int^x \frac{1}{H}\frac{\partial S}{\partial x}dx = -\int^H \frac{1}{H}\frac{\partial S}{\partial H}dH. \tag{76}$$

To illustrate, first make the small amplitude approximation. Then $\omega^* \approx \omega_0$, $H \approx h$, so that the dispersion relation becomes $\omega_0^2 = gk\tanh(kh)$ and yields $k = k(h)$, $D = 0$. The constant wave action flux condition reduces to

$$c_g a^2 = c_{g0} a_0^2, \tag{77}$$

where subscript "0" indicates the values at the depth $h = h_0$ offshore. The expression (76) can be written as follows. At first, we consider the total derivative of D. Now D can be considered as $D = D(k(h), h)$, so

$$\frac{dD}{dh} = \frac{\partial D}{\partial k}\frac{\partial k}{\partial h} + \frac{\partial D}{\partial h} = 0. \tag{78}$$

Because

$$\frac{\partial \overline{L}^{(\mathrm{w})}}{\partial h} = \frac{E^*}{2}\frac{\partial D}{\partial h}, \quad \frac{\partial \overline{L}^{(\mathrm{w})}}{\partial k} = \frac{E^*}{2}\frac{\partial D}{\partial k} = -F_0, \tag{79}$$

using (78) we obtain

$$\frac{\partial \overline{L}^{(\mathrm{w})}}{\partial h} = F_0 \frac{\partial k}{\partial h}. \tag{80}$$

Using (80), $\frac{1}{h}\frac{\partial S}{\partial h}$ becomes

$$\begin{aligned}
\frac{1}{h}\frac{\partial S}{\partial h} &= \frac{1}{h}\left\{ kF_0 - \frac{\partial}{\partial h}\left(h\frac{\partial \overline{L}^{(\mathrm{w})}}{\partial h}\right)\right\}, \\
&= \frac{1}{h}\left\{ F_0\frac{\partial k}{\partial h} - \frac{\partial \overline{L}^{(\mathrm{w})}}{\partial h} - h\frac{\partial^2 \overline{L}^{(\mathrm{w})}}{\partial h^2}\right\}, \\
&= -\frac{\partial^2 \overline{L}^{(\mathrm{w})}}{\partial h^2}.
\end{aligned}$$

Then, (76) becomes

$$\begin{aligned}
g\overline{\zeta} &= -\int^h \frac{1}{h}\frac{\partial S}{\partial h} dh, \\
&= \int^h \frac{\partial^2 \overline{L}^{(\mathrm{w})}}{\partial h^2} dh, \\
&= \frac{\partial \overline{L}^{(\mathrm{w})}}{\partial h}, \\
&= -\frac{1}{2}\left(\frac{1 - \tanh^2(kh)}{\tanh(kh)}\right) kE^*, \\
&= -\frac{1}{2}\left(\frac{-\sinh^2(kh) + \cosh^2(kh)}{\sinh(kh)\cosh(kh)}\right) kE^*, \\
&= -\frac{kE^*}{\sinh(2kh)},
\end{aligned} \tag{81}$$

so that,

$$\overline{\zeta} = -\frac{ka^2}{2\sinh(2kh)}, \tag{82}$$

where $\overline{\zeta}_0 = 0$. This is always negative, and so is a set-down. In shallow water $kh \to 0$, $c_g \approx (gh)^{1/2}$, and

$$\frac{k}{k_0} \approx \left(\frac{h_0}{h}\right)^{1/2}, \quad \frac{a}{a_0} \approx \left(\frac{h_0}{h}\right)^{1/4}, \tag{83}$$

so that, (82) can be approximated by

$$\overline{\zeta} \approx -\frac{a^2}{4h} = \frac{a_0^2 h_0^{1/2}}{4h^{3/2}}. \tag{84}$$

Since this small-amplitude theory predicts infinite amplitudes as $h \to 0$, we must consider nonlinear effects. One option is to impose an empirical wave-breaking condition $a/h = 0.44$ (Thornton and Guza, 1982, 1983), which defines the depth $h = h_b$, beyond which there is a surf zone. Here, we shall examine nonlinear effects in $h > h_b$ in the shallow water approximation $kH \to 0$. Then the Lagrangian (72) becomes

$$\overline{L}^{(\mathrm{w})} \approx \frac{DE^*}{2} - \frac{9E^{*2}}{16gk^2H^4}, \quad D \approx \frac{\omega^{*2}}{gHk^2}\left(1 + \frac{k^2H^2}{3}\right) - 1. \tag{85}$$

This Lagrangian is only valid when $ak \ll k^3H^3$, that is, for a very small Stokes number (Stokes, 1847). Using the linear shallow water expressions (83) we require that $S_0 = a_0/k_0^2h_0^3 \ll (h/h_0)^{9/4}$, which must fail as $h \to 0$. Hence, we infer that in shallow water we need to use a new theory, valid for Stokes number of order unity, so we consider the Korteweg-de Vries model next.

The Korteweg-de Vries (KdV) equation for weakly nonlinear long water waves, propagating on a constant undisturbed mean depth H, is given by (Mei, 1983, chapter 11.5.3)

$$\frac{\partial\zeta}{\partial t} + c_0\frac{\partial\zeta}{\partial x} + \frac{3c_0}{2H}\zeta\frac{\partial\zeta}{\partial x} + \frac{c_0H^2}{6}\frac{\partial^3\zeta}{\partial x^3} = 0, \quad c_0 = (gH)^{1/2}. \tag{86}$$

The KdV balance has linear dispersion, represented by $H^3\frac{\partial^3\zeta}{\partial x^3}$, balanced by nonlinearity, represented by $\zeta\frac{\partial\zeta}{\partial x}$. To leading order, the waves propagate unchanged in form with the linear long wave speed $c_0 = (gH)^{1/2}$. Nonlinearity leads to wave steepening, opposed by wave dispersion, resulting in the KdV balance and the well-known solitary wave

$$\zeta = a_s\mathrm{sech}^2[\kappa(x - ct)], \quad \frac{c}{c_0} - 1 = \frac{a_s}{2H} = \frac{2\kappa^2H^2}{3}. \tag{87}$$

The periodic wave solution of KdV equation (86) is

$$\zeta = 2a\left[b(m) + cn^2(\gamma\theta; m)\right], \quad \omega = -\frac{\partial\theta}{\partial t}, \quad k = -\frac{\partial\theta}{\partial x}, \tag{88}$$

$$b = \frac{1-m}{m}\frac{E(m)}{mK(m)}, \quad \frac{a}{H} = \frac{2}{3}m\gamma^2(kH)^2, \quad \gamma = \frac{K(m)}{\pi}, \tag{89}$$

$$c = \frac{\omega}{k} = c_0 \left\{ 1 + \frac{a}{H} \left[\frac{2-m}{m} - \frac{3E(m)}{mK(m)} \right] \right\}, \tag{90}$$

Here $cn(x;m)$ is the elliptic function of modulus m where $0 < m < 1$, and $K(m), E(m)$ are elliptic integrals of the first and second kind. The amplitude is a and the mean value is 0. As $m \to 1$, this solution becomes a solitary wave, since then $b \to 0$ and $cn^2(x) \to \mathrm{sech}^2(x)$. As $m \to 0$, $\gamma \to 1/2$, and it reduces to sinusoidal waves of small amplitude $a \sim m$. This cnoidal wave (88) contains two free parameters; we take these to be the amplitude a and the wavenumber k.

We now use the cnoidal wave expression (88) to evaluate the averaged Lagrangian (12), incorporating a mean current U,

$$\overline{L}^{(\mathrm{w})} = \left(\frac{c^{*2}}{gH} - 1 \right) G(m) \frac{E^*}{2} + O(E^{*2}), \quad E^* = \frac{ga^2}{2} \tag{91}$$

where

$$G(m) = 8(< cn^4(\gamma\theta; m) > - b^2), \tag{92}$$

or

$$G(m) = \frac{8[EK(4-2m) - 3E^2 - K^2(1-m)]}{3K^2m^2}. \tag{93}$$

To leading order the phase speed $c^* = W = (gH)^{1/2}$, while the wave action density, wave action flux and radiation stress now become, to leading order

$$A = \frac{\partial \overline{L}^{(\mathrm{w})}}{\partial \omega} = \frac{G(m)E^*}{\omega^*}, \quad F = -\frac{\partial \overline{\omega}^{(\mathrm{w})}}{\partial k} = (U + c^*)A, \tag{94}$$

$$S = \frac{3\omega^* A}{2} = \frac{3G(m)E^*}{2}. \tag{95}$$

As before, we now seek the steady solutions ($\frac{\partial}{\partial t} = 0$), so again $\omega = \omega_0$ is the constant wave frequency, and to leading order $kh^{1/2} = k_0 h_0^{1/2}$ is constant. Next $F = F_0$ is the constant wave action flux, implying that, to leading order in wave amplitude,

$$h^{1/2} G(m) a^2 = \text{constant}. \tag{96}$$

Then using the expression (89) we find that $a \propto mK^2k^2h^3$ and so finally we get that

$$\tilde{G}(m) = K^4 m^2 G(m) = \text{constant} \cdot h^{-9/2}. \tag{97}$$

The wave amplitude determined from (96, 97) is shown in Figure 4. As $m \to 0$, $G \propto 1$, $\tilde{G} \propto m^2$, and so $m \propto h^{-9/4}$, $a \propto h^{-1/4}$ which is the linear Green's law (Green, 1837) result. But, as $m \to 1$, $G \propto K^{-1}$, $\tilde{G} \propto K^3$, $a \propto h^{-1}$.

Wave set-up is found from (69, 95) and is given by

$$g\overline{\zeta} = -\frac{1}{h}\frac{\partial S}{\partial x}, \quad S = \frac{3\omega A}{2} = \frac{3G(m)E^*}{2}. \tag{98}$$

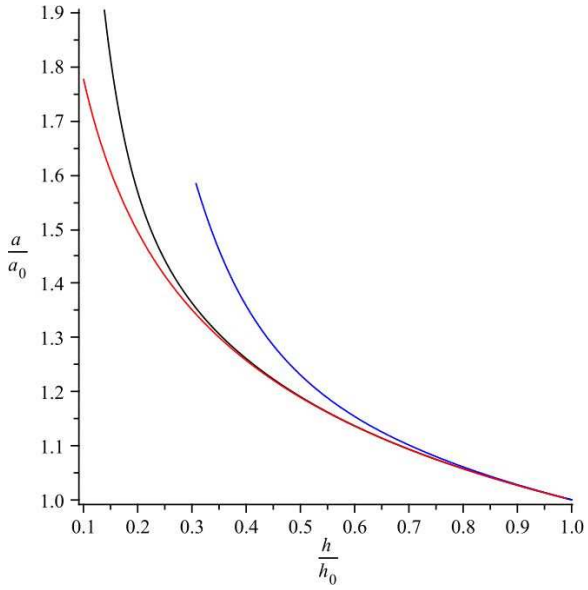


Figure 4: The wave amplitude is determined from (96, 97). The plots are for an initial modulus $m_0 = 0.1, 0.5$ (black, blue), while the linear solution $\zeta \propto h^{-1/4}$ is the red curve.

But since the wave frequency $\omega = kc_0$, $c_0 = (gh)^{1/2}$ and the wave action flux $c_0 A$ are conserved (see (96)), we readily find that

$$\overline{\zeta} = -\frac{a^2 G(m)}{4h} = -\frac{a_0 2h_0^{1/2} G(m_0)}{4h^{3/2}}, \tag{99}$$

This is just the linear law again, and is independent of how the wave amplitude varies. Note that for $a_0/h_0 \ll 1$, $m_0 \approx 0$, $G(m_0) \approx 1$.

References

[1] G. A. El, R. H. J. Grimshaw, and A. M. Kamchatnov, *Evolution of solitary waves and undular bores in shallow-water ows over a gradual slope with bottom friction*, J. Fluid Mech., 585 (2007), pp. 213–244.

[2] G. Green, *On the motion of waves in a variable canal of small depth and width*, Trans. Cambridge Philos. Soc., 6 (1837), pp. 457–462.

[3] R. H. J. Grimshaw, *Wave action and wave-mean flow interaction, with application to stratied shear flows*, Ann. Rev. Fluid Mech., 16 (1984), pp. 11–44.

[4] C. C. Mei, *The Applied Dynamics of Ocean Surface Waves*, Wiley-Interscience, New York, 1983.

[5] R. Miche, *Mouvements ondulatoires des mers en profundeur constante on décroisante form limte de la houle lors ds son déferlement. application aux digues marities*, Annales des Ponts et Chaussées, 114 (1944), pp. 25–78, 131–164, 270–292, 369–406.

[6] G. G. Stokes, *On the theory of oscillatory waves*, Trans. Cambridge Philos. Soc., 8 (1847), pp. 441–455.

[7] E. B. Thornton and R. T. Guza, *Energy saturation and phase speeds measured on a natural beach*, Journal of Geophysical Research, 84 (1982), pp. 9499–9508.

[8] ——, *Transformation of wave height distribution*, Journal of Geophysical Research, 88 (1983), pp. 5295–5938.

[9] G. B. Whitham, *Non-linear dispersion of water waves*, J. Fluid Mech., 27 (1967), pp. 399–412.

[10] ——, *Linear and Nonlinear Waves*, Wiley-Interscience, New York, 1974.

Lecture 19: Wave-Mean Flow Interaction, Part II: General Theory

Lecturer: Roger Grimshaw. Write-up: Michael Bates

25 June, 2009

1 Introduction

Suppose that we have a non-dissipative system that obeys a variational principle with a Lagrangian density, $L(\phi, \phi_t, \phi_{x_i}; t, x)$, where the field values are elements of the vector valued field $\phi(t, x_i)$, t is time and the spatial variables are x_i with $i = 1, 2, \ldots$. The governing equation of such a system is the Euler-Lagrange equation,

$$\frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \phi_t}\right) + \frac{\partial}{\partial x_i}\left(\frac{\partial L}{\partial \phi_{x_i}}\right) - \frac{\partial L}{\partial \phi} = 0\,, \tag{1}$$

where we have implied summation over the index, i. Using the calculus of variations, we can recover conservation laws which correspond to the symmetries of the Lagrangian. Energy conservation corresponds to a time symmetry, $\partial_t L = 0$, and momentum conservation corresponds to space symmetry, $\partial_{x_i} L = 0$. Wave action conservation corresponds to a phase symmetry.

If we seek a solution to describe waves superimposed on a mean flow, we introduce a phase parameter, θ, such that,

$$\phi(t, x_i, \theta + 2\pi) = \phi(t, x_i, \theta)\,. \tag{2}$$

As an example, small amplitude waves might have,

$$\phi(t, x_i) \approx a \sin(k_i x_i - \omega t + \theta)\,, \tag{3}$$

where k_i is the wavenumber vector, ω is the wave frequency and a is the wave amplitude.

Finally, we define a phase average by using the angle bracket notation,

$$\langle \cdot \rangle = \frac{1}{2\pi}\int_0^{2\pi} (\cdot) \mathrm{d}\theta\,. \tag{4}$$

Another common notation for denoting averages is the bar notation, $\bar{\cdot}$, which we will use interchangeably with the angle bracket notation, $\langle \cdot \rangle$. Note that all averaged quantities are independent of θ. Thus, we may write the entire function as a linear combination of the mean flow, $\langle \phi \rangle$, and a wave field, $\hat{\phi}$,

$$\phi = \langle \phi \rangle + \hat{\phi}\,, \tag{5}$$

where by definition, $\langle \hat{\phi} \rangle \equiv 0$.

2 Wave action

A useful consequence of the Euler-Lagrange equation (1) is that if we take any field with the same dimensions as ϕ, then we get the equation [4],

$$\frac{\partial}{\partial t}\left(\psi\frac{\partial L}{\partial \phi_t}\right)+\frac{\partial}{\partial x_i}\left(\psi\frac{\partial L}{\partial \phi_{x_i}}\right)=\psi_t\frac{\partial L}{\partial \psi_t}+\psi_{x_i}\frac{\partial L}{\partial \phi_{x_i}}+\psi\frac{\partial L}{\partial \phi}\,. \tag{6}$$

If we put $\psi=\phi_\theta$ and phase average we get,

$$\frac{\partial}{\partial t}\left\langle\hat{\phi}_\theta\frac{\partial L}{\partial \phi_t}\right\rangle+\frac{\partial}{\partial x_i}\left\langle\hat{\phi}_\theta\frac{\partial L}{\partial \phi_{x_i}}\right\rangle=0\,, \tag{7}$$

noting that the Lagrangian has no intrinsic dependence on the parameter θ. If we define the following quantities,

$$A=\langle\hat{\phi}_\theta\frac{\partial L}{\partial \phi_t}\rangle,\quad B_i=\langle\hat{\phi}_\theta\frac{\partial L}{\partial \phi_{x_i}}\rangle\,, \tag{8}$$

we can see that the corresponding phase symmetry is,

$$\frac{\partial A}{\partial t}+\frac{\partial B_i}{\partial x_i}=0\,. \tag{9}$$

A is the **wave action density** and B_i is the **wave action flux**. Physically, the wave action is the wave energy divided by the intrinsic wave frequency (which we shall define later). So, when a wave's frequency decreases (increases), the wave gains (loses) energy from (to) the mean flow in order to conserve wave action. Being wave quantities, if A and $\boldsymbol{B}$ are zero, then there are no waves. As such, wave action is a good measure of wave activity. Equation (9) is a conservation law in all unforced, non-dissipative systems (as is the case with the system we are considering). Formally, the wave action conservation law is valid without restriction on amplitude or on the relative time and space scales of the waves (with respect to the mean flow).

If we now suppose that the mean flow, background medium and wave parameters are slowly varying, we may write the wave field as,

$$\hat{\phi}\sim\hat{\phi}(S(t,x_i)+\theta;t,x_i)\,. \tag{10}$$

As we are assuming that S varies much more rapidly than the wave field's explicit dependence on t or x_i, the explicit derivatives $\partial\hat{\phi}/\partial x_i$ and $\partial\hat{\phi}/\partial t$ are small terms. Therefore, we can write

$$\hat{\phi}_t\sim\hat{\phi}_\theta\frac{\partial S}{\partial t}\quad\hat{\phi}_{x_i}\sim\hat{\phi}_\theta\frac{\partial S}{\partial x_i}\,, \tag{11}$$

noting that,

$$\frac{\partial S}{\partial t}=-\omega\quad\frac{\partial S}{\partial x_i}=\kappa_i\,, \tag{12}$$

where ω is the local frequency and κ_i is the local wavenumber. Substituting equations (12) into equations (11), we get

$$\hat{\phi}_t\sim-\omega\hat{\phi}_\theta\quad\hat{\phi}_{x_i}\sim\kappa_{x_i}\hat{\phi}_\theta\,. \tag{13}$$

Substitution into equations (8) yields

$$A = \left\langle \hat{\phi}_\theta \frac{\partial L}{\partial(-\omega\hat{\phi}_\theta)} \right\rangle \quad B = \left\langle \hat{\phi}_\theta \frac{\partial L}{\partial(\kappa_i\hat{\phi}_\theta)} \right\rangle . \tag{14}$$

Both A and B are $O(a^2)$, where a is the wave amplitude in the small amplitude limit, although, these equations are formally valid without any restriction on amplitude. It can be shown that

$$\left\langle \hat{\phi}_\theta \frac{\partial L}{\partial(-\omega\hat{\phi}_\theta)} \right\rangle = \hat{\phi}_\theta \hat{\phi}_\theta^{-1} \frac{\partial \langle L \rangle}{\partial(-\omega)}$$

$$\Rightarrow A = -\frac{\partial \bar{L}}{\partial \omega} , \tag{15}$$

where we note that ω is already an averaged quantity. Similarly for B_i, we get,

$$B_i = \frac{\partial \bar{L}}{\partial \kappa_i} . \tag{16}$$

where again we note that κ_i is an averaged quantity. Here $\bar{L} \equiv \langle L \rangle$, is the averaged Lagrangian and, similarly to ω and κ_i, is a slowly varying function of t and x_i.

Finally for slowly varying waves, we also require the equation for conservation of wave action,

$$\frac{\partial \kappa_i}{\partial t} + \frac{\partial \omega}{\partial x_i} = 0 \tag{17}$$

3 Linearised waves

We are now in a position where we can decompose the Lagrangian as follows,

$$L = L_0(\bar{\phi}_t, \bar{\phi}_{x_i}, \bar{\phi}; t, x_i) + L_1(\hat{\phi}_t, \hat{\phi}_{x_i}, \hat{\phi}; t, x_i) , \tag{18}$$

where $\bar{\phi} \equiv \langle \phi \rangle$. Here $L_0 = L(\bar{\phi}, \bar{\phi}_t, \bar{\phi}_{x_i}; t, x_i)$ is the Lagrangian for the mean flow and $L_1 = L(\hat{\phi}, \hat{\phi}_t, \hat{\phi}_{x_i}; tx_i)$ is the Lagrangian for the wave field. While the Lagrangian for the wave field depends on the Lagrangian for the mean flow, the dependence is not explicit but instead is present implicitly via the explicit dependence on t and x_i.

We may further simplify equation (10) by assuming a small amplitude,

$$\hat{\phi}(t, x_i) \approx a(t, x_i) \sin(S(t, x_i) + \theta) . \tag{19}$$

If we substitute into equation (18) and then average, we find that the average of the Lagrangian wave field is, to leading order, given by

$$\bar{L}_1 \approx D(\omega^*, \kappa_i; t, x_i) a^2 . \tag{20}$$

We note that there are no linear terms in a, as these terms average to zero, furthermore, the error is $O(a^4)$. In equation (20), we have extracted the dependence of the wave field on the mean field, where U_i is the **mean velocity field** and ω^* is the **intrinsic frequency**

given by $\omega^* = \omega - U_i\kappa_i$, while the rest of the fields remain suppressed in the explicit x_i and t dependencies. In the linearised approximation, the mean fields are known, with the result that $\bar{L}_1$ depends only on the mean velocity U_i through the intrinsic frequency, ω^*, which follows from Galilean invariance.

If we substitute equation (20) into our expressions for A and B, equations (15) and (16), we obtain,

$$A = -\frac{\partial D}{\partial \omega^*} a^2 \tag{21}$$

$$B = \frac{\partial D}{\partial \kappa_i} a^2 \,, \tag{22}$$

where we note that $(\partial D/\partial\omega^*)(\partial\omega^*/\partial\omega) = \partial D/\partial\omega^*$ as U_i is slowly varying. Furthermore, we note that the variation of $\bar{L}$ is independent of a, $\partial\bar{L}_1/\partial a = 0$. As a result of this latter relation, we find that the dispersion relation is,

$$\begin{aligned} &D(\omega^*, \kappa_i; t, x_i) 2a = 0 \\ &\therefore D(\omega^*, \kappa_i; t, x_i) = 0 \,. \end{aligned} \tag{23}$$

It follows, upon differentiation with respect to κ_i (remembering the definition of the intrinsic frequency) that,

$$\frac{\partial D}{\partial \omega^*} \boldsymbol{c}_{g_i} + \frac{\partial D}{\partial \kappa_i} = 0 \,, \tag{24}$$

where $\boldsymbol{c}_{g_i} = \partial\omega/\partial\kappa_i = U_i + \boldsymbol{c}_{g_i}^*$ is the **group velocity** and $\boldsymbol{c}_{g_i}^* = \partial\omega^*/\partial\kappa_i$ is the intrinsic group velocity. Use of equations (21) and (22) gives us the relation $B_i = \boldsymbol{c}_{g_i} A$, which, upon substitution into the wave action equation (9), yields the result,

$$\frac{\partial A}{\partial t} + \frac{\partial (\boldsymbol{c}_{g_i} A)}{\partial x_i} = 0 \,. \tag{25}$$

4 Energy and momentum

In order to provide a more physical interpretation of wave action, we consider the conservation laws for energy and momentum. Supposing that we have a full Lagrangian system governed by equation (1), $L(\phi, \phi_{x_s}; x_s)$ where $s = 0, 1, 2, 3$ and we note that $x_0 = t$ is a time-like variable. Recalling equation (6), we write,

$$\frac{\partial}{\partial x_s}\left(\psi \frac{\partial L}{\partial \phi_{x_s}}\right) = \psi_{x_s} \frac{\partial L}{\partial \phi_{x_r}} + \psi \frac{\partial L}{\partial \phi} \,.$$

If we now put $\psi = \phi_{x_r}$, we find a conservation law,

$$\begin{aligned} \frac{\partial}{\partial x_s}\left(\phi_{x_r} \frac{\partial L}{\partial \phi_{x_s}}\right) &= \frac{\partial \phi_{x_r}}{\partial x_s} \frac{\partial L}{\partial \phi_{x_r}} + \frac{\partial \phi}{\partial x_r} \frac{\partial L}{\partial \phi} \\ \frac{\partial}{\partial x_s}\left(\phi_{x_r} \frac{\partial L}{\partial \phi_{x_s}}\right) &= \frac{\mathrm{d}L}{\mathrm{d}x_r} - \left(\frac{\partial L}{\partial x_r}\right)_e \,, \end{aligned} \tag{26}$$

where $(\cdot)_e$ indicates that the derivative is taken keeping ϕ and ϕ_{x_r} fixed while,

$$\frac{\mathrm{d}L}{\mathrm{d}x_r} \equiv \frac{\partial L}{\partial x_r} + \phi_{x_r}\frac{\partial L}{\partial \phi} + \frac{\partial L}{\partial \phi_{x_s}}\frac{\partial \phi_{x_s}}{\partial \phi_{x_r}} \tag{27}$$

may be recognised as the total derivative. Furthermore, we note that $\frac{\mathrm{d}L}{\mathrm{d}x_r} = \delta_{rs}\frac{\mathrm{d}L}{\mathrm{d}x_r} = \frac{\partial}{\partial x_r}\delta_{rs}L$ in which δ_{rs} is the Kronecker delta symbol. This gives us the result

$$\begin{aligned}\frac{\partial}{\partial x_s}\left(\phi_{x_r}\frac{\partial L}{\partial \phi_{x_s}} - L\delta_{rs}\right) &= \frac{\partial L}{\partial x_r}\\ \frac{\partial T_{rs}}{\partial x_s} &= -\frac{\partial L}{\partial x_r}\,,\end{aligned} \tag{28}$$

where $T_{rs} = \phi_{x_r}\frac{\partial L}{\partial \phi_{x_s}} - L\delta_{rs}$ can be identified as the energy-momentum tensor from classical physics.

While the exact components of the tensor depend on the problem being studied, we can identify T_{00} as the energy density, T_{0j} as the energy flux and T_{i0} as the momentum density and T_{ij} as the corresponding fluxes. We may apply the averaging operator to the conservation law, equation (28), which gives us the averaged total energy $\langle T_{00}\rangle$ and averaged total momentum, $\langle T_{i0}\rangle$. These are not, on their own, particularly useful, as they contain both the mean and wave fields. Typically, the wave field is $O(a^2)$, and as such, is small relative to the mean field contribution.

We follow the averaging procedure of [2], by putting $\psi = \hat{\phi}_{x_s}$, recalling that $\phi = \bar{\phi} + \hat{\phi}$, and essentially follow the same procedure as for the full wave field to yield,

$$\frac{\partial T_{rs}}{\partial x_s} = -\frac{\partial \bar{L}_1}{\partial x_r} \tag{29}$$

where now

$$T_{rs} = \left\langle \hat{\phi}_{x_r}\frac{\partial L_1}{\partial \hat{\phi}_{x_s}} - \delta_{rs}L_1 \right\rangle . \tag{30}$$

We may now identify T_{00} as the pseudoenergy density, T_{0j} as the pseudoenergy flux and T_{i0} as the pseudomomentum density and T_{ij} is the pseudomomentum flux. Recalling that $\phi = \bar{\phi} + \hat{\phi}$, we can see that we are able to replace ϕ with $\hat{\phi}$ throughout as there is a linear relation between the two, with the value for $\bar{\phi}$ and its derivatives being slowly varying. We also recall that the total Lagrangian is a linear combination of the average and wave Lagrangians, $L = L_0 + L_1$. However, we note that L_0 is not a function of $\hat{\phi}$, and as such terms like $\partial L_0/\partial\hat{\phi} = 0$. These approximations make T_{rs} an $O(a^2)$ wave property (where we recall that a is the wave amplitude). Note, however, that equation (30) is not a conservation law unless L_1 is independent of x_s. As a consequence, the mean flow, $\bar{\phi}$ is also required to be independent of x_s.

If we put $\psi = \hat{\phi}_\theta$, we regain equation (8). If a suitable ergodic principle exists, we may identify the phase with a particular coordinate, $\theta = x_s$. This allows us to identify $T_{s0} = A$ and $T_{si} = B_i$. We note that the diagonal term T_{ss} is thus absent from the conservation law, equation (30). As such, we can now say that wave action is pseudoenergy for time averaging and is pseudomomentum for space averaging.

Generally speaking, wave energy is not as useful a quantity as wave action, since wave energy is not generally conserved. If we suppose that the mean flow consists of a mean velocity U_i, and a vector valued mean field, λ, that we require to satisfy the equation,

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \Lambda_{ij}\lambda\frac{\partial U_i}{\partial x_j} = 0\,, \tag{31}$$

where $\frac{\mathrm{d}}{\mathrm{d}t} \equiv \frac{\partial}{\partial t} + U_i\frac{\partial}{\partial x_i}$ is the material derivative and $\mathbf{\Lambda}$ is a dyadic of basis vectors. Physically, λ incorporates quantities such as mean depth, mean density or even mean magnetic field. We can now follow [3] and define the wave energy E as the pseudoenergy in a reference frame moving with the mean flow,

$$E = T_{00} + U_i T_{i0} \tag{32}$$

$$E = \left\langle \frac{\mathrm{d}\hat{\phi}}{\mathrm{d}t}\frac{\partial L_1}{\partial \hat{\phi}_t} - L_1 \right\rangle\,, \tag{33}$$

where we note that $T_{00} = \langle \frac{\partial\hat{\phi}}{\partial t}\frac{\partial L_1}{\partial\hat{\phi}_t} - L_1\rangle$ and $U_iT_{i0} = \langle U_i\hat{\phi}_{x_i}\frac{\partial L_1}{\partial\hat{\phi}_t}\rangle$ from equation (30). Similarly, the wave energy flux is given by,

$$F_i = \left\langle \frac{\mathrm{d}\hat{\phi}}{\mathrm{d}t}\frac{\partial L_1}{\partial \hat{\phi}_{x_i}} - U_iL_1 \right\rangle\,. \tag{34}$$

Here, we note that $L_1 = L_1(\hat{\phi}, \hat{\phi}_t, \hat{\phi}_{x_i}; U_i, \lambda; x_i, t)$. If we suppose that the dependence of E and $\boldsymbol{F}$ is solely through the material derivative, then it can be shown that

$$\frac{\partial E}{\partial t} + \frac{\partial F_i}{\partial x_i} = -R_{ij}\frac{\partial U_i}{\partial x_j} - \left(\frac{\mathrm{d}\bar{L}_1}{\mathrm{d}t}\right)_e \tag{35}$$

where $(\ldots)_e$ indicates the explicit derivative with respect to t and x_i while the wave field, $\hat{\phi}$ and the mean fields, U_i, λ are held constant and $\boldsymbol{R}$ is the **radiation stress tensor**,

$$R_{ij} = -T_{ij} + U_jT_{i0} - \Lambda_{ij}\lambda\frac{\partial\bar{L}_1}{\partial\lambda}\,. \tag{36}$$

The final piece of the puzzle is an equation for the mean flow, which is gained by variation of the mean field. In order to achieve this, we apply the averaging operator, equation (4), to the Lagrangian, equation (1), subject to the constraint described by equation (31),

$$\frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial U_i}\right) + \frac{\partial}{\partial x_j}\left(U_i\frac{\partial L_0}{\partial U_i} - \Lambda_{ij}\lambda\frac{\partial L_0}{\partial\lambda} + L_0\delta_{ij}\right) - \left(\frac{\partial L_0}{\partial x_i}\right)_e = -\frac{\partial R_ij}{\partial x_j} + \left(\frac{\partial\bar{L}_1}{\partial x_i}\right)_e\,. \tag{37}$$

This equation may also be derived using Whitham's variational principle. When $\mathbf{\Lambda}$ is isotropic, $\Lambda_{ij} = M\delta_{ij}$, then there is a mean pressure Q, such that equation (37) becomes,

$$-\frac{\partial R_{ij}}{\partial x_j} = -\frac{\partial T_{i0}}{\partial t} - \frac{\partial(U_jT_{i0})}{\partial x_j} + \frac{\partial Q}{\partial x_i}\,. \tag{38}$$

5 Slowly varying waves

In the circumstances where we have a slowly varying, almost-plane wave field (where the dependence on phase $S(x_i)$ is rapidly varying compared to the explicit dependence of the field on x_i), which can be reasonably approximated by equation (10). Recall that under such circumstances the frequency is given by $\omega = -\partial S/\partial t$ and wavenumber is given by $\kappa_i = \partial S/\partial x_i$. We also get useful reductions for pseudoenergy,

$$T_{00} \approx \omega A - \bar{L}_1 \tag{39}$$
$$T_{0i} \approx \omega B_i \,, \tag{40}$$

for pseudomomentum,

$$T_{i0} \approx -\kappa_i A \tag{41}$$
$$T_{ij} \approx -\kappa_i B_j - \bar{L}_1 \delta_{ij} \,, \tag{42}$$

for wave energy,

$$E \approx \omega^* A - \bar{L}_1 \tag{43}$$
$$F \approx \omega^* (B_i - U_i A) \,, \tag{44}$$

and the radiation stress tensor reduces to

$$R_{ij} \approx \kappa_i (B_j - U_j A) + \bar{L}_1 \delta_{ij} - \Lambda_{ij} \lambda \frac{\partial \bar{L}_1}{\partial \lambda} \,. \tag{45}$$

For linearised waves, we can achieve further results by recalling the dispersion relation, equation (23), which implies that the intrinsic frequency is $\omega^* = \Omega(\kappa_i; \lambda; x_i, t)$, and $\omega = \kappa_i U_i + \omega^*$. We also recall the wave action equation, equation (25) with the group velocity given by $\boldsymbol{c}_{g_i} = U_i + \partial\Omega/\partial\kappa_i$. Noting, that for linearised waves $\bar{L}_1 = 0$, which implies that the wave energy is given by $E = \omega^* A$ and the pseudoenergy is give by $T_{00} = \omega A$, giving new expressions for the wave energy equation (35) and the radiation stress equation (36),

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x_i}([U_i + \boldsymbol{c}_{g_i}^*]E) = -R_{ij}\frac{\partial U_i}{\partial x_j} + \frac{E}{\omega^*}\left(\frac{\mathrm{d}\Omega}{\mathrm{d}t}\right)_e \tag{46}$$
$$R_{ij} = A\left(\kappa_i \boldsymbol{c}_{g_j} + \Lambda_{ij}\lambda\frac{\partial \Omega}{\partial \lambda}\right) . \tag{47}$$

It is worth noting that there are some linear, almost plane waves for which this approximation does not hold, with one of the more notable exceptions being Rossby Waves.

6 Extensions

The theory outlined above can be extended to modal waves. That is, waves that are confined to a wave guide, such that the direction of propagation is in a reduced set of spatial dimensions and a modal structure in the remaining set of dimensions. Examples of such waves are water waves and internal waves.

The general theory outlined above can still be used, although we need to combine the Lagrangian averaging with integration across the waveguide.

In order to apply the theory outlined in the previous sections to fluid flows, we need to identify a suitable Lagrangian. Sometimes it is possible to find it directly from the Eulerian formulation, however, it is usually most convenient to find the Lagrangian using a Lagrangian formulation of the equations of motion.

In order to consider finite amplitude waves, it is best to use **generalised Lagrangian mean theory** (GLM) developed by [1, 2]. In GLM theory, we define the particle displacements from a mean position that moves with the mean velocity U_i. As such, x_i are Lagrangian variables moving with the Lagrangian mean velocity U_i, relative to which the particle displacements are defined as ξ_i. The Eulerian variables thus become $x_i' = x_i + \xi_i$ and the Eulerian velocity becomes

$$u_i' = U_i + \frac{\mathrm{d}\xi_i}{\mathrm{d}t}\,, \tag{48}$$

where we note that the material derivative is $\frac{\mathrm{d}}{\mathrm{d}t} \equiv \frac{\partial}{\partial t} + U_i \frac{\partial}{\partial x_i}$, and we note that $\langle \xi_i \rangle = 0$.

References

[1] D. G. Andrews and M. E. McIntyre, *An exact theory of nonlinear waves on a lagrangian mean flow*, Journal of Fluid Mechanics, 89 (1978), pp. 609–646.

[2] ——, *On wave action and its relatives*, Journal of Fluid Mechanics, 89 (1978), pp. 647–664.

[3] F. P. Bretherton and C. J. R. Garett, *Wavetrains in inhomogeneous moving media*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences., 302 (1968), pp. 529–554.

[4] W. D. Hayes, *Conservation of action and modal wave action*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences., 320 (1970), pp. 187–208.

Lecture 20: The explosive instability due to 3-wave or 4-wave mixing.

Lecturer: Harvey Segur. Write-up: Hélène Scolan and Alireza Mashayekhi

June 26, 2009

1 Introduction

This lecture builds on the foundations of Lecture 13 where we introduced the triad interaction process. We first recall the main properties of triad interactions, before discussing one of the important cases alluded to in Lecture 13, namely that of the explosive instability.

The dispersion relation $\omega(\mathbf{k})$ for linear waves propagating in any physical system can be obtained by linearization of the governing equations (we consider a non-dissipative system here). A resonant triad exists if $\omega(\mathbf{k})$ admits 3 pairs $\{\mathbf{k}, \omega(\mathbf{k})\}$ verifying :

$$\mathbf{k}_1 \pm \mathbf{k}_2 \pm \mathbf{k}_3 = 0 \quad \text{and} \quad \omega(\mathbf{k}_1) \pm \omega(\mathbf{k}_2) \pm \omega(\mathbf{k}_3) = 0 \ . \tag{1}$$

If this is not the case, the system follows the 4-wave equations describing a resonant quartet.

1.1 3-wave equations

In the case of a triad interaction, the 3-wave amplitude equations are obtained by using the method of multiple scales. For a single triad $\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\}$, the solution is written as:

$$u(\mathbf{x}, t; \epsilon) = \epsilon \left[\sum_{m=1}^{3} A_m(\epsilon\mathbf{x}, \epsilon t) \exp\{i\mathbf{k}_m \cdot \mathbf{x} - i\omega_m t\} + cc \right] + O(\epsilon^2) \ , \tag{2}$$

where the amplitude depends on the slow spatial and time variables. When (2) is substituted back into the governing nonlinear equations, a compatibility condition yields the amplitude evolution equations (see Lecture 13):

$$\frac{\partial A_m}{\partial \tau} + \mathbf{c}_m \cdot \nabla A_m = i\delta_m A_n^* A_l^* \ , \tag{3}$$

with $m, n, l = \{1, 2, 3\}$ and by assumption m, n and l are all different. These three coupled equations have been studied for capillary-gravity waves or in nonlinear optics using χ_2 material. But as seen in Lecture 13, the 3-wave resonance is impossible for pure gravity waves.

1.2 4-wave equations

When 3-resonance is impossible, we are forced to consider resonant quartets with $\mathbf{k}_1 \pm \mathbf{k}_2 \pm \mathbf{k}_3 \pm \mathbf{k}_4 = 0$ and $\omega(\mathbf{k}_1) \pm \omega(\mathbf{k}_2) \pm \omega(\mathbf{k}_3) \pm \omega(\mathbf{k}_4) = 0$.

A special case always exists where :

$$\{\mathbf{k} + \delta\mathbf{k}\} + \{\mathbf{k} - \delta\mathbf{k}\} - \mathbf{k} = \mathbf{k} \ , \tag{4}$$

$$\{\omega(\mathbf{k}) + \delta\omega\} + \{\omega(\mathbf{k}) - \delta\omega\} - \omega(\mathbf{k}) = \omega(\mathbf{k}), \tag{5}$$

i.e. the resonant quartet describes the interaction between waves with very close wave numbers and frequencies. In this case, as shown in Lecture 14, the equation for the slowly varying complex amplitude naturally recovers the nonlinear Schrödinger equation ('NLS') for the one wave-field (in 1-D):

$$i\left(\frac{\partial A}{\partial \tau} + c\frac{\partial A}{\partial X}\right) + \epsilon\left[\alpha\frac{\partial^2 A}{\partial X^2} + \gamma|A|^2 A\right] = 0, \tag{6}$$

where α is a real-valued constant obtained from $\omega(k)$, and γ is the interaction coefficient and is real (for a non-dissipative problem).

For more complex interactions, with waves modes interacting with themselves and with other modes, the evolution of the amplitudes are described by coupled NLS equations. In the case of 4-wave mixing in 1-D, we have [2] :

$$i\left(\frac{\partial A_m}{\partial \tau} + c_m\frac{\partial A_m}{\partial X}\right) + \epsilon\left[\alpha_m\frac{\partial^2 A}{\partial X^2} + A_m\sum_{n=1}^{4}\gamma_{mn}|A_n|^2 + \delta_m A_p^* A_q^* A_r^*\right] = 0, \tag{7}$$

with $\{m, p, q, r\} \in \{1, 2, 3, 4\}$ and $p, q, r \neq m$, c_m is the group velocity, $\{\gamma_{mn}\}$ are interaction coefficients and $\{\delta_m\}$ are real-valued coefficients of the 4-wave mixing terms.

These NLS equations are obtained for example in the case of gravity-driven surface waves and in nonlinear optics using χ_3 material.

1.3 Explosive instability

As discussed in Lecture 13 in the case of the 3-wave equations, Coppi, Rosenbluth & Sudan in 1969 [3] showed that if $\{\delta_1, \delta_2, \delta_3\}$ in equation (3) all have the same sign, then A_1, A_2, A_3 can all blow up at the same time, everywhere in space. This is called the *explosive instability*, appropriately describing the waves blowing up everywhere to infinity in finite time (as $(\tau - \tau_o)^{-1}$). This blow-up occurs even with no spatial structure (ie $A_m = A_m(\tau)$) so there is no wave collapse or self-focussing.

B. Safdi and H.Segur [7] found in 2007 the same possibility of explosive instability even in the case of the 4-wave equations where the 3-wave resonance is impossible. Both cases are now studied in more detail.

2 Explosive instability in ODEs

2.1 3-wave mixing

As seen in Lecture 13, the simplest model of 3-wave mixing ignores the spatial dependence of the interacting modes and the set of equations described in (3) reduces to three coupled

ordinary differential equations, with $\{\delta_1, \delta_2, \delta_3\}$ known and real-valued:

$$A_1'(\tau) = i\delta_1 A_2^* A_3^* \quad A_2'(\tau) = i\delta_2 A_1^* A_3^* \quad A_3'(\tau) = i\delta_3 A_1^* A_2^*. \tag{8}$$

It can be shown (see below) that if the δ_i have different sign, the solutions are bounded for all time.

Conditions for explosive instability. The following are necessary and sufficient conditions for blow-up in finite time:

- At least two of $\{A_1(0), A_2(0), A_3(0)\}$ are non-zero. Indeed, when at least two of them are zero at time $\tau = 0$ then the right-hand-sides are zero and nothing changes.
- $\{\delta_1, \delta_2, \delta_3\}$ all have the same sign and are non-zero [3].

These necessary condition are related to the existence of the constants of motion J_1, J_2 associated with the problem written in its Hamiltonian form (see Lecture 13):

$$J_1 = \frac{|A_1|^2}{\delta_1} - \frac{|A_3|^2}{\delta_3} \quad J_2 = \frac{|A_2|^2}{\delta_2} - \frac{|A_3|^2}{\delta_3}. \tag{9}$$

We see that if $\text{sign}(\delta_m) \neq \text{sign}(\delta_n)$ for any $m \neq n$, there is no blow-up. For instance, if $\delta_1 < 0$ and $\delta_3 > 0$, then $J_1 < 0$ and is constant, so both $|A_1|^2$ and $|A_3|^2$ are bounded and neither can blow up. But since J_2 is also constant then $|A_2|^2$ is bounded as well, so none of the wave fields can blow up. Thus it is necessary that $\{\delta_1, \delta_2, \delta_3\}$ all have the same sign and are non-zero for the amplitudes to blow-up.

Let us now suppose that $\{\delta_1, \delta_2, \delta_3\}$ all have the same sign and are non-zero. One shows by direct substitution into (8) that a special 3-parameter family of singular solutions to the governing ODEs is:

$$A_m(\tau) = \sqrt{\frac{\delta_m}{\delta_1\delta_2\delta_3}}\frac{e^{i\theta_m}}{\tau_o - \tau}, \quad \theta_1 + \theta_2 + \theta_3 = \frac{\pi}{2} + 2\pi N \text{ , for } m = 1, 2, 3, \tag{10}$$

where $\{\tau_o, \theta_1, \theta_2\}$ are 3 free real-valued parameters. This expression readily exhibits blow-up in the finite time τ_o. To find a general solution of the ODEs, we can assume that this special solution is the first term in a Laurent series in the neighbourhood of the pole τ_o and write:

$$A_m(\tau) = \sqrt{\frac{\delta_m}{\delta_1\delta_2\delta_3}}\frac{e^{i\theta_m}}{\tau_o - \tau}\left[1 + \alpha_m(\tau_o - \tau) + \beta_m(\tau_o - \tau)^2 + \gamma_m(\tau_o - \tau)^3 + O((\tau_o - \tau)^4\right]. \tag{11}$$

By substituting this ansatz into the original ODEs, the complex coefficients can be determined order by order to yield:

$$\alpha_m = 0, \quad \Im(\beta_m) = 0, \quad \Re(\beta_1 + \beta_2 + \beta_3) = 0, \quad \Re(\gamma_m) = 0$$

$$\Im(\gamma_1) = \Im(\gamma_2) = \Im(\gamma_3) = \gamma.$$

Thus, the general solution of the equation can be written:

$$A_m(\tau) = \sqrt{\frac{\delta_m}{\delta_1\delta_2\delta_3}}\frac{e^{i\theta_m}}{\tau_o-\tau}\left[1+\beta_m(\tau_o-\tau)^2+i\gamma(\tau_o-\tau)^3+O((\tau_o-\tau)^4\right] , \qquad (12)$$

where $\{\tau_o, \theta_1, \theta_2, \beta_1, \beta_2, \gamma\}$ are 6 free real-valued constants, and $\beta_1+\beta_2+\beta_3 = 0$. Note how every nontrivial solution of the ODEs near $\tau = \tau_o$ blows up at $\tau = \tau_o$.

Mathematically, the series converges absolutely if:

1. $\beta_1 = \beta_2 = 0, |\frac{\gamma(\tau_o-\tau)^3}{3}| < 1$ or
2. $|\beta_n| \le B, |\gamma| \le B^{3/2}, |\tau_o-\tau|^2 B < 1$.

Physical interpretation of blow-up. Recall that the theory of non-linear resonant interactions is a weakly nonlinear theory and assumes that the 3-wave equations evolve on the long time-scale ($t = O(\epsilon^{-1})$). In finite-time blow-up, both of these model assumptions break down as $\tau \to \tau_0$, and strongly nonlinear interactions must be taken into account. For instance in the case of vorticity waves, the outcome of the explosive instability can be wave breaking or intense vortex formation [8]. Note that additional information can be obtained from the model before it breaks down completely. For example, it can reveal that significant energy is transferred into the wave modes from a background source (pump wave in nonlinear optics, shear flows in fluid dynamics, etc..). Indication to the presence of such a background source is that all four δ coefficients should have the same sign.

2.2 4-wave mixing

The equivalent 4-wave mixing equations without spatial dependence can be written:

$$iA_m' + A_m\sum_{n=1}^{4}\gamma_{\mathrm{mn}}|A_n|^2 + \delta_m A_p^* A_q^* A_r^* = 0. \qquad (13)$$

These are four coupled complex ODEs with known real-valued coefficients.

The necessary and sufficient conditions for blow-up in finite time in this case are the following:

- At least three of $\{A_1(0), A_2(0), A_3(0), A_4(0)\}$ are non-zero.
- $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ all have the same sign and non-zero.
- $\left|\sum_{m,n=1}^{4}\gamma_{\mathrm{mn}}|\delta_n|\right| < 4\sqrt{\delta_1\delta_2\delta_3\delta_4}$.

In the same way as in the 3-wave mixing case, it can be found [7] that the general solution of (13) is an 8-parameter family of solutions that all blow up.

Open problem. So far there is no known physical example of an explosive instability caused by 4-wave mixing. As with 3-wave mixing, an explosive instability in a 4-wave system requires a background source of energy, from which the waves modes extract energy to grow in intensity together.

2.3 Effect of damping on the blow-up

It is interesting to introduce a damping component in the governing amplitude equations to see if the blow-up can be stopped by the damping. If we add damping in (8) for example we have:

$$A'_m(\tau) = i\delta_m A^*_p A^*_q - \nu_m A_m \quad \text{with} \quad \nu_m \geq 0. \tag{14}$$

Different situations occur depending on whether the ν_m coefficients are all the same (uniform damping) or differ (non-uniform damping).

2.3.1 Uniform damping

In the simple case of uniform damping $\nu_1 = \nu_2 = \nu_3 = \nu$, it can be shown [6] that the change of variables $T = \frac{1-e^{-\nu\tau}}{\nu}$, $A_m(\tau) = e^{-\nu\tau}\alpha_m(\tau)$ (with $m = 1, 2, 3$) transforms the equation (14) into the non-damping equation (16). This shows that uniform damping does not always stop the blow-up but instead introduces a threshold for the initial amplitudes below which blow-up is suppressed. However, if $\{\delta_1, \delta_2, \delta_3\}$ all have the same sign and $|A_m(0)| \geq \sqrt{\frac{\delta_m}{\delta_1\delta_2\delta_3}}\frac{\nu}{1-e^{-\nu\tau_0}}$, then the solutions still blow-up in finite time. A similar result can be found for the 4-wave equations.

2.3.2 Non-uniform damping

In this case, the general Laurent series for the solution is too restrictive since the singularity is not a pole, so one seeks instead a solution with:

$$A_m(\tau) = \sqrt{\frac{\delta_m}{\delta_1\delta_2\delta_3}}\frac{e^{i\theta_m}}{\tau_o - \tau}\left[1 + \alpha_m(\tau_o - \tau) + b_m(\tau_o - \tau)^2\ln(\tau_o - \tau) + \beta_m(\tau_o - \tau)^2 + ...\right] , \tag{15}$$

where the coefficients $\{\alpha_m\}$ and $\{b_m\}$ are real and depend on $\{\nu_n\}$, while the coefficients $\{\beta_m\}$ are real with $\{\beta_1, \beta_2\}$ free.

These ODEs are no longer completely integrable and the singularity is no longer a pole but blow-up persists for a 6-parameter family of solutions. The damping does not prevent blow-up, but as in the uniform case, introduces a threshold for blow-up.

3 Blow-ups in PDEs

The general set of amplitude equations for triad interactions are the following partial differential equations:

$$\frac{\partial A_m}{\partial \tau} + \mathbf{c}_m \cdot \nabla A_m = i\delta_m A^*_n A^*_l , \tag{16}$$

(for $m = 1, 2, 3$). Zakharov and Manakov [9] found that the 3-wave PDEs are completely integrable. Kaup [4] solved these equations in 1-D on whole line through numerical and analytical work to learn about blow-up. But what happens in the case of periodic boundary conditions or in more than one spatial dimension remains unknown.

Here, we use an alternative approach based on the general ODE solutions derived above:

$$A_m(\tau) = \sqrt{\frac{\delta_m}{\delta_1\delta_2\delta_3}}\frac{e^{i\theta_m}}{\tau_o - \tau}\left[1 + \beta_m(\tau_o - \tau)^2 + i\gamma(\tau_o - \tau)^3 + O((\tau_o - \tau)^4\right] , \tag{17}$$

where $\{\tau_o, \theta_1, \theta_2, \beta_1, \beta_2, \gamma\}$ are real-valued (and $\beta_1 + \beta_2 + \beta_3 = 0$). To take into account the spatial dependency, we consider a variation of the parameters, looking for $\beta_1 = \beta_1(\mathbf{x})$, $\beta_2 = \beta_2(\mathbf{x})$ and $\gamma = \gamma(\mathbf{x})$:

$$A_m(\mathbf{x}, \tau) = \sqrt{\frac{\delta_m}{\delta_1 \delta_2 \delta_3}} \frac{e^{i\theta_m}}{\tau_o - \tau} \left[1 + \beta_m(\mathbf{x})(\tau_o - \tau)^2 + \{g_m(\mathbf{x}) + i\gamma(\mathbf{x})\}(\tau_o - \tau)^3 + O((\tau_o - \tau)^4\right], \tag{18}$$

with $\{\beta_1(\mathbf{x}), \beta_2(\mathbf{x}), \gamma(\mathbf{x})\}$ real-valued and arbitrary while the $\{g_m(\mathbf{x})\}$ are real-valued and known (in terms of first derivatives of the $\beta_m(\mathbf{x})$ functions). Note that again we must also have $\beta_1 + \beta_2 + \beta_3 = 0$.

If we allow an additional spatial dependency for θ_1 and θ_2, then one finds that the solutions of the PDEs can be written:

$$\begin{aligned} A_m(\mathbf{x}, \tau) = \sqrt{\frac{\delta_m}{\delta_1 \delta_2 \delta_3}} \frac{e^{i\theta_m(\mathbf{x})}}{\tau_o - \tau} \big[1 &+ i a_m(\mathbf{x})(\tau_o - \tau) \\ &+ \{\beta_m(\mathbf{x}) + i b_m(\mathbf{x})\}(\tau_o - \tau)^2 \\ &+ \{g_m(\mathbf{x}) + i\gamma(\mathbf{x})\}(\tau_o - \tau)^3 + ...\big], \end{aligned} \tag{19}$$

where the functions $a_m(\mathbf{x})$ involve first derivatives of $\{\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \theta_3(\mathbf{x})\}$, the functions $b_m(\mathbf{x})$ involve first and second derivatives of the same functions, and finally the functions $g_m(\mathbf{x})$ also involve first and second derivatives of the $\{\theta_m(\mathbf{x})\}$ as well as first derivatives of $\beta_1(\mathbf{x})$ and $\beta_2(\mathbf{x})$. Thus this family of formal solutions of the PDEs admits five free functions of $\mathbf{x}$. These functions must be infinitely differentiable, but are otherwise arbitrary, and can be members of any desired function space.

A (formal) general solution of (16), with six functions that must be real-valued and smooth, but are otherwise arbitrary, is:

$$\begin{aligned} A_m(\mathbf{x}, \tau) = \sqrt{\frac{\delta_m}{\delta_1 \delta_2 \delta_3}} \frac{\rho_m(\mathbf{x}) e^{i\theta_m(\mathbf{x})}}{\tau_o(\mathbf{x}) - \tau} \big[1 &+ i a_m(\mathbf{x})(\tau_o - \tau) \\ &+ \{\beta_m(\mathbf{x}) + i b_m(\mathbf{x})\}(\tau_o - \tau)^2 \\ &+ \{g_m(\mathbf{x}) + i\gamma(\mathbf{x})\}(\tau_o - \tau)^3 + ...\big], \end{aligned} \tag{20}$$

with the ρ_m functions verifiying:

$$\rho_1^2(\mathbf{x}) = (1 - \mathbf{c}_2 \cdot \nabla\tau_0(\mathbf{x})) \cdot (1 - \mathbf{c}_3 \cdot \nabla\tau_0(\mathbf{x})), \tag{21}$$

$$\rho_2^2(\mathbf{x}) = (1 - \mathbf{c}_3 \cdot \nabla\tau_0(\mathbf{x})) \cdot (1 - \mathbf{c}_1 \cdot \nabla\tau_0(\mathbf{x})), \tag{22}$$

$$\rho_3^2(\mathbf{x}) = (1 - \mathbf{c}_1 \cdot \nabla\tau_0(\mathbf{x})) \cdot (1 - \mathbf{c}_2 \cdot \nabla\tau_0(\mathbf{x})). \tag{23}$$

There are still open questions regarding the convergence of the series in this case, and the constraints on the free functions for blow-up.

References

[1] M. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*, SIAM Philadelphia, 1981.

[2] D. Benney and A. Newell, *The propagation of nonlinear wave envelopes*, J. Math. Phys, (1967).

[3] B. Coppi, M. Rosenbluth, and R. N. Sudan, *Nonlinear interactions of positive and negative energy modes in rarefied plasmas.*, Annals of Physics, 55 (1969), pp. 207–270.

[4] D. Kaup and A. Newell, *An exact solution for a derivative nonlinear schrödinger equation*, Journal of Mathematical Physics, 19 (1978), pp. 798–801.

[5] J. Manley and H. Rowe, *Some general properties of nonlinear elements-part i. general energy relations*, Proceedings of the IRE, 44 (1956), pp. 904–913.

[6] J. W. Miles, *Nonlinear Faraday resonance*, Journal of Fluid Mechanics, 146 (1984), pp. 285–302.

[7] B. R. Safdi and H. Segur, *Explosive instability due to four-wave mixing*, Physical Review Letters, 99 (2007), p. 245004.

[8] V. Shrira, V. Voronovich, and N. Kozhelupova, *Explosive Instability of Vorticity Waves*, JJ. Phys. Oceanogr., 27 (1997), p. 542???554.

[9] V. E. Zakharov and S. V. Manakov, *The theory of resonance interaction of wave packets in nonlinear media*, Soviet Journal of Experimental and Theoretical Physics, 42 (1975), p. 842.

Lecture 21: Potpourri

Lecturer: H. Segur. Write-up: Ali Mashayek and Hélène Scolan

June 26, 2009

1 Introduction

In this last lecture of the 2009 GFD series, a collection of topics is presented including:
A. Near-Shore processes such as "Wave Shoaling" and"Rip Currents".
B. Rogue waves.
C. Wave breaking in shallow water.

2 Wave shoaling

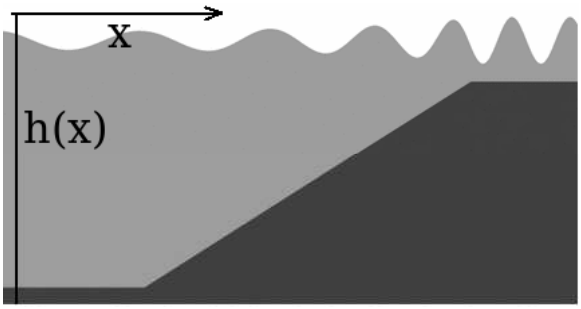


Figure 1: Wave shoaling

Wave shoaling refers to the change in the amplitude of surface waves as they travel in a water layer of decreasing depth (e.g. as they approach the coast). According to the 2-D shallow-water wave equation with x-dependent depth, we have (see Lecture 8)

$$\partial_t^2 \eta = \nabla \cdot \{g\, h(x) \nabla \eta\}, \tag{1}$$

$$c(x) = \sqrt{g\, h(x)}, \tag{2}$$

where η is the displacement of the free surface, $c(x)$ is the local phase speed and $h(x)$ is the local depth of the water layer. As a long wave in shallow water travels into even shallower water (as shown in Figure 1), the front of the wave feels the decrease in the depth sooner than the back of the wave. Correspondingly, the phase speed of the wave front decreases while the back is still traveling at a faster speed. This phenomenon compresses the wave as it moves to the shallower region, as shown in the Figure. This compression also naturally

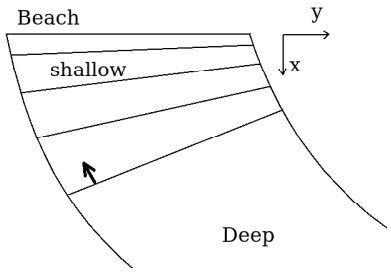


Figure 2: Wave crests turning towards the beach as they approach it.

leads to an increase in the wave amplitude, the combination of these two effects being referred to as wave shoaling.

The dependence of phase speed on the water depth also explains why wave crests line up parallel to the coast as they approach a beach. Consider a sloping beach, such as the one shown in Figure 2, and incoming waves with crests making an arbitrary angle with the coast. As they move into the shallow water region, waves begin to feel the effect of the bottom boundary and react by adjusting their phase speed: the part of the wave which is closer to the coast (the right sides of the waves shown in the Figure) slows down sooner than the other end of the wave crest (left side). Consequently, the left side of the wave catches up with the right side and the wave crest progressively turns towards the coast. This phenomenon owes its existence to the inhomogeneity of the medium (due to change in depth) and is similar to wave refraction in optics (i.e. the bending of light due to density change along a ray-path). Figure 3 illustrates this process near the coast of Duck, NC.

Figure 3: Waves near the coast at Duck, NC (1991). Photo by C. Miller. The photo was taken at a time when Hurricane Grace was passing 100 miles off-shore.

3 Rip currents

A rip current is a narrow jet which carries water away from shore. They typically form in the region of breaking waves, and extend somewhat beyond the breaking region. They are observed on ocean and sea beaches, and even on beaches of large lakes. Figure 4 shows rip

currents formed in Rosarita beach, Baja California. These currents can be very dangerous, dragging swimmers (and non-swimmers who are standing in water) away from the beach and drowning them as they attempt to fight the current. In this scenario, the best action is to try getting out of the current by swimming in the cross-current direction.

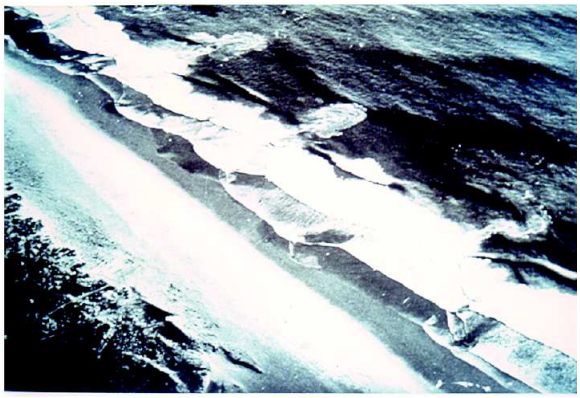


Figure 4: Rip currents along the coast of Rosarita beach, Baja California.

Various theories have been proposed to explain the mechanisms responsible for the formation and evolution of rip currents. The standard explanation is that the backwash water associated with wave breaking needs to find its way back into the sea. Although a part of it returns through undercurrents, water is also pushed sideways along the shoreline looking for an exit into the sea. Thus, a rip current can form in a trench between sandbars as shown schematically in Figure 5. Several issues have been raised regarding this sandbar

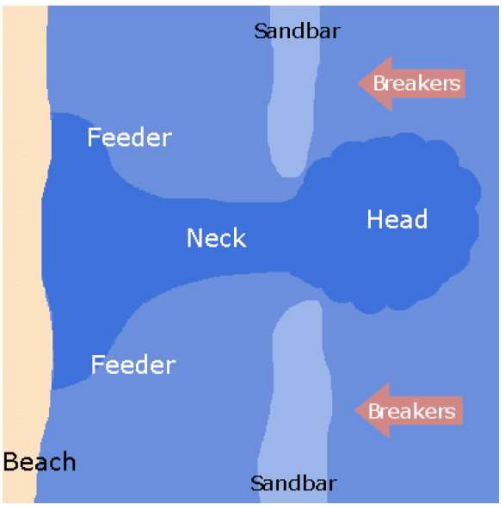


Figure 5: A rip current formed between two sandbars. Figure from http://www.ripcurrents.noaa.gov/science.html

theory. Observations have shown that these currents can appear in a spatially periodic fashion along a coastline. Figure 6 shows one such example in Sand City, California. As

there is no a priori reason for the existence of a periodic array of sandbars in the shallow depth along the shoreline, the sandbar theory alone cannot explain the pattern seen in the Figure. Another problem with the sandbar theory is the temporal evolution of the rip currents. Some shorelines are characterized by permanent rip currents which form at a fixed location such as a break in a reef or other hard structures. Meanwhile, some currents are also observed to migrate along a stretch of coastline. Some rip currents are persistent, lasting for many days or months in one location and some are ephemeral, forming quickly and lingering for a few hours or days before dissipating. All these facts are somehow in contradiction with the sandbar theory.

Figure 6: Rip currents along the coast of Sand City, California.

An alternative explanation for the rip currents might be the following. Recall our description of KP solutions of genus 2 in water of uniform depth (see Lecture 7) and the associated experimental results of Hammack *et al.*(1995) [4] (shown in Figure 7). In these solutions (and experimental results) some areas have smaller velocities (see for example the regions marked by arrows in Figure 7). If such a pattern reaches the shore and breaks, the circled regions would be good candidates for the backwash water to return to the ocean and form rip currents. Some experiments by Hammack, Scheffner & Segur (1991) [5] have shown the formation of rip currents in these regions, visualising them using dye injection.

The KP solutions (of genus 2) used in the rip current experiment of Hammack *et al.*(1995) [4] were symmetric, and propagated normal to the sloping beach. More general KP solutions (of genus 2) are not symmetric and would therefore not propagate normal to the beach (as shown in lower panel of Figure 7). This raises the interesting question of whether such waves could generate slowly migrating rip currents. One can also ask the question of "How far can the incoming waves deviate from symmetry before the rip currents disappear?"

To answer these questions and explain rip currents in a more systematic manner one has to create a suitable mathematical model to describe them. Such model should allow for variable depth (uniform slope will do), wave breaking and return flows, and finally 3-D motion in order to describe the currents properly. Once a good mathematical model exists,

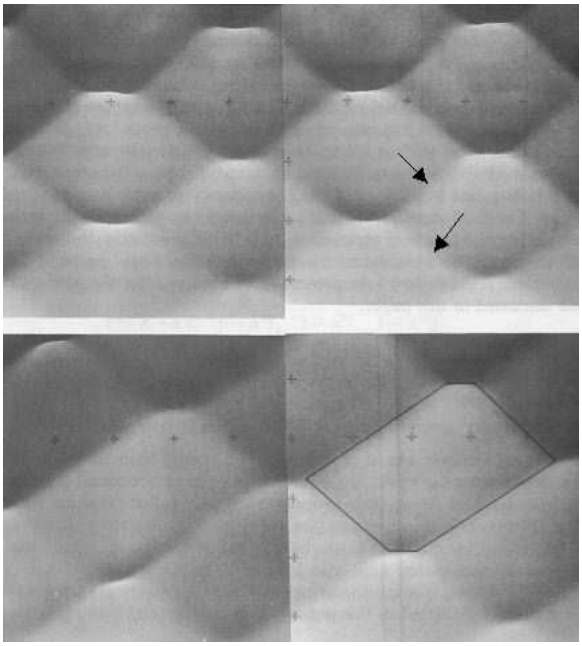


Figure 7: Two 2-dimensional surface patterns of nearly permanent form, propagating in shallow water.(Hammack *et al.*, 1995), [4]

Harry Yeh's tank (at Oregon State University) would be a suitable place to test the theory.

4 Rogue waves

Rogue waves are large ocean surface waves which appear spontaneously and are usually unexpectedly large compared with the typical wave amplitude in their environment. Figure 8 shows an example of a rogue wave breaking over a supertanker in a storm off Durban, S.A. in 1980. These waves are also referred to as freak waves, monster waves, killer waves, or extreme waves. They are extremely dangerous due to their size, and are thought to be responsible for many ship accidents and disappearances in the oceans. Figure 9 shows a map of ship accidents for the period between 1995 and 1999; from Lloyd's Marine Information Service (LMIS) database, two ships per week are lost at sea due to heavy weather. This highlights the importance of understanding the physics behind the formation of rogue waves, with the hope of predicting their occurrence (e.g. location and time).

Because of their rarity, rogue waves are almost impossible to measure in a systematic manner. Their existence was originally inferred from the damages inflicted to ships. Today, one of the only solid scientific measurement of a rogue wave is the sea-surface height data from the Draupner oil platform, in the North Sea on January 1, 1995, showing the passage of an 18.5m wave (see Figure 10). The damages inflicted on the platform during this event confirmed the validity of the measurement.

There seems to be no obvious reason behind the much larger amplitude of the rogue wave compared with the nearby waves, thus raising the conceptual question " Is a rogue wave a rare event from a known population or is it an element of an entirely different population?"

Figure 8: Rogue wave breaking over a supertanker in a storm off Durban, S.A. in 1980. The mast seen starboard stands 25 m above the mean sea level. The mean wave height at the time was 5-10 m. Photo: Phillipe Lijour, first mate.

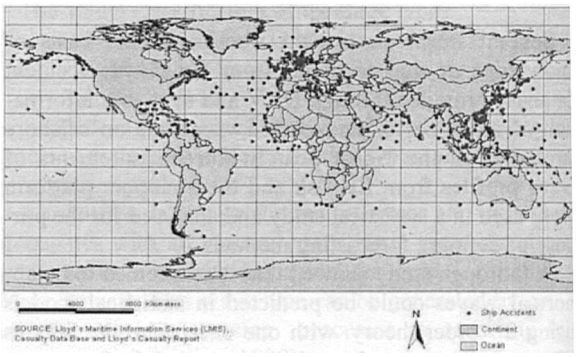


Figure 9: World-wide ship accidents.

Several possible mechanisms have been proposed to explain the occurrence of these waves. One possibility, brought up by Smith (1976) [7] and supported by White & Fornberg (1998) [8] and Baschek (2005) [2], is that rogue waves occur as a result of the interaction between internal or surface waves and underlying currents. A second possibility could be wave breaking: satellite radar measurements have shown that extreme waves events can result from wave breaking. Two other candidates are the geometric focusing and frequency focusing of wave energies at a single point, leading to enhanced wave amplitudes and the formation of a rogue wave. A more general candidate is that the rogue waves are formed due to strongly nonlinear wave dynamics. Studies such as Bateman, Swan & Taylor (2003) [3] try to show that rare extreme events such as rogue waves can occur, and may in fact not be that rare.

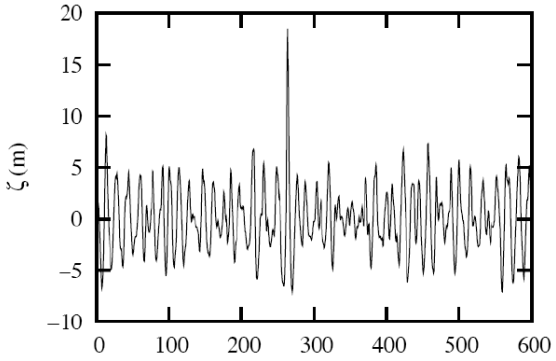


Figure 10: Waves measured at the Draupner oil platform in the North Sea, Jan. 1, 1995. Peak wave height: 18.5 m. Standard deviation of wave record: about 3 m.

In short, very little is known about the rogue waves. From an observational point of view, what information needs to be gathered to learn more about these waves and how that information can be sought, both remain to be determined. From a theoretical point of view, the general and fundamental question is whether one should try to learn more about the tail of our known distribution of ocean waves or whether one should instead look for a new kind of mechanism.

5 Waves breaking in shallow water

In general, a wave usually breaks once its amplitude grows to a critical level beyond which some process causes a large fraction of the wave energy to be dissipated. The physical models discussed thus far generally become invalid (the underlying assumptions break down) in this limit, particularly those which assume linear or weakly nonlinear behavior.

The shallow-water equations, for example, are hyperbolic, implying that waves may break in shallow water, and they do. In this case, the question is how should these equations be modified to describe more realistically wave breaking in shallow waters. Perhaps a dissipative term might help modeling "shock waves" or a dispersive term might help modeling "collisionless shocks" etc,. Both internal and surface waves can break, close or far from the coastlines. The water surface waves breaking at the coastline is perhaps the most familiar scenario. As the waves approach the shore, they refract and adjust themselves parallel to the shoreline, and their amplitude grows due to the shoaling effect explained earlier. Once their amplitude is sufficiently large, the waves overturn and break. However, note that waves can also break in the mid-ocean where refraction and shoaling effects are not present.

There are some basic types of breaking water waves like spilling, collapsing, surging and plunging [1]. As an example, Figure 11 shows a "plunging breaking" which occurs as the crest of the wave curls over and crashes down into the base of the wave. Plunging breaking is a dramatic form of break, usually associated with a sudden depth change (in the mid- or coastal ocean) or a steep sea floor near the shore.

Dramatic wave breaking events can have catastrophic consequences in the coastal areas.

Figure 11: A "plunging breaker". Photo from Clark Little/SWNS.

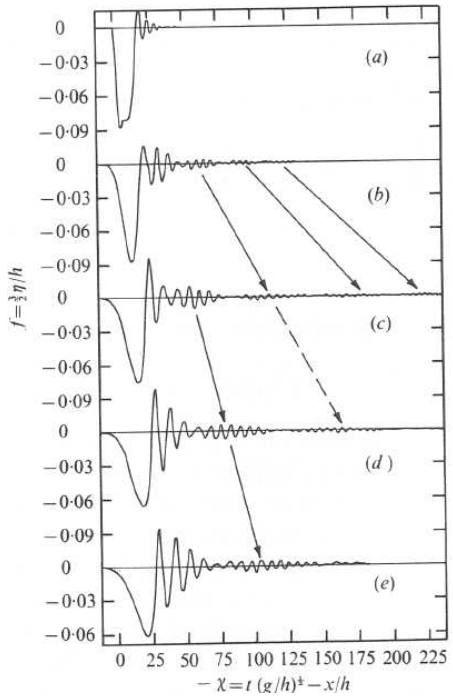


Figure 12: Hammacks experiments in shallow water, from Hammack *et al.*(1978) [6]

Recall Hammack's experiments in shallow water which led to formation of an "Undular bore" (shown in Figure 12, see also Lecture 5). As the bore travels from right to left (in the Figure), the coast would experience a huge suction of the water into the ocean (which can attract curious people into the ocean following the lost water) followed by a train of large waves which break down at the shore and lead to tragedies such as the 2004 tsunami in Thailand, as discussed in Lecture 7. In summary, although the shallow-water equations are similar to the equations of gas dynamics in 2-D, breaking water waves seem to be more complicated than ordinary shock waves in gas dynamics and how to model wave breaking properly remains an open question.

References

[1] *Shore Protection Manual*, U. S. Govt. Print. Off., 1973.

[2] B. Baschek, *Wave-current interaction in tidal fronts*, 14th 'Aha Huliko'a Winter Workshop 2005: Rogue Waves, ed. P. Muller and C. Garrett., (2005).

[3] W. Bateman, C. Swan, and P. Taylor, *On the calculation of the water particle kinematics arising in a directionally spread wavefield*, J. of Computational Physics, (2003).

[4] J. Hammack, D. Mccallister, N. Scheffner, and H. Segur, *Two-dimensional periodic waves in shallow water. part 2. asymmetric waves*, J. Fluid Mechanics, 285 (1995).

[5] J. Hammack, N. Scheffner, and H. Segur, *A note on the generation and narrowness of rip currents*, J. Geophys. Res., 96 (1991).

[6] J. L. Hammack and H. Segur, *The korteweg de vries equation and water waves. part 3. oscillatory waves*, J. Fluid Mechanics, 84 (1978).

[7] R. Smith, *Giant waves*, J. Fluid Mechanics, (1976).

[8] B. White and B. Fornberg, *On the chance of freak waves at sea*, J. Fluid Mechanics, (1998).

Resonant triad interactions on an extended equatorial β-plane

Michael Bates Advisor: Roger Grimshaw

June-August, 2009

1 Introduction and Motivation

Due to the low value of the vertical component of the Earth's rotation the tropics possess dynamical properties that are quite different in nature to other regions of higher latitude. One property is that the equator acts as a wave guide due to the changing in sign of the Coriolis parameter.

Linear equatorial wave theory (Matsuno, 1966) has shown that equatorially trapped waves (particularly Rossby and Kelvin waves) are important waves in determining large scale climatic processes such as the El Niño-Southern Oscillation (ENSO; e.g. Philander, 1990; Clarke, 2008). Other factors that determine properties like ENSO periodicity and strength are stochastic processes (Kleeman, 2008). There are, however, nonlinear equatorial processes that are well observed and are thought to impact on phenomena such as ENSO, but many of these processes remain relatively poorly understood from a theoretical point of view.

In the rest of this section, we shall examine previous studies on nonlinear equatorial waves. We shall also review a class of waves that are highly nonlinear in nature and have been observed for several decades now, however, recent insights by linear resonance theory have shown to give insights into their dynamics. The end of this section briefly reviews the properties of the equatorial β-plane, and the consequences for theoretical models of using such an approximation.

Section 2 examines the hydrostatic, inviscid, Boussinesq and rigid-lid momentum and continuity equations on a sphere and then transforms them to the Mercator projection. In section 3 we examine the linear shallow water equations in geopotential coordinates, but, we use some inspiration from the Mercator projection by using a tanh profile for the coriolis parameter (the "extended β-plane"), rather than the traditional linear β-plane approximation.

The modal equations and dispersion relations for the extended β-plane are then derived and investigated in section 4. Finally, section 5 examines the effects of weak nonlinear resonance using a multiscale expansion to derive the coupling coefficints for a system of three baroclinic Rossby waves interacting. Finally, a summary and a discussion of potential future works is presented in section 6.

1.1 Previous Studies of Equatorial Nonlinear Wave Dynamics

There is a large literature on linear equatorial wave dynamics. The nonlinear literature, on the other hand, is somewhat smaller. Furthermore, most of the studies have used either a single layer, or reduced gravity model with only a handfull of studies including baroclinic waves. Herewith, we shall give a brief overview of the classes of nonlinearities and waves that have been studies.

Boyd (1980b) studied the nonlinear Kelvin wave. The linear Kelvin wave is a dispersionless wave, and as such, nonlinear effects are very small. Using the method of strained coordinates, he showed how nonlinearity can cause frontogenesis, an alteration in the phase speed and breaking. Boyd (1980a) examined how long equatorial Rossby waves can be described by the Korteweg-de Vries equation (Korteweg and de Vries, 1895) or the modified Korteweg-de Vries equation. Boyd (1983) examines how highly dispersive waves (short Rossby waves, the Yanai wave and inertia-gravity waves) can be described by the nonlinear Schrödinger equation and propagate as a solitary wave packets of permanent form.

More recently Le Sommer et al. (2004) have shown that there is a dynamical split between fast nonlinear waves (fast Yanai and inertia-gravity waves) and slow nonlinear waves (slow Rossby and Kelvin waves).

Ripa (1983a,b) examines nonlinear resonance in a one layer reduced gravity model in the equatorial β-plane using the method of Ripa (1981). In those studies, the various types of interactions are classified into 19 categories and the properties of waves in the various categories are investigated.

Recently, it has been shown in that nonlinear resonance between two equatorially trapped baroclinic Rossby waves and one free barotropic Rossby wave are possible (Reznik and Zeitlin, 2006, 2007a,b).

1.2 An example of observed Nonlinear Equatorial Waves

Tropical Instability Waves (TIWs; Düing et al., 1975; Legeckis, 1977) are an equatorial wave in which the dynamics are still relatively poorly understood, despite the fact that they are a dominant feature of the monthly variability of the equatorial Pacific and Atlantic. In addition, TIWs are climatologically important, as it is thought that TIWs are closely associated with dynamics of the El Niño-Southern Oscillation (Yu and Liu, 2003)

TIWs are seen as cusps on the front north equatorial front (the front that delineates the equatorial cold tongue; see figure 1). TIWs are seen on the southern front as well, but the amplitudes are typically much smaller than on the northern side. It has been shown in the Pacific that there are two distinct periods of 17 and 33 days (Lyman et al., 2007), although the shorter period wave does not appear to exist in the Atlantic (von Schuckmann et al., 2008).

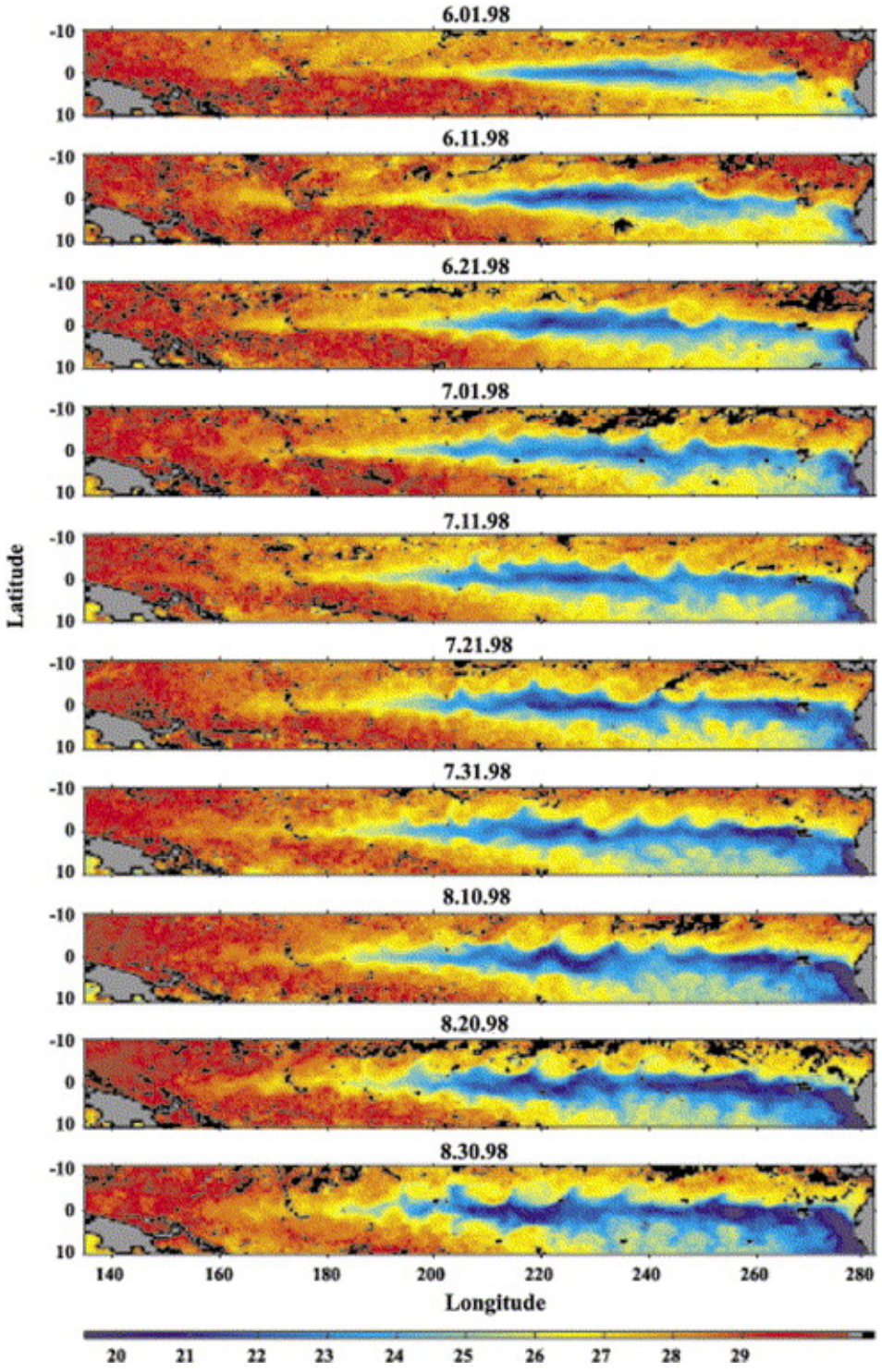


Figure 1: Sea Surface temperature signature of tropical instability waves,seen in imagery from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager, from Willett et al. (2006)

Various process studies have indicated that the waves are caused by instability in the intense zonal mean flow (for instance Philander, 1976, 1978; Proehl, 1996). Lawrence and Angell (2000) suggested that modification of the phase of TIWs by Rossby waves helps to explain some of the meridional asymmetry displayed by TIWs. More recently, it has been suggested that TIWs could be described as a linear unstable resonance between two baroclinic equatorial Rossby waves (Lyman et al., 2005). While linear analysis has offered some insights into the dynamics of TIWs, it is thought that these waves are nonlinear in nature (Philander, 1978; Kennan and Flament, 2000).

1.3 The Equatorial β-plane

The equatorial β-plane has been used extensively in the study of equatorial wave dynamics. The equatorial β-plane is not universally valid, particularly for barotropic waves (Boyd, 1985).

While the validity of using linear equations for the barotropic waves is somewhat limited, previous studies examining nonlinear resonance at the equator have avoided purely baroclinic systems. The reason stated by Reznik and Zeitlin (2007a) is that because of the rapid decay of the parabolic cylinder function (the function that describes the meridional structure of many equatorially trapped waves) for baroclinic waves, the nonlinear interaction of baroclinic waves is also small. On the other hand, the large meridional extent of barotropic waves allows for greater interaction.

2 Equations on a Sphere and the Mercator Projection

2.1 Equations on a Sphere

We begin by noting that in spherical coordinates, where θ is latitude, ϕ is longitude and r is the radial distance from the centre of the Earth, the divergence operator and gradient operators are (see chapter 4.12 of Gill, 1982),

$$\nabla \cdot \boldsymbol{F} \equiv \frac{1}{\cos\theta}\left\{\frac{\partial}{\partial\phi}\left(\frac{F_\phi}{r}\right)+\frac{\partial}{\partial\theta}\left(\frac{F_\theta\cos\theta}{r}\right)+\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2F_r\cos\theta\right)\right\} \tag{1}$$

$$\nabla\varrho \equiv \left(\frac{\partial_\phi\varrho}{r\cos\theta},\frac{\partial_\theta\varrho}{r},\partial_r\varrho\right), \tag{2}$$

where $\boldsymbol{F}$ is some arbitrary vector and ϱ is an arbitrary scalar. As the radial length scales are much smaller than the lateral length scales, we may treat r as a constant, a_r, unless we are taking a derivative with respect to r. The divergence and gradient operators become (to a good approximation for the oceans),

$$\nabla \cdot \boldsymbol{F} \approx \frac{\partial_\phi F_\phi}{a_r\cos\theta}+\frac{\partial_\theta(F_\theta\cos\theta)}{a_r\cos\theta}+\partial_r F_r \tag{3}$$

$$\nabla\varrho \approx \left(\frac{\partial_\phi\varrho}{a_r\cos\theta},\frac{\partial_\theta\varrho}{a_r},\partial_r\varrho\right). \tag{4}$$

We can thus say that the inviscid, Boussinesq, hydrostatic lateral momentum equation in spherical coordinates is,

$$\boldsymbol{u}_{,t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} + \left[f_0 + \frac{u}{a_r \cos\theta} \right] \boldsymbol{e}_r \wedge \boldsymbol{u} \sin\theta = -\frac{\nabla p}{\rho}\,, \tag{5a}$$

where we have ignored the horizontal component of the Earth's rotation and the radial metric terms, both of which follow naturally from assuming that the radial length scale is much smaller than the lateral. Here, $\boldsymbol{u} = (u, v)$ is the zonal and meridional velocity and the gradient operator only operates in two dimensions in the momentum equation and $\boldsymbol{e}$ indicates a unit vector. Furthermore, we note that f_0 is the value of the radial component of the Earth's rotation vector (the Coriolis parameter) at the North Pole. We also note that $\nabla \boldsymbol{u} \equiv (\nabla u, \nabla v)$, where ∇ is that defined in equation (2). The terms are, from left to right, the rate of change of velocity, the advective term, the Coriolis term, a metric term and the pressure gradient term. As we are considering an incompressible fluid, the continuity equation becomes,

$$\nabla \cdot (\boldsymbol{u}, w) = 0\,, \tag{5b}$$

where w is the radial velocity and the gradient operator is acting in all three directions. Written in their full form, the lateral momentum equations and the continuity equation is,

$$u_{,t} + \frac{u u_{,\phi}}{a_r \cos\theta} + \frac{v u_{,\theta}}{a_r} - v f_0 \sin\theta - \frac{uv \tan\theta}{a_r} = -\frac{p_{,\phi}}{\rho a_r \cos\theta} \tag{6a}$$

$$v_{,t} + \frac{u v_{,\phi}}{a_r \cos\theta} + \frac{v v_{,\theta}}{a_r} + u f_0 \sin\theta - \frac{u^2 \tan\theta}{a_r} = -\frac{p_{,\theta}}{\rho a_r} \tag{6b}$$

$$\frac{u_\phi}{a_r \cos\theta} + \frac{(v \cos\theta)_{,\theta}}{a_r \cos\theta} + w_{,r} = 0\,, \tag{6c}$$

where we have used the comma notation to denote partial derivatives (see Aris, 1962).

2.2 The Mercator Projection

We now set about transforming the equations from spherical coordinates to Mercator coordinates. Firstly, we note that the unit vectors for each coordinate system are equivalent,

$$\boldsymbol{e}_x = \boldsymbol{e}_\phi \tag{7a}$$

$$\boldsymbol{e}_y = \boldsymbol{e}_\theta \tag{7b}$$

$$\boldsymbol{e}_z = \boldsymbol{e}_r\,, \tag{7c}$$

Here, x is the zonal distance from a reference $x = 0$ and y is the meridional distance from $y = 0$ (the equator), such that $-\infty < y < \infty$ while, $-\pi/2 < \theta < \pi/2$. In addition, z is the radial distance from the bottom of the ocean, a_r. Spherical coordinates and Mercator coordinates are thus related by,

$$x = a_r \phi \tag{8a}$$

$$y = a_r \tanh^{-1}(\sin\theta) \tag{8b}$$

$$z = r - a_r\,, \tag{8c}$$

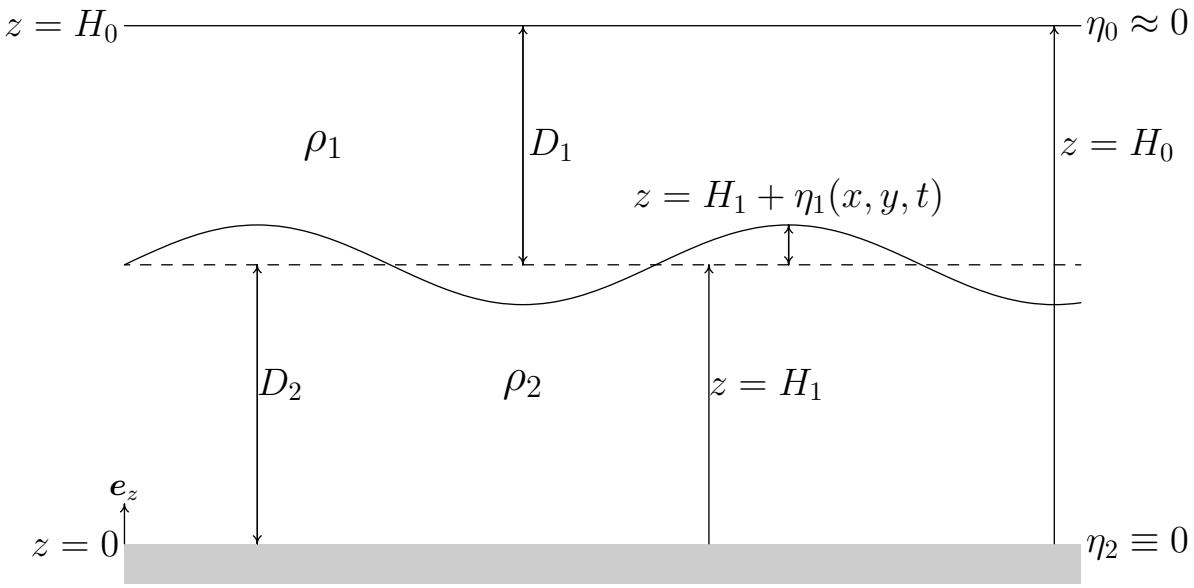


Figure 2: The shaded region is the (constant depth) ocean floor, the dashed line indicates mean sea level and the solid line represents the free surface. H is distance from mean sea level to the ocean floor, and $\eta(x, y, t)$ is the deviation of the free surface from mean sea level.

which essentially projects the sphere onto a cylinder of height $2a_r$. Applying this transform to the spherical equations (6) yields,

$$u_{,t} + \frac{uu_{,x}}{\operatorname{sech}(y/a_r)} + \frac{vu_{,y}}{\operatorname{sech}(y/a_r)} - vf_0 \tanh(y/a_r) - \frac{uv \tanh(y/a_r)}{a_r \operatorname{sech}(y/a_r)} = -\frac{p_{,x}}{\rho \operatorname{sech}(y/a_r)} \tag{9a}$$

$$v_{,t} + \frac{uv_{,x}}{\operatorname{sech}(y/a_r)} + \frac{vv_{,y}}{\operatorname{sech}(y/a_r)} + uf_0 \tanh(y/a_r) - \frac{u^2 \tanh(y/a_r)}{a_r \operatorname{sech}(y/a_r)} = -\frac{p_{,y}}{\rho \operatorname{sech}(y/a_r)} \tag{9b}$$

$$[u \operatorname{sech}(y/a_r)]_{,x} + [v \operatorname{sech}(y/a_r)]_{,y} + w_{,z} \operatorname{sech}^2(y/a_r) = 0 \tag{9c}$$

where we note from equations (8) that $x_{,\phi} = a_r$, $y_{,\theta} = \frac{a_r}{\cos\theta}$, $z_{,r} = 1$ $\sin\theta = \tanh(y/a_r)$ and $\cos\theta = \operatorname{sech}(y/a_r)$. If we now make the substitution $\mathcal{U} = u \operatorname{sech}(y/a_r)$ and $\mathcal{V} = v \operatorname{sech}(y/a_r)$ (which is equivalent to the $\cos\theta$ substitution made by Longuet-Higgins, 1965),

$$\operatorname{sech}^2(y/a_r)\mathcal{U}_{,t} + \mathcal{U}\mathcal{U}_{,x} + \mathcal{V}\mathcal{U}_{,y} - \operatorname{sech}^2(y/a_r)\mathcal{V}f_0 \tanh(y/a_r) = -\frac{p_{,x}}{\rho} \operatorname{sech}^2(y/a_r) \tag{10a}$$

$$\operatorname{sech}^2(y/a_r)\mathcal{V}_{,t} + \mathcal{U}\mathcal{V}_{,x} + \mathcal{V}\mathcal{V}_{,y} + \frac{(\mathcal{U}^2 + \mathcal{V}^2)\tanh(y/a_r)}{a_r} + \operatorname{sech}^2(y/a_r)\mathcal{U}f_0 \tanh(y/a_r) = -\frac{p_{,y}}{\rho} \operatorname{sech}^2(y/a_r) \tag{10b}$$

$$\mathcal{U}_{,x} + \mathcal{V}_{,y} + \operatorname{sech}^2(y/a_r)w_{,z} = 0 \tag{10c}$$

If we now examine these equations for a two layer fluid (which is illustrated in figure 2), where we label each layer with $i = 1, 2$, 1 being the top layer and 2 being the lower layer

and η_0 being the free surface. In both layers we use the hydrostatic approximation,

$$p_{,z} = -\rho g\,, \tag{11}$$

If we also employ the rigid lid approximation, we write the pressure for the top layer as $P = rho_1 g\eta_0$ and we ignore the deviations of the surface elsewhere in the equations, $\eta_0 = 0$. Furthermore, linearising gives a more simple expression for the momentum equation of the top layer,

$$\mathcal{U}_{1,t} - \mathcal{V}_1 f_0 \tanh(y/a_r) = -P_{,x} \tag{12a}$$

$$\mathcal{V}_{1,t} + \mathcal{U}_1 f_0 \tanh(y/a_r) = -P_{,y}\,. \tag{12b}$$

In the bottom layer we make the Boussinesq approximation in addition to the same approximations to the top layer,

$$\mathcal{U}_{2,t} - \mathcal{V}_2 f_0 \tanh(y/a_r) = -(P_{,x} + g'\eta_{1,x}) \tag{12c}$$

$$\mathcal{V}_{2,t} + \mathcal{U}_2 f_0 \tanh(y/a_r) = -(P_{,y} + g'\eta_{1,y})\,. \tag{12d}$$

Here, the Boussinesq approximation means that we have assumed that $\rho_1/\rho_2 \approx 1$. Note that we have written $g' = g(\rho_2 - \rho_1)/\rho_2$ to represent the "reduced gravity". Integration of the continuity equation across each layer, assuming that the layers are immiscible, the bottom is impermeable, $\eta_2 = 0$ and that deviations of the interface between the fluids from its resting state are small $w(\eta_i) \approx \partial_t \eta_i$, we obtain the conservation equation for each layer,

$$-\operatorname{sech}^2(y/a_r)\eta_{1,t} + (D_1\mathcal{U}_1)_{,x} + (D_1\mathcal{V}_1)_{,y} = 0 \tag{13a}$$

$$\operatorname{sech}^2(y/a_r)\eta_{1,t} + (D_2\mathcal{U}_2)_{,x} + (D_2\mathcal{V}_2)_{,y} = 0\,. \tag{13b}$$

To arrive at these expressions, we note that the rigid lid approximation implies that, $\eta_{0,t} \ll \eta_{1,t}$ and $\eta_0 \ll D_1$; and the small amplitude assumption implies that $\eta_1 \ll D_1$ and $\eta_1 \ll D_2$.

It is important to note that one of the important factors in the validity of the linearization, is that the equatorial Rossby radius of deformation (Gill and Clarke, 1974) must be small when compared to the radius of the Earth, $\mathrm{Ro} \ll a_r$. We note that this is true for baroclinic modes, whose wave speed is 0.5–3.0ms^{-1}, however, it is only marginally true for the barotropic modes, whose wave speed is $\sim 200\mathrm{ms}^{-1}$ (chapter 11.5 of Gill, 1982), and as such, the barotropic results must be interpreted very carefully.

If we now define a mean velocity $\mathscr{U}$ and a shear velocity $\mathscr{R} = (\mathcal{R}, \mathcal{S})$,

$$H_0\mathscr{U} = D_1(\mathcal{U}_1, \mathcal{V}_1) + D_2(\mathcal{U}_2, \mathcal{V}_2) \tag{14a}$$

$$\mathscr{R} = (\mathcal{U}_2 - \mathcal{U}_1, \mathcal{V}_2 - \mathcal{V}_1)\,. \tag{14b}$$

We may combine equations (12) and (13),

$$\mathscr{U}_{,t} + f_0 \tanh(y/a_r)\boldsymbol{e}_z \wedge \mathscr{U} + \frac{\nabla(P + g'\eta_1)}{H_0} = 0 \tag{15a}$$

$$\mathscr{R}_{,t} + f_0 \tanh(y/a_r)\boldsymbol{e}_z \wedge \mathscr{R} + c^2\nabla h = 0 \tag{15b}$$

$$\nabla \cdot \mathscr{U} = 0 \tag{15c}$$

$$\operatorname{sech}^2(y/a_r)h_{,t} + \nabla \cdot \mathscr{R} = 0 \tag{15d}$$

where $c^2 = g'D_1D_2/H_0$, $h = H_0\eta_1/(D_1D_2)$. We note that c is the baroclinic phase speed. Employment of the rigid lid approximation implies that the mean velocity field is non-divergent, and as such, we may define a stream function, ψ,

$$\mathscr{U} = \left(-\frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial x}\right). \tag{16}$$

We now assume that the solutions to equations (15) are separable, having the form

$$\psi(x,y,t) = \Psi(y)\left[Ae^{i\varphi} + A^*e^{-i\varphi}\right] \tag{17a}$$
$$\mathcal{R}(x,y,t) = \Xi(y)\left[iAe^{i\varphi} - iA^*e^{-i\varphi}\right] \tag{17b}$$
$$\mathcal{S}(x,y,t) = \Phi(y)\left[Ae^{i\varphi} + A^*e^{-i\varphi}\right] \tag{17c}$$
$$h(x,y,t) = G(y)\left[iAe^{i\varphi} - iA^*e^{-i\varphi}\right], \tag{17d}$$

where $\varphi = kx - \omega t$, noting that k is the wavenumber and ω is the frequency. Here A is the amplitude of the wave and A^* is the complex conjugate of the amplitude.

Using the definition of the streamfunction, equation (16), and taking a combination of equations (15a) and (15c), we obtain an equation for the evolution of the streamfunction,

$$\nabla^2\psi_{,t} + f'(y)\psi_{,x} = 0, \tag{18}$$

where $f(y) = f_0\tanh(y/a_r)$. We note that the barotropic equation is uncoupled from the baroclinic equations.

To find the baroclinic modal equation we may take various combinations of equations (15b) and (15d) to eliminate $\mathcal{R}$, and then applying the solutions in equation (17), we find two modal equations,

$$-\omega^2\operatorname{sech}^2(y/a_r)G + kf(y)\Phi + c^2k^2G - \omega\Phi_{,y} = 0 \tag{19a}$$
$$-k\omega\Phi + f(y)\omega\operatorname{sech}^2(y/a_r)G - f(y)\Phi_{,y} - kc^2G_{,y} = 0 \tag{19b}$$

Rearranging equation (19a) in terms of $G(y)$, we get,

$$G = \frac{\omega\Phi_{,y} - kf\Phi}{c^2k^2 - \omega^2\operatorname{sech}^2(y/a_r)}. \tag{20}$$

Differentiating with respect to y, we find

$$G_{,y} = \frac{k(f'\Phi + f\Phi_{,y}) - \omega\Phi_{,yy}}{\omega^2\operatorname{sech}^2(y/a_r) - c^2k^2} + \frac{1}{a_r}(kf(y)\Phi - \omega\Phi_{,y})\frac{2\omega^2\tanh(y/a_r)\operatorname{sech}^2(y/a_r)}{(\omega^2\operatorname{sech}^2(y/a_r) - c^2k^2)^2}. \tag{21}$$

We note that under the assumption that the wave length must be much smaller than the radius of the Earth, the second term in equation (21) will be small relative to other terms, and as such, we shall ignore it. Substitution of equations (20) and (21) into equation (19b) yields a modal equation in Φ,

$$\Phi_{,yy} + \left\{\operatorname{sech}^2(y/a_r)\left[\frac{\omega^2}{c^2} - \frac{f^2(y)}{c^2}\right] - k^2 - \frac{kf'(y)}{\omega}\right\}\Phi = 0. \tag{22}$$

We wish to nondimensionalize the barotropic and baroclinic modal equations by introducing the length and time scales,

$$L_* = \left(\frac{c}{\beta}\right)^{\frac{1}{2}} \tag{23a}$$

$$T_* = \frac{1}{\beta L_*}\,. \tag{23b}$$

where we have noted that the equatorial Rossby radius of deformation is an appropriate length scale. Associated with these length scales are the dimensionless variables,

$$\omega = \frac{\tilde{\omega}}{T_*} \tag{24a}$$

$$k = \frac{\tilde{k}}{L_*} \tag{24b}$$

$$f_0 = \beta L_* \tilde{f}_0 \tag{24c}$$

We also note that as we are considering the case close to the equator (see chapter 3.17 of Pedlosky, 1987),

$$a_r \approx \frac{f_0}{\beta}\,. \tag{25}$$

Substitution into equation (18) for the barotropic case and equation (22) for the baroclinic case yield, respectively,

$$\Psi_{,\tilde{y}\tilde{y}} - \left\{\tilde{k}^2 + \frac{\tilde{k}\tilde{f}'(\tilde{y})}{\tilde{\omega}}\right\}\Psi = 0 \tag{26a}$$

$$\Phi_{,\tilde{y}\tilde{y}} + \left\{\mathrm{sech}^2(\tilde{y}/\tilde{f}_0)\left[\tilde{\omega}^2 - \tilde{f}(\tilde{y})\right] - \tilde{k}^2 - \frac{\tilde{k}\tilde{f}'(\tilde{y})}{\tilde{\omega}}\right\}\Phi = 0\,. \tag{26b}$$

Using the Mercator projection as a motivation for a tanh profile for the Coriolis parameter, we examine the more familiar two layer shallow water equations in geopotential coordinates.

3 The β-Plane and "Extended" β-Plane

We begin by stating the inviscid, linear shallow water equations for a two layer fluid in geopotential coordinates,

$$\boldsymbol{u}_{i,t} + f(y)\boldsymbol{e}_z \wedge \boldsymbol{u}_i + \frac{\nabla p_i}{\rho_i} = 0 \tag{27}$$

$$\eta_{i-1,t} - \eta_{i,t} + \nabla\cdot([D_i + \eta_{i-1} - \eta_i]\boldsymbol{u}_i) = 0\,,\,, \tag{28}$$

for $i = 1, 2$. Here, the pressure, p_i, is given by,

$$p_1 = P \tag{29a}$$

$$p_2 = P + g'\eta_1\,. \tag{29b}$$

Similarly with section 2.2 we employ the rigid lid approximation, which implies that $P = g\rho_1\eta_0$ but that $\eta_0 = 0$ anywhere else it appears in an equation. We also utilise the Boussinesq approximation, assuming that $\rho_1/\rho_2 \approx 1$.

Again, as in section 2.2, we define a mean velocity, $\boldsymbol{U} = (U, V)$ and a shear velocity, $\boldsymbol{R} = (R, S)$,

$$H_0\boldsymbol{U} = D_1\boldsymbol{u}_1 + D_2\boldsymbol{u}_2 \tag{30a}$$

$$\boldsymbol{R} = \boldsymbol{u}_2 - \boldsymbol{u}_1\,. \tag{30b}$$

Applying these definitions to the momentum equation (27), and continuity equation (28) in conjunction with the rigid lid approximation and small amplitude assumption gives,

$$\boldsymbol{U}_{,t} + f(y)\boldsymbol{e}_z \wedge \boldsymbol{U} + \nabla Q = 0 \tag{31a}$$

$$\boldsymbol{R}_{,t} + f(y)\boldsymbol{e}_z \wedge \boldsymbol{R} + c^2\nabla h = 0 \tag{31b}$$

$$\nabla \cdot \boldsymbol{U} = 0 \tag{31c}$$

$$h_{,t} + \nabla \cdot \boldsymbol{R} = 0\,, \tag{31d}$$

where, $Q = P + g'\eta_1$, $h = H_0\eta_1/(D_1D_2)$ and $c^2 = g'D_1D_2/H_0$. We may take various combinations of equations (31) and assume solutions of the same form as in equations (17) to get a barotropic equation for Ψ,

$$\Psi_{,yy} - \left\{k^2 + \frac{kf'(y)}{\omega}\right\}\Psi = 0 \tag{32}$$

and a baroclinic equation for Φ

$$\Phi_{,yy} + \left\{\frac{\omega^2}{c^2} - \frac{f^2(y)}{c^2} - k^2 - \frac{kf'(y)}{\omega}\right\}\Phi = 0\,. \tag{33}$$

Through the use of equations (31), we can find expressions for the modal functions for Ξ and G,

$$G(y) = \frac{\omega\Phi_{,y} - kf(y)\Phi}{c^2k^2 - \omega^2} \tag{34a}$$

$$\Xi(y) = \frac{c^2k\Phi_{,y} - \omega f(y)\Phi}{c^2k^2 - \omega^2}\,. \tag{34b}$$

We may non-dimensionalize the modal equations (32) and (33) using the definitions in equations (23),

$$\Psi_{,\tilde{y}\tilde{y}} - \left\{\tilde{k}^2 + \frac{\tilde{f}'(\tilde{y})\tilde{k}}{\tilde{\omega}}\right\}\Psi = 0 \tag{35}$$

$$\Phi_{,\tilde{y}\tilde{y}} + \left\{\tilde{\omega}^2 - \tilde{f}(\tilde{y})^2 - \tilde{k}^2 - \frac{\tilde{k}\tilde{f}'(\tilde{y})}{\tilde{\omega}}\right\}\Phi = 0\,. \tag{36}$$

Furthermore, using the scaling

$$G = \frac{\widetilde{G}}{c}\,, \tag{37}$$

we may non-dimensionalize the modal expressions for h and R,

$$\widetilde{G}(\tilde{y}) = \frac{\tilde{\omega}\Phi_{,\tilde{y}} - \tilde{k}\tilde{f}(\tilde{y})\Phi}{\tilde{k}^2 - \tilde{\omega}^2} \tag{38a}$$

$$\Xi(y) = \frac{\tilde{k}\Phi_{,\tilde{y}} - \tilde{\omega}\tilde{f}(\tilde{y})\Phi}{\tilde{k}^2 - \tilde{\omega}^2}\,. \tag{38b}$$

The traditional β-plane approximation locally approximates the effects of the curvature of the Earth by assuming a plane surface (thus ignoring the geometric effects), but allows the coriolis parameter to vary linearly with latitude. The traditional equatorial β-plane is generally written as $f(y) = \beta y$. In dimensionless form the coriolis parameter is $\tilde{f}(\tilde{y}) = \tilde{y}$, $\tilde{f}'(\tilde{y}) = 1$. We can see that in this case, we may assume a solution of the form $\tilde{\psi} = e^{i(\tilde{k}\tilde{x}+\tilde{l}\tilde{y}-\tilde{\omega}\tilde{t})}$, which gives a modal equation,

$$\tilde{\omega} = \frac{-\tilde{k}}{\tilde{k}^2 + \tilde{l}^2}\,. \tag{39}$$

Importantly, the modal equation for the β-plane has no trapped wave solutions, unlike the modal equation (26a) for the Mercator projetion. We note also that with the β-plane approximation, the solution for the meridional modal function is a parabolic cylinder function (see Miller, 1965), yielding a dispersion relation of the form,

$$\tilde{\omega}_m^3 - (\tilde{k}^2 + 2m + 1)\tilde{\omega}_m - \tilde{k} = 0 \quad m = 0, 1, 2, \ldots\,. \tag{40}$$

Both equations are common in the literature and specifically, are in agreement with Reznik and Zeitlin (2006, 2007a,b), as is equations (38) when the β-plane approximation is substituted. We further note that there is an additional solution for the trapped waves, corresponding to $\Phi \equiv 0$, which yields a dispersion relation of $\tilde{\omega} = \tilde{k}$. This solution is the equatorial Kelvin wave (Matsuno, 1966).

Another special case of equation (40) is that when $m = 0$. In this instance, there are three branches. One is spurious, yielding the unphysical relation $\tilde{\omega} = -\tilde{k}$, which is unbounded. The other two branches are described by $\tilde{\omega}(\tilde{\omega} - k) = 1$. This dispersion relation corresponds to the Yanai wave (Matsuno, 1966).

In addition to the traditional β-plane approximation, there is also the δ-plane approximation (Yang, 1987) which is a second order Taylor expansion of the Coriolis parameter. Our aim in the present study is to use the full expression for the coriolis parameter, but, to ignore the metric terms (and thus, ignore geometric effects).

4 Meridional Structure and Dispersion Relations

We begin by deriving the properties of the system discussed in section 3.

4.1 Solving for the Meridional structure

Taking the barotropic modal equation (35), and using our chosen coriolis profile, $\tilde{f}(\tilde{y}) = \tanh(\widetilde{Y})$, where $\widetilde{Y} = \tilde{y}/\tilde{f}_0$, we find that the modal equation becomes,

$$\Psi_{,\widetilde{Y}\widetilde{Y}} + \left\{\alpha_{\mathrm{T}} + \gamma_{\mathrm{T}}(1-\lambda^2)\right\}\Psi = 0 \tag{41}$$

where, $\lambda = \tanh(\widetilde{Y})$ and

$$\alpha_{\mathrm{T}} = -\tilde{k}^2\tilde{f}_0^2 \tag{42a}$$

$$\gamma_{\mathrm{T}} = -\tilde{f}_0^2\tilde{k}/\tilde{\omega}\,. \tag{42b}$$

If we use the chain rule, $\Psi_{,\widetilde{Y}} = \Psi_{,\lambda}\lambda_{,\widetilde{Y}}$, noting that $\lambda_{,\widetilde{Y}} = 1-\lambda^2$, we transform equation (41) to,

$$\left\{(1-\lambda^2)\Psi_{,\lambda}\right\}_{,\lambda} + \left(\gamma_{\mathrm{T}} + \frac{\alpha_{\mathrm{T}}}{1-\lambda^2}\right)\Psi = 0\,. \tag{43}$$

Using the same procedure, the baroclinic modal equation (36) becomes,

$$\left\{(1-\lambda^2)\Phi_{,\lambda}\right\}_{,\lambda} + \left(\gamma_{\mathrm{C}} + \frac{\alpha_{\mathrm{C}}}{1-\lambda^2}\right)\Phi = 0\,, \tag{44}$$

where,

$$\alpha_{\mathrm{C}} = \tilde{f}_0^2(-\tilde{f}_0^2 - \tilde{k}^2 + \tilde{\omega}^2) \tag{45a}$$

$$\gamma_{\mathrm{C}} = -\tilde{f}_0^2(\tilde{f}_0^2 + \tilde{k}/\tilde{\omega})\,, \tag{45b}$$

noting that the subscript $_{\mathrm{T}}$ indicates barotropic and subscript $_{\mathrm{C}}$ indicates baroclinic. Note that subsequent statements regarding α and γ that do not explicitly indicate whether it is discussing the barotropic or baroclinic case are equally valid for the barotropic and baroclinic variables.

With the chosen form of the coriolis parameter, $\tilde{f}(\tilde{y})$, we now have the same differential equation for the barotropic and baroclinic equations, which are second order ordinary differential equations, namely the associated Legendre equation.

In the case where $\alpha < 0$ we have discrete eigenvalues given by (Drazin and Johnson, 1989)

$$\alpha_m = -\left\{(\gamma + 1/4)^{\frac{1}{2}} - (m + 1/2)\right\}^2 \quad m = 0, 1, 2, \ldots, N\,. \tag{46}$$

We note that from its definition, equation (42a), that α_{T} is always less than zero, and as such the barotropic modes are always trapped. On the other hand, α_{C} is not necessarily zero, and it has trapped modes only when $\tilde{\omega}^2 < \tilde{f}_0^2 + \tilde{k}^2$ (see section 4.3).

The solutions for the batotropic case are proportional to associated Legendre functions (see, for example Stegun, 1965),

$$\Psi(\lambda) = C_\Psi P_\nu^\mu(\lambda) + D_\Psi Q_\nu^\mu(\lambda)\,, \tag{47}$$

where $P_\nu^\mu(\lambda)$ and $Q_\nu^\mu(\lambda)$ are the associated Legendre functions of the first and second kinds respectively. C_Ψ and D_Ψ are constants of integration. Noting that the associated Legendre function of the second kind, $Q_\nu^\mu(\lambda)$, goes unbounded as $\lambda \to 1$, we thus say $D_\Psi = 0$, thus leaving,

$$\Psi(\lambda) = C_\Psi P_\nu^\mu(\lambda)\,. \tag{48}$$

Analogously for $\Phi(\widetilde{Y})$,

$$\Phi(\lambda) = C_\Phi P_\nu^\mu(\lambda)\,. \tag{49}$$

The integer N is determined by the criterion in the bracket $\{\cdot\}$, on the right hand side of equation (46), that is, $N = [(\gamma + 1/4)^{\frac{1}{2}} - 1/2] + 1$. In this instance the square brackets $[\cdot]$ denote the integral part, which take a real number greater than zero as an argument and rounds it down to the nearest integer. If the integral part is already an integer, then the $+1$ is omitted from the expression. The eigenfunctions for the associated Legendre functions, equations (48) and (49), are given by $\mu = (\gamma + 1/4)^{\frac{1}{2}} - (m + 1/2)$ and exist if there is a root, $\nu(\nu + 1) = \gamma$.

We note from equation (42b) that for $m > 0$, $\gamma_{\mathrm{T}} > 0 \,\forall\, \tilde{k}$. On the other hand, we note from equation (45b) that γ_{C} is not necessarily greater than zero. This point shall be elaborated upon in section 4.3.

4.2 Barotropic dispersion relations

Substitution of α_{T} and γ_{T} into the eigenvalue relation, equation (46), gives the barotropic dispersion relation,

$$\tilde{\omega} = -\frac{\tilde{k}\tilde{f}_0^2}{\tilde{k}^2\tilde{f}_0^2 + |\tilde{k}|\tilde{f}_0(m + \frac{1}{2}) + m(m+1)}\,. \tag{50}$$

The first four modes for the barotropic waves are plotted in figure 3. We can see that they attain a maximum value in frequency-wavenumber space. For reasons that will become clear later, we wish to find the maximum frequency and its relative size compared to $\tilde{f}_0$.

We note that the maximum occurs at a turning point in wavenumber-frequency space. As such, in order to find the maximum frequency, we need to find where the gradient of the dispersion relation, equation (50), is zero (which, incidentally, corresponds to a zero group velocity). Thus, the maximum value of $\tilde{\omega}$ occurs when,

$$\tilde{k}\tilde{f}_0 = -\sqrt{m(m+1)}\,, \tag{51}$$

where we have chosen the negative root for $\tilde{k}$ based on the anti-symmetric nature of $\tilde{\omega}$.

Substitution of equation (51) into the dispersion relation gives the maximum frequency, $\tilde{\omega}_{\max}$, as a function of $\tilde{f}_0$. A more useful measure, however, is the ratio of the maximum frequency to $\tilde{f}_0$,

$$\frac{\tilde{\omega}_{\max}}{\tilde{f}_0} = \frac{1}{2\sqrt{m(m+1)} + (m + \frac{1}{2})} \tag{52}$$

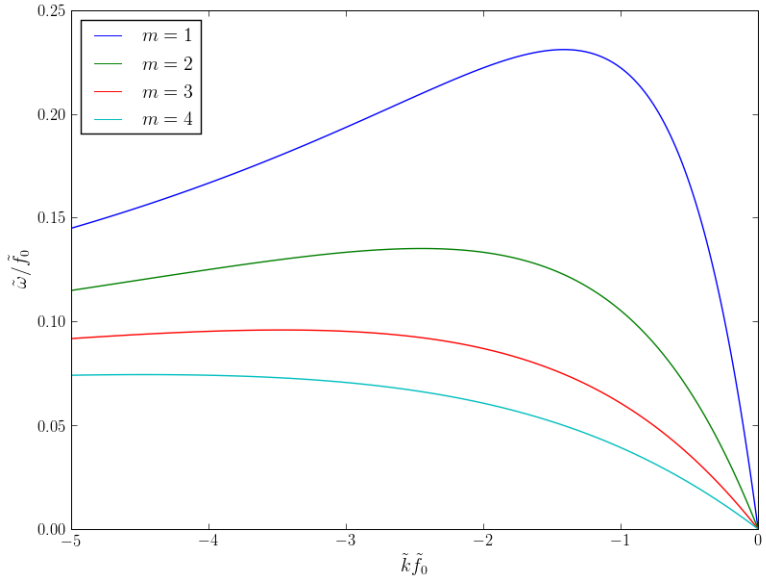


Figure 3: The first four meridional modes for the barotropic dispersion relation, equation (50). Note, that $\tilde{f}_0$ has been scaled out of the plot by making $\tilde{\omega}/\tilde{f}_0$ the dependent variable and $\tilde{k}\tilde{f}_0$ the independent variable.

We can thus see that ratio of the maximum frequency for barotropic modes to $\tilde{f}_0$ remains constant and that the ratio $\tilde{\omega}_{\max}/\tilde{f}_0$ decreases with increasing mode number.

Interestingly, we note through the use of the symmetry condition, $\tilde{k}(\omega) = -\tilde{k}(-\tilde{\omega})$, that the phase speed for the modes considered must be negative (i.e. westward).

4.3 Baroclinic dispersion relation

For the baroclinic case, we firstly consider the case for $\alpha_{\mathrm{C}} < 0$, necessitating that $-k\!\!\!/\omega > f_0^2$, which, similarly to the barotropic case, indicates that the phase speed is negative and is also bounded in absolute value. Furthermore, by the definition of α_{C}, we also require that,

$$\omega^2 - f_0^2 < k^2 \,, \tag{53}$$

The discrete dispersion relation – from equation (46) – is therefore given by,

$$\tilde{k}^2 + \tilde{f}_0^2 - \tilde{\omega}^2 = \left\{ \left(\tilde{f}_0^2 - \frac{\tilde{k}}{\tilde{\omega}} + \frac{1}{4\tilde{f}_0^2} \right)^{\frac{1}{2}} - \frac{m + 1\!\!\!/2}{\tilde{f}_0} \right\}^2 . \tag{54}$$

As $\tilde{f}_0 \to \infty$, we see that equation (54) reduces to the dispersion relation of Reznik and Zeitlin (2006, 2007a,b), equation (40) when we expand for large $\tilde{f}_0$. When equation (53) is not satisfied, this implies that there is a continuous spectrum in wavenumber-frequency

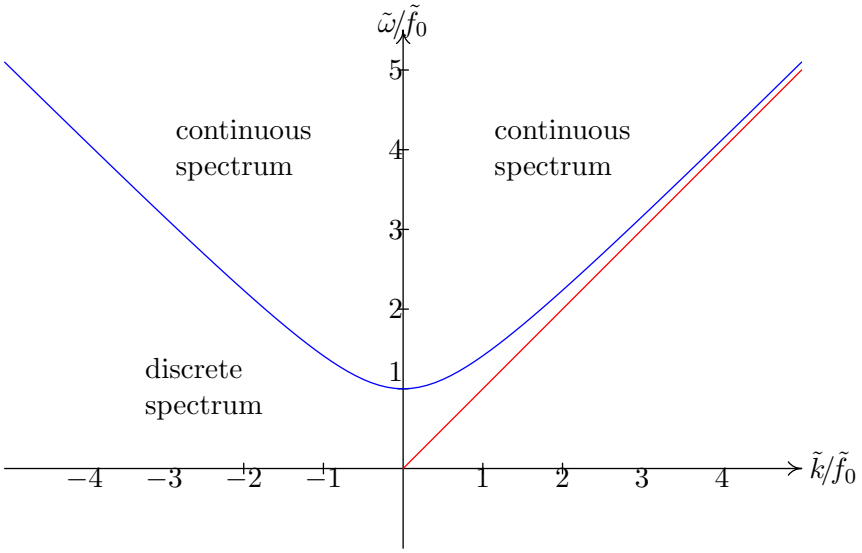


Figure 4: The blue line is $\tilde{\omega}/\tilde{f}_0 = \sqrt{\tilde{k}^2/\tilde{f}_0^2 + 1}$, which deliniates the continuous spectrum of frequency-wavenumber solutions from the discrete spectrum and the red line is $\tilde{\omega} = \tilde{k}$ which is the dispersion relation for equatorial Kelvin waves.

space.

Solutions for the discrete eigenvalues of the barotropic and baroclinic modes may all be found using associated Legendre functions. More general solutions to the associated Legendre equation can be found with the use of hypergeometric functions (see, for instance Oberhettinger, 1965). Again, following Drazin and Johnson (1989), we find that the solution for equation (43) is given by,

$$\Phi(\widetilde{Y}) = a2^{ip}[\mathrm{sech}(\widetilde{Y})]^{-ip} F(\check{a}, \check{b}; \check{c}; [\tanh(\widetilde{Y})]/2)\,, \tag{55}$$

where $a = a(\alpha_{\mathrm{C}})$ is a constant of integration, $p^2 = \alpha_{\mathrm{C}}$, $\check{a} = \frac{1}{2} - ip + (\gamma_{\mathrm{C}} + \frac{1}{4})^{\frac{1}{2}}$, $\check{b} = \frac{1}{2} - ik - (\gamma^c + \frac{1}{4})^{\frac{1}{2}}$, $\check{c} = 1 - ip$ and F is the hypergeometric function. Specifically, we note that for this type of function, the solutions for $\alpha_{\mathrm{C}} > 0$ yields a continuous spectrum for the baroclinic modes.

Similarly to the equatorial β-plane, we get a solution for $\Phi \equiv 0$, which again corresponds to an equatorial Kelvin wave, of the form $G(y) = \mathrm{sech}^{f_0^2}(Y)$. Again similarly to the β-plane, $m = 0$ has a spurious branch for $\tilde{\omega} = -\tilde{k}$ and the remaining solution is the Yanai wave. Finally, the modes that correspond to $m = 1, 2, \ldots$ are the trapped low-frequency Rossby waves. We can see that the high frequency branch of the dispersion relation for $m = 1, 2, \ldots$ actually lie within the continuous spectrum. Figure 4 shows the regions in frequency-wavenumber space where discrete and continuous solutions exist.

4.4 Resonance Conditions

We wish to investigate two different classes of interaction. The first is where we have three waves that lie within the discrete spectrum, and the second is where we have two waves in the discrete spectrum interacting with a third wave that lies in the continuous spectrum (recall, baroclinic waves that lie in the continuous spectrum satisfy the relation $\tilde{\omega}^2 > \tilde{k}^2 + \tilde{f}_0^2$). Most of the discussion in the remainder of this section pertains to the latter scenario, as there are some additional constraints that the waves from the discrete spectrum must satisfy in order to form part of a resonant triad.

Resonant triads (that is, the non-linear interaction between three waves) have resonance conditions of the form,

$$k_1 + k_2 + k_3 = 0 \tag{56a}$$

$$\omega_1 + \omega_2 + \omega_3 = 0\,. \tag{56b}$$

If we consider the case where one of the waves in the triad lies in the continuous spectrum, we note that the frequency of the continuous spectrum has a minimum value of $\tilde{f}_0$, we need to investigate the circumstances under which the resonance conditions will be satisfied with two waves from the discrete spectrum (which we label waves 1 and 2), and one from the continuous spectrum (which we label wave 3). A necessary (but not sufficient), condition for resonance to occur is for $\tilde{\omega}_{1\max} + \tilde{\omega}_{2\max} > \tilde{f}_0$, where $\tilde{\omega}_{\max}$ is the maximum frequency of a particular mode. As can be seen from equation (52), a combination of any two barotropic Rossby waves ($m \geq 1$) will not be able to satisfy this condition. However, if a Rossby wave is interacting with a Yanai wave ($m = 0$) or Kelvin wave ($m = -1$), then there may be circumstances where this condition may be satisfied.

The operations required to obtain an equivalent expression relating the maximum frequency to $\tilde{f}_0$ for the baroclinic dispersion relation, equation (54) requires more subtle analysis. We begin by taking the long wave limit, $\tilde{k} \to 0$. We define the dimensionless phase speed as $\tilde{c} = \tilde{\omega}/\tilde{k}$. We will examine two different limits,

- $\tilde{\omega} \to 0$ as $\tilde{c} \to \tilde{c}_{(0)}$ (the long Rossby limit), and
- $\tilde{c} \to 0$ as $\tilde{\omega} \to \tilde{\omega}_{(0)}$ (the long inertia-gravity limit);

The long Rossby limiting case yields

$$\tilde{f}_0 = \left(\tilde{f}_0^2 - \frac{1}{\tilde{c}_{(0)}} + \frac{1}{4\tilde{f}_0^2}\right)^{\frac{1}{2}} - \frac{m+\frac{1}{2}}{\tilde{f}_0} \tag{57a}$$

$$\Rightarrow \lim_{\tilde{\omega}\to 0} \tilde{c}_{(0)} = \frac{-\tilde{f}_0^2}{\tilde{f}_0^2(2m+1) + m(m+1)}\,. \tag{57b}$$

The equation for $\tilde{c}_{(0)}$ tells us that the long wave speed is negative for all modes $m \geq 0$ and positive for $m = -1$. Furthermore, the phase speed decreases in absolute value with increasing mode number, m.

The second case to consider yields the equation,

$$\sqrt{\tilde{f}_0^2 - (\tilde{\omega}_{(0)})^2} = \left(\tilde{f}_0^2 + \frac{1}{4\tilde{f}_0^2}\right)^{\frac{1}{2}} - \frac{m+\frac{1}{2}}{\tilde{f}_0} \tag{58a}$$

$$\Rightarrow \lim_{\tilde{c}\to 0}(\tilde{\omega}_{(0)})^2 = (2m+1)\left(1+\frac{1}{4\tilde{f}_0^4}\right)^{\frac{1}{2}} - \frac{(2m+1)^2+1}{4\tilde{f}_0^2}\,. \tag{58b}$$

The left hand side of equation (58a) tells us that the frequency must always be such that $0 < \tilde{\omega}_{(0)} < \tilde{f}_0$. To satisfy this requirement, the right hand side of equation (58a) must be

$$\begin{aligned}(m+\frac{1}{2})\tilde{f}_0^{-2} &< \sqrt{\tilde{f}_0^2 + (4\tilde{f}_0^2)^{-1}} \\ \Rightarrow m(m+1) &< \tilde{f}_0^4\,,\end{aligned} \tag{59}$$

in the long wave limit. We can see from equation (59) that the Yanai wave has a solution for all $\tilde{f}_0 > 0$. On the other hand, the inertia-gravity waves (with $m > 0$ and $\tilde{\omega}_{(0)} > 0$) are only trapped for $\tilde{k} = 0$ if $\tilde{f}_0^4$ is large enough. A corollary of this is that for some finite $\tilde{f}_0$ only low mode waves are trapped.

We also wish to examine whether any trapped modes can intercept the continuous spectrum. We determine this by recognising that the left hand side of the dispersion relation, equation (54), is also the boundary between the continuous and discrete spectra (the blue curve in figure 4). Equating each side to zero yields intersection conditions,

$$\tilde{\omega}^2 = \tilde{k}^2 + \tilde{f}_0^2 \tag{60}$$

$$\frac{\tilde{k}}{\tilde{\omega}} = -\frac{m(m+1)}{\tilde{f}_0^2} + \tilde{f}_0^2\,. \tag{61}$$

If we set $m = 0$ to search for a Yanai wave, we note that the two curves intercept at $\tilde{k}_0^2 = \tilde{f}_0^6(1-\tilde{f}_0^4)^{-1}$, where $\tilde{k}_0$ is the wavenumber of intersection for the $m = 0$ wave. Thus, the dispersion relation for a Yanai wave intercepts the boundary between the continuous and discrete spectra at $\tilde{k}_0$ when $0 < \tilde{f}_0 < 1$, having a discrete dispersion relation for $-\infty < \tilde{k} < \tilde{k}_0$ and a continuous spectrum for $\tilde{k} > \tilde{k}_0$. Conversely for $\tilde{f}_0 > 1$ they do not intercept the Yanai wave has a discrete dispersion relation for $-\infty < \tilde{k} < \infty$.

Using similar reasoning, we find that inertia-gravity waves cross from discrete relations to continuous if $\{m(m+1)\}^{\frac{1}{2}} < \tilde{f}_0^2 < m+1$, there is a solution $\tilde{k} = \tilde{k}_m > 0$ so that the trapped (discrete) baroclinic modes exist only for $-\infty < \tilde{k} < \tilde{k}_m$ and that $\tilde{k}_0 > \tilde{k}_1 > \tilde{k}_2 > \ldots$. Otherwise, if $\tilde{f}_0^2 > m+1$, the trapped waves exist for all $\tilde{k}$. If, however, $m < \tilde{f}_0^2 < \{m(m+1)\}^{\frac{1}{2}}$, there is a solution such that $k = \tilde{k}_m < 0$ where the inertia gravity wave crosses from the discrete spectrum to the continuous spectrum. Otherwise if $\tilde{f}_0^2 < m$, there are no trapped inertia-gravity modes.

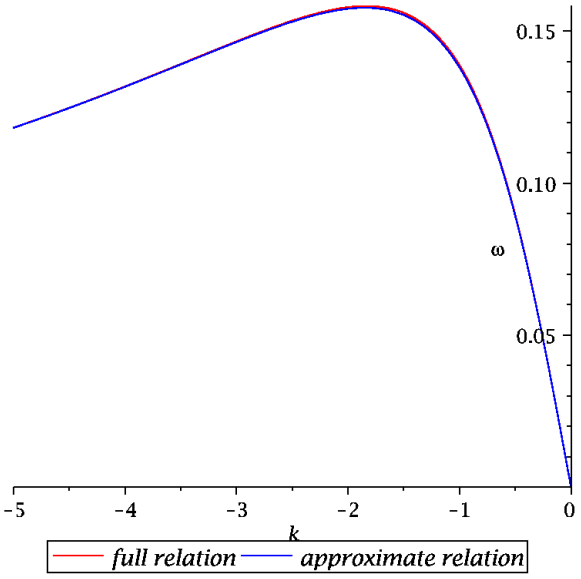


Figure 5: The red line is a plot of the full baroclinic dispersion relation, equation (54) and the blue line is the low frequency approximate dispersion relation, equation (62). The values used were $m = 1$ and $\tilde{f}_0 = 1$, and was made using the implicit plotting capabilities of Maple.

On the other hand, baroclinic Rossby waves remain trapped for all $\tilde{k}$, and as such, we may justifiably use a low frequency approximation, i.e. that the $\tilde{\omega}^2$ terms are small compared with the other terms in equation (54). This yields a dispersion relation,

$$\tilde{\omega} \approx \frac{-\tilde{k}\tilde{f}_0^2}{\tilde{k}^2\tilde{f}_0^2 + \tilde{f}_0(2m+1)\sqrt{\tilde{k}^2+\tilde{f}_0^2} + m(m+1)}\,. \tag{62}$$

Figure 5 shows a plot of the full discrete dispersion relation, equation (54), and the approximate low frequency dispersion relation for $m = 1$ and $f_0 = 1$. The difference between the two curves is quite small, and is of similar magnitude for higher modes. This confirms that the low frequency relation is an appropriate approximation to use for this study. A similarly small error is present for the first five trapped Rossby modes are plotted in figure 6 using the low frequency approximation.

As in section 4.2 we wish to ascertain under what circumstances we may obtain triad interactions between the discrete and continuous spectra. As we have a good approximate dispersion relation for Rossby waves that is explicit in $\tilde{\omega}$, we again begin by trying to find the maximum frequency as a function of $\tilde{f}_0$, we differentiate equation (62),

$$\frac{\mathrm{d}\omega}{\mathrm{d}\tilde{k}} = -\frac{\tilde{f}_0}{\tilde{k}^2\tilde{f}_0^2 + \tilde{f}_0(2m+1)\sqrt{\tilde{k}^2+\tilde{f}_0^2} + m(m+1)} + \tilde{k}\tilde{f}_0\frac{2\tilde{k}\tilde{f}_0^2 + (\tilde{f}_0\tilde{k}(2m+1))(\tilde{k}^2+\tilde{f}_0^2)^{-1}}{[\tilde{k}^2\tilde{f}_0^2 + \tilde{f}_0(2m+1)\sqrt{\tilde{k}^2+\tilde{f}_0^2} + m(m+1)]^2}\,. \tag{63}$$

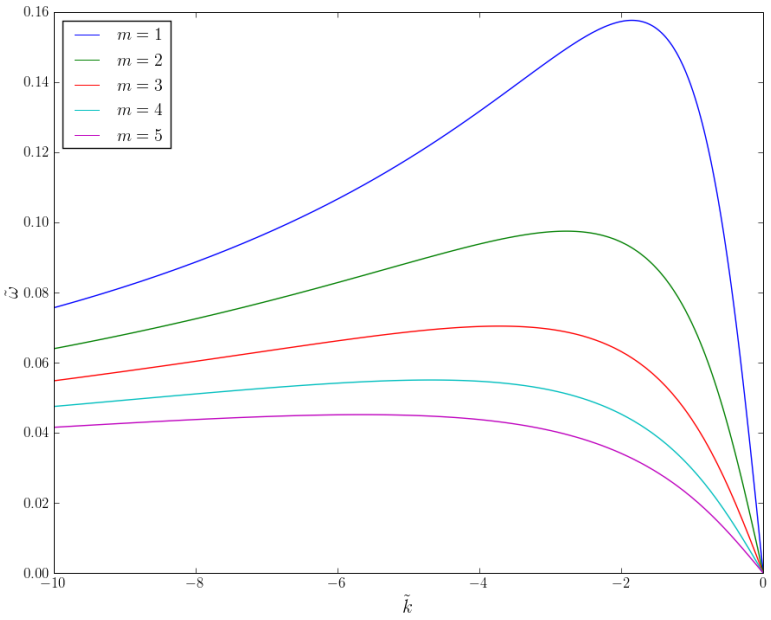


Figure 6: A dispersion relation for the first five low baroclinic frequency, trapped modes with $\tilde{f}_0 = 1$.

Setting the equation to zero and isolating $\tilde{k} = \tilde{k}_{max}$, and substituting back into the dispersion relation allows us to plot the ratio of the maximum frequency, $\tilde{\omega}_{max}$, to $\tilde{f}_0$ as well as the value of $\tilde{k}$ where this occurs, $\tilde{k}_{max}$, as a function of $\tilde{f}_0$. Such a plot is shown in figure 7 for $m = 1$.

We note that for both the barotropic case (see section 4.2) and the baroclinic case, we are unable to find resonant triads consisting of two discrete Rossby waves and one wave from the continuous spectrum. Again, it is necessary to use at least one Yanai or Kelvin wave in order to satisfy the resonance conditions, equations (56). The remainder of this report shall focus on weakly nonlinear resonant interactions between three discrete baroclinic Rossby waves. The properties of interactions involving Yanai or Kelvin waves, as well as the continuous spectrum are potential avenues of further research.

5 Resonant Triads

We now wish to examine "near-linear theory," for waves of moderately small amplitude. Small amplitude expansions to linear wave theory allow us to examine the interactions, and energy sharing between different Fourier components (see chapter 14 of Whitham, 1974).

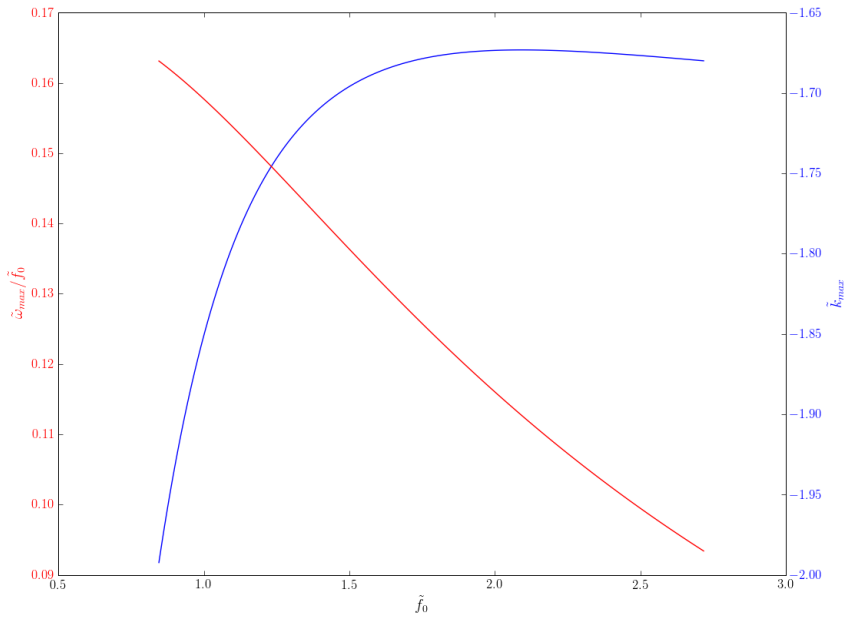


Figure 7: The ratio of the value of $\tilde{k}$ where the maximum frequency occurs (blue line and right axis) and the value of the ratio of the maximum value of $\tilde{\omega}$ to $\tilde{f}_0$ (red line and left axis), as a function of $\tilde{f}_0$, using the low frequency approximation for the dispersion relation, equation (62).

5.1 Nonlinear Equations

We need to examine a non-linear model to identify the important terms for the wave interactions. We use the baroclinic momentum and continuity equations (3.6a-d) of Benilov and Reznik (1996), which are modified here for an extended $f(y)$ profile,

$$\nabla^2\psi_{,t}+J(\psi,\nabla^2\psi)+\frac{D_1D_2}{H_0^2}(\partial_{xx}-\partial_{yy})\left[\left(1+\frac{D_1-D_2}{H_0}h-\frac{D_1D_2}{H_0^2}h^2\right)RS\right]$$
$$-\frac{D_1D_2}{H_0^2}\partial_{xy}\left[\left(1+\frac{D_1-D_2}{H_0}h-\frac{D_1-D_2}{H_0^2}h^2\right)(R^2-S^2)\right]+f'(y)\psi_{,x}=0 \quad (64a)$$

$$\boldsymbol{R}_{,t}+J(\psi,\boldsymbol{R})-\boldsymbol{R}\cdot\nabla(\boldsymbol{e}_z\wedge\nabla\psi)$$
$$-\frac{2D_1D_2}{H_0^2}h\boldsymbol{R}\cdot\nabla\boldsymbol{R}-\frac{D_1D_2}{H_0^2}\boldsymbol{R}\boldsymbol{R}\cdot\nabla h+c^2\nabla h=-f(y)\boldsymbol{e}_z\wedge\boldsymbol{R} \quad (64b)$$

$$h_{,t}+J(\psi,h)+\nabla\cdot\boldsymbol{R}+((D_1-D_2)H_0^{-1}-D_1D_2H_0^{-2})\nabla\cdot(h\boldsymbol{R})-\nabla\cdot\left(h^2\boldsymbol{R}\right)=0 \quad (64c)$$

where $J(A,B)=A_{,x}B_{,y}-A_{,y}B_{,x}$ is the Jacobi operator. We also recall that $h=H_0\eta_1/D_1D_2$ and $c^2=g'D_1D_2H_0^{-1}$. From these equations, we obtain the non-dimensional equations using

the length and time scales defined in equations (23),

$$\widetilde{\nabla}^2\tilde{\psi}_{,\tilde{t}} + \tilde{f}'(\tilde{y})\tilde{\psi}_{,\tilde{x}} = -J(\tilde{\psi},\widetilde{\nabla}^2\tilde{\psi}) - \varpi(\partial_{\tilde{x}\tilde{x}} - \partial_{\tilde{y}\tilde{y}})[(1+q\tilde{h}-\varpi\tilde{h}^2)\widetilde{R}\widetilde{S}] - \partial_{\tilde{x}\tilde{y}}[(1+qh-\varpi\tilde{h}^2)(\widetilde{R}^2-\widetilde{S}^2)] \tag{65a}$$

$$\widetilde{\boldsymbol{R}}_{,\tilde{t}} + \widetilde{\nabla}\tilde{h} + \tilde{f}(\tilde{y})\boldsymbol{e}_z \wedge \widetilde{\boldsymbol{R}} = -J(\tilde{\psi},\widetilde{\boldsymbol{R}}) + \widetilde{\boldsymbol{R}}\cdot\widetilde{\nabla}(\boldsymbol{e}_z\wedge\nabla\tilde{\psi}) - \varpi q\widetilde{\boldsymbol{R}}\cdot\widetilde{\nabla}\widetilde{\boldsymbol{R}} + \varpi\left[2\tilde{h}\widetilde{\boldsymbol{R}}\cdot\widetilde{\nabla}\widetilde{\boldsymbol{R}} + \widetilde{\boldsymbol{R}}\widetilde{\boldsymbol{R}}\cdot\widetilde{\nabla}\tilde{h}\right] \tag{65b}$$

$$\tilde{h}_{,\tilde{t}} + \nabla\cdot\widetilde{\boldsymbol{R}} = -J(\tilde{\psi},\tilde{h}) - \widetilde{\nabla}\cdot(\tilde{h}\widetilde{\boldsymbol{R}}) + \varpi\widetilde{\nabla}\cdot(\tilde{h}^2\widetilde{\boldsymbol{R}})\,, \tag{65c}$$

where, $\varpi = \widetilde{D}_1\widetilde{D}_2/\widetilde{H}_0^2$ and $q = (\widetilde{D}_1 - \widetilde{D}_2)/\widetilde{H}_0$.

5.2 Solutions

Let us now consider an asymptotic, multitimescale solution (Bretherton, 1964) to the above equations , such that,

$$\tilde{\psi} = \epsilon^1\tilde{\psi}^{(1)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \epsilon^2\tilde{\psi}^{(2)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \ldots \tag{66a}$$

$$\widetilde{R} = \epsilon^1\widetilde{R}^{(1)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \epsilon^2\widetilde{R}^{(2)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \ldots \tag{66b}$$

$$\widetilde{S} = \epsilon^1\widetilde{S}^{(1)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \epsilon^2\widetilde{S}^{(2)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \ldots \tag{66c}$$

$$\tilde{h} = \epsilon^1\tilde{h}^{(1)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \epsilon^2\tilde{h}^{(2)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) + \ldots\,, \tag{66d}$$

where, ϵ is a small parameter and $\tilde{\tau} = \epsilon\tilde{t}$ is a slow time variable. Recalling equations (17), we assume that the solutions take the form,

$$\tilde{\psi}^{(1)}(\tilde{x},\tilde{y},\tilde{t},\tilde{\tau}) = \sum_{j=1}^{3}\Psi_j^{(1)}(\tilde{y})\left[A_j(\tilde{\tau})e^{i\tilde{\varphi}_j} + A_j^*(\tilde{\tau})e^{-i\tilde{\varphi}_j}\right] \tag{67a}$$

$$\widetilde{R}^{(1)}(\tilde{x},\tilde{y},\tilde{\tau}) = \sum_{j=1}^{3}\Xi_j^{(1)}(\tilde{y})\left[iA_j(\tilde{\tau})e^{i\tilde{\varphi}_j} - iA_j^*(\tilde{\tau})e^{-i\tilde{\varphi}_j}\right] \tag{67b}$$

$$\widetilde{S}^{(1)}(\tilde{x},\tilde{y},\tilde{\tau}) = \sum_{j=1}^{3}\Phi_j^{(1)}(\tilde{y})\left[A_j(\tilde{\tau})e^{i\tilde{\varphi}_j} + A_j^*(\tilde{\tau})e^{-i\tilde{\varphi}_j}\right] \tag{67c}$$

$$\tilde{h}^{(1)}(\tilde{x},\tilde{y},\tilde{\tau}) = \sum_{j=1}^{3}\widetilde{G}_j^{(1)}(\tilde{y})\left[iA_j(\tilde{\tau})e^{i\tilde{\varphi}_j} - iA_j^*(\tilde{\tau})e^{-i\tilde{\varphi}_j}\right]\,, \tag{67d}$$

where a superscripted asterisk, * , indicates a complex conjugate, $\tilde{\varphi} \equiv \tilde{k}\tilde{x} - \tilde{\omega}\tilde{t}$ and $j = 1, 2, 3$ identifies the wave in the triad. Here, we have assumed that the amplitude of the waves is slowly varying. We substitute the first two orders of our expansion into equations (65a) to (65c) and equate powers of ϵ (Luke, 1966). For the barotrpic equations, we equate terms that are multiplied by ϵ,

$$\begin{aligned}\epsilon\widetilde{\nabla}^2\tilde{\psi}^{(1)}_{,\tilde{t}} + \epsilon\tilde{f}'(\tilde{y})\tilde{\psi}^{(1)}_{,\tilde{x}} &= 0\\ \widetilde{\nabla}^2\tilde{\psi}^{(1)}_{,\tilde{t}} + \tilde{f}'(\tilde{y})\tilde{\psi}^{(1)}_{,\tilde{x}} &= 0\,,\end{aligned} \tag{68}$$

which may be recognised as the linear equation (18). Doing the same for $\tilde{h}$ and $\widetilde{\boldsymbol{R}}$ yields,

$$\tilde{h}^{(1)}_{,\tilde{t}} + \nabla \cdot \widetilde{\boldsymbol{R}}^{(1)} = 0 \tag{69a}$$

$$\widetilde{\boldsymbol{R}}^{(1)}_{,\tilde{t}} + \widetilde{\nabla}\tilde{h}^{(1)} + \tilde{f}(\tilde{y})\boldsymbol{e}_z \wedge \widetilde{\boldsymbol{R}}^{(1)} = 0 \tag{69b}$$

which can be recognised as the linear equations (31b) and (31d) respectively.

Similarly, we may equate terms that are multiplied by ϵ^2,

$$\widetilde{\nabla}^2\tilde{\psi}^{(2)}_{,\tilde{t}} + \tilde{f}'(\tilde{y})\tilde{\psi}^{(2)}_{,\tilde{x}} = -\widetilde{\nabla}^2\tilde{\psi}^{(1)}_{,\tau} + N^{\psi} \tag{70a}$$

$$\widetilde{R}^{(2)}_{,\tilde{t}} + \tilde{h}^{(2)}_{,\tilde{x}} - \tilde{f}(\tilde{y})\widetilde{S}^{(2)} = -\widetilde{R}^{(1)}_{,\tau} + N^{R} \tag{70b}$$

$$\widetilde{S}^{(2)}_{,\tilde{t}} + \tilde{h}^{(2)}_{,\tilde{y}} + \tilde{f}(\tilde{y})\widetilde{R}^{(2)} = -\widetilde{S}^{(1)}_{,\tau} + N^{S} \tag{70c}$$

$$\tilde{h}^{(2)}_{,\tilde{t}} + \nabla \cdot \widetilde{\boldsymbol{R}}^{(2)} = -\tilde{h}^{(1)}_{,\tau} + N^{h} \tag{70d}$$

where we note that via use of the chain rule $\widetilde{\nabla}^2\tilde{\psi}^{(1)}_{,\tau}\tau_{,t} = \widetilde{\nabla}^2\tilde{\psi}^{(1)}_{,\tau}\epsilon$, and similarly for $\widetilde{R}^{(1)}_{,\tau}$, $\widetilde{S}^{(1)}_{,\tau}$ and $\tilde{h}^{(1)}_{,\tau}$. Here, the nonlinear terms N^{ψ}, $N^{\boldsymbol{R}}$ and N^{h} are given by,

$$N^{\psi} = -J(\tilde{\psi}^{(1)}, \widetilde{\nabla}^2\tilde{\psi}^{(1)}) - \varpi(\partial_{\tilde{x}\tilde{x}} - \partial_{\tilde{y}\tilde{y}})\widetilde{R}^{(1)}\widetilde{S}^{(1)} - \partial_{\tilde{x}\tilde{y}}(\widetilde{R}^{(1)\,2} - \widetilde{S}^{(1)\,2}) \tag{71a}$$

$$N^{R} = -J(\tilde{\psi}^{(1)}, \widetilde{R}^{(1)}) + \widetilde{\boldsymbol{R}}^{(1)} \cdot \widetilde{\nabla}\tilde{\psi}^{(1)}_{,\tilde{x}} - \varpi q\widetilde{\boldsymbol{R}}^{(1)} \cdot \widetilde{\nabla}\widetilde{R}^{(1)} \tag{71b}$$

$$N^{S} = -J(\tilde{\psi}^{(1)}, \widetilde{S}^{(1)}) + \widetilde{\boldsymbol{R}}^{(1)} \cdot \widetilde{\nabla}\tilde{\psi}^{(1)}_{,\tilde{y}} - \varpi q\widetilde{\boldsymbol{R}}^{(1)} \cdot \widetilde{\nabla}\widetilde{S}^{(1)} \tag{71c}$$

$$N^{h} = -J(\tilde{\psi}^{(1)}, \tilde{h}^{(1)}) - \widetilde{\nabla} \cdot (\tilde{h}^{(1)}\widetilde{\boldsymbol{R}}^{(1)}) \,. \tag{71d}$$

If we apply the resonance conditions, equations (56), to these nonlinear equations, we note that both the left hand side and the right hand sides have terms that are proportional $\exp(i\tilde{\varphi}_j)$. For example, if we apply the resonance condition for the $j = 1$ wave, then terms that are not discarded are those that are proportional to $\exp(i\tilde{\varphi}_1)$ and $\exp(-i[\tilde{\varphi}_2 + \tilde{\varphi}_3])$. In order to find a compatibility condition we substitute our assumed barotropic streamfunction solution, equation (67a) into equation (70a) to yield,

$$i\tilde{\omega}_j\Psi^{(2)}_{j,\tilde{y}\tilde{y}} + (\tilde{k}_j\tilde{f}' - \tilde{\omega}_j\tilde{k}_j^2)i\Psi^{(2)}_j = \mathcal{F}^{\Psi}_j \,, \tag{72}$$

where $\mathcal{F}^{\Psi}_j$ may thought of as a forcing function arising from the nonlinear wave interactions and is given by,

$$\mathcal{F}^{\Psi}_j = (\tilde{k}_j^2\Psi^{(1)}_j - \Psi^{(1)}_{j,\tilde{y}\tilde{y}})A_{j,\tilde{\tau}} + \mathcal{N}^{\psi}_j \,, \tag{73}$$

in which $\mathcal{N}^{\psi}_j$ are the non-linear terms from equation (71a) that are proportional to $\exp(i\tilde{\varphi}_j)$.

Similarly, combining equations (70b) to (70d), we find a forced modal equation for the second order effects,

$$i\tilde{\omega}_j\Phi^{(2)}_{j,\tilde{y}\tilde{y}} + \left\{-\tilde{\omega}_j\tilde{f}^2 - \tilde{\omega}_j\tilde{k}_j^2 + \tilde{\omega}_j^3 - \tilde{k}_j\tilde{f}'\right\} i\Phi^{(2)}_j = \mathcal{F}^{\Phi}_j \,, \tag{74}$$

where $\mathcal{F}_j^\Phi$ is the baroclinic forcing term due to interaction with other waves and is given by,

$$\begin{aligned}\mathcal{F}_j^\Phi =&\Phi_j^{(1)}(\tilde{\omega}_j^2 - \tilde{k}_j^2)A_{j,\tilde{\tau}} + (\partial_{\tilde{t}\tilde{t}} - \partial_{\tilde{x}\tilde{x}})\mathcal{N}_j^S A_2^* A_3^* \\ &+ (\tilde{k}_j \Xi_{j,\tilde{y}}^{(1)} + \tilde{f}\tilde{\omega}_j \Xi_j^{(1)})A_{j,\tilde{\tau}} + (\partial_{\tilde{x}\tilde{y}} - \tilde{f}\partial_{\tilde{t}})\mathcal{N}_j^R A_2^* A_3^* \\ &+ (\tilde{f}\tilde{k}_j \widetilde{G}_j^{(1)} + \tilde{\omega}_j \widetilde{G}_{j,\tilde{y}}^{(1)})A_{j,\tilde{\tau}} + (\tilde{f}\partial_{\tilde{x}} - \partial_{\tilde{t}\tilde{y}})\mathcal{N}_j^h A_2^* A_3^* \,, \end{aligned} \tag{75}$$

where $\mathcal{N}^R$, $\mathcal{N}^S$ and $\mathcal{N}^h$ are the nonlinear terms that are proportional to $\exp(i\tilde{\varphi}_j)$.

Equations (72) and (74) have compatibility conditions,

$$\int_{-\infty}^{\infty} \mathcal{F}_j^\Psi \Psi_j \mathrm{d}\tilde{y} = 0 \tag{76a}$$

$$\int_{-\infty}^{\infty} \mathcal{F}_j^\Phi \Phi_j \mathrm{d}\tilde{y} = 0 \,, \tag{76b}$$

where Ψ_j and Φ_j are functions that satisfy the modal equations (43) and (44), respectively. Although there appear to be two compatibility conditions, there is indeed only one case required, as for a barotropic mode $\Phi_j(\tilde{y}) \equiv 0$ and for the baroclinic case, $\Psi_j(\tilde{y}) \equiv 0$.

In the next section, we examine form of $\mathcal{N}$ for three trapped Rossby waves, which will eventually allow us to find the equations for the evolution of the wave amplitude.

5.3 Triad Interactions of Three Baroclinic Rossby Waves

We wish to examine the interaction of three baroclinic waves. Firstly, we use the graphical technique (for example, Simmons, 1969) to illustrate the possibility of resonant interactions. The approximate baroclinic Rossby wave dispersion relation, equation (62), is plotted in figure 8. The origin of the $m = 1, 2, 4, 5$ waves (blue curves) is translated to an arbitrary point along the $m = 3$ baroclinic wave (the red line) while $m = 1, 2, 4, 5$ are also plotted in black. As can be seen, there are many possible intersections for just the first five modes, and indeed further possibilities may be found by translating the origin of the blue lines to a different place on the red line, or, considering a different mode to $m = 3$, or indeed, considering more modes than just the first five.

We note that in the purely baroclinic case, there will be no barotropic terms, as $\tilde{\psi}^{(1)} \equiv 0$. This allows us to expand the nonlinear terms in the forcing function, equation (75). To illustrate, we begin by considering the wave labelled 1. We examine the non-linear terms, equation (71), paying particular attention to those terms that are proportional to $\exp(-i[\tilde{\varphi}_2 + \tilde{\varphi}_3])$,

$$N_1^R = i\varpi q \left(\tilde{k}_1 \Xi_2^{(1)} \Xi_3^{(1)} + \Phi_2^{(1)} \Xi_{3,\tilde{y}}^{(1)} + \Xi_{2,\tilde{y}}^{(1)} \Phi^{(1)}\right) e^{-i(\tilde{\varphi}_2 + \tilde{\varphi}_3)} + \dots \tag{77a}$$

$$N_1^S = \varpi q \left(\tilde{k}_2 \Phi_2^{(1)} \Xi_3^{(1)} + \tilde{k}_3 \Xi_2^{(1)} \Phi_3^{(1)} - \partial_{\tilde{y}}[\Phi_2^{(1)} \Phi_3^{(1)}]\right) e^{-i(\tilde{\varphi}_2 + \tilde{\varphi}_3)} + \dots \tag{77b}$$

$$N_1^h = i \left(\partial_{\tilde{y}}[\Phi_2 \widetilde{G}_3 - \widetilde{G}_2 \Phi_3] + \tilde{k}_1[\Xi_2 \widetilde{G}_3 - \widetilde{G}_2 \Xi_3]\right) e^{-i(\tilde{\varphi}_2 + \tilde{\varphi}_3)} + \dots \,, \tag{77c}$$

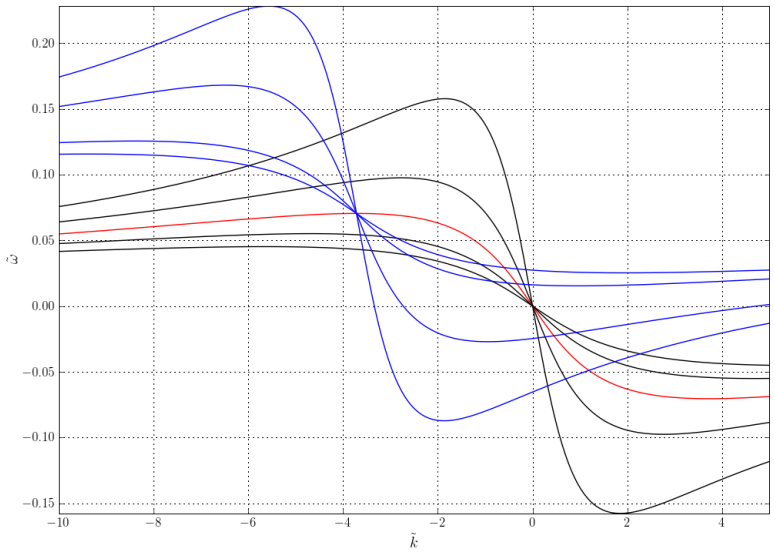


Figure 8: A rich array of three wave resonant interactions can be found in the system under consideration. The approximate low frequency dispersion relation for baroclinic Rossby waves with $\tilde{f}_0 = 1$ is plotted for $m = 1, 2, 4, 5$ (black lines) and $m = 3$ (red line). The origin for the $m = 1, 2, 4, 5$ waves is translated along the $m = 3$ relation to an arbitrary point. Intersection of the blue lines with black lines indicate resonant triads.

where ... represents terms that are not proportional to $e^{-i\tilde{\varphi}_1}$, and are thus ignored. The terms shown are $\mathcal{N}_1^R$, $\mathcal{N}_1^S$ and $\mathcal{N}_1^h$ in equation (75) for the forcing term. In order to find the forcing term, we need to operate on the nonlinear terms,

$$\mathcal{N}_1^R = \varpi q(k_1\partial_{\tilde{y}} + \tilde{f}\tilde{\omega}_1)\left(\tilde{k}_1\Xi_2^{(1)}\Xi_3^{(1)} + \Phi_2^{(1)}\Xi_{3,\tilde{y}}^{(1)} + \Xi_{2,\tilde{y}}^{(1)}\Phi^{(1)}\right) \tag{78a}$$

$$\mathcal{N}_1^S = \varpi q(\tilde{\omega}_1^2 - \tilde{k}_1^2)\left(\tilde{k}_2\Phi_2^{(1)}\Xi_3^{(1)} + \tilde{k}_3\Xi_2^{(1)}\Phi_3^{(1)} - \partial_{\tilde{y}}[\Phi_2^{(1)}\Phi_3^{(1)}]\right) \tag{78b}$$

$$\mathcal{N}_1^h = (\tilde{\omega}_1\partial_{\tilde{y}} + \tilde{f}\tilde{k}_1)\left(\partial_{\tilde{y}}[\Phi_2\widetilde{G}_3 - \widetilde{G}_2\Phi_3] + \tilde{k}_1[\Xi_2\widetilde{G}_3 - \widetilde{G}_2\Xi_3]\right) \tag{78c}$$

where

$$\mathscr{N}_1^R = -(\partial_{\tilde{x}\tilde{y}} - \tilde{f}\partial_{\tilde{t}})\mathcal{N}_1^R\,, \quad \mathscr{N}_1^S = -(\partial_{\tilde{t}\tilde{t}} - \partial_{\tilde{x}\tilde{x}})\mathcal{N}_1^S\,, \quad \mathscr{N}_1^h = -(\tilde{f}\partial_{\tilde{x}} - \partial_{\tilde{t}\tilde{y}})\mathcal{N}_1^h\,.$$

We can see that the equivalent expressions for $\mathscr{N}_2$ and $\mathscr{N}_3$ can be found by swapping indicies.

As the governing equation for the nonlinear interactions, equation (74), is a forced modal equation, we use the compatibility condition – noting that $\Phi_1^{(1)}$ satisfies the modal equation (44) – to yield an equation for the time evolution of the amplitude,

$$A_{1,\tilde{\tau}}\int_{-\infty}^{\infty}(\tilde{\omega}_1^2 - \tilde{k}_1^2)\left(\Phi_1^{(1)}\right)^2 + \Phi_1^{(1)}(\tilde{f}\tilde{\omega}_1 + \tilde{k}_1\partial_{\tilde{y}})\Xi_1^{(1)} + \Phi_1^{(1)}(\tilde{f}\tilde{k}_1 + \tilde{\omega}_1\partial_{\tilde{y}})\widetilde{G}_1^{(1)}\mathrm{d}\tilde{y}$$
$$= A_2^*A_3^*\int_{-\infty}^{\infty}\Phi_1^{(1)}\mathscr{N}_1^R + \Phi_1^{(1)}\mathscr{N}_1^S + \Phi_1^{(1)}\mathscr{N}_1^h\mathrm{d}\tilde{y}\,. \tag{79}$$

Before being able to solve such an equation, we use the relations $\widetilde{G}$ and Ξ in terms of Φ, equations (38). It is known that Φ is an associated Legendre function, equation (49). Substituting, we find that,

$$\begin{aligned}&(\tilde{\omega}_1^2-\tilde{k}_1^2)\left(\Phi_1\right)^2+\Phi_1(\tilde{f}\tilde{\omega}_1+\tilde{k}_1\partial_{\tilde{y}})\Xi_1+\Phi_1(\tilde{f}\tilde{k}_1+\tilde{\omega}_1\partial_{\tilde{y}})\widetilde{G}_1\\&=\Phi_1\Phi_{1,\tilde{y}\tilde{y}}\frac{\tilde{k}_1^2+\tilde{\omega}_1^2}{\tilde{k}_1^2-\tilde{\omega}_1^2}-\Phi_1(\Phi_1\tilde{f})_{,\tilde{y}}\frac{2\tilde{k}_1\tilde{\omega}_1}{\tilde{k}_1^2-\tilde{\omega}_1^2}-\Phi_1^2\tilde{f}^2\frac{\tilde{k}_1^2+\tilde{\omega}_1^2}{\tilde{k}_1^2-\tilde{\omega}_1^2}+\tilde{f}\Phi_{1,\tilde{y}}^2\frac{\tilde{k}_1\tilde{\omega}_1}{\tilde{k}_1^2-\tilde{\omega}_1^2}-\Phi_1^2(\tilde{k}_1^2-\tilde{\omega}_1^2)\,,\end{aligned}\tag{80}$$

where we have dropped the superscript, (1), for brevity, as all modal functions are first order. In order to garner information about the time evolution of the amplitude of the waves, we need to find out information about the integrals in equation (79). We begin with the left hand side. Firstly we note that from the approximate baroclinic dispersion relation, equation (62), $|\tilde{k}|>|\tilde{\omega}|$ for $|\tilde{k}|>0$ and as such, $\tilde{k}^2-\tilde{\omega}^2>0$. We also note from the dispersion relation that $\tilde{\omega}\tilde{k}<0$ for $|\tilde{k}|>0$.

We use integration by parts and the property that $\Phi(\tilde{y})\to 0$ as $\tilde{y}\to\pm\infty$, to note the following,

$$\begin{aligned}\int_{-\infty}^{\infty}\Phi\Phi_{,\tilde{y}\tilde{y}}\mathrm{d}\tilde{y}&=\Phi\Phi_{,\tilde{y}}\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}\Phi_{,\tilde{y}}^2\mathrm{d}\tilde{y}\\&=-\int_{-\infty}^{\infty}\Phi_{,\tilde{y}}^2\mathrm{d}\tilde{y}\end{aligned}\tag{81a}$$

$$\begin{aligned}\frac{1}{2}\int_{-\infty}^{\infty}\tilde{f}\Phi_{,\tilde{y}}^2&=\frac{1}{2}\left(\tilde{f}\Phi^2\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}\Phi^2\tilde{f}'\mathrm{d}\tilde{y}\right)\\&=-\frac{1}{2}\int_{-\infty}^{\infty}\Phi^2\tilde{f}'\mathrm{d}\tilde{y}\end{aligned}\tag{81b}$$

$$\begin{aligned}\int_{-\infty}^{\infty}\Phi(\Phi\tilde{f})_{,\tilde{y}}\mathrm{d}\tilde{y}&=\Phi^2\tilde{f}\Big|_{-\infty}^{\infty}-\frac{1}{2}\int_{-\infty}^{\infty}\tilde{f}\Phi_{,\tilde{y}}^2\mathrm{d}\tilde{y}\\&=\frac{1}{2}\int_{-\infty}^{\infty}\Phi^2\tilde{f}'\mathrm{d}\tilde{y}\,,\end{aligned}\tag{81c}$$

We say that Υ_1 is equivalent to equation (80) and use the results of equations (81) to obtain,

$$\int_{-\infty}^{\infty}\Upsilon_1\mathrm{d}\tilde{y}=-\frac{1}{\tilde{k}_1^2-\tilde{\omega}_1^2}\int_{-\infty}^{\infty}(\tilde{k}_1^2+\tilde{\omega}_1^2)\Phi_{,\tilde{y}}^2+\frac{3\tilde{k}_1\tilde{\omega}_1}{2}\tilde{f}'\Phi_1^2+\Phi_1^2\tilde{f}^2(\tilde{k}_1^2+\tilde{\omega}_1^2)+\Phi_1^2(\tilde{k}_1^2-\tilde{\omega}_1^2)^2\mathrm{d}\tilde{y}\,.\tag{82}$$

We furthermore note that one property of the associated Legendre function is that it is either symmetric or anti-symmetric. In the present case it is the meridional mode number, m, that determines whether the modal function is symmetric or anti-symmetric – see equation (46). We also note that $\tilde{f}(\tilde{y})$ is an anti-symmetric function and that $\tilde{f}'(\tilde{y})>0$ is a positive, symmetric function. Using these properties we find the following for the various terms in

equations (80) and (81),

$$\int_{-\infty}^{\infty} \Phi_{,\tilde{y}}^2 \mathrm{d}\tilde{y} > 0 \tag{83a}$$

$$\int_{-\infty}^{\infty} \Phi^2 \tilde{f}' \mathrm{d}\tilde{y} > 0 \tag{83b}$$

$$\int_{-\infty}^{\infty} \tilde{f}^2 \Phi^2 \mathrm{d}\tilde{y} > 0 \tag{83c}$$

$$\int_{-\infty}^{\infty} \Phi^2 \mathrm{d}\tilde{y} > 0\,, \tag{83d}$$

Use of these results tells us that the integral of Υ_1, equation (82) is negative.

With respect to the right hand side of equation (79) we find, after much algebra, that,

$$\begin{aligned}\Phi_1 \mathscr{N}_1^R = -\varpi q \Bigg\{ & \frac{\tilde{k}_1 \Phi_1 (\tilde{\omega}_1 \tilde{f} + \tilde{k}_1 \partial_{\tilde{y}}) \left[\tilde{k}_2 \tilde{k}_3 \Phi_{2,\tilde{y}} \Phi_{3,\tilde{y}} - \tilde{k}_2 \tilde{\omega}_3 \tilde{f} \Phi_{2,\tilde{y}} \Phi_3 - \tilde{\omega}_2 \tilde{k}_3 \tilde{f} \Phi_2 \Phi_{3,\tilde{y}} + \tilde{\omega}_2 \tilde{\omega}_3 \tilde{f}^2 \Phi_2 \Phi_3 \right]}{(\tilde{k}_2^2 - \tilde{\omega}_2^2)(\tilde{k}_3^2 - \tilde{\omega}_3^2)} \\ & + \Phi_1 \frac{\tilde{k}_1 \tilde{k}_2 (\Phi_{2,\tilde{y}\tilde{y}} \Phi_3)_{,\tilde{y}} - \tilde{k}_1 \tilde{\omega}_2 [(\Phi_2 \tilde{f}' \Phi_3)_{,\tilde{y}} + (\tilde{f} \Phi_{2,\tilde{y}} \Phi_3)_{,\tilde{y}}] + \tilde{\omega}_1 \tilde{f} [\tilde{k}_2 \Phi_{2,\tilde{y}\tilde{y}} \Phi_3 - \tilde{\omega}_2 \Phi_3 (\tilde{f} \Phi_2)_{,\tilde{y}}]}{\tilde{k}_2^2 - \tilde{\omega}_2^2} \\ & + \Phi_1 \frac{\tilde{k}_1 \tilde{k}_3 (\Phi_2 \Phi_{3,\tilde{y}\tilde{y}})_{,\tilde{y}} - \tilde{k}_1 \tilde{\omega}_3 [(\Phi_2 \tilde{f}' \Phi_3)_{,\tilde{y}} + (\tilde{f} \Phi_2 \Phi_{3,\tilde{y}})_{,\tilde{y}}] + \tilde{\omega}_1 \tilde{f} [\tilde{k}_3 \Phi_2 \Phi_{3,\tilde{y}\tilde{y}} - \tilde{\omega}_3 \Phi_2 (\tilde{f} \Phi_3)_{,\tilde{y}}]}{\tilde{k}_3^2 - \tilde{\omega}_3^2} \Bigg\}\end{aligned} \tag{84a}$$

$$\begin{aligned}\Phi_1 \mathscr{N}_1^S = \varpi q (\tilde{k}_1^2 - \tilde{\omega}^2) \Bigg\{ & \Phi_1 (\Phi_2 \Phi_3)_{,\tilde{y}} - \tilde{k}_2 \tilde{k}_3 \left(\frac{\Phi_1 \Phi_{2,\tilde{y}} \Phi_3}{\tilde{k}_2^2 - \tilde{\omega}_2^2} + \frac{\Phi_1 \Phi_2 \Phi_{3,\tilde{y}}}{\tilde{k}_3^2 - \tilde{\omega}_3^2} \right) \\ & + \tilde{f} \Phi_1 \Phi_2 \Phi_3 \left(\frac{\tilde{k}_3 \tilde{\omega}_2}{\tilde{k}_2^2 - \tilde{\omega}_2^2} + \frac{\tilde{k}_2 \tilde{\omega}_3}{\tilde{k}_3^2 - \tilde{\omega}_3^2} \right) \Bigg\}\end{aligned} \tag{84b}$$

$$\begin{aligned}\Phi_1 \mathscr{N}_1^h = & \frac{\Phi_1 (\tilde{f} \tilde{k}_1 + \tilde{\omega}_1 \partial_{\tilde{y}}) [\tilde{k}_3 (\tilde{f} \Phi_2 \Phi_3)_{,\tilde{y}} - \tilde{\omega}_3 (\Phi_2 \Phi_{3,\tilde{y}})_{,\tilde{y}}]}{\tilde{k}_3^2 - \tilde{\omega}_3^2} \\ & - \frac{\Phi_1 (\tilde{f} \tilde{k}_1 + \tilde{\omega}_1 \partial_{\tilde{y}}) [\tilde{k}_2 (\tilde{f} \Phi_2 \Phi_3)_{,\tilde{y}} - \tilde{\omega}_2 (\Phi_{2,\tilde{y}} \Phi_3)_{,\tilde{y}}]}{\tilde{k}_2^2 - \tilde{\omega}_2^2} \\ & + \frac{\Phi_1 \tilde{k}_1 (\tilde{f} \tilde{k}_1 + \tilde{\omega}_1 \partial_{\tilde{y}}) [(\tilde{k}_2 \tilde{k}_3 - \tilde{\omega}_2 \tilde{\omega}_3)(\tilde{f} \Phi_{2,\tilde{y}} \Phi - \tilde{f} \Phi_2 \Phi_{3,\tilde{y}}) + (\tilde{k}_2 \tilde{\omega}_3 - \tilde{k}_3 \tilde{\omega}_3)(\tilde{f}^2 \Phi_2 \Phi_3 - \Phi_{2,\tilde{y}} \Phi_{3,\tilde{y}})]}{(\tilde{k}_2^2 - \tilde{\omega}_2^2)(\tilde{k}_3^2 - \tilde{\omega}_3^2)}\,.\end{aligned} \tag{84c}$$

The sign of the integrals of equation (84) are not as easy to determine as that for equation (82). We can simplify these equations, again through the use of integration by parts and the property of the modal functions that $\Phi \to 0$ as $\tilde{y} \to \pm\infty$,

$$\int_{-\infty}^{\infty} \Phi_1 \partial_{\tilde{y}} F \mathrm{d}\tilde{y} = -\int_{-\infty}^{\infty} \Phi_{1,\tilde{y}} F \mathrm{d}\tilde{y}\,, \tag{85}$$

where $F = F(\tilde{y})$. When integrating equations (84), we can thus replace all terms of the form $\Phi_1 \partial_{\tilde{y}} F$ with $-\Phi_{1,\tilde{y}} F$, making the equations somewhat more transparent. Unfortunately, there was insufficient time to fully examine what conditions determine the sign of the integrals of equations (84).

We can thus write down the coupling coefficient for time evolution equation (79) of the amplitude of wave 1,

$$\delta_1 = \frac{\int_{-\infty}^{\infty} \Phi_1 \mathscr{N}_1^R + \Phi_1 \mathscr{N}_1^S + \Phi_1 \mathscr{N}_1^h \mathrm{d}\tilde{y}}{\int_{-\infty}^{\infty} \Upsilon_1 \mathrm{d}\tilde{y}} \,. \tag{86}$$

The coupling coefficients δ_2 and δ_3 can now be trivially gained by changing the labels appropriately in equation (86), which gives the standard, ordinary coupled differential equations for the time evolution of amplitude of a weakly non-linear triad interaction,

$$A_{1,\tilde{\tau}} = \delta_1 A_2^* A_3^* \tag{87a}$$

$$A_{2,\tilde{\tau}} = \delta_2 A_1^* A_3^* \tag{87b}$$

$$A_{3,\tilde{\tau}} = \delta_3 A_1^* A_2^* \,. \tag{87c}$$

The sign of the coupling coefficients is important, as, if they are single signed, then we shall encounter explosive instability (Coppi et al., 1969), which is an unphysical circumstance in the situation we are considering. We have already seen that the sign the integral of Υ_1, equation (82), is fixed. Thus, the sign of the coupling coefficients is determined by the sign of the sum of the integrals of equations (84), which is the numerator in equation (86).

In practise, it is not particularly practical to attempt to evaluate the coupling coefficient, equation (86), analytically and a future possibility would be to numerically integrate these equations.

6 Conclusion and Discussion

The governing equations for the equations on a Mercator projection centred about the equator were derived. We took inspiration from these equations to examine an "extended β-plane" in which we use the standard Cartesian, geopotential coordinates except employing the use of a tanh profile for the coriolis parameter, rather than the usual linear profile of the standard equatorial β-plane.

The properties of the solutions of such a system were investigated, with the finding that both a discrete and continuous spectrum exist for baroclinic waves, while only the discrete spectrum exists for barotropic waves. There exist many avenues for potential future study of these systems, a few of which shall now be discussed.

6.1 Potential Future Work

As with any interesting line of enquiry, we are left with more questions than answers. A brief list of potential future works, based upon this project is presented.

- Examine the properties of the Mercator projection including the metric terms, and compare with the solutions on a sphere, as described by Longuet-Higgins (1964, 1965).
- Examine the properties of interactions between a Rossby wave and a Yanai or Kelvin and the continuous spectrum.
- The derivation and numerical evaluation of the Manly-Rowe relations for the system of 3 baroclinic Rossby waves considered in section 5.3. Such an evaluation would help to ascertain whether any observed phenomena may be explained by such interactions (viz-a-viz the motivation given in section 1).
- It may be beneficial, and easier to conduct manipulations and further studies into these phenomena if the system were described using a Lagrangian or Hamiltonian method (for example, Ripa, 1981).
- The re-derivation of the system with a background shear (which is a more realistic set up for near equatorial oceanic dynamics).

Acknowledgements

I wish to acknowledge the large amounts of assistance given to me by Roger Grimshaw, without whom, this project would not have gotten very far. I also received help from Harvey Segur and Oliver Bühler, who showed an interest in the project and gave very sensible and helpful comments. I wish to thank Woods Hole Oceanographic Institute, as well as this year's programme directors, Oliver Bühler and Karl Helfrich for organising a great programme and giving me the opportunity to participate. I would like to acknowledge the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems, whose financial assistance made possible the trip from Australia to the US and back again. Finally, I would like to say thank you to all the participants in the GFD programme, both faculty and students, without whom, the programme would not be the same. In particular George Veronis, who took someone who watches cricket and got them playing (and enjoying) softball.

References

Aris, R., 1962. Vectors, Tensors and the Basic Equations of Fluid Mechanics. Dover Publications.

Benilov, E. S., Reznik, G. M., 1996. The complete classification of large-amplitude geostrophic flows in a two-layer fluid. Geophysical and Astrophysical Fluid Dynamics 82 (1–2), 1–22.

Boyd, J. P., 1980a. Equatorial solitary waves. part 1: Rossby solitons. Journal of Physical Oceanography 10 (11), 1699–1717.

Boyd, J. P., 1980b. The nonlinear equatorial Kelvin wave. Journal of Physical Oceanography 10 (1), 1–11.

Boyd, J. P., 1983. Equatorial solitary waves. part 2: Envelope solitons. Journal of Physical Oceanography 13 (3), 428–449.

Boyd, J. P., 1985. Barotropic equatorial waves: The nonuniformity of the equatorial beta-plane. Journal of the Atmospheric Sciences 42 (18), 1965–1967.

Bretherton, F. P., 1964. Resonant interactions between waves. the case of discrete oscillations. Journal of Fluid Mechanics 20 (3), 457–479.

Clarke, A. J., 2008. An Introduction to the Dynamics of El Niño & the Southern Oscillation. Academic Press.

Coppi, B., Rosenbluth, M. N., Sudan, R. N., 1969. Nonlinear interactions of positive and negative energy modes in rarefied plasmas. Annals of Physics 55, 207–247.

Drazin, P. G., Johnson, R. S., 1989. Solitons: an introduction. Cambridge University Press.

Düing, W., Hisard, P., Katz, E., Meincke, J., Miller, L., Moroshkin, K. V., Philander, G., Ribnikov, A. A., Voigt, K., Weisberg, R., 1975. Meanders and long waves in the equatorial Atlantic. Nature 275 (5524), 280–284.

Gill, A. E., 1982. Atmosphere-Ocean Dynamics. Vol. 30 of International Geophysics Series. Academic Press.

Gill, A. E., Clarke, A. J., 1974. Wind-induced upwelling, coastal currents and sea-level changes. Deep-Sea Research 21, 325–345.

Kennan, S. C., Flament, P. J., 2000. Observations of a tropical instability vortex. Journal of Physical Oceanography 30 (9), 2277–2301.

Kleeman, R., 2008. Stochastic theories for the irregularity of ENSO. Proceedings of the Royal Society A 366 (1875), 2511–2526.

Korteweg, D. J., de Vries, G., 1895. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. Philosophical Magazine 39, 422–443.

Lawrence, S. P., Angell, J. P., 2000. Evidence for Rossby wave control of tropical instability waves in the pacific ocean. Geophysical Research Letters 27 (15), 2257–2260.

Le Sommer, J., Reznik, G. M., Zeitlin, V., 2004. Nonlinear geostrophic adjustment of long-wave disturbances in the shallow-water model on the equatorial beta-plane. Journal of Fluid Mechanics 515, 135–170, DOI: 10.1017/S0022112004000229.

Legeckis, R., 1977. Long waves in the eastern equatorial Pacific Ocean: A view from a geostationary satellite. Science 197, 1197–1181.

Longuet-Higgins, M. S., 1964. Waves on a rotating sphere. Proceedings of the Royal Society A 279 (1379), 446–473.

Longuet-Higgins, M. S., 1965. Waves on a rotating sphere 2. Proceedings of the Royal Society A 284 (1396), 40–68.

Luke, J. C., 1966. A perturbation method for nonlinear dispersive wave problems. Proceedings of the Royal Society A 292 (1430), 403–412.

Lyman, J. M., Chelton, D. B., deSzoeke, R. A., Samelson, R. M., 2005. Tropical instability waves as a resonance between equatorial Rossby waves. Journal of Physical Oceanography 35 (2), 232–254.

Lyman, J. M., Johnson, G. C., Kessler, W. S., 2007. Distinct 17- and 33-day tropical instability waves in subsurface observations. Journal of Physical Oceanography 37 (4), 855–872, DOI:10.1175/JPO3023.1.

Matsuno, T., 1966. Quasi-geostrophic motions in the equatorial area. Journal of the Meteorological Society of Japan 44 (1), 25–42.

Miller, J. C. P., 1965. Parabolic cylinder functions. In: Abramowitz, M., Stegun, I. A. (Eds.), Handbook of Mathematical Functions. Dover Publications, New York, Ch. 19, pp. 685–720.

Oberhettinger, F., 1965. Hypergeometric functions. In: Abramowitz, M., Stegun, I. A. (Eds.), Handbook of Mathematical Functions. Dover Publications, New York, Ch. 15, pp. 555–566.

Pedlosky, J., 1987. Geophysical Fluid Dynamics. Springer-Verlag.

Philander, S. G. H., 1976. Instabilities of zonal equatorial currents. Journal of Geophysical Research 81 (21), 3725–3735.

Philander, S. G. H., 1978. Instabilities of zonal equatorial currents, 2. Journal of Geophysical Research 83 (C7), 3679–3682.

Philander, S. G. H., 1990. El Niño, La Niña, and the Southern Oscillation. Vol. 46 of International Geophysics Series. Academic Press.

Proehl, J. A., 1996. Linear stability of equatorial zonal flows. Journal of Physical Oceanography 26 (4), 601–621.

Reznik, G. M., Zeitlin, V., 2006. Resonant excitation of Rossby waves in the equatorial waveguide and their nonlinear evolution. Physical Review Letters 96 (3), 4, doi/10.1103/PhysRevLett.96.034502.

Reznik, G. M., Zeitlin, V., 2007a. Interaction of free Rossby waves with semi-transparent equatorial waveguide. Part 1. Wave triads. Physica D 226, 55–79.

Reznik, G. M., Zeitlin, V., 2007b. Resonant excitation and nonlinear evolution of waves in the equatorial waveguide in the presence of the mean current. Physical Review Letters 99 (6), 4, doi/10.1103/PhysRevLett.99.064501.

Ripa, P., 1981. On the theory of nonlinear wave-wave interactions among geophysical waves. Journal of Fluid Mechanics 103, 87–115.

Ripa, P., 1983a. Weak interactions of equatorial waves in a one-layer model. part I: General properties. Journal of Physical Oceanography 13 (7), 1208–1226.

Ripa, P., 1983b. Weak interactions of equatorial waves in a one-layer model. part II: Applications. Journal of Physical Oceanography 13 (7), 1227–1240.

Simmons, W. F., 1969. A variational method for weak resonant wave interactions. Proceedings of the Royal Society A 309, 551–577.

Stegun, I. A., 1965. Legendre functions. In: Abramowitz, M., Stegun, I. A. (Eds.), Handbook of Mathematical Functions. Dover Publications, New York, Ch. 8, pp. 331–354.

von Schuckmann, K., Brandt, P., Eden, C., 2008. Generation of tropical instability waves in the Atlantic Ocean. Journal of Geophysical Research 113 (C08034), 12, doi:10.1029/2007JC004712.

Whitham, G. B., 1974. Linear and Nonlinear Waves. John Wiley and Sons, New York.

Willett, C. S., Leben, R. R., Lavin, M. F., 2006. Eddies and tropical instability waves in the eastern tropical Pacific: A review. Progress in Oceanography 69 (2–4), 218–238.

Yang, H., 1987. Evolution of a Rossby wave packet in barotropic flows with asymmetric basic current, topography and δ-effect. Journal of Atmospheric Sciences 44 (16), 2267–2276.

Yu, J.-Y., Liu, W. T., 2003. A linear relationship between ENSO intensity and tropical instability wave activity in the eastern pacific ocean. Geophysical Research Letters 30 (14), 4, doi:10.1029/2003GL017176.

Degradation of the internal tide over long bumpy topography

Erinna Chen

1 Introduction

Internal waves in the ocean carry a significant amount of energy and are the main cause of diapycnal mixing in the ocean. Barotropic tidal currents flowing over topography generate a large number of these internal waves, also known as internal tides. From TOPEX/Poseidon altimetry data, it is estimated that up to 1 TW of power is transferred away from the barotropic tides in the deep ocean, primarily by internal tides [7].

The generation of the internal tides over topography has been extensively studied (e.g. [5], [2], [9], [3]) and it is suggested that in an ocean of finite-depth, the primary waves generated are those of low baroclinic modes [17]. However, mechanisms for internal-tide-driven mixing and energy dissipation such as shear instability, wave-wave interactions and parametric subharmonic instability are most effective at high (i.e. greater than wavenumber 10) modes. The transition from the low-mode internal tides to higher mode waves where mixing can occur is poorly understood. It is important to understand where these processes occur in that different parameterizations of internal wave dissipation may significantly alter the behavior of ocean circulation in models [8].

Low-mode internal tides have been observed to propagate over O(1000 km) distances from their generation site [6], [15]. It is likely that over this distance, the internal tides should encounter varying bottom topography that would scatter the low mode waves into higher modenumbers. Topographic scattering in the deep ocean has been studied extensively in two main contexts: generation of the internal tide and scattering of internal waves over finite-length topography (e.g. [1], [12], [13], [16]). However, little work has looked at internal wave propagation over topography with multiple bumps, random topography, or topography with long extent.

In this work, we study the degradation of the internal tide first by extending methods previously utilized to understand the conversion of the barotropic tide and also by using a ray tracing method. We primarily studied the behavior of the internal tide propagating over well-behaved "model" topographies; however, there are ways to extend this work to include more random topography.

2 Governing equations

We begin by considering the dynamics of an incompressible, hydrostatic, stratified fluid governed by the linear Boussinesq equations in two-dimensions (x and z):

$$u_t + P_x = 0, \tag{1}$$
$$w_t + P_z = b, \tag{2}$$
$$b_t + N^2 w = 0, \tag{3}$$
$$u_x + w_z = 0, \tag{4}$$

where N is the buoyancy frequency defined as

$$N^2 = -g\frac{d\ln\bar{\rho}_0}{dz}.$$

We will only consider constant bouyancy frequency. We employ subscript notation throughout to represent partial differentiation.

The momentum equations can be combined by differentiating (1) with respect to z, (2) with respect to x, and then adding, resulting in a vorticity equation:

$$(u_z - w_x)_t + b_x = 0. \tag{5}$$

By expressing the velocities u and w in terms of a streamfunction ψ where

$$u = \psi_z,\ w = -\psi_x, \tag{6}$$

the vorticity and buoyancy equations ((5), (3), respectively) can be combined into a single wave equation:

$$\left(\nabla^2\psi\right)_{tt} + N^2\psi_{xx} = 0 \tag{7}$$

2.1 Stable solutions to the wave equation

We now seek stable solutions to equation (7) of the form

$$\psi = \hat{\psi}(x,z)e^{-i\omega t}. \tag{8}$$

By substituting (8) into the wave equation (7), the streamfunction now satisfies

$$\hat{\psi_{zz}} = \left(\frac{N^2}{\omega^2} - 1\right)\hat{\psi_{xx}}. \tag{9}$$

This equation can be scaled and non-dimensionalized via the transform

$$x \to \frac{H}{\pi}\sqrt{\frac{N^2-\omega^2}{\omega^2}}x',\ z \to \frac{H}{\pi}z',$$

resulting in a scaled, non-dimensional form for the wave equation

$$\hat{\psi_{x'x'}} = \hat{\psi_{z'z'}}. \tag{10}$$

From this point forth, we will only be considering waves in their scaled, non-dimensional form and therefore, we drop the "prime" and "hat" notation.

While we have neglected to include the effects of rotation in deriving (10), rotation can easily be included and only appears in the scaling. The scaling is such that the wave crests propagate at 45°.

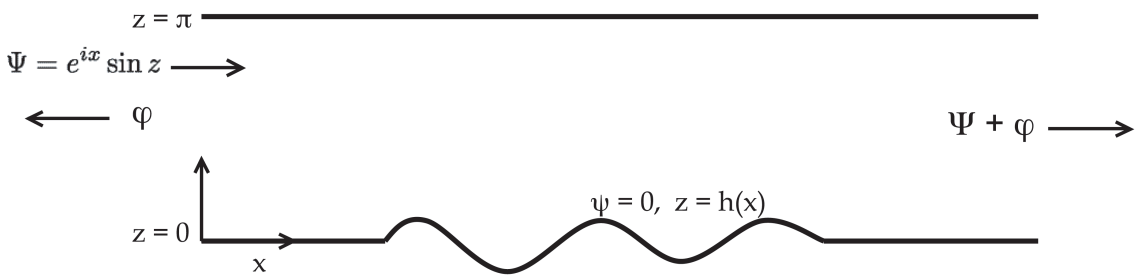


Figure 1: Schematic of problem setup. The ocean has finite depth, and in scaled coordinates z extends from 0 to π. The radiation conditions require that the scattered wave component φ propagate outwards. The internal tide Ψ is a mode-1 wave that enters from the left of the topography and propagates rightward over the topography. The topography is given as $z = h(x)$ where $h(x) = 0$ in the left and right far-fields.

3 Scattering of the internal tide over weak topography

The scattering of internal waves over topography has been studied extensively in the context of barotropic tidal conversion (e.g. [5], [9], [3], [10]). While some of these previous calculations can account for topography that is steep (i.e. slopes greater than 1) or has finite to large-amplitude, the most analytically tractable problem involves topography in the so-called "weak topography" limit. For the weak topography limit to be appropriate, the amplitude of the topography must be much small than the vertical scales of the waves and the topographic slopes are much less than the slope of a tidal beam ($s = 1$ in our reference frame). If this is the case, the boundary condition can be applied at the flat surface, instead of on the topography. We adopt a method of analysis similar to [5], but extend this for the internal tide propagating over weak topography. One notable difference in our setup is that while typically the barotropic tide is assumed to be an infinite source of energy, the internal tide has finite energy and will degrade as it propagates.

We begin by approximating the internal tide as a single, mode-1 wave given by

$$\Psi = e^{ix} \sin z. \tag{11}$$

The internal tide has finite amplitude (and energy), and our goal is to understand how this amplitude decays as a function of the bottom topography. Figure 1 is a schematic of the problem. The internal tide is assumed to propagate from left to right in the domain. The topography is given by $h(x)$, where $h(x) = 0$ in the far-fields (both right and left).

Throughout the domain, the streamfunction ψ can be expressed as

$$\psi = \Psi + \varphi, \tag{12}$$

where Ψ represents the internal tide and φ represents the scattered waves due to interaction with the topography. The scattered waves are subject to radiation conditions in the far-field, that is, on the left and right boundaries of the domain, the scattered waves must propagate

outwards. In addition, the streamfunction ψ is subject to the boundary conditions

$$\psi = 0 \text{ at } z = h(x), \tag{13}$$

$$\psi = 0 \text{ at } z = \pi. \tag{14}$$

Thus, this requires that along the bottom boundary

$$\varphi = -\Psi = -e^{ix}\sin(h(x)) \text{ at } z = h(x). \tag{15}$$

For weak topography, we can expand this boundary condition around $z = 0$, resulting in the $O(1)$ boundary condition

$$\varphi = -e^{ix}h(x) \text{ at } z = 0. \tag{16}$$

In order to understand how the mode-1 internal tide is scattered to higher modenumbers, we express the scattered wave component of the streamfunction as a Fourier series,

$$\varphi = \sum_{n=1}^{\infty} \varphi_n \sin nz. \tag{17}$$

The scattering coefficients φ_n can be obtained first by carrying out a Galerkin projection of φ_{zz} onto $\sin nz$.

$$\int_0^{\pi} \varphi_{zz} \sin nz \; dz = -nhe^{ix} - \frac{n^2\pi}{2}\varphi_n. \tag{18}$$

By integrating the left hand side of equation (18) and substituting for φ_{zz} using the wave equation (10), we obtain

$$\frac{d^2\varphi_n}{dx^2} + n^2\varphi_n = \frac{-2}{\pi}nhe^{ix}. \tag{19}$$

This differential equation can be solved using a Green's function

$$\frac{d^2 g_n}{dx^2} + n^2 g_n = \delta(x). \tag{20}$$

subject to the radiation conditions. Thus, g_n is given by

$$g_n = \frac{1}{2in}^{\,in|x|} \tag{21}$$

and the scattering coefficients φ_n are

$$\varphi_n(x) = \frac{-2n}{\pi}\int_{-\infty}^{\infty} g_n(x-x')h(x')e^{ix'}dx' = \frac{i}{\pi}\int_{-\infty}^{\infty} e^{in|x-x'|+ix'}h(x')dx'. \tag{22}$$

While equation (22) gives us information about the scattered wavefield over the entire domain, we are particularly interested in the scattered waves that propagate away from the topography. In the far fields, the scattered waves must satisfy the radiation conditions

$$\text{(left) } x \to -\infty : \varphi \to \sum_{n=1}^{\infty} a_n e^{-inx}\sin nz, \tag{23}$$

$$\text{(right) } x \to +\infty : \varphi \to \sum_{n=1}^{\infty} b_n e^{inx}\sin nz. \tag{24}$$

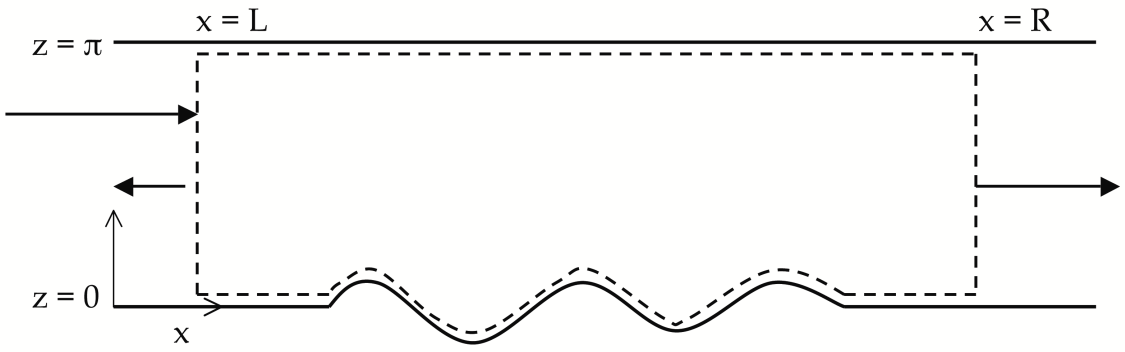


Figure 2: Energy flux is constant through the domain indicated by the dashed region. A finite amount of energy enters the domain via the internal tide on the left side of the topography and leaves the domain via waves in both directions.

By combining (22)-(24), we obtain the scattering coefficients in the far-fields

$$\text{(left) } x \to -\infty : \varphi_{Ln} = \frac{i}{\pi} \int_{-\infty}^{\infty} e^{i(n+1)x'} h(x') dx', \tag{25}$$

$$\text{(right) } x \to +\infty : \varphi_{Rn} = \frac{i}{\pi} \int_{-\infty}^{\infty} e^{-i(n-1)x'} h(x') dx'. \tag{26}$$

This solution is known in other contexts as Bragg scattering. Knowing the scattering coefficients now allows us to calculate the amount of energy converted from the internal tide to higher mode numbers.

3.1 Scattered wave energy flux

The full streamfunction ψ satisfies

$$\overrightarrow{\nabla} \cdot \overrightarrow{\mathcal{J}} = 0, \tag{27}$$

where $\mathcal{J}$ represents the flux and is defined as

$$\overrightarrow{\mathcal{J}} = i \left[(\psi \psi_x^* - \psi^* \psi_x), -(\psi \psi_z^* - \psi^* \psi_z) \right]. \tag{28}$$

Integrating over the domain $x = [L, R]$ and $z = [0, \pi]$ as illustrated in figure 2, we find that

$$\int_0^\pi (\psi \psi_x^* - \psi^* \psi_x)\, dz \Big|_{x=R} - \int_0^\pi (\psi \psi_x^* - \psi^* \psi_x)\, dz \Big|_{x=L} = 0, \tag{29}$$

that is the horizontal energy flux through the domain is zero.

If we now substitute $\psi = \Psi + \varphi_L$ at $x = L$ and $\psi = \Psi + \varphi_R$ at $x = R$ into equation (29), we find that the energy loss from the internal tide is a function of the scattered wave coefficients,

$$\Delta E_{IT} = -\frac{1}{2} \sum_{n=1}^{\infty} n \left(|\varphi_{Rn}|^2 + |\varphi_{Ln}|^2 \right). \tag{30}$$

3.2 Scattered wave solutions for model topographies

For the topographies

$$h(x) = \frac{1}{2}\epsilon \cos kx$$
$$h(x) = \frac{1}{2}\epsilon(\cos kx + 1),$$

the integrals for the scattering coefficients (25, 26) can be evaluated analytically. Assuming the topography is located only on the interval $[-b\pi, b\pi]$, vanishes at $x = -b\pi$ and $x = b\pi$, and is zero elsewhere, then the scattering coefficients are

$$\varphi_{Ln} = \frac{i\epsilon}{4\pi} \int_{-b\pi}^{b\pi} e^{i(n+1)x'} \left(e^{ikx'} + e^{-ikx'} + 2\alpha \right) dx' \tag{31}$$

$$\varphi_{Rn} = \frac{i\epsilon}{4\pi} \int_{-b\pi}^{b\pi} e^{-i(n-1)x'} \left(e^{ikx'} + e^{-ikx'} + 2\alpha \right) dx' \tag{32}$$

where $\alpha = 0$ or 1.

For the model topographies, this solution suggests that there are a limited number of scattered waves. To the right of the topography, the scattered waves occur as a $n = 1$ mode and a $n = k + 1$ mode. To the left of the topography, there is a single scattered wave, the $n = k - 1$ mode. The amplitudes of these waves depend on the length of the topography, b, and the height of the topography, ϵ,

$$\begin{aligned} |\varphi_{R1}| &= \alpha\epsilon b \\ |\varphi_{R,k+1}| &= \frac{1}{2}\epsilon b \\ |\varphi_{L,k-1}| &= \frac{1}{2}\epsilon b. \end{aligned} \tag{33}$$

From (30), the energy drained from the internal tide is the sum of the product of the wavenumbers and their corresponding wave amplitude,

$$\Delta E_{IT} = -\frac{1}{2}\left(\alpha\epsilon b + \frac{1}{2}(k+1)\epsilon b + \frac{1}{2}(k-1)\epsilon b \right). \tag{34}$$

However, the amount of energy drained from the internal tide is not finite in this solution. For fixed ϵ, the drained energy grows unbounded with the length of the topography, b.

4 Scattered wave solutions using a Green's function approach

Equation (34) suggests that our solution for the scattered wave coefficients has a limited range of validity. This may be due to an inconsistency in the boundary condition expansion. To $O(\epsilon)$ the bottom boundary condition is

$$-\epsilon h e^{ix} = \varphi(x, 0) + \epsilon h \varphi_z(x, 0). \tag{35}$$

Previously, the solution in §3 did not include the second term on the right hand side.

To rectify this problem, we now incorporate the fully non-linear bottom boundary condition (15) by utilizing a method previously used in calculations of wave scattering off finite and supercritical topography (e.g. [3], [14]). We define a Green's function as

$$G(x-x',z,z') = \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}e^{in|x-x'|}\sin nz\sin nz' \tag{36}$$

The scattered component of the streamfunction is given by

$$\varphi = \int_{-a}^{a}\gamma(x')G(x-x',z,h(x'))dx' \tag{37}$$

where the topography lies between $[-a,a]$ and $\gamma(x)$ is the source density function representing the topography. $\gamma(x)$ is found through the solution of a matrix problem incorporating the bottom boundary condition

$$-e^{ix}\sin h(x) = \int_{-a}^{a}\gamma(x')G(x-x',h(x),h(x'))dx'. \tag{38}$$

Once we have the source density function, the full solution for the scattered wave field is given by

$$\varphi(x,z) = \int_{-a}^{a}\gamma(x')\sum_{n=1}^{\infty}\frac{1}{\pi n}e^{in|x-x'|}\sin nz\sin nh(x')dx' \tag{39}$$

The scattered wave field to the left of the topography is

$$\varphi_L(x,z) = \sum_{n=1}^{\infty}e^{-inx}\sin nz\frac{1}{\pi n}\int_{-a}^{a}\gamma(x')e^{inx'}\sin nh(x')dx'. \tag{40}$$

The corresponding modal scattering coefficients are thus

$$\varphi_{Ln}(x) = e^{-inx}\frac{1}{\pi n}\int_{-a}^{a}\gamma(x')e^{inx'}\sin nh(x')dx'. \tag{41}$$

Similarly, to the right of the topography, the scattered wave field is

$$\varphi_R(x,z) = \sum_{n=1}^{\infty}e^{inx}\sin nz\frac{1}{\pi n}\int_{-a}^{a}\gamma(x')e^{-inx'}\sin nh(x')dx', \tag{42}$$

and the corresponding scattering coefficients are

$$\varphi_{Rn}(x) = e^{inx}\frac{1}{\pi n}\int_{-a}^{a}\gamma(x')e^{-inx'}\sin nh(x')dx'. \tag{43}$$

For this formulation, the scattering coefficients must be solved for numerically. The numerical code we employ takes an input for the topography $h(x)$, solves for the source function $\gamma(x)$ and then solves for the scattering coefficients. The Green's function defined in (36) has the limitation that the topography must lie above the flat bottom ($z = 0$). However, by using the Green's function approach, the topography is not limited to the weak topography regime and thus, may be used to model more "realistic" topography than our model topography.

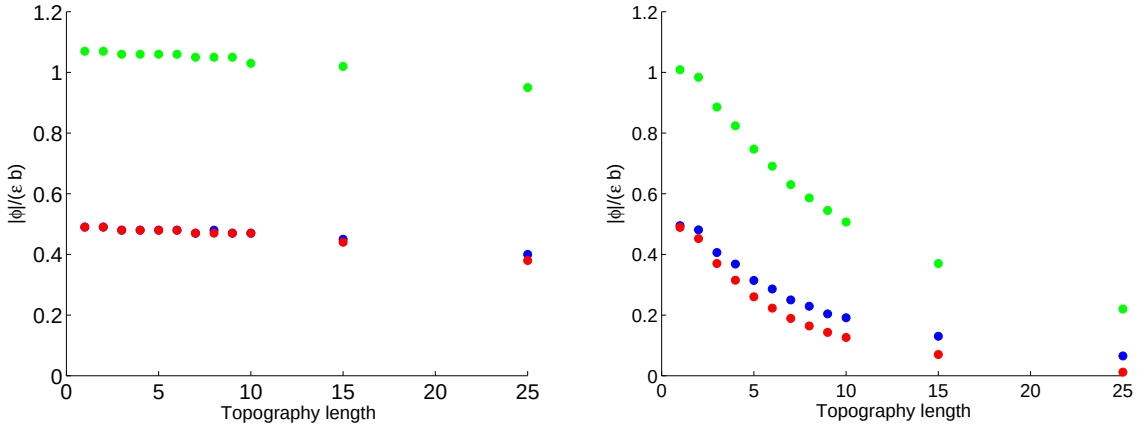


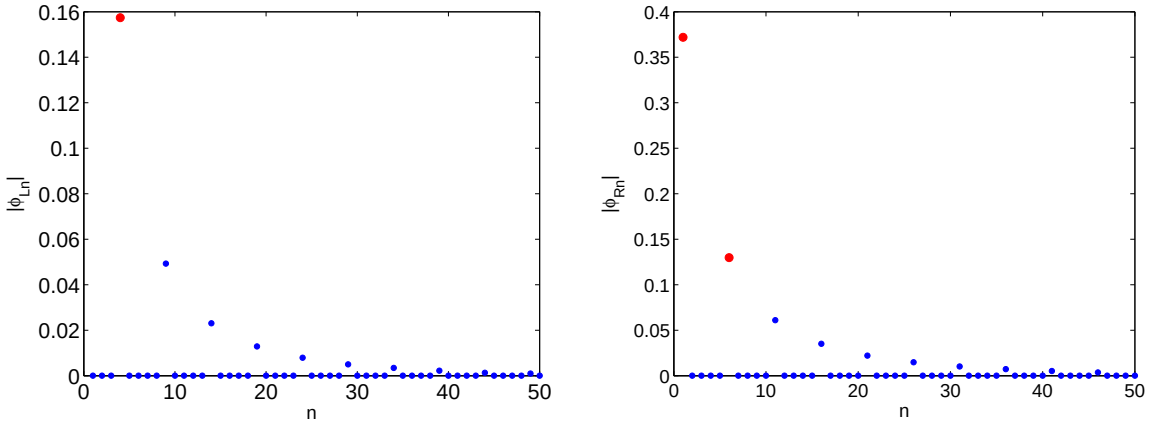
Figure 3: Scattering coefficients solved for using the Green's function method for the predicted modes from (33), φ_{R1}(green), $\varphi_{R,k+1}$(red), $\varphi_{L,k-1}$(blue), shown with $k = 5$. Solution from (33) suggests that $\frac{\varphi_n}{\epsilon b}$ should remain constant, independent of topography length. (a) Coefficients for $\epsilon = 0.01$ (b) Coefficients for $\epsilon = 0.1$

4.1 Numerical solutions for the wave equation

For the model topography $h(x) = \frac{1}{2}(\cos kx + 1)$, the scattered wave solution found in §3.2 says that there are only three scattered waves and that the parameter $\frac{\varphi_n}{\epsilon b}$ should remain constant. As stated before, the range of validity for this solution must be limited because the internal tide has finite energy. Figures 3 and 4 show results of solutions from the Green's function approach for variable length topography.

Figure 3 shows the scattering coefficients for the three predicted modes from (33), i.e. the modes φ_{R1}, $\varphi_{R,k+1}$, and $\varphi_{L,k-1}$. For figure 3a, the topography is extremely weak, ϵ equals 0.01. In this case, the prediction that $\frac{\varphi_n}{\epsilon b}$ remains constant seems to hold even for fairly long topography. However, for topography of higher amplitude (figure 3b), the Green's function solution diverges significantly from the prediction even for short topographic lengths (i.e. 2-3 wavelengths).

Figure 4 shows the full spectrum of the scattering coefficients for topography with $\epsilon = 0.1$. The peaks of the spectrum are the three modes predicted from the linear solution (shown in red); however, there is a significant amount of energy also contained in higher harmonic modes. The harmonic modes are the $nk - 1$ modes on the left of the topography and the $nk + 1$ modes on the right of the topography, where $n = 2, 3, \ldots$. The divergence of the actual solution and the predicted solution in figure 3b is likely due to the presence of these higher-order harmonics. As the topography gets longer, more energy from the internal tide is lost to these harmonics.



(a) (b)

Figure 4: Scattering coefficients solved for using the Green's function method for $h(x) = \frac{1}{2}(0.1)(\cos 5x + 1)$. Red dots indicate modes predicted from (33). (a) Scattering coefficients to the left of the topography (b) Scattering coefficients to the right of the topography

5 Multiscale asymptotic analysis

While the Green's function method may allow us to incorporate the fully nonlinear bottom boundary condition, but solving for the source function is computationally expensive. The scattered wave solution given in §3.2 suggests that the range of validity for our solution may be extended if we introduce another length scale.

In the classic fashion, we introduce a long horizontal scale

$$X = \epsilon x \tag{44}$$

such that

$$\frac{\partial}{\partial x} \rightarrow \frac{\partial}{\partial x} + \epsilon \frac{\partial}{\partial X}. \tag{45}$$

The streamfunction can be written in expanded form as

$$\psi = \psi^{(0)}(x, z, X) + \epsilon \psi^{(1)}(x, z, X) + O(\epsilon^2) \tag{46}$$

where the superscripts represent the order of the expansion. Plugging (46) into the wave equation, the $O(1)$ equation on ψ is

$$\psi^{(0)}_{xx} - \psi^{(0)}_{zz} = 0. \tag{47}$$

Thus to zeroth order, the streamfunction satisfies the wave equation. In order to solve for $\psi^{(0)}(x, z, X)$, we must expand to next order in ϵ. To $O(\epsilon)$, ψ satisfies

$$\psi^{(1)}_{xx} - \psi^{(1)}_{zz} = -2\psi^{(0)}_{xX}. \tag{48}$$

We prescribe that the solvability condition on ψ is that the $O(1)$ solution is periodic on the fast length scale x. We then expand $\psi^{(0)}$ as a Fourier series of the form

$$\psi^{(0)}(x,z,X) = \sum_{n=1}^{\infty} \left(T_n(X)e^{inx} + R_n(X)e^{-inx}\right) \sin nz \tag{49}$$

where T_n are the transmitted waves (equivalent to φ_{Rn} on the far right) and R_n are the reflected waves (equivalent to φ_{Ln} on the far left). This expansion of $\psi^{(0)}$ satisfies the $O(1)$ boundary condition on the bottom

$$\psi^{(0)}(x,0,X) = 0. \tag{50}$$

Substituting for $\psi^{(0)}$ in (48) gives

$$\psi^{(1)}_{xx} - \psi^{(1)}_{zz} = -2i \sum_{n=1}^{\infty} n \left(T_n(X)e^{inx} + R_n(X)e^{-inx}\right) \sin nz. \tag{51}$$

The coefficients T_n and R_n are a function of the bottom topography and can be solved for by manipulating equation (51) and applying the bottom boundary condition. To solve for T_n, we multiply both sides of (51) by $e^{-inx}\sin nz$ and integrate over the domain. Because of orthogonality and periodicity, the right hand side of the equation reduces to

$$\int_{-\pi}^{\pi}\int_{0}^{\pi} -2i \sum_{m=1}^{\infty} m \left(T_m(X)e^{imx} + R_n(X)e^{-imx}\right) \sin mz \sin nz \; e^{-inx} dz \; dx = -i2\pi^2 n T_{nX}. \tag{52}$$

The left hand side simplifies to

$$\int_0^{\pi}\int_{-\pi}^{\pi} \psi^{(1)}_{xx} - \psi^{(1)}_{zz} dx dz = -\int_{-\pi}^{\pi} n\psi^{(1)}(x,X,0)e^{-inx} dx \tag{53}$$

The $O(\epsilon)$ boundary condition that must be satisfied is

$$\psi^{(1)}(x,0,X) = -h(x)\psi^{(0)}_z(x,0,X). \tag{54}$$

By combining (52)-(54), we obtain the equation for the evolution of the transmitted wave coefficients,

$$T_{nX} = \frac{i}{2\pi^2}\int_{-\pi}^{\pi} h(x)e^{-inx} \sum_{m=1}^{\infty} \left(T_m e^{imx} + R_m e^{-imx}\right) dx. \tag{55}$$

Similarly, the reflected wave coefficients, R_n, can be obtained by multiplying (51) by $e^{inx}\sin nz$ and integrating over the domain. The reflected wave coefficients satisfy

$$R_{nX} = \frac{i}{2\pi^2}\int_{-\pi}^{\pi} h(x)e^{inx} \sum_{m=1}^{\infty} \left(T_m e^{imx} + R_m e^{-imx}\right) dx. \tag{56}$$

Equations (55) and (56) suggest that the long-length scale evolution of the transmitted and reflected waves are coupled to each other by their short-scale behavior over the topography.

5.1 Solutions for model topographies

In the case of our model topographies, the evolution equations (55) and (56) take on a form that is readily solvable.

5.1.1 General solutions for $k > 0$

Substituting our model topography,

$$h(x) = \frac{1}{2}(\cos kx + \alpha) \text{ where } \alpha = 0 \text{ or } 1,$$

into (55) and (56), the transmitted waves and reflected waves are given by

$$T_{nX} = \frac{i}{4\pi}\left[2n\alpha T_n + (n-k)T_{n-k} + (n+k)T_{n+k} + (k-n)R_{k-n}\right], \tag{57}$$

$$R_{nX} = \frac{i}{4\pi}\left[2n\alpha R_n + (n-k)R_{n-k} + (n+k)R_{n+k} + (k-n)T_{k-n}\right]. \tag{58}$$

The evolution equations (57) and (58) predict the presence of the higher-order harmonics seen in figure 4. Since the only incoming wave is $n = 1$ internal tide, for small X, the only coupling occurs between the $n = 1$ and itself, the $n = 1$ wave and a transmitted $n = k + 1$ wave, and the $n = 1$ wave and a reflected $n = k-1$ wave, as predicted by the linear solution. However, over long topographic length scales, these incited waves will couple to higher order wave numbers and energy will be transferred from the internal tide to other mode numbers by this coupling.

In general, the evolution equations (57) and (58) are a set of coupled linear ordinary differential equations and can be solved by finding the eigenvalues of a system-defining matrix. All the eigenvalues for this matrix are complex, thus the solution is stable. We do not tackle this problem here, but instead focus on a single value for k, $k = 1$.

5.1.2 Analytical solutions for $k = 1$ using a generating function

In the $k = 1$ case, the transmitted and reflected waves are decoupled from each other. Therefore, since there are no reflected waves to begin with, the long length scale evolution of the system only consists of transmitted waves. Furthermore, the evolution equation for the transmitted waves can be solved analytically using a generating function.

We define a generating function, E, as

$$E(X, s) = \sum_{n=1}^{\infty} T_n(X)s^n. \tag{59}$$

If E has an analytical solution, then the coefficients for the transmitted waves are simply

$$T_n(X) = \left.\frac{\partial^{(n)}E}{\partial s^{(n)}}\right|_{s=0}. \tag{60}$$

For $k = 1$, the long length scale evolution of the transmitted waves is given by

$$T_{nX} = \frac{i}{4\pi}\left[2\alpha n T_n + (n-1)T_{n-1} + (n+1)T_{n+1}\right] \tag{61}$$

By multiplying by s^n and summing over n, equation (61) becomes

$$\sum_{n=1}^{\infty} T_{nX} s^n = \frac{i}{4\pi} \sum_{n=1}^{\infty} s^n \left[2nT_n + (n-1)T_{n-1} + (n+1)T_{n+1} \right], \tag{62}$$

and the evolution of the transmitted waves can be expressed in terms of the generating function,

$$E_X = \frac{i}{4\pi} \left(E_s(1 + 2\alpha s + s^2) - T_1 \right). \tag{63}$$

Now to solve for E, we transform E into a function $F(s, X)$,

$$E \to F + \frac{i}{4\pi} \int_0^X T_1(\tilde{X}) d\tilde{X} \tag{64}$$

such that equation (63) can be written as

$$F_X = \frac{i}{4\pi}(1 + 2\alpha s + s^2) F_s. \tag{65}$$

Because

$$\frac{\partial^{(n)} E}{\partial s^{(n)}} = \frac{\partial^{(n)} F}{\partial s^{(n)}}, \tag{66}$$

we only need to solve for F to determine the behavior of the transmitted waves. By a change of variables, the solution to equation (65) is given by

$$F(s, X) = F_0(\zeta - X) \tag{67}$$

where

$$\frac{\partial s}{\partial \zeta} = \frac{i}{4\pi}(1 + 2\alpha s + s^2). \tag{68}$$

The boundary condition at the far left of the domain ($X = 0$) is that the only transmitted wave is the internal tide (i.e. $T_n(0) = 0$ for $n \neq 1$),

$$F(s, 0) = \sum_{n=1}^{\infty} s^n T_n(0) = s. \tag{69}$$

F can be obtained by solving for s in equation (68).

Solution for $h(x) = \frac{1}{2}\cos x$

For the topography $h(x) = \frac{1}{2}\cos x$, $\alpha = 0$ and equation (68) becomes

$$\frac{\partial s}{\partial \zeta} = \frac{i}{4\pi}(1 + s^2). \tag{70}$$

The solution to equation (70) is

$$s = \tan\left(\frac{i\zeta}{4\pi}\right) \text{ or } \zeta = -4\pi i \tan^{-1} s. \tag{71}$$

Substituting the above solution into the left boundary condition (69), we obtain

$$F(s,0) = s = F_0(\zeta) = \tan\left(\frac{i\zeta}{4\pi}\right) \tag{72}$$

and thus,

$$F_0(\zeta - X) = \tan\left(\frac{i(\zeta - X)}{4\pi}\right). \tag{73}$$

We obtain F by substituting for ζ in equation (73),

$$F(s,X) = \frac{s - i\tanh\frac{X}{4\pi}}{1 + is\tanh\frac{X}{4\pi}}. \tag{74}$$

By combining equations (60) and (66), we obtain the transmitted wave coefficients,

$$T_n(X) = (-i)^{(n-1)}\left(1 - \tanh^2\left(\frac{X}{4\pi}\right)\right)\tanh^{(n-1)}\left(\frac{X}{4\pi}\right) \tag{75}$$

For increasing X, the transmitted wave coefficients decay exponentially.

Solution for $h(x) = \frac{1}{2}(\cos x + 1)$

For the topography $h(x) = \frac{1}{2}(\cos x + 1)$, s is given by

$$\frac{\partial s}{\partial \zeta} = \frac{i}{4\pi}(1 + 2s + s^2) = \frac{i}{4\pi}(1+s)^2, \tag{76}$$

whose solution is

$$s = \frac{4\pi i}{\zeta} - 1 \text{ or } \zeta = \frac{4\pi i}{1+s}. \tag{77}$$

Satisfying the boundary condition on F,

$$F_0(\zeta) = \frac{4\pi i}{\zeta} - 1. \tag{78}$$

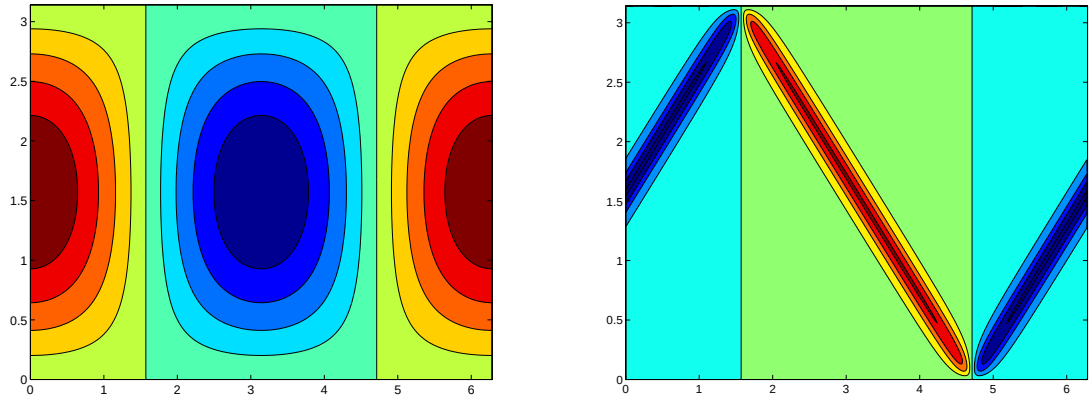
By substituting $\zeta - X$ for ζ in (78) and for ζ from (77), we obtain F

$$F(s,X) = \left(\frac{1}{1+s} + \frac{iX}{4\pi}\right)^{-1} - 1. \tag{79}$$

The transmitted wave coefficients from equation (60) are

$$T_n(X) = \frac{(-1)^{n-1}\left(\frac{iX}{4\pi}\right)^{n-1}}{\left(1 + \frac{iX}{4\pi}\right)^{n+1}}. \tag{80}$$

As X increases, the transmitted waves decay algebraically, not exponentially, as in the case for the topography $h(x) = \frac{1}{2}\cos x$. The choice of α was initially arbitrary, but the behavior of the solutions seem to diverge significantly. However, in both cases, for large values of X, the waves begin to focus in narrow bands, (see figure 5). The behavior of these solutions for large X is better understood when we move from solutions derived from the streamfunction construction to solutions of the wave equation using the method of characteristics (otherwise known as ray tracing).



(a) (b)

Figure 5: Contours of the real part of the streamfunction ψ for $h(x) = \frac{1}{2}\cos x$. (a) Streamfunction to the left of the topography. (b) Streamfunction to the right of the topography after 25 topographic wavelengths.

6 Solutions using the method of characteristics

Because the wave equation (7) is hyperbolic, it can also be solved using the method of characteristics. The general solution to the wave equation is

$$\psi = f(x - z) + g(x + z). \tag{81}$$

We define upward characteristics as lines of constant $x - z$ and downward characteristics as lines of constant $x + z$, i.e. characteristics occur along 45° angles (see figure 6). Along these characteristics, we can define Riemann invariants

$$\begin{aligned} R^+ &= \psi_z - \psi_x = u + w, \\ R^- &= \psi_z + \psi_x = u - w, \end{aligned} \tag{82}$$

where $R^+(R^-)$ represents the Riemann invariant on an upward (downward) characteristic.

Characteristics carry a Riemann invariant until they encounter the top or bottom boundary. Along flat boundaries, the boundary condition is such that there is no vertical velocity, $w = 0$. From the definition of the Riemann invariants (82), at a flat boundary (grey region in figure 6), an upward (or downward) characteristic will simply reflect from the boundary, and carry the same Riemann invariant along with it,

$$R^+(x_0, z = \pi) = R^-(x_0, z = \pi). \tag{83}$$

For subcritical topography, when downward characteristics encounter topography (blue region in figure 6) they reflect forwards, but the Riemann invariant is altered. The bottom boundary condition is that there is no normal velocity to the topography,

$$\frac{w}{u} = h_x. \tag{84}$$

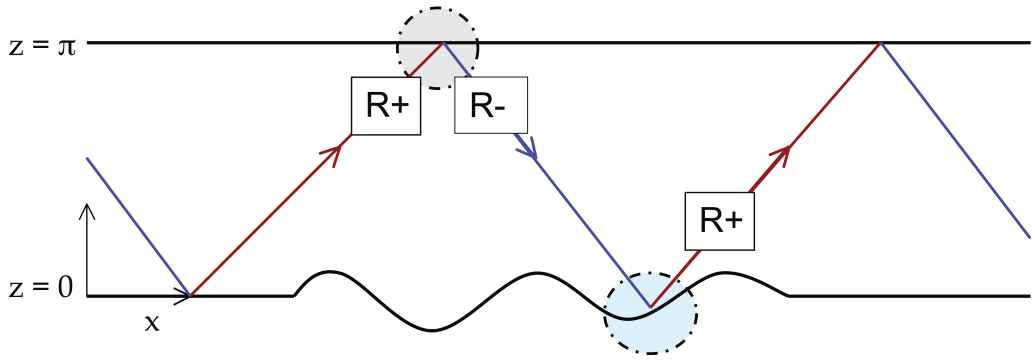


Figure 6: Schematic of tracing characteristics. There is a Riemann invariant $R^+(R^-)$ on upward(downward) characteristics. At a flat boundary, indicated by the grey region, the value of the Riemann invariant is unaffected by reflection ($R^- = R^+$). When a characteristic reflects off topography, e.g. the blue region, the Riemann invariant carried by the characteristic is changed according to equation (85).

The Riemann invariant carried by the reflected upward-going characteristic is then

$$R^+ = R^- \frac{(1+h_x)}{(1-h_x)}. \tag{85}$$

In our scaled setup, characteristics that do not encounter topography are 2π-periodic. A characteristic that originates at $(x_0, 0)$ will reflect at the top boundary at $(x_0 + \pi, \pi)$ and return to the bottom boundary at $(x_0 + 2\pi, 0)$. A characteristic that does encounter topography will have its wavelength changed as seen in figure 7. The wavelengths of characteristics that hit the topography where $h > 0$ will shorten, while wavelenths of those characteristics that hit the topography where $h < 0$ will lengthen.

6.1 Solutions for the wave field using Riemann invariants

Characteristics in our reference frame are 2π-periodic, and thus the field determined by the Riemann invariants are also 2π-periodic. If we define a set of Riemann invariants on an interval of length 2π, the wave field can be determined everywhere. The slopes of the topography must remain subcritical for this to be the case; the mapping of characteristics and the definition of the interval over which Riemann invariants must be defined becomes more difficult for slopes that are supercritical (i.e. characteristics can reflect backwards) [12],[13].

Before the topography, the wave field consists only of the internal tide. We define a web of characteristics as a set of upward characteristics uniformly spaced along a flat bottom boundary. The wave field consists of only the internal tide, so the Riemann invariants along the bottom boundary are

$$R^+(x_0) = e^{-ix_0} \tag{86}$$

where x_0 is the origin of the upward characteristic and $x_0 \in [0, 2\pi]$. The value of the Riemann invariants will be altered according to (85) where h_x is evaluated at the point x

such that

$$h(x) = -x + x_0 + 2\pi. \tag{87}$$

The Riemann invariants to the right of the topography are

$$R^+(x_0)_R = \prod_{i=1}^{b} \frac{(1 + h_x(x_i))}{(1 - h_x(x_i))} R^+(x_0)_L, \tag{88}$$

where b is the number of reflections off the bottom (also the number of topographic wavelengths), x_i is the location of the reflection and $x_0 \in [0, 2\pi]$. In the weak topography limit, we can expand equations (85) and (87). The Riemann invariants after the topography are then given by

$$R^+(x_0)_R = (1 + 2\epsilon b h(x_0)) e^{-i(x_0 + 2\epsilon h(x_0))}. \tag{89}$$

Equation (89) recovers the weak topography solution from §3.2.

6.2 Fixed points for model topographies with $k = 1$

We can now determine the behavior of characterstics for our model topographies when $k = 1$. Characteristics that hit the bottom boundary where $h = 0$ will be unchanged in wavelength. Since $\cos x$ is 2π-periodic, these characteristics will reflect at the same relative place on the topography. For those that do not reflect at z=0, the wavelength of the characteristic will change.

Thus, we can characterize the zeros of the topography as either stable or unstable fixed points. Figure 8 shows the behavior of the fixed points for the model topographies, where green indicates a stable fixed point and red indicates and unstable fixed point. For the topography $h(x) = \frac{1}{2}\cos\ x$ in figure 8a, characteristics move towards zeros on positive slopes and move away from zeros on negative slopes. For a stable fixed point located at $(x_s, 0)$, characteristics reflecting at $x_1 > x_s$ will be shortened because $h(x_1) > 0$ and the subsequent bottom reflection will occur at $x_1 > x_2 > x_s$. For multiple reflections, the characteristics will be shortened until the characteristic arrives at $x = x_s$. Characteristics reflecting at $x_1 < x_s$ will be lengthened such that a subsequent reflection will occur at $x_1 < x_2 < x_s$. For an unstable fixed point, x_{us}, characteristics reflecting from $x_1 > x_{us}$ will be lengthened so that $x_1 < x_2$ since $h(x_1) < 0$. Thus, characteristics will move away from the unstable fixed point.

For the topography $h(x) = \frac{1}{2}(\cos\ x + 1)$ in figure 8b, the fixed points are degenerate (two fixed points become a single fixed point). Because the topography is always greater than zero, characteristics not reflecting from a fixed point always shorten. As a result, characteristics will move towards the fixed point on positive slopes and away from the fixed point on negative slopes.

We now interpret the results of §5.1.2 in the context of fixed points of the topography. If we begin to the left of the topography with a web of characteristics that is uniformly distributed along the flat bottom as shown in figure 9a, when these characteristics reflect off topography, their lengths will be altered and the uniform web will become distorted. Subsequent bounces will continue to warp the web such that characteristics will move away from the unstable fixed points and move towards the stable fixed points. After a large number of reflections (equivalent to large X in §5.1.2), the characteristics will be located

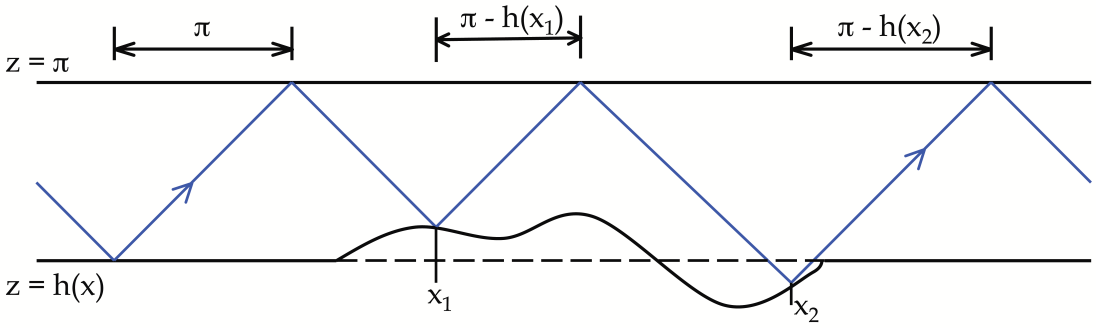


Figure 7: Behavior of characteristics when reflecting off topography. For reflections off the flat bottom ($z = 0$), the wavelength of the characteristics is 2π. When there is positive topography ($h(x) > 0$), the characteristics will shorten by $h(x)$ where x is the reflection point (shown here as x_1). For negative topography, the characteristics will lengthen by $|h(x)|$, (shown as x_2).

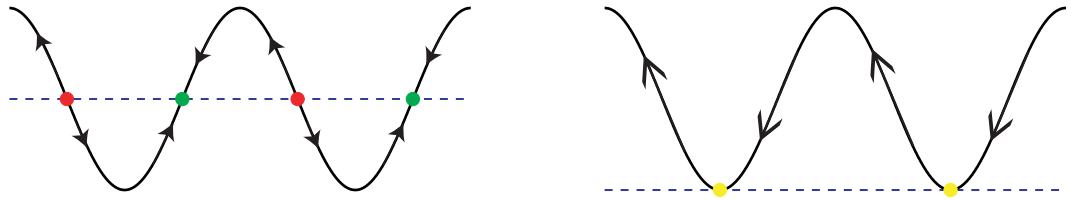
in a narrow band around the fixed point as in figure 9b, and the field carried by these characteristics will focus into the narrow band seen in figure 5b.

Recall that the asymptotic solutions for large X differed between the two model topographies; the solution for $h(x) = \frac{1}{2}\cos x$ decayed exponentially while the solution for $h(x) = \frac{1}{2}(\cos x + 1)$ decayed algebraically. The decay behavior is dictated by how rapidly characteristics converge on the fixed points. In the case of the topography $h(x) = \frac{1}{2}\cos x$, characteristics are never more than a distance of π away from a stable fixed point. However, for the topography $h(x) = \frac{1}{2}(\cos x + 1)$, characteristics can be located a distance of up to 2π away from the nearest stable fixed point and as result, characteristics converge on the stable fixed point more slowly.

6.3 Characteristic mapping and solutions for model topographies with $k \neq 1$

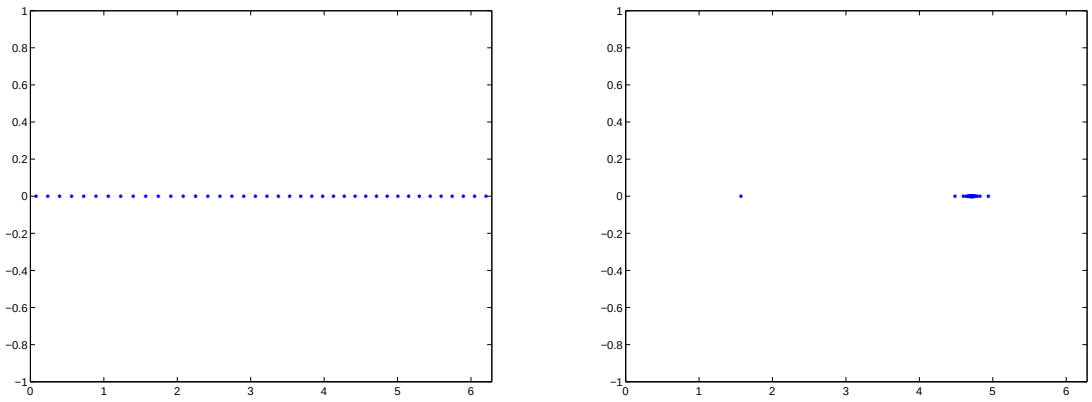
Up to this point, we have looked primarily at solutions for the model topographies with $k = 1$. With $k \neq 1$, as seen from the evolution equations, the wave field has reflected waves that propagate leftwards. The Riemann invariants before the topography are not given simply by (86); they also include a (unknown) reflected component that must be solved for as well. Solving for the scattered wave components is not a trivial task. The transmitted waves and reflected waves are coupled; however, for low values of k, the effect of this coupling on the transmitted waves should be minimal. Therefore, the behavior of the transmitted waves should behave similarly to the $k = 1$ case. The characteristics should converge on stable fixed points.

To find the stable fixed points, we can use a 1-D map (figure 10a) to determine the location of the next reflection point given a starting point of x_0. In the map shown in figure 10a, the blue curve plots $x = h(x)$ and the green curve plots $x = x$. The procedure to determine the fixed points are:



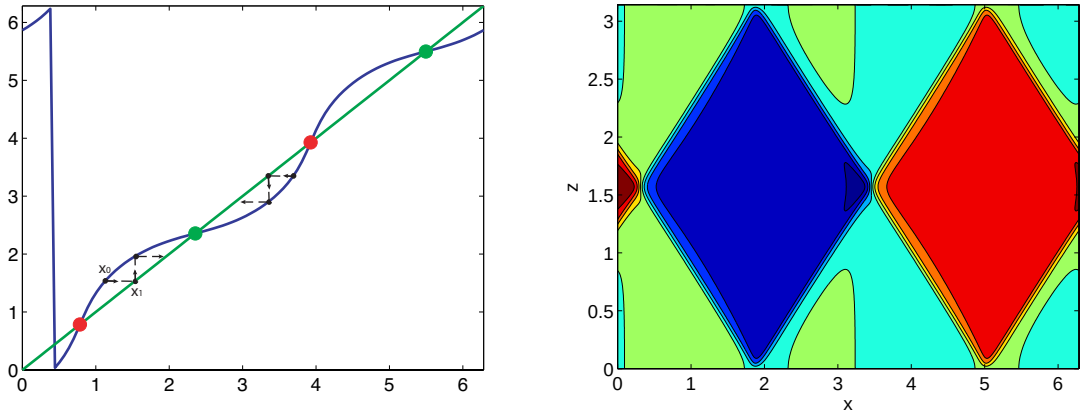
(a) (b)

Figure 8: Behavior of fixed points for (a) $h(x) = \frac{1}{2}\cos x$ and (b) $h(x) = \frac{1}{2}(\cos x + 1)$. The blue dotted line indicates $z = 0$. For the topography in (a), on positive slopes, zeros are stable fixed points (indicated in green). Characteristics will shorten when reflecting to the right of the topography ($h(x) > 0$) and will lengthen when reflecting to the left ($h(x) < 0$). The red points indicate unstable fixed points, and characteristics will move away from them. For the topography in (b), the stable and unstable fixed points collapse onto each other and form a degenerate fixed point (indicated in yellow). Characteristics will focus from the right of the fixed point and diverge to the left.



(a) (b)

Figure 9: (a) Origination points for a uniform web of characteristics, defined on the interval $[0, 2\pi]$ to the left of the topography. (b) Web of characteristics after 25 reflections off the topography $h(x) = \frac{1}{2}\cos x$. Characteristics have focused around the stable fixed point.



(a) (b)

Figure 10: Behavior of transmitted waves for the topography $h(x) = \frac{1}{2}(\cos 2x + 1)$. (a) 1-D map used to determine fixed points of the topography. (b) Contours of the streamfunction for the transmitted waves on the right side of the topography calculated by the Green's function approach. Waves in (b) are focused around two stable fixed points as predicted by (a).

- A characteristic will begin on the topography on the blue curve at x_0.
- The next reflection point, x_1, can be found by moving along a horizontal line to the green curve.
- The topographic height at x_1 is found by moving along a vertical line to the blue curve.
- A fixed point is a point that maps onto itself. Fixed points occur when the curves $x = h(x)$ and $x = x$ intersect. Characteristics will move towards the stable fixed points and away from the unstable fixed points.

The map for $k = 2$ is shown in figure 10a. The map predicts that the characteristics will converge upon two stable fixed points. Figure 10b shows the numerical solution to the transmitted wave field for $k = 2$ from the Green's function approach. The solution converges onto two stable fixed points.

7 Conclusions

Internal waves propagating in the ocean are likely to encounter long stretches of topography where energy is drained from the low-mode internal tides to and transferred to higher mode numbers. In this work, we primarily looked at model topographies that are unlikely to be found in the ocean. We found that for these special topographies, energy can be transferred rapidly to higher mode number waves. Additionally, characteristics for these topographies

converged onto stable fixed points for topography with multiple reflections. This may be seen as "periodic wave attractors" similar to wave attractors in basins (e.g. [11]).

In the ocean, the bottom topography is not regular and this wave-focusing mechanism is unlikely to be seen. However, we have developed tools that may be utilized for realistic ocean topography. Although computationally expensive, the Green's function approach can provide solutions to the scattering problem over more complex topography. Further work utilizing the method of characteristics may also be valuable to understanding the degradations of the internal tide over random topography.

Acknowledgments

I would like to thank Neil Balmforth and Oliver B$\acute{}$uhler for their countless amounts of patience and assistance throughout the summer. I would also like to thank Bill Young and Joe Keller for their helpful discussions. No GFD summer would be complete without softball, so special thanks to George Veronis for coaching our team. Lastly, thanks to all the fellows and staff for a wonderful summer.

References

[1] P. Baines, *The reflexion of internal/inertial waves from bumpy surfaces*, J. Fluid Mech., 46 (1971), pp. 273–291.

[2] ——, *The generation of internal tides by flat-bump topography*, Deep-Sea Res., 20 (1973), pp. 179–205.

[3] N. J. Balmforth, G. R. Ierley, and W. R. Young, *Tidal Conversion by Subcritical Topography*, J. Phys. Oceanogr., 32 (2002), pp. 2900–2914.

[4] T. H. Bell, Jr., *Statistical features of sea-floor topography*, Deep-Sea Res., 22 (1975), pp. 883–892.

[5] ——, *Topographically Generated Internal Waves in the Open Ocean*, J. Geophys. Res., 80 (1975), pp. 320–327.

[6] P. F. Cummins, J. Y. Cherniawsky, and M. G. G. Foreman, *North Pacific internal tides from the Aleutian Ridge: Altimeter observations and modeling*, J. Mar. Res., 59 (2001), pp. 167–191.

[7] G. D. Egbert and R. D. Ray, *Estimates of M_2 tidal energy dissipation from TOPEX/Poseidon altimeter data*, J. Geophys. Res., 106 (2001), pp. 22,475–22,502.

[8] S. R. Jayne, *The Impact of Abyssal Mixing Parameterizations in an Ocean General Circulation Model*, J. Phys. Oceanogr., 39 (2009), pp. 1756–1775.

[9] S. G. Llewellyn Smith and W. R. Young, *Conversion of the Barotropic Tide*, J. Phys. Oceanogr., 32 (2002), pp. 1554–1566.

[10] ——, *Tidal conversion at a very steep ridge*, J. Fluid Mech., 495 (2003), pp. 175–191.

[11] L. R. M. Maas, D. Benielli, J. Sommeria, and F.-P. A. Lam, *Observation of an internal wave attractor in a confined stable stratified fluid*, Nature, 388 (1997), pp. 557–561.

[12] P. Müller and X. Liu, *Scattering of Internal Waves at Finite Topography in Two Dimensions. Part I: Theory and Case Studies*, J. Phys. Oceanogr., 30 (2000), pp. 532–549.

[13] ——, *Scattering of Internal Waves at Finite Topography in Two Dimensions. Part II: Spectral Calculations and Boundary Mixing*, J. Phys. Oceanogr., 30 (2000), pp. 550–563.

[14] F. Petrelis, S. Llewellyn Smith, and W. R. Young, *Tidal Conversion at a Submarine Ridge*, J. Phys. Oceanogr., 36 (2006), pp. 1053–1071.

[15] L. Rainville and R. Pinkel, *Propagation of Low-Mode Internal Waves through the Ocean*, J. Phys. Oceanogr., 36 (2006), pp. 1220–1236.

[16] H. Sandstrom, *On topographic generation and coupling of internal waves*, Geophys. Astrophys. Fluid Dyn., 7 (1975), pp. 231–270.

[17] L. St. Laurent and C. Garrett, *The Role of Internal Tides in Mixing the Deep Ocean*, J. Phys. Oceanogr., 32 (2002), pp. 2882–2899.

Quantum vortices in a glass of Bose-Einstein condensate

Nicolas Grisouard

1 Introduction

Bose-Einstein condensation occurs for bosons when, for a temperature low enough (typically a few kelvins for a liquid condensate and a few nanokelvins for a gaseous condensate), a macroscopic fraction[1] of the bosons gathers in the quantum state of lowest energy, forming a Bose-Einstein condensate (BEC). In this case, the behavior of the fluid can be described in the case of a dilute gas by the nonlinear Schrödinger equation (NLS), also called the Gross-Pitaevskii equation. It can be shown that the dynamics of the NLS is very close to shallow water dynamics with surface tension. One of the striking features of BECs is the possibility of point vortices to appear, which dynamics is explained in some the oldest textbooks of fluid dynamics (see *e.g.* [5]). Surprinsingly, although vortices in BECs have been extensively studied, some of the most simple *Gendankenexperiments* have yet to be studied. In [6] for example, the (apparently) simple problem of the motion of a single vortex along a plane boundary in a semi-infinite, uniform BEC is studied, eleven years after the first experimental realisation of a BEC ([3]) and eighty-two years after Einstein's translation of the letter Bose wrote him about what would later be called Bose-Einstein statistics ([1]). An opportunity is then to be seized by classical fluid dynamicists, who can make the study of quantum fluid dynamics benefit from their experience. The idea of the present project, unlike [6] who use an almost all-analytical approach, is to develop a numerical code to integrate the NLS in polar coordinates and to perform a few numerical experiments about vortices in a two-dimensional circular box (a "glass"). In this report, we will first present some aspects of quantum fluid dynamics for geophysical fluid dynamicists to understand the analogies between the two fields. The code written during the project will then be presented in Section 3. In Section 4, a few numerical experiments will be presented to validate the code, study various simple configurations and see what the code can do. In a last section, conclusions will be drawn and possible extensions of this work will be proposed.

2 Two-dimensional Bose-Einstein condensates

2.1 The nonlinear Schrödinger equation

In a dilute gas of N particles, each particle barely interacts with the others and in first approximation, its wave function does not feel the presence of the other particles. In this case, it is possible to approximate the total wave function Ψ as the product of all the

[1] to be understood as a fraction not involving negative powers of ten

individual wave functions $\widetilde{\psi}_i$:

$$\Psi(\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, t) = \prod_{i=1}^{N} \widetilde{\psi}_i(\boldsymbol{r_i}, t).$$

As the particles are all identical and in the same fundamental state, all $\widetilde{\psi}_i$ have the same form which we can write as

$$\widetilde{\psi}_i(\boldsymbol{r_i}, t) = \frac{\psi(\boldsymbol{r_i}, t)}{\sqrt{N}}.$$

The reason why we chose this particular scaling is that the normalization relation, which is that the probability of finding the i^{th} particle in the whole physical domain $\mathcal{D}$ is equal to 1, can be written as

$$\int_S |\psi|^2 \, dS = N. \tag{1}$$

Therefore, $|\psi|^2$ can be considered as the particle density.

We mentioned earlier that the particles barely interact with each other. In a dilute gas, this approximation is accurate but let us refine this model by introducing the fact that these particles indeed never interact except when their positions exactly coincide (collisions), in which case they feel a repulsive interaction. The hamiltonian of the system is therefore

$$H = \sum_{i=1}^{N} \left(-\frac{\hbar^2}{2m} \boldsymbol{\nabla}_i^2 + V(\boldsymbol{r}_i) \right) + \frac{1}{2} \sum_{i<j} U_0 \delta(\boldsymbol{r}_i - \boldsymbol{r}_j),$$

where $\hbar$ is the reduced Planck constant[2], m is the mass of a single particle, $V(\boldsymbol{r}_i)$ is the external energy potential and $\boldsymbol{\nabla_i}$ is the gradient operator relative to the set of coordinates $\boldsymbol{r_i}$. Integration of $\psi^* H \psi$ across the whole domain (where * means complex conjugate) gives the energy of the system:

$$E = \int_S \left(\frac{\hbar^2}{2m} |\boldsymbol{\nabla}\psi|^2 + V(\boldsymbol{r})|\psi|^2 + \frac{1}{2} U_0 |\psi|^4 \right) dx \, dy, \tag{2}$$

The wave function minimizes E under the constraint of Equation (1) and satisfies the NLS:

$$i\hbar \frac{\partial \psi}{\partial t} = \left[-\frac{\hbar^2}{2m} \boldsymbol{\nabla}^2 \psi + V(x, y) \right] \psi + U_0 |\psi|^2 \psi, \tag{3}$$

which entirely describes the dynamics of the BEC. The first term of the right member is the usual Schrödinger equation and the last term is the one introduced by the interaction terms, which is also the nonlinear term. In our case, we will consider a two-dimensional circular glass of radius R associated with a set of polar coordinates (r, θ) which origin lies in the center, in other words:

$$V(r \geq R, \theta) = +\infty, \qquad V(r < R, \theta) = V \text{ (constant)}.$$

In this case, the probability of finding a particle outside the glass is strictly equal to zero and we will therefore only focus on the interior of it.

[2] $\hbar \simeq 1.054571628 \times 10^{-34}$ J.s

2.2 The Madelung transformation

Before looking at the behavior of the NLS solutions, it might be useful for the reader who is more used to geophysical fluid dynamics to see how the dynamics of the NLS equation can seem familiar. ψ written in its polar form is

$$\psi(\boldsymbol{r},t) = \sqrt{n(\boldsymbol{r},t)}e^{i\alpha(\boldsymbol{r},t)}.$$

In this case, n and α are real, n is positive and can be easily interpreted as the particle density, as implied by Equation (1):

$$n(\boldsymbol{r},t) = |\psi(\boldsymbol{r},t)|^2.$$

Now multiplying Equation (3) by ψ^* and adding the result to its complex conjugate, one finds that n indeed satisfies a conservation law:

$$\frac{\partial n}{\partial t} = \frac{i\hbar}{2m}(\psi^*\boldsymbol{\nabla}^2\psi - \psi\boldsymbol{\nabla}^2\psi^*) = -\boldsymbol{\nabla}\cdot\boldsymbol{j},$$

with $\boldsymbol{j} = -(i\hbar/2m)(\psi^*\boldsymbol{\nabla}\psi - \psi\boldsymbol{\nabla}\psi)$ is the probability current also written as

$$\boldsymbol{j} = \frac{\hbar}{m}\Im(\psi^*\boldsymbol{\nabla}\psi) = \frac{\hbar}{m}n\boldsymbol{\nabla}\alpha.$$

A velocity vector can be defined such that $\boldsymbol{j} = n\boldsymbol{u}$, or

$$\boldsymbol{u} = \frac{\hbar}{m}\boldsymbol{\nabla}\alpha = \frac{\hbar}{m}\frac{\Im(\psi^*\boldsymbol{\nabla}\psi)}{n}.$$

As $\boldsymbol{u}$ is a gradient, we have $\boldsymbol{\nabla}\times\boldsymbol{u} = 0$, *i.e.* the flow is irrotationnal. Equation (4) is then

$$\partial_t n + \boldsymbol{\nabla}\cdot(n\boldsymbol{u}) = 0, \tag{4}$$

which is formally identical to the continuity equation in fluid dynamics. It is also possible to derive an equation similar to the Euler equations, although we will restrict ourselves to just mentioning it in the general case $\boldsymbol{\nabla}V \neq 0$:

$$\frac{D\boldsymbol{u}}{Dt} = -\frac{1}{m}\boldsymbol{\nabla}V - U_0\boldsymbol{\nabla}n + \frac{\hbar^2}{2m^2}\boldsymbol{\nabla}\left\{\frac{\boldsymbol{\nabla}^2(\sqrt{n})}{\sqrt{n}}\right\}. \tag{5}$$

The similarity to the Euler equations in classical fluid dynamics is made clearer when writing it in the following form:

$$\frac{D\boldsymbol{u}}{Dt} = -\boldsymbol{\nabla}\Phi - g\boldsymbol{\nabla}\eta + \frac{\sigma}{\rho}\boldsymbol{\nabla}\left\{\frac{\boldsymbol{\nabla}^2\eta}{(1+|\boldsymbol{\nabla}\eta|^2)^{3/2}}\right\},$$

where Φ is the potential of the external conservative forces acting on the fluid other than gravity, η the free-surface height, ρ the density and σ the surface tension. The term $-\boldsymbol{\nabla}\Phi$ is related to $-\boldsymbol{\nabla}V$ in Equation (5). The hydrostatic pressure gradient $-g\boldsymbol{\nabla}\eta$ is related to $-U_0\boldsymbol{\nabla}n$ which would be absent of Equation (5) if U_0 would be zero: the action of the nonlinear effects is then to avoid high concentrations of particles and "flatten" the particle

distribution, like pressure does for the height of a fluid surface. One can also notice that now, n is not related to ρ but to η. Finally, the last term of Equation (5), although not formally identical, can be related to the surface tension term of the last equation. Effects on the dynamics of the distribution of the particle density that are similar to those induced by surface tension at the surface of a fluid will be involved throughout this report, the most striking feature being the existence of density waves that behave like surface capillary-gravity waves in a classical fluid. As a remark, the flow can be compressible.

2.3 Non-dimensional problem

Before writing the NLS in a non-dimensional form, we first have to define characteristic quantities.

2.3.1 A characteristic length: the healing length

We have noticed in Section 2.2 that the nonlinear effects tend to homogenize the particle distribution. But the external potential is such that the wave function vanishes at the boundary of the domain. In the steady state, the particle density must therefore experience an evolution between the boundary condition and the interior of the BEC, this evolution being given by the steady NLS. For simplicity reasons, we consider here a plane, infinite wall which boundary lies along the y axis. We then have $\psi \equiv 0$ for $x \leq 0$ and the steady NLS is now

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V\psi + U_0\psi^3 = 0. \tag{6}$$

This ODE is somehow similar to the following ODE:

$$\frac{1}{2}f'' + f - f^3 = 0,$$

which solution is the tanh function. In order to put Equation (6) in a similar form, we write $\psi = \sqrt{n_0}\hat{\psi}$ where n_0 is the background density away from the boundary and $\hat{\psi}$ is the dimensionless wave function. We also set $V = -U_0 n_0$, the external potential being now a direct result of all the individual interactions acting as a mean-field. Equation (6) becomes

$$-\frac{\hbar^2}{2mU_0n_0}\frac{d^2\hat{\psi}}{dx^2} + \hat{\psi} - \hat{\psi}^3 = 0.$$

The solution to this equation is then $\hat{\psi} = \tanh\left(x/(\sqrt{2}\xi)\right)$ where

$$\xi = \frac{\hbar}{\sqrt{2mn_0U_0}}. \tag{7}$$

ξ is the healing length and will be our characteristic length. As shown in Figure 1, $\xi\sqrt{2}$ is the typical distance on which the density relaxes from zero at the boundary to its background value n_0. We see again an action of the nonlinear term: without it, the ODE to solve would simply be the equation of an harmonic oscillator and the solutions would be the eigenmodes one finds for free particles trapped in a box, exhibiting oscillatory behaviors. We see as in Section 2.2 that the particle density is flattened by the extra term brought by the interactions between particles.

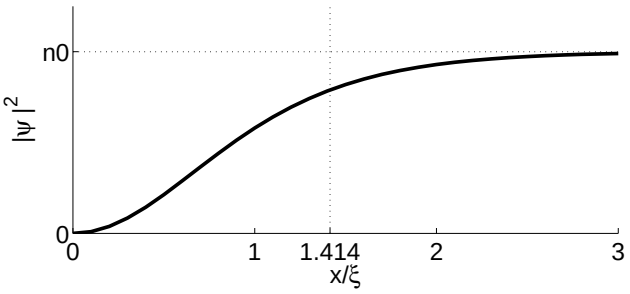


Figure 1: Particle density at rest from the boundary in $x = 0$ to its background density n_0.

2.3.2 Non-dimensional variables, non-dimensional NLS and invariants

We can now define a full set of non-dimensional variables by scaling length by ξ, time by $m\xi^2/\hbar$ and wave function by $\sqrt{n_0}$. By doing so and keeping $V = -n_0 U_0$ in the interior of the box, we get the non-dimensional version of the NLS:

$$\frac{\partial \psi}{\partial t} = \frac{i}{2}\boldsymbol{\nabla}^2\psi + \frac{i}{2}\left(1 - |\psi|^2\right)\psi. \tag{8}$$

We dropped the hat for the non-dimensional wave function. Scaling energy by $n_0\hbar^2/m$, Equation (2) can be written in a non-dimensional form:

$$E = \frac{1}{2}\int_S \left(|\boldsymbol{\nabla}\psi|^2 - |\psi|^2 + \frac{1}{2}|\psi|^4\right) dS.$$

In our two-dimensional, polar symmetry, the energy can be written as

$$E = \frac{1}{2}\int_0^{2\pi}\int_0^R \left(r\left|\frac{\partial\psi}{\partial r}\right|^2 + \frac{1}{r}\left|\frac{\partial\psi}{\partial\theta}\right|^2 - r|\psi|^2 + \frac{r}{2}|\psi|^4\right) dr\, d\theta.$$

We have derived two invariants so far, the number of particles N and the total energy E. A third invariant, the total angular momentum L, can be defined as

$$\boldsymbol{L} = \int_S \psi^*(-i\,\boldsymbol{r}\times\boldsymbol{\nabla})\psi\, dS$$

The only non-zero component of $\boldsymbol{L}$ being the one normal to the (r,θ) plane, we have

$$L = -i\int_0^{2\pi}\int_0^R \psi^*\frac{\partial\psi}{\partial\theta}\, r\, dr\, d\theta,$$

which is our third invariant.

Finally, the velocity can be made non-dimensional by scaling it by $\hbar/(m\xi)$. Its non-dimensional definition then simply becomes

$$\boldsymbol{u} = \boldsymbol{\nabla}\alpha.$$

2.4 Solutions of the nonlinear Schrödinger equation

2.4.1 Waves

Let us study the wave spectrum of the NLS for small perturbations relative to an uniform basic state of density $n_0 = 1$ with no mean velocity of the BEC:

$$\psi = 1 + \phi \quad \text{with} \quad |\phi| \ll 1.$$

Equation (8) becomes

$$\begin{aligned}\frac{\partial \phi}{\partial t} &= \frac{i}{2}\boldsymbol{\nabla}^2\phi - \frac{i}{2}\left(\phi + \phi^* + |\phi|^2\right)(1+\phi) \\ &= \frac{i}{2}\boldsymbol{\nabla}^2\phi - i\frac{\phi+\phi^*}{2} + O(|\phi|^2).\end{aligned}$$

Neglecting high-order terms and taking the complex conjugate of the previous expression gives the following system

$$\begin{cases} \dfrac{\partial \phi}{\partial t} = \dfrac{i}{2}\boldsymbol{\nabla}^2\phi - i\dfrac{\phi+\phi^*}{2}, \\ \dfrac{\partial \phi^*}{\partial t} = -\dfrac{i}{2}\boldsymbol{\nabla}^2\phi^* + i\dfrac{\phi^*+\phi}{2}. \end{cases}$$

We can rewrite the previous system as a linear matrix equation

$$\frac{\partial}{\partial t}\begin{bmatrix} \phi \\ \phi^* \end{bmatrix} = M \begin{bmatrix} \phi \\ \phi^* \end{bmatrix} \qquad \text{with} \quad M = \frac{i}{2}\begin{bmatrix} \boldsymbol{\nabla}^2 - 1 & -1 \\ 1 & 1 - \boldsymbol{\nabla}^2 \end{bmatrix}. \tag{9}$$

Let the solution be assumed to oscillate at frequency ω and let $[u, v]e^{-i\omega t}$ be the eigenvector of M associated with the eigenvalue $-i\omega$. As we will see in Section 3, the normal modes to use in polar coordinates are $u, v \sim e^{in\theta}J_n(kr)$ and as[3] $\boldsymbol{\nabla}^2(e^{in\theta}J_n(kr)) = -k^2 e^{in\theta}J_n(kr)$, Equation (9) is now

$$(M + i\omega)\begin{bmatrix} u \\ v \end{bmatrix} = \frac{i}{2}\begin{bmatrix} -k^2 - 1 + 2\omega & -1 \\ 1 & 1 + k^2 + 2\omega \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = 0. \tag{10}$$

For this system to have a non-trivial solution, its determinant has to be zero, which yields the dispersion relation

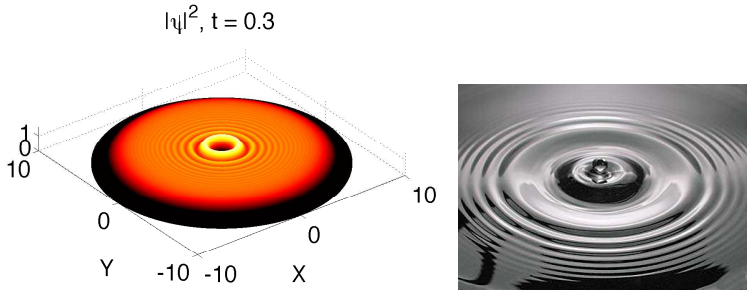
$$\omega^2 = \frac{k^2}{2} + \frac{k^4}{4}.$$

This dispersion relation is formally similar to the dispersion relation of gravity-capillary waves in shallow water which reads (in dimensional units)

$$\omega^2 = gH|\boldsymbol{k}|^2 + \frac{\sigma H}{\rho}|\boldsymbol{k}|^4,$$

with H the mean water height. The first term in Equation (2.4.1) is a non-dispersive term due to the nonlinear term of the NLS. In the long-wave approximation ($k \to 0$), it is dominant and the dispersion relation is similar to the dispersion relation of long gravity waves or of acoustic waves. The second term becomes significant for shorter waves and in such a case, the waves disperse like gravity-capillary waves with short waves traveling faster than long waves. A visual comparison can easily be made in Figure 2.

[3] in polar coordinates, this equation is in fact the Bessel equation of order n multiplied by $e^{in\theta}$



(a) Waves in a BEC (b) Water waves

Figure 2: Density waves in a BEC (a) compared with gravity-capillary waves in (deep) water (b). Due to the dispersion relation, short waves travel faster than long waves. The initial condition for the wave function used in picture (a) is $\psi = e^{i\theta}\tanh((R-r)/\sqrt{2})$ in the code described in Section 3 and the waves are created by the transient.

2.4.2 Vortices

We noticed in Section 2.2 that the flow was intrinsically irrotational. Writing that, we neglected possible singularities that could make the flow rotational on these points and we will see on an example how and why it is the case. Let us suppose that locally, the wave function can be written as

$$\psi = x + iy = r\exp(i\theta). \tag{11}$$

In this case, the wave function has a zero in $r = 0$ where the density vanishes. The velocity is then

$$\boldsymbol{u} = \boldsymbol{\nabla}\theta = \frac{1}{r}\hat{\boldsymbol{\theta}}.$$

In classical fluid dynamics, this velocity field would be the one induced by a point vortex inducing a divergence of the azimuthal velocity when $r \to 0$, which would be regularized giving some spatial extent to the core of the vortex. Here, this singularity is a fundamental feature of the field and it is required that these zeros are points and not lines or surfaces, as a result of the cancellation of both real and imaginary parts. It is also required that these two components cancel along lines and that these lines cross only on points. Take for example $\psi = x + ix$: this function admits a line of zeros defined by $x = 0$ where both real and imaginary parts cancel but no velocity is induced by it. If we take the simple example described by Equation (11), use Stoke's theorem to rewrite the circulation and compute the circulation of the velocity field around the origin, we get:

$$\Gamma = \oint_{\mathcal{C}} \boldsymbol{u}\cdot\boldsymbol{dr} = \iint_S (\boldsymbol{\nabla}\times\boldsymbol{\nabla}\theta)\cdot d\boldsymbol{S} = 2\pi.$$

Two things have to be noticed here. First, that the circulation is quantized. In general, for any vortex, it can only be an element of $\{2\pi p,\ p\in\mathbb{Z}\}$ ($\{2\pi p\hbar/m,\ p\in\mathbb{Z}\}$ in dimensional units). Second, that although Equation (11) looks harmless, a singularity is hidden at the origin as the phase is multivalued, which makes the quantity $\boldsymbol{\nabla}\times\boldsymbol{\nabla}\theta$ infinite in just one point.

One can wonder how far Equation (11) is a realistic expression for a vortex. A more refined but still approximate expression for the density structure around a steady point vortex is $|\psi|^2(r) = r^2/(r^2+1)$ (see *e.g.* [2]), so Equation (11) is a good approximation in the center of this vortex. One might also want to push the robustness of this feature and see what happens when the density around the vortex is perturbed. Let us consider a simple perturbation for the wave function:

$$\psi(x,y) = x + i(a_1 x + a_2 y). \tag{12}$$

The corresponding density is

$$n(x,y) = (1+a_1^2)\,x^2 + 2\,a_1 a_2\,xy + b_2\,y^2$$

and the lines of constant density are now ellipses. In polar coordinates ($x = r\cos\theta$, $y = r\sin\theta$), the wave function reads:

$$\psi(r,\theta) = r(\cos\theta + i(a_1\cos\theta + a_2\sin\theta))$$

and its phase is

$$\alpha = \arctan(a_1 + a_2\tan\theta),$$

which does not depend on r. This means that the radial velocity will be zero and the velocity will then be purely azimuthal, just like for the undisturbed vortex. As far as we know, this peculiar feature has surprisingly never been noticed. The azimuthal velocity is then

$$\begin{aligned} u_\theta(r,\theta) &= \frac{1}{r}\frac{\partial\alpha}{\partial\theta} = \frac{1}{r}\frac{a_2^2}{\cos^2\theta + (a_1\cos\theta + a_2\cos\theta)^2} \\ &= \frac{a_2^2\,r}{n(r,\theta)}. \end{aligned}$$

For a given r, the azimuthal velocity is maximum where the density is minimum, which is due to the conservation of the number of particles.

We have reviewed all the physical phenomena involved in this study. We will now describe the code that has been developed in order to perform simple numerical experiments.

3 The code

3.1 Operator splitting method

Operator splitting methods [4] are commonly used to solve the NLS, mostly because they are simple to implement in that case. The general idea is that the resolution of the NLS is splitted in two steps, starting from a time t:

[4] see *e.g.* [7]

Step 1: integration of $\psi(r,\theta,t)$ through equation

$$\frac{\partial \psi}{\partial t} = \frac{i\left(1-|\psi|^2\right)}{2}\psi$$

over a time step dt. As the variation of ψ is equal to $iK\psi$ where K is real, $|\psi|$ remains unchanged under this operation and the integration is then exact. After this step, we get an intermediate ψ_{int}:

$$\psi_{int}(r,\theta) = \psi(r,\theta,t)\exp\left(\frac{i(1-|\psi(r,\theta,t)|^2)dt}{2}\right).$$

Step 2: integration of $\psi_{int}(r,\theta)$ through equation

$$\frac{\partial \psi}{\partial t} = \frac{i}{2}\boldsymbol{\nabla}^2\psi \tag{13}$$

over a time step dt. If the integration is performed in the spectral space, this step is also exact and gives $\psi(r,\theta,t+dt)$. It will be explained in more detail in the next section.

As a remark, both steps are *individually* exact and energy preserving, although this will be discussed for the second step, but the combination of the two is not, therefore we need to choose dt small enough. A time step of $dt = 5\times10^{-3}$ ensures the stability of all simulations presented here.

3.2 Integrating the Laplacian

We here solve Equation (13) with ψ_{int} as initial condition for the wave function. In polar coordinates, this function can be decomposed as

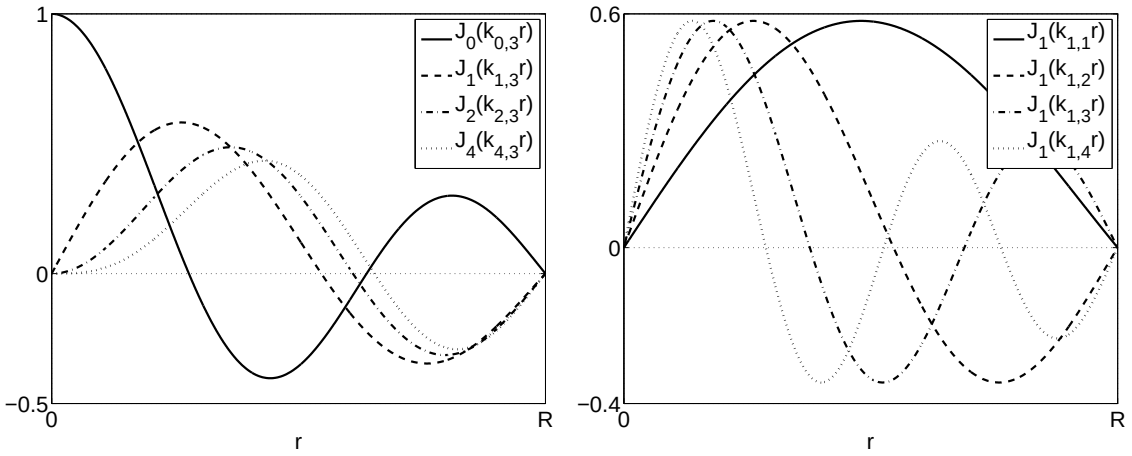
$$\psi_{int} = \sum_{q=-\infty}^{+\infty}\sum_{j=1}^{+\infty}(a_{q,j})_{int}\, e^{iq\theta}\, J_q(k_{q,j}r). \tag{14}$$

Here, q is the index of the angular mode, j the index of the radial mode, J_q the q^{th}-order Bessel function of the first kind and $k_{q,j}$ a coefficient scaling J_q such that $\forall(q,j)$, $J_q(k_{q,j}R) = 0$ (which ensures that $\psi_{int}(r=R,\theta)\equiv 0$). The index j is therefore the number of zeros in the interval $[0,R]$ of the considered Bessel function (see Figure 3 for a visualization of various $J_q(k_{q,j}r)$'s). Finally, the set of $a_{q,j}$ is the set of weights that will characterize ψ_{int} and therefore depend on time. It turns out that applying the Laplacian to Equation (14) gives the Bessel equation and as in Section 2.4.1, we get:

$$\boldsymbol{\nabla}^2\psi_{int} = -\sum_{q=-\infty}^{+\infty}\sum_{j=1}^{+\infty}k_{q,j}^2(a_{q,j})_{int}\, e^{iq\theta}J_q(k_{q,j}r).$$

Plugging this in Equation (13) and isolating every mode gives:

$$\frac{da_{q,j}}{dt} = -\frac{ik_{q,j}^2}{2}a_{q,j} \qquad \forall\ (q,j)$$



(a) (b)

Figure 3: (a) First four orders of the Bessel functions of the first kind, each of them scaled in order to have three zeros in $[0, R]$. (b) First four radial scalings of the Bessel function of the first kind of order 1.

and therefore, the integration from $(a_{q,j})_{int}$ to $a_{q,j}(t + dt)$ in a time step dt is exactly:

$$a_{q,j}(t + dt) = (a_{q,j})_{int} \exp\left(-\frac{ik_{q,j}^2 dt}{2}\right).$$

The recipe is then a classical one: go to the spectral space, integrate every coefficient $a_{q,j}(t)$ and go back to the physical space. It turns out that going to the spectral space and back is the most difficult part here. To isolate an azimuthal mode q, a Fourier transform of ψ_{int} in the azimuthal direction is performed, thanks to the orthogonality relation

$$\int_0^{2\pi} e^{-il\theta}\, e^{iq\theta} d\theta = 2\pi\delta_{ql},$$

δ_{ql} being the Kronecker symbol. Numerically, a FFT in the azimuthal direction is performed and no dealiaising is done. The output of the transform is then, for each $q \in \mathbb{Z}$:

$$\mathcal{F}[\psi_{int}]_q(r) = 2\pi \sum_{j=1}^{\infty} a_{q,j}(t) J_q(k_{q,j} r).$$

The radial modes described by the index j must now be isolated: the Fourier transform has isolated one order for the Bessel functions of the first kind, but not the different ways of scaling these function such that they would be zero in $r = R$. In other words, one has to isolate every function that is plotted in Figure 3(b), which is done using the following orthogonality relation:

$$\int_0^1 x J_q(\zeta_{q,j} x) J_q(\zeta_{q,p} x) dx = \frac{\delta_{jp}}{2} J_{q+1}^2(\zeta_{q,j}),$$

where $\zeta_{q,j}$ is the j^{th} zero of $J_q(x)$. By definition, $\zeta_{q,j} = k_{q,j}R$ and applying the variable change $r = Rx$, one gets the relation that is actually used for each j:

$$\int_0^R r J_q(k_{q,j}r) J_q(k_{q,p}r) dr = \frac{R^2 \delta_{jp}}{2} J_{q+1}^2(k_{q,j}R).$$

The corresponding operation performed on $\mathcal{F}[\psi_{int}]_q$ is a Hankel transform of order q, which is mathematically defined, for a function $f(x)$, as:

$$\mathcal{H}_q[f](k) = \int_0^{+\infty} x f(x) J_q(kx) dx.$$

In our case, we use a quasi-discrete Hankel transform (QDHT). The algorithm, exposed in detail and with clarity in [4], needs the input function to be evaluated at values of the radius $r_{q,j} = k_{q,j}R/k_{q,Y+1}$, where Y is the number radial modes that will be computed. This grid is specific to each q and we have to perform the QDHT for each one of them. It means that starting from an arbitrary ψ evaluated on a given grid, once this function has experienced a Fourier transform in the angular dimension, each projection on the azimuthal modes has to be interpolated on the grid $\{r_{q,j}\}$ with $1 \leq j \leq Y$ used by the QDHT of order q, which is a source of inaccuracy. The Matlab `spline` interpolation method has been chosen. Y has been chosen to be equal to half the number of points in the radial direction in order for the interpolation to be computed quite easily, the idea being that even for the highest radial modes, the oscillations of the Bessel function can be described by a few points between two zeros. This latter choice has not been the subject of any extensive study and shall be handled with care and suspicion.

We have isolated every (q, j) couples and are now in the spectral space. Before going back to the physical space, we just have to multiply every $(a_{q,j})_{int}$ by the propagator $\exp(-ik_{q,j}^2 dt/2)$. The next step is then to perform inverse quasi-discrete Hankel transforms (iQDHT) for each q, then re-interpolate the results back on the initial regular grid and finally perform an inverse FFT. We are now in the physical space with the function ψ evaluated at time $t + dt$.

3.3 Summary of the algorithm

Starting from $\psi(t)$, the different steps to get $\psi(t + dt)$ are:

1. multiply ψ by $\exp(i(1 - |\psi|^2)dt/2) \Rightarrow \psi_{int}$
2. perform a FFT of ψ_{int} along the azimuthal direction $\Rightarrow \mathcal{F}[\psi_{int}]_q$,
3. for each q, interpolate $\mathcal{F}[\psi_{int}]_q$ on the intermediate grid defined by $r_{q,j} = k_{q,j}R/k_{q,Y+1}$ with $1 \leq j \leq Y$,
4. for each q, perform the QDHT of order $m \Rightarrow \mathcal{H}_q[\mathcal{F}[\psi_{int}]_q]$,
5. for each (q, j), multiply $\mathcal{H}_q[\mathcal{F}[\psi_{int}]_q]$ by $\exp(-ik_{q,j}^2 dt/2) \Rightarrow \mathcal{H}_q[\mathcal{F}[\psi(t + dt)]_q]$,
6. for each q, perform the iQDHT of order q,
7. for each q, interpolate the previous results on the regular grid $\Rightarrow \mathcal{F}[\psi(t + dt)]_q$,
8. perform an iFFT of $\mathcal{F}[\psi(t + dt)]_q$ along the azimuthal modes direction $\Rightarrow \psi(t + dt)$.

4 Numerical experiments

4.1 General features of the settings

In all numerical experiments that are presented in this report, the radius of the glass is $R = 10$, the number of points in the radial and azimuthal directions are $n_r = n_\theta = 256$ and the number of radial modes is $Y = 128$, according to the rule we mentioned in Section 3.2. Initial conditions will all involve vortices put in a background particle density approximating a steady density distribution, as described in Section 2.3.1 for a different geometry. The function, approximating what the density distribution should be in the glass, is indeed inspired by the latter distribution, namely:

$$\psi_{bg}(r) = \tanh\left(\frac{R-r}{\sqrt{2}}\right).$$

The wave function of an isolated vortex of circulation $\Gamma = 2p\pi$ ($p \in \mathbb{Z}$) located in (r_0, θ_0) is approximated by:

$$\psi^v_{\{r_0;\theta_0;p\}}(r,\theta) = \frac{re^{ip\theta} - r_0 e^{ip\theta_0}}{\sqrt{1+|re^{ip\theta} - r_0 e^{ip\theta_0}|^2}}.$$

It is now possible to generate various configurations involving vortices in a glass. For example, if one wants to generate an initial condition with one vortex of circulation $\Gamma = -2\pi$ located in $(R/2, 0)$ and one vortex of circulation $\Gamma = +4\pi$ located in $(R/8, \pi/2)$, the initial condition to set is:

$$\psi_0 = \psi_{bg} \times \psi^v_{\{R/2;0;-1\}} \times \psi^v_{\{R/8;\pi/2;2\}}.$$

Both definitions of ψ_{bg} and ψ^v are *approximations* of the background and vortices wave functions. The definition of ψ_{bg} will be more realistic for large values of R for instance, as it is inspired by a cartesian configuration with an infinitely long, plane wall. This will result in the generation of transients in the first moments of the simulations that can pollute the expected evolution of the system.

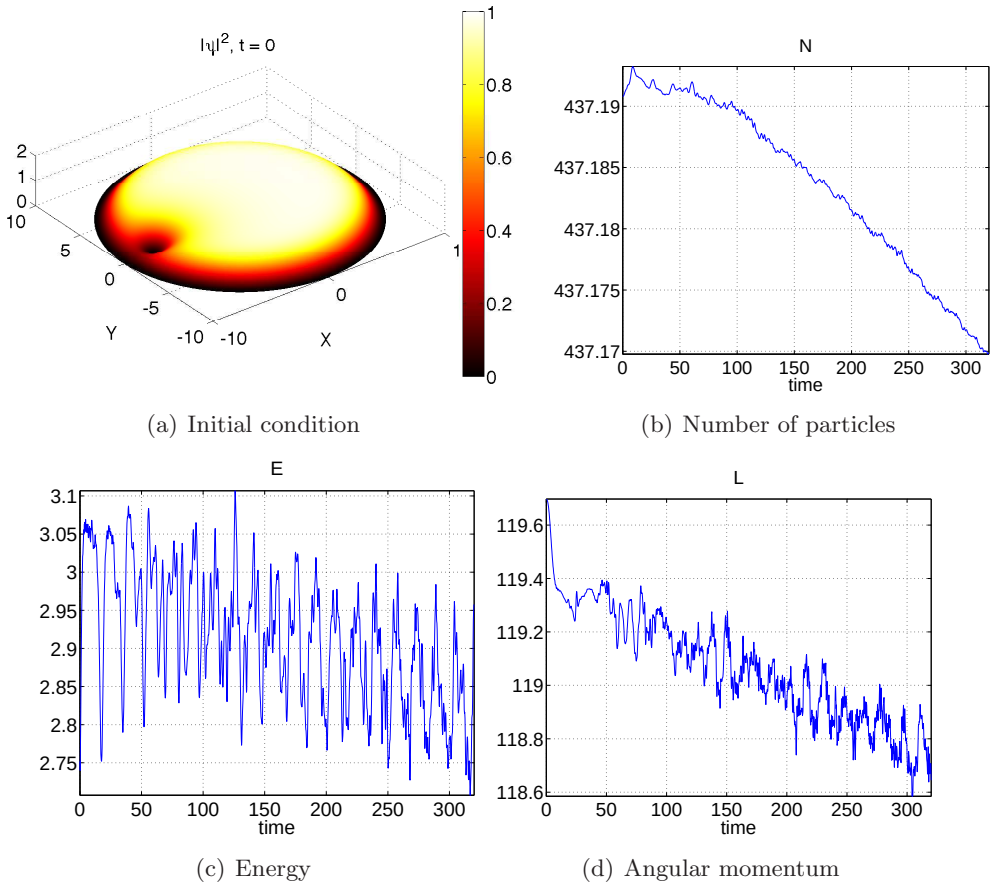
4.2 Conservation of the invariants

We will now check on a simple example if the three invariants N, E and L described in Section 2.3.2 are indeed conserved by our code. A property of the QDHT is to preserve the invariants but the use of the interpolations introduces errors, therefore the need to control if these errors are noticeable or not. We will set up an initial condition with one vortex:

$$\psi_0 = \psi_{bg} \times \psi^v_{\{3R/4;0;1\}},$$

as can be shown in Figure 4(a). Figure 4 in general shows how conserved are these quantities over a long period of time, a duration of more than 300 being three to four times longer than the experiments that will be shown later. We can see that N is only very slightly decreasing and that L and E both have a tendency to slightly decrease also, although they are subjected to oscillations. A possible explanation for these oscillations are that both of these quantities involve derivatives which are computed with the Matlab function `gradient` based on a rustic 2^{nd}-order finite difference scheme. E is especially oscillatory and is in the

same time the function that includes the most derivatives, which is another hint that these oscillations might indeed be due to a rustic post-processing rather than a problem in the code.



(a) Initial condition

(b) Number of particles

(c) Energy

(d) Angular momentum

Figure 4: Conservation of the three invariants N, E and L on an example with initial condition shown in (a). Although all curves seem to decrease, their value don't change much, as shown by the values displayed on the vertical axes.

Computing the invariants is a way to validate the accuracy of the code as to our knowledge, no analytical solution of simple problems exist in this geometry. We will now run numerical experiments to study simple problems that could be found in textbooks of classical fluid dynamics but that are uncharted territory in quantum fluid dynamics.

4.3 Vortex generation

We will see on a simple experiment that pair of vortices can be generated. It is hard to define an initial condition that would directly generate two vortices, so the following procedure has been applied: when two counter-rotating vortices are close to each other, they get closer and merge (see [2]). One can imagine then making two vortices merge, reversing the velocity field and plugging the resulting wave function in the NLS will later produce two

counter-rotating vortices. It is not that trivial as the NLS dynamics is not reversible (the time average of the field produced by any initial condition will eventually be proportional to ψ_{bg}). The procedure is then the following:

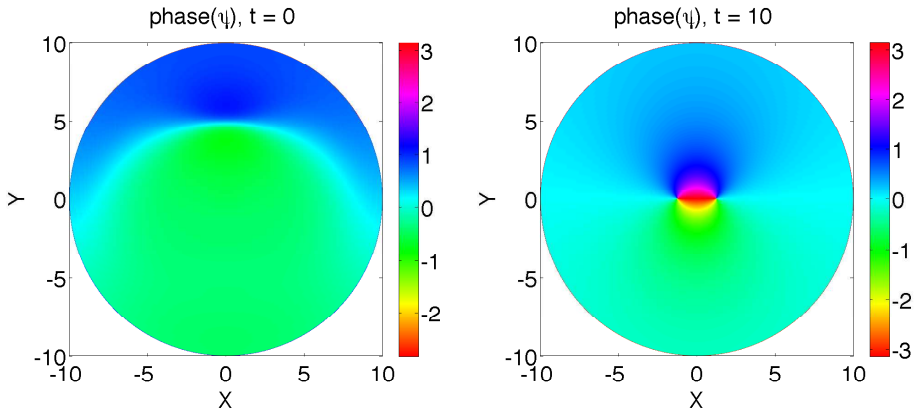
1. take an initial condition with two counter-rotating vortices:

$$\psi_0 = \psi_{bg} \times \psi^v_{\{R/8:0;1\}} \times \psi^v_{\{R/8;\pi;-1\}},$$

   and run the simulation for a duration of 10. As described in [2], the two vortices get closer, merge and disappear,

2. take the complex conjugate of the resulting wave function. By doing so, $\sqrt{n(t)}e^{i\alpha(t)} \mapsto \sqrt{n(t)}e^{-i\alpha(t)}$ and therefore $\boldsymbol{u} = \boldsymbol{\nabla}\alpha \mapsto \boldsymbol{u}' = -\boldsymbol{\nabla}\alpha = -\boldsymbol{u}$,

3. run again the simulation with the previous wave function as an initial condition for a duration of 20. Two vortices, almost exactly identical to the ones defined at the beginning of the first step of this procedure except with opposite circulation, appear and disappear again (see Figure 5).



(a) (b)

Figure 5: Phase of the wave function for the generation of vortices. (a) Initial condition of step 2 in Section 4.3. (b) Vortices having been generated. We recall that the velocity is deduced from the phase by taking its gradient and that vortices are multi-valued points of the phase.

The vortices disappear shortly after having been generated. Nonetheless, we will see that the trajectory of a vortex can be modified, namely by a wave field in the numerical experiments shown in the rest of the present document. Therefore, a strong wave field could be able to separate two vortices and bring them far enough from each other such that they could behave like independent vortices. In the following studies, we will then focus on numerical experiments about one or two vortices of positive circulation[5].

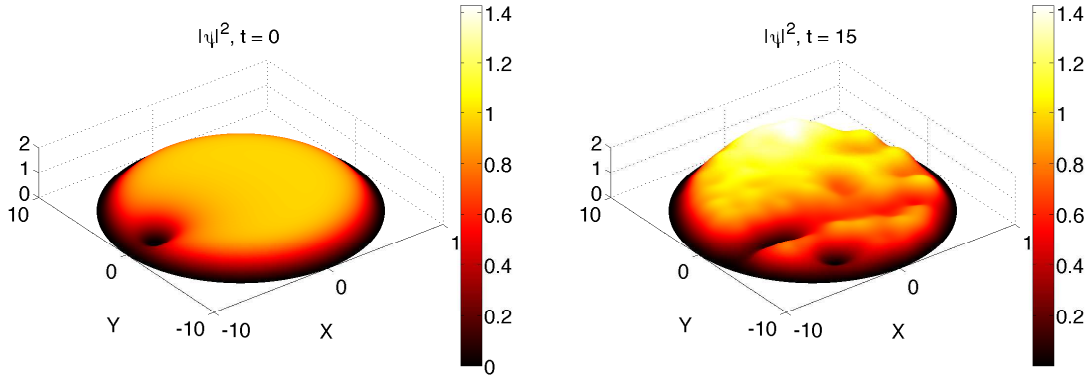
[5] although considering isolated vortices in such a small glass ($R = 10$) is still unrealistic!

4.4 Initial vortex with double circulation

We design here an experiment with one vortex of double circulation at a distance $r_0 = 3R/4$ from the center:

$$\psi_0 = \psi_{bg} \times \psi^v_{\{3R/4;\pi;2\}}$$

(cf. Figure 6(a)) and let it run for a duration of 15. We see in Figure 6(b) that this vortex of double circulation has split into two vortices of simple circulation. Vortices of circulation higher than 2π in absolute value are therefore very likely to be unstable and in the next experiments, we will only consider vortices of simple circulation.



(a) Initial density

(b) Density after a little while

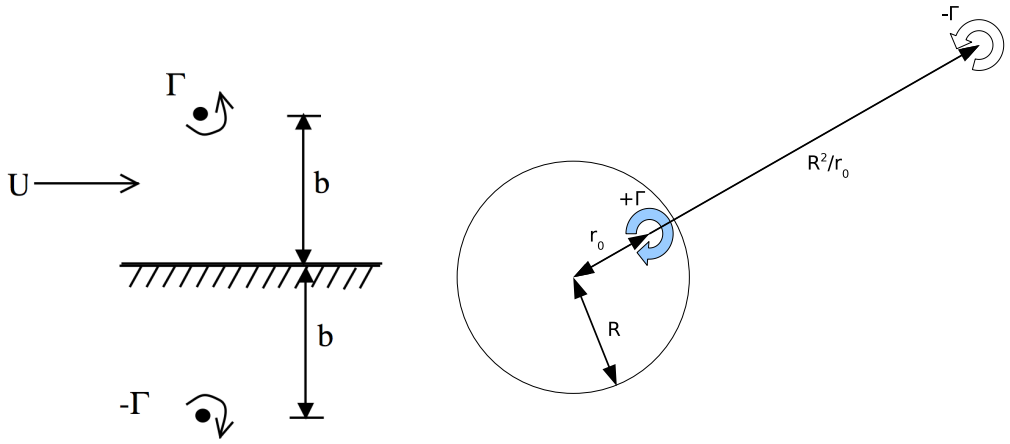
Figure 6: Initial and final density fields for an initial vortex of double circulation (a) which splits into two vortices of single circulation (b).

4.5 Motion of a single vortex

4.5.1 Method of images

In classical fluid dynamics, the motion of a point vortex near a wall is described by the method of images. In the case of a plane wall, the velocity field is the same as if there was no wall but instead a vortex of equal and opposite circulation and symmetrical position with respect to the wall, as sketched in Figure 7(a). In this case, the velocity induced by the virtual vortex and with which the vortex moves is parallel to the wall and equal to $\Gamma/4\pi b$, b being the distance to the wall. In the case of a vortex in a circular glass, the virtual vortex still has equal and opposite circulation but is now situated at a distance R^2/r_0 from the center of the glass, r_0 being the distance of the real vortex to the center (see Figure 7(b)). The velocity induced by it is again parallel to the wall and equal to $\Gamma r_0/2\pi(R^2 - r_0^2)$. The time for a vortex to loop around the entire box is then

$$\tau_{loop} = \frac{4\pi^2(R^2 - r_0^2)}{\Gamma}.$$



(a) Image vortices, plane wall (©2003 MIT, Dpt. of Ocean Engineering)

(b) Image vortices, circular glass

Figure 7: Sketch of the method of images for a vortex near a boundary in two different cases. In Figure (b), the blue vortex is real, the white one is virtual.

As explained in [6] in the case of a plane wall, this model has to include the boundary layer near the wall, which shifts the position of the wall. A way to understand it (rather than to prove it) is to consider a volume sketched in Figure 8, of arbitrary length and delimited by the wall and a line that is $\sqrt{2}$ away from the wall. The particle flux from the two lateral boundaries cancel each other as the system is invariant by translation along the wall and the particle flux through the wall is zero. The conservation of the number of particles tells that the particle flux has to be zero through the boundary facing the interior of the BEC. Within the healing length, the velocity is along the wall as the area is too confined to exhibit too complex structures, which means that there is no particle crossing this boundary either. The free-slip boundary condition that is necessary for the method of image vortices to be applied is therefore active *a healing length away* (times $\sqrt{2}$) from the wall. The correction in our case is then that the effective radius of the glass is $R' = R - \sqrt{2}$ and that the time for a vortex to loop around the glass is

$$\tau'_{loop} = \frac{4\pi^2((R-\sqrt{2})^2 - r_0^2)}{\Gamma}.$$

4.5.2 Numerical experiments

A series of numerical experiments has been performed with the same type of initial conditions as in Section 4.2, all starting from $\theta_0 = \pi$ where the initial distance to the wall has been varied. As displayed in Figure 9, the time for a vortex to make a loop is in general agreement with τ'_{loop}, whereas there is a systematic discrepancy with τ_{loop}. The model with correction is not valid in two cases. The first case is if the initial vortex position is too close to the wall, more precisely between the wall and the circle that is $\sqrt{2}$ away from the

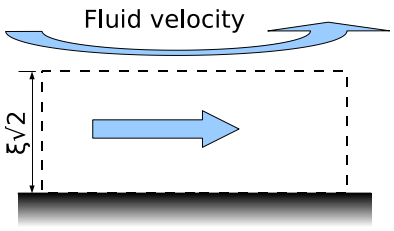


Figure 8: Sketch describing the effect of the boundary layer within the volume materialized by the dashed frame.

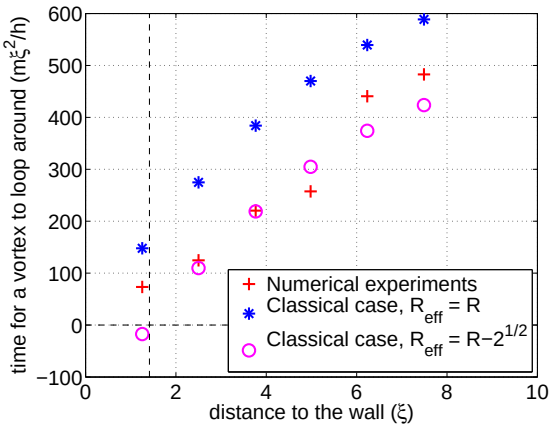
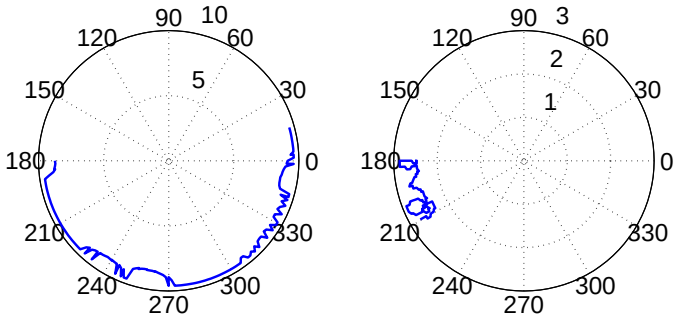


Figure 9: Measure of the time for a vortex to loop around the glass compared with the result provided by the image method with and without correction of the healing length for the effective radius R_{eff}. The vertical dashed line marks the virtual wall at $\sqrt{2}$.

wall, where the model with correction obviously doesn't apply. What happens, as shown in Figure 10(a) is that the vortex can even get "swallowed" in the wall if it is closer to it than $\sqrt{2}$. The second case is when the vortex is too close to the center. It then moves very slowly and is easily pushed by the wave field that has been generated mostly during the transient of the initial moments. As displayed in Figure 10(b), the course of the vortex is then erratic and again can't be described by the image vortex model.



(a) $r_0 = R/4$ (b) $r_0 = 7R/8$

Figure 10: Trajectories of (a) a vortex, within the healing length, getting swallowed in the wall and (b) a vortex, far from the wall, being pushed by impulses due to the waves.

All experiments have been run for a duration of 40 and the time for a vortex to make a loop has been deduced from the angular velocity averaged over that period. An experiment starting from $r_0 = 3\pi/4$ has been run for a duration of 1570 (more than 12 loops), in order to determine if the vortex would get closer to the wall as it would propagate. This would mean that the vortex would lose energy, probably to the wave field. Even over such a long period of time, no shift in the distance to the wall has been observed.

4.6 Motion of two co-rotating vortices

In the case of two co-rotating vortices separated by a distance l, the velocity induced by one vortex at the location of the second one is $\Gamma/(2\pi l)$ and it then takes $\tau_{swap} = \pi^2 l^2/\Gamma$ for the two vortices to swap positions. An experiment is set up with two vortices of single circulation located symmetrically around the center of the box, as defined by

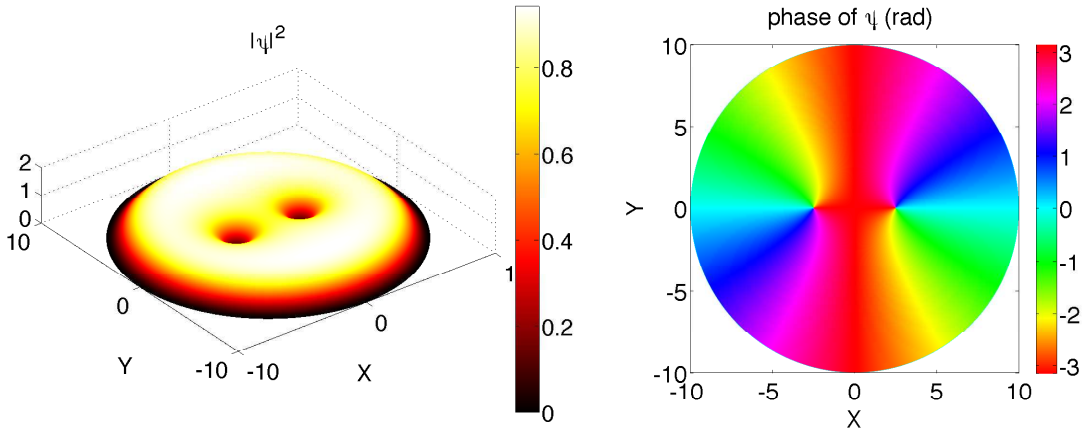
$$\psi_0 = \psi_{bg}(r,\theta) \times \psi^v_{\{R/4;0;1\}} \times \psi^v_{\{R/4;\pi;1\}},$$

as shown in Figure 11. In this case, the effect of the wall on each of these vortices has to be taken into account. Including the velocity induced by the other real vortex and the velocity induced by the virtual vortex of the corrected image method, the velocity of a vortex is

$$U = \frac{r_0}{(R-\sqrt{2})^2 - r_0^2} + \frac{1}{2r_0}$$

and the time for two vortices to swap positions is now

$$\tau'_{swap} = \frac{2\pi\, r_0^2 \left((R-\sqrt{2})^2 - r_0^2\right)}{(R-\sqrt{2})^2 + r_0^2} \approx 33.$$



(a) Initial density

(b) Initial phase

Figure 11: Initial condition with two co-rotating vortices.

We run enough iterations for the two vortices to swap positions and just measure this duration. The measured time is $\tau'_{swap} = 53 \pm 1$, which is 1.6 times longer than the predicted duration. This discrepancy has not been explained yet, but one can notice that the vortices are quite close from each other and that the combined effects of what is going on in the depleted areas surrounding them might have to be taken into account.

5 Conclusion

The NLS equation governing quantum fluid dynamics can be transformed via the Madelung transformation in a set of equations resembling the set of shallow water, irrotational, compressible equations with surface tension. Quantum fluids support density waves behaving like capillary-gravity waves in shallow water and vortices are to be found where both real and imaginary parts of the wave function vanish on isolated points, making the phase of the wave function multi-valued and therefore singular. A novel method of integrating the NLS in polar coordinates, based on an operator-splitting method and the combination of FFTs in the azimuthal directions and quasi-discrete Hankel transforms in the radial direction, has been developed and partially validated. This allowed to do some physics in a quantum glass: we have seen that pair of vortices can be generated, justifying the existence of vortices in a non-rotating BEC and that to limit our study to vortices of single circulation $\Gamma = \pm 2\pi$ was fully relevant as vortices of double (and more) circulation were unstable and would split into vortices of simple circulation. We then saw that the motion of a single vortex has to take into account the presence of the depleted surface between the wall and the interior of the BEC which reduces the effective diameter of the glass. Finally, we saw that the motion of two co-rotating vortices that are close to each other noticeably departs from the classical case.

The possible extensions of this work are numerous and can be sorted into two categories. The first category involves the dynamics of BECs and more especially the interaction with

waves. We have seen here that it can modify the trajectory of a vortex and we infer that it can spearate freshly created vortices. We also infer that vortices can be peeled off a boundary by a wave field, as seen more of less convincingly in Figure 10(a) and that a vortex can lose energy by creating waves, which would lead to its location getting closer to the wall. To check the latter is especially challenging as it would require at least two things: first, to generate cleaner initial conditions for the shape of the wave function in a steady circular BEC and a cleaner profile for the shape of a vortex. Indeed, the approximateness of our ψ_{bg} and $\psi^v_{\{r_0;\theta_0;p\}}$ is the source of transients and of generation of waves, making it difficult to see if waves are generated when a vortex moves. Second, to see if the vortex will get closer after a while requires to run simulations for a while and then to be confident about the robustness of the code over long integrations, which brings us to the second category of extensions of the work, the numerics developed here: with the chosen time step, the code is stable but its duration can probably be optimized. The choice of n_r, n_θ and Y has not been optimized also and the exact effects of the interpolations on the precision and speed of the code remains unknown. De-aliaising has not been taken care of and in general, the performances of the method compared with other methods in polar coordinates are completely unknown.

Acknowledgements The staff of this 2009 GFD summer school is acknowledged, especially the two main lecturers Harvey Segur and Roger Grimshaw and the organisers Oliver Bühler, Karl Helfrich and Janet Fields. The company of the other fellows and undergraduates gravitating around Walsh Cottage made this summer an exceptionnal personal experience and I sincerely hope to be able to enjoy it for many years to come[6]. I also thank Pascale for her delicious brunches on Crooked Pond, George for sacrificing his patience to our softball skills[7] and Marcel Leutenegger for providing precious help in explaining how to use his Matlab Hankel transform toolbox. Last but not least, Oliver is warmly thanked again, this time for his supervision of this project, both efficient and full of humour.

References

[1] S. N. BOSE, *Plancks Gesetz und Lichtquantenhypothese*, Zeitschrift fuer Physik, 26 (1924), pp. 178–181. German translation of Bose's paper on Planck's law by Abert Einstein.

[2] O. BUHLER, *Wave–vortex interactions in fluids and superfluids.* In prep.

[3] K. B. DAVIS, M. O. MEWES, M. R. ANDREWS, N. J. VAN DRUTEN, D. S. DURFEE, D. M. KURN, AND W. KETTERLE, *Bose-einstein condensation in a gas of sodium atoms*, Phys. Rev. Lett., 75 (1995), pp. 3969–3973.

[4] M. GUIZAR-SICAIROS AND J. C. GUTIÉRREZ-VEGA, *Computation of quasi-discrete hankel transforms of integer order for propagating optical wave fields*, J. Opt. Soc. Am. A, 21 (2004), pp. 53–58.

[6] Long live the 6$^{\text{th}}$ Republic.

[7] Which existence has not yet been proved as we write.

[5] H. Lamb, *Hydrodynamics*, Cambridge Mathematical Library, Cambridge University Press, 6 ed., 1932, ch. VII, pp. 202–248.

[6] P. Mason, N. G. Berloff, and A. L. Fetter, *Motion of a vortex line near the boundary of a semi-infinite uniform condensate*, Physical Review A (Atomic, Molecular, and Optical Physics), 74 (2006), p. 043611.

[7] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes: the art of scientific computing*, Cambridge University Press, New York, NY, USA, 3 ed., 2007, ch. 20.3.3, pp. 1052–1053.

Ray Theory of Nonlinear Water Waves

Andong He

September 29, 2009

1 Introduction.

In past decades there has been growing interest in propagation of nonlinear waves over a variable bottom and internal waves in a variable medium, which leads to a nonlinear partial differential equation of KdV type with variable coefficients (vc-KdV). The study of vc-KdV equation could enhance our understanding of many problems in oceanography and plasma, such as tsunami([10]) and dust ion acoustic waves([7]). A methodology in linear theory, the so-called ray method, was extended to nonlinear wave theory in Cartesian geometry by Shen&Keller[8] and in cylindrical geometry by Shen&Shen[9]. The wave amplitude satisfies the vc-KdV equation along each ray which is determined by the Eikonal equation. The solution is valid when the wavelength is small compared to other horizontal scale lengths. Since the canonical Korteweg-de Vries equation describes the propagation of long waves in shallow water, in some sense this method combines features of both the short wave theory and long wave theory.

In section 2 a vc-KdV equation is derived by applying the ray method in Cartesian coordinate system. Section 3 is dedicated to seek solutions to Eikonal equation in the case that the topography is independent of one spatial variable. Rays in plane waves and circular waves propagating over different bottoms are calculated and plotted. A variable coefficient KdV equation in cylindrical geometry (vc-cy-KdV) is given in section 4, which can be transformed into a vc-KdV by a suitable change of variables. Exact solutions to the vc-cy-KdV equation and conditions under which we are able to find them explicitly are discussed. We also obtain approximate solutions using the method developed by Grimshaw[4], and study the large-time behavior of the amplitude of both cnoidal waves and solitary waves. These two types of waves share the same critical value of the bottom slope for the amplitude to blow up or vanish far away from the source of waves. Finally in section 5 we discuss some directions in which future work could advance.

2 Formulation of the problem.

Let us consider an inviscid, incompressible fluid over an uneven bottom $z^* = -h^*(x^*, y^*)$, as sketched in figure 1. The free surface is denoted by $z^* = \eta^*(x^*, y^*, z^*, t^*)$ where η^* is an unknown function. Denote by ρ, P and $\vec{u}$ the fluid density, pressure and velocity respectively. We can define dimensionless variables as follows:

$$
\begin{aligned}
&h = \frac{h^*}{H}, \quad \eta = \frac{\eta^*}{H}, \quad \epsilon = (\frac{H}{L})^{\frac{2}{3}}, \quad \rho = \frac{\rho^*}{D} \\
&(x, y, z) = (\frac{\epsilon^{\frac{3}{2}}}{H}x^*, \frac{\epsilon^{\frac{3}{2}}}{H}y^*, \frac{z^*}{H}), \quad t = \epsilon^{\frac{3}{2}}(\frac{g}{H})^{\frac{1}{2}}t^*, \\
&P = \frac{P^*}{gHD}, \quad \vec{u} = (u, v, w) = (gH)^{-\frac{1}{2}}(u^*, v^*, \epsilon^{-\frac{1}{2}}w^*),
\end{aligned} \tag{2.1}
$$

where H, L, D are scales for vertical length, horizontal length and density respectively.

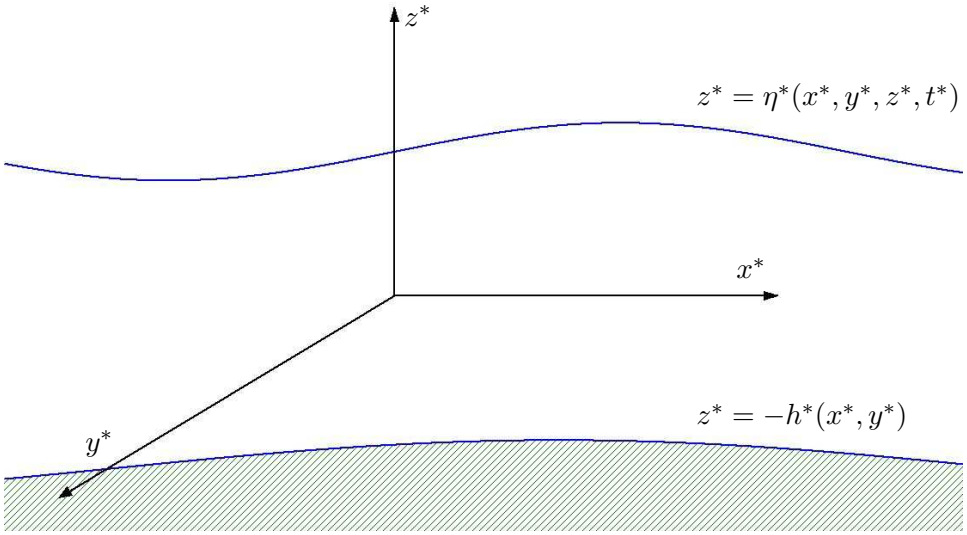


Figure 1: Flow region is between the rigid bottom $z^* = -h^*(x^*, y^*)$ and the free surface $z^* = \eta^*(x^*, y^*, z^*, t^*)$.

Using these dimensionless variables, we can represent the equation of motion, equation of continuity and boundary conditions on the rigid bottom and the free surface in the following way:

- Equation of motion: $\rho^*(\frac{\partial \vec{u}^*}{\partial t^*} + \vec{u}^* \cdot \nabla^* \vec{u}^*) = -\nabla^* P^*$ becomes

$$\epsilon[\rho(u_t + uu_x + vu_y) + P_x] + \rho w u_z = 0, \tag{2.2}$$

$$\epsilon[\rho(v_t + uv_x + vv_y) + P_y] + \rho w v_z = 0, \tag{2.3}$$

$$\epsilon^2\rho(w_t + uw_x + vw_y) + \epsilon\rho w w_z + P_z + \rho = 0. \tag{2.4}$$

- Equation of continuity $\nabla^* \cdot \vec{u}^* = 0$ becomes

$$\epsilon(u_x + v_y) + w_z = 0. \tag{2.5}$$

- Two boundary conditions $P^* = const$, $\frac{D^*}{D^* t^*}(z^* - \eta^*) = 0$ at $z^* = \eta^*$ and $\vec{u}^* \cdot \nabla^*(z^* +$

$h^*) = 0$ at $z^* = -h^*$ becomes

$$P = C, \quad \text{at } z = \eta, \tag{2.6a}$$

$$\epsilon(\eta_t + u\eta_x + v\eta_y) - w = 0, \quad \text{at } z = \eta, \tag{2.6b}$$

$$\epsilon(uh_x + vh_y) + w = 0, \quad \text{at } z = -h. \tag{2.6c}$$

Here ∇^* is the dimensional gradient operator and $\frac{D^*}{D^*t^*}$ is the dimensional material derivative $\frac{D}{Dt}$.

2.1 Asymptotic expansion and Eikonal equation.

To find wave-like solutions to (2.2)-(2.6) we introduce a *phase function* $S(x, y, t)$ and a *"fast-phase"* variable $\xi = \epsilon^{-1}S(x, y, t)$. Assume that u, v, w, P and η all have the form of asymptotic expansion

$$G(t, x, y, z, \xi) = G_0(t, x, y, z, \xi) + \epsilon G_1(t, x, y, z, \xi) + O(\epsilon^2). \tag{2.7}$$

By noting that operators $\frac{\partial}{\partial t}$, $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ become $\frac{\partial}{\partial t}+\epsilon^{-1}S_t\frac{\partial}{\partial \xi}$, $\frac{\partial}{\partial x}+\epsilon^{-1}S_x\frac{\partial}{\partial \xi}$ and $\frac{\partial}{\partial y}+\epsilon^{-1}S_y\frac{\partial}{\partial \xi}$ respectively, we can expand equations (2.2)-(2.6). Equating $O(1)$ terms to zero yields the static-flow solution; and equating $O(\epsilon)$ terms to zero yields

$$\rho u_{1\xi}S_t + P_{1\xi}S_x = 0, \tag{2.8}$$

$$\rho v_{1\xi}S_t + P_{1\xi}S_y = 0, \tag{2.9}$$

$$P_{1z} = 0, \tag{2.10}$$

$$u_{1\xi}S_x + v_{1\xi}S_y + w_{1z} = 0, \tag{2.11}$$

$$P_1 = \rho\eta_1, \quad \text{at } z = 0, \tag{2.12}$$

$$w_1 = \eta_{1\xi}S_t, \quad \text{at } z = 0, \tag{2.13}$$

$$w_1 = 0, \quad \text{at } z = -h(x, y). \tag{2.14}$$

To solve equations (2.8)-(2.14) for unknowns $u_1, v_1, w_1, P_1, \eta_1$ and S, we firstly solve (2.8) and (2.9) for $u_{1\xi}$ and $v_{1\xi}$ to get

$$u_{1\xi} = -\frac{P_{1\xi}S_x}{\rho S_t}, \quad v_{1\xi} = -\frac{P_{1\xi}S_y}{\rho S_t}. \tag{2.15}$$

Substituting (2.15) into (2.11) yields

$$w_{1z} = \frac{P_{1\xi}(S_x^2 + S_y^2)}{\rho S_t}. \tag{2.16}$$

From (2.10) and (2.16) we know w_{1z} is independent of z, so using boundary condition (2.14) we have

$$w_1 = \frac{P_{1\xi}(S_x^2 + S_y^2)}{\rho S_t}(z + h). \tag{2.17}$$

Equations (2.12) and (2.13) implies

$$w_1 = \frac{P_{1\xi} S_t}{\rho}, \quad \text{at } z = 0. \tag{2.18}$$

Now combining (2.17) and (2.18) and eliminating w_1 we get the *Eikonal* equation

$$\frac{S_t^2}{S_x^2 + S_y^2} = h(x, y). \tag{2.19}$$

By solving the Eikonal equation with proper boundary conditions we can determine the phase function $S(x, y, t)$, which is postponed to section 3. Once the surface elevation η_1 and the phase function S are determined, P_1 can be found by (2.12), and w_1 by (2.17). From (2.15) we can obtain special solutions u_1, v_1 as

$$u_1 = -\frac{P_1 S_x}{\rho S_t}, \quad v_1 = -\frac{P_1 S_y}{\rho S_t}. \tag{2.20}$$

Arbitrary functions independent of ξ can be added to the right-hand sides of (2.20). We choose them to be zeros so that the fluid velocity has the same direction as rays.

2.2 Amplitude equation.

To determine η_1, we need to proceed further by equating coefficients of $O(\epsilon^2)$ terms in equations (2.2)-(2.6):

$$S_x P_{2\xi} + \rho S_t u_{2\xi} = -(\rho u_{1t} + \rho u_1 S_x u_{1\xi} + \rho v_1 S_y u_{1\xi} + P_{1x}), \tag{2.21}$$

$$S_y P_{2\xi} + \rho S_t v_{2\xi} = -(\rho v_{1t} + \rho u_1 S_x v_{1\xi} + \rho v_1 S_y v_{1\xi} + P_{1y}), \tag{2.22}$$

$$P_{2z} = -\rho S_t w_{1\xi}, \tag{2.23}$$

$$w_{2z} + S_x u_{2\xi} + S_y v_{2\xi} = -u_{1x} - v_{1y}, \tag{2.24}$$

$$P_2 - \rho \eta_2 = 0, \quad \text{at } z = 0, \tag{2.25a}$$

$$w_2 - S_t \eta_{2\xi} = \eta_{1t} + u_1 S_x \eta_{1\xi} + v_1 S_y \eta_{1\xi} - w_{1z} \eta_1, \quad \text{at } z = 0, \tag{2.25b}$$

$$w_2 = -u_1 h_x - v_1 h_y, \quad \text{at } z = -h(x, y). \tag{2.25c}$$

Equations (2.21)-(2.24) are a system of equations of u_2, v_2, w_2, η_2 and P_2 whose right-hand sides are all known except for exactly one quantity (η_1, w_1, or P_1). By straightforward but a little tedious calculations we can eliminate u_2, v_2 and η_2 to get a compatibility condition under which a solution exists(see appendix in [8] for details). This compatibility condition turns out to be in a following KdV-like form of a nonlinear partial differential equation

$$\widehat{A}_t + \frac{J_t}{2J}\widehat{A} + (\frac{\omega h^2}{2} - \frac{\omega}{\rho^2 h})\widehat{A}\widehat{A}_\xi + \frac{\omega^3 h}{6}\widehat{A}_{\xi\xi\xi} = 0, \tag{2.26}$$

where $\widehat{A} = \rho\eta_1$, ω (angular frequency) will be defined in section 3, and J is the Jacobian from the ray coordinates $(\tau, \gamma_1, \gamma_2)$ to the time-Cartesian coordinate system (t, x, y).

Denote $A = J^{1/2}\widehat{A}$, then (2.26) becomes

$$A_t + f_1(t)AA_\xi + f_2(t)A_{\xi\xi\xi} = 0, \tag{2.27}$$

where $f_1(t) = J^{-1/2}(\frac{\omega h^2}{2} - \frac{\omega}{\rho^2 h})$, and $f_2(t) = \frac{\omega^3 h}{6}$.

Equation (2.27) is called the variable coefficient Korteweg-de Vries equation(vc-Kdv) and was derived by Shen&Keller[8]. It was shown by Joshi [5] that it is completely integrable if and only if coefficient functions f_1 and f_2 satisfy

$$f_2(t) = f_1(t)(c_1 \int^t f_1(s)ds + c_2), \tag{2.28}$$

for arbitrary constants c_1 and c_2. Under this condition, Lax pair and infinitely many conservation laws of (2.27) can be found(see e.g. Fan[2] and Zhang[12]).

3 Ray tracing.

Eikonal equation (2.19) can be solved by using the method of characteristics. By doing so, we obtain the following system of ordinary differential equations

$$\dot{x}(\tau) = -\frac{S_x}{\sqrt{S_x^2 + S_y^2}}\ \theta, \tag{3.1a}$$

$$\dot{y}(\tau) = -\frac{S_y}{\sqrt{S_x^2 + S_y^2}}\ \theta, \tag{3.1b}$$

$$\dot{t}(\tau) = 1, \tag{3.1c}$$

and

$$\dot{S}_x(\tau) = \sqrt{S_x^2 + S_y^2}\ \theta_x, \tag{3.2a}$$

$$\dot{S}_y(\tau) = \sqrt{S_x^2 + S_y^2}\ \theta_y, \tag{3.2b}$$

$$\dot{S}_t(\tau) = 0, \tag{3.2c}$$

and very importantly,

$$\dot{S}(\tau) = 0, \tag{3.3}$$

where $\theta(x, y) = -h^{\frac{1}{2}}(x, y)$, and *dot* is the total derivative with respect to τ.

System of equations (3.1) determines a two-parameter family of space-time curves, which are called *rays*; (3.2) determine the change of gradient of S along a ray; and (3.3) indicates that S remains constant along a ray. In following discussion we often do not distinguish ∂t and $\partial\tau$ due to (3.1c) unless special attention is drawn. It follows from (3.2c) that on each ray $S_t(\tau)$ is also a constant. We will only consider in this paper the case when S_t is the same constant for all rays. We call $\omega = -S_t$ the *angular frequency*.

We can parameterize rays by using parameters γ_1 and γ_2 which are constants along a ray. They can be chosen to be, for instance,

$$\begin{aligned} \gamma_1 &= S(x,y,t), \\ \gamma_2 &= y - y(\tau(x)), \end{aligned} \tag{3.4}$$

then $\frac{\partial\gamma_1}{\partial x} = S_x$, $\frac{\partial\gamma_1}{\partial y} = S_y$, $\frac{\partial\gamma_2}{\partial x} = -\frac{\dot{y}}{\dot{x}}$, and $\frac{\partial\gamma_2}{\partial y} = 1$. The Jacobian J from the ray coordinates $(\tau, \gamma_1, \gamma_2)$ to the time-Cartesian coordinate (t, x, y) can be calculated as

$$J = \left| \frac{\partial\gamma_1}{\partial x}\frac{\partial\gamma_2}{\partial y} - \frac{\partial\gamma_1}{\partial y}\frac{\partial\gamma_2}{\partial x} \right|^{-1} = \frac{S_x}{S_x^2 + S_y^2} = \frac{S_x h}{S_t^2}. \tag{3.5}$$

In general the system of ordinary differential equations (3.1)-(3.3) is difficult to solve since they are coupled with each other. But if the topography is only dependent on one variable they can be solved analytically. Assume $h_y = 0$, then it follows from (3.2b) that $S_y(\tau) = S_y(0) = k_2$. We substitute $\theta_x = \frac{\dot{\theta}}{\dot{x}}$ into (3.2a) and use (3.1a) to get

$$\frac{S_x \dot{S}_x}{S_x^2 + k_2^2} = -\frac{\dot{\theta}}{\theta}$$

Integrating it from 0 to τ gives us

$$S_x(\tau) = [h_0(k_1^2 + k_2^2)h^{-1} - k_2^2]^{\frac{1}{2}}, \tag{3.6}$$

where $k_1 = S_x(0)$ and $h_0 = h(0)$. The location of rays can be obtained by substituting $S_x(\tau)$ into (3.1a) and (3.1b):

$$\dot{x}(\tau) = \sqrt{h(x(\tau)) - \frac{k_2^2}{h_0(k_1^2 + k_2^2)}h^2(x(\tau))}, \tag{3.7a}$$

$$\dot{y}(\tau) = \frac{k_2}{\sqrt{h_0(k_1^2 + k_2^2)}}h(x(\tau)). \tag{3.7b}$$

Equation (3.7a) is a first-order ordinary differential equation thus can be solved analytically for a large scope of bottom functions $h(x)$; then (3.7b) is solved readily. The value of S at any point P will be given by its value at a boundary point Q from which the ray emits and passes through P.

Next we will consider several physical scenarios (various topography & various forms of waves) when (3.7) is solved with suitable boundary conditions and rays are calculated explicitly.

3.1 Plane waves.

Let us suppose that a sinusoidal plane wave is generated at $x = -\infty$ and propagating towards the shore at $x = l$. The angle between the wavefronts and positive $y-$axis is φ.

The bottom function is selected in the following way so that (3.2a) and (3.2b) can be solved easily,

$$h(x) = \begin{cases} \alpha^2 l^2, & x \le 0; \\ \alpha^2 (l-x)^2, & x \ge 0. \end{cases} \quad \text{with } \alpha > 0. \tag{3.8}$$

A phase function representing sinusoidal plane waves is $S(x, y, t) = k_1 x + k_2 y - \omega t$ with $\frac{k_2}{k_1} = \arctan(\varphi)$. Parameters k_1, k_2 and ω are related by the dispersion relation $\frac{\omega^2}{k_1^2+k_2^2} = h_{loc}$, where h_{loc} is the local depth of water. The boundary conditions for the system are prescribed on $\Gamma_1 = \{(x, y, z) | x = 0, y \in \mathbb{R}, t > 0\}$ as

$$S_x(0) = k_1, \quad S_y(0) = k_2, \quad S_t(0) = -\omega, \quad S(0, y, t) = k_2 y - \omega t. \tag{3.9}$$

It immediately follows from (3.2) that $S_y(\tau) = k_2$, $S_t(\tau) = -\omega$, and

$$S_x(\tau) = \frac{1}{2c_1}(c_1^2 e^{\alpha\tau} - k_2^2 e^{-\alpha\tau}), \tag{3.10}$$

where the constant c_1 is defined as $c_1 = \sqrt{k_1^2 + k_2^2} + k_2$. The location of characteristics can be readily calculated

$$\begin{aligned} \dot{x} &= \alpha(l-x)\frac{c_1^2 e^{\alpha\tau} - k_2^2 e^{-\alpha\tau}}{c_1^2 e^{\alpha\tau} + k_2^2 e^{-\alpha\tau}} \\ \dot{y} &= \frac{\alpha l(c_1^2 + k_2^2)}{2c_1 k_2}[\operatorname{sech}(\log\frac{c_1}{k_2} + \alpha\tau)]^2 \\ \dot{t} &= 1. \end{aligned} \tag{3.11}$$

Using initial conditions $x(0) = 0$, $y(0) = y_0$ and $t(0) = t_0$ we can integrate (3.11) to obtain

$$\begin{aligned} x(\tau) &= l(1 - \frac{c_1^2 + k_2^2}{c_1^2 e^{\alpha\tau} + k_2^2 e^{-\alpha\tau}}) \\ y(\tau) &= y_0 + \frac{l(c_1^2 + k_2^2)}{2c_1 k_2}(\frac{c_1^2 e^{\alpha\tau} - k_2^2 e^{-\alpha\tau}}{c_1^2 e^{\alpha\tau} + k_2^2 e^{-\alpha\tau}} - \frac{c_1^2 - k_2^2}{c_1^2 + k_2^2}), \\ t(\tau) &= t_0 + \tau, \end{aligned} \tag{3.12}$$

which are plotted in figure 2(a).

To find the phase function S at any point (x^*, y^*, t^*), we need to find y_0, t_0 and τ such that $x(\tau) = x^*$, $y(\tau) = y^*$, and $t(\tau) = t^*$. Then the phase function is

$$S(x^*, y^*, t^*) = S(0, y_0, t_0) = k_2 y_0 - \omega t_0.$$

One can easily see in (3.12) that

$$\lim_{\tau\to\infty} x(\tau) = l, \quad \text{and} \quad \lim_{\tau\to\infty} y(\tau) = y_0 + (\sqrt{k_1^2 + k_2^2} + k_1)^{-1} k_2 l,$$

indicating that the rays will approach the shoreline asymptotically.

3.2 Waves generated by a point oscillator.

Suppose that there is a point source oscillator at the origin, generating sinusoidal waves. In this case all information on the phase function are only available on a straight line $\Gamma_2 = \{x = y = 0, t \geq 0\}$, where all rays are emitted. Therefore in order to determine value of the phase function it is necessary to find both the point on $t-$axis and the angle from which a ray emits. Since in a small neighborhood of the origin the phase function can be written in the form of $S = kr - \omega t + o(r)$, the boundary condition for phase function on $t-$axis is $S = -\omega t$. For a ray leaving the origin with an angle ϕ to the positive $x-$axis, its initial condition is $(S_x, S_y, S_t)|_{x=y=0} = (k\cos\phi, k\sin\phi, -w)$.

The analysis leading to the result in previous section can be applied here, with k_1, k_2 replaced by $k\cos\phi$, $k\sin\phi$ respectively. A straightforward calculation shows that the rays are given by

$$\begin{aligned} x(\tau) &= l - l(\cos^2(\frac{\phi}{2})e^{\alpha\tau} + \sin^2(\frac{\phi}{2})e^{-\alpha\tau})^{-1}, \\ y(\tau) &= \frac{l}{\sin\phi}(\frac{\cos^2(\frac{\phi}{2})e^{\alpha\tau} - \sin^2(\frac{\phi}{2})e^{-\alpha\tau}}{\cos^2(\frac{\phi}{2})e^{\alpha\tau} + \sin^2(\frac{\phi}{2})e^{-\alpha\tau}} - \cos\phi), \\ t(\tau) &= t_0 + \tau, \end{aligned} \tag{3.13}$$

for $\phi \in (\frac{-\pi}{2}, \frac{\pi}{2})\backslash\{0\}$. Notice that $\lim_{\tau\to\infty} x(\tau) = l$, showing that rays approach the shore asymptotically which is similar to the plane-wave case. When $\phi \in (\frac{\pi}{2}, \frac{3\pi}{2})$, waves are traveling in a medium with constant water depth, thus all rays in this region are propagating radially. Rays given by (3.13) are plotted in figure2(b).

To find the value of S at any point (x^*, y^*, t^*), we need to find ϕ, t_0 and τ such that $x(\tau) = x^*$, $y(\tau) = y^*$ and $t(\tau) = t^*$. Then the phase function is

$$S(x^*, y^*, t^*) = S(0, 0, t_0) = -\omega t_0.$$

3.3 Reflected rays in plane waves.

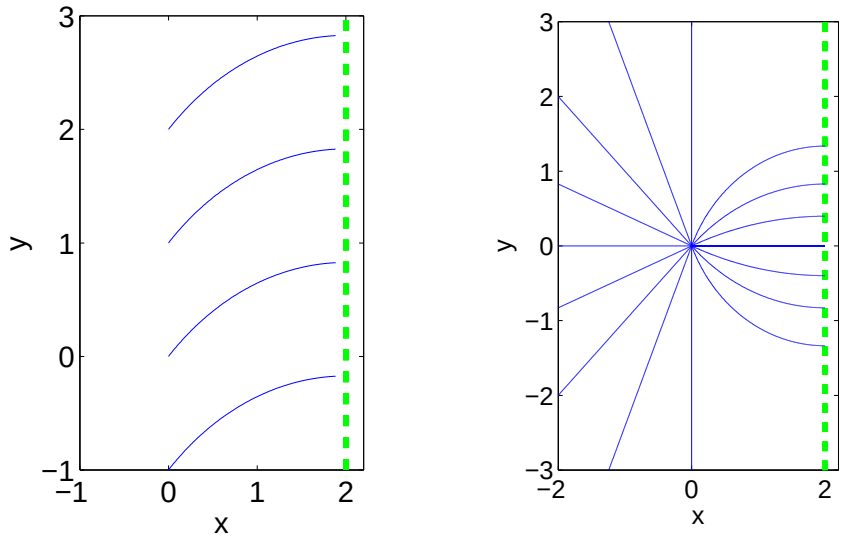
The fact that rays approach to shoreline asymptotically without being reflected is due to the nature of varying topography. Indeed it can be easily shown that the rays will not be reflected back for a bottom function $h(x) \sim (x - l)^{-\beta}$ if and only if $\beta \geq 2$.

In the case of a piecewise linear topography:

$$h(x) = \begin{cases} \alpha^2 l, & x \leq 0; \\ \alpha^2(l - x), & x \geq 0. \end{cases} \qquad \text{with } \alpha > 0, \tag{3.14}$$

we can solve equations (3.7a) and (3.7b) to get

$$\begin{aligned} x(\tau) &= l - \frac{h_0(k_1^2 + k_2^2)}{2k_2^2}[1 - \sin(\frac{k_2\alpha}{\sqrt{h_0(k_1^2 + k_2^2)}}\tau + \arcsin(\frac{k_1^2 - k_2^2}{k_1^2 + k_2^2}))], \\ y(\tau) &= y_0 - \frac{k_1 l}{k_2} + \frac{(k_1^2 + k_2^2)l}{2k_2^2}[\frac{k_2\alpha}{\sqrt{l(k_1^2 + k_2^2)}}\tau + \cos(\frac{k_2\alpha}{\sqrt{l(k_1^2 + k_2^2)}}\tau + \arcsin(\frac{k_1^2 - k_2^2}{k_1^2 + k_2^2}))], \end{aligned} \tag{3.15}$$



(a) (b)

Figure 2: Rays of (a) plane waves propagating from $x = -\infty$ and (b) circular waves generated by a point source at the origin, moving toward a shoreline. The bottom function is chosen to be that in (3.8) with $\alpha = 1$ and $l = 2$. Other parameters are: (a) $k_1 = k_2 = 1, y_0 = -1, 0, 1, 2$; (b) $k = 1, \phi = \frac{n\pi}{8} (n = 0, ..., 15)$.

for $\tau \leq \tau_M = \arccos(\frac{2k_1k_2}{k_1^2+k_2^2})$. Here τ_M is the time when rays hit the shoreline and bounce back.

At the initial stage of traveling along the beach, rays propagate in an apparently analogous way as that in figure 2a, and tend to be perpendicular to the shore. At some finite time they reach the beach where the water depth is zero. Therefore current theory breaks down in the vicinity of the shore and a more careful examination is needed to correctly describe the process taking place therein. Nevertheless, this difficulty does not hinder us from obtaining a solution to the Eikonal equation if we assign to S on $x = l$ the value carried by incoming rays and use them as the boundary conditions for reflected rays. Both incoming rays and reflected rays are plotted in figure 3.

4 Solutions to vc-cy-KdV equation.

The corresponding KdV-like equation in the cylindrical coordinate system can be obtained by a similar approach (Shen&Shen[9]); namely,

$$\eta_{1t} + (2\omega^{-1}rJ)^{-1}\frac{d\omega^{-1}rJ}{dt} + \frac{3\omega\eta_1\eta_{1\xi}}{2h} + \frac{1}{6}\omega^3 h\eta_{1\xi\xi\xi} = 0, \tag{4.1}$$

where r stands for the radial component; η_1, ω, J and h are analogously defined in the cylindrical coordinate system.

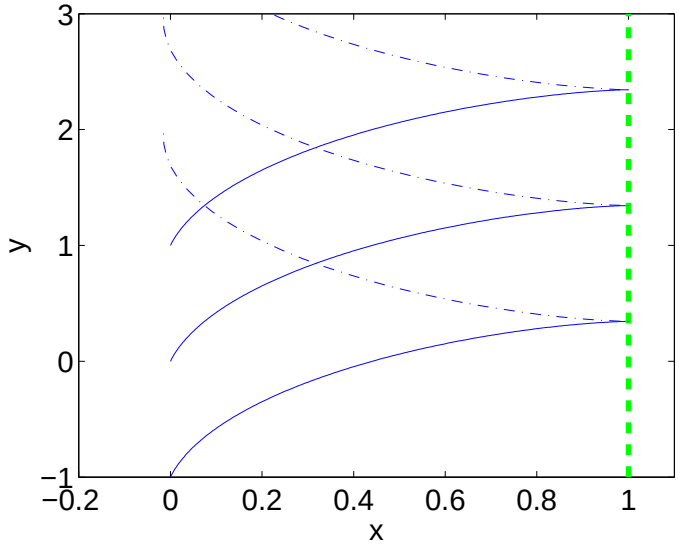


Figure 3: Rays of plane waves (solid curves) traveling over a beach with constant slope are reflected by the beach (dot-dashed curves), with $\alpha = 1, l = 1, k_1 = 1, k_2 = 8$, and $y_0 = -1, 0, 1$.

Assume now that h is a function of r only and all other variables do not depend on the azimuthal variable ϕ, then the phase function is

$$S(r, \phi, t) = -t + \int_{r_0}^{r} h^{-\frac{1}{2}} dr, \tag{4.2}$$

for some fixed radius r_0. If we introduce a function $A(t, r, \xi)$ such that $\eta_1 = Ah(r)$, then (4.1) can be written as a vc-cl-KdV equation

$$A_r + (\frac{3}{2}h^{-\frac{1}{2}})AA_\xi + (\frac{1}{6}h^{\frac{1}{2}})A_{\xi\xi\xi} + (\frac{5}{4}h_r h^{-1})A + \frac{A}{2r} = 0, \tag{4.3}$$

Equation (4.3) is the counterpart of (2.27) in the cylindrical coordinate system under the assumption of azimuthal independence.

By applying a change of variables

$$A(r, \xi) = \frac{2\sqrt{6}}{3} h^{-5/4} r^{-1/2} B(\sigma, \xi), \quad \text{and} \quad \sigma = \frac{1}{6}\int^{r} h^{1/2} dr, \tag{4.4}$$

equation (4.3) becomes

$$B_\sigma + f(\sigma) B B_\xi + B_{\xi\xi\xi} = 0, \tag{4.5}$$

where

$$f(\sigma) = 6h^{-9/4}(\int^{\sigma} h^{-1/2} d\sigma)^{-1/2} = 6\sqrt{6} r^{-1/2} h(r)^{-9/4}. \tag{4.6}$$

Equation (4.5) is in the form of (2.27), except that the "time" variable σ now has a different scale from the actual time variable t.

4.1 Exact solutions

To look for exact solutions to (4.5), we apply the following change of variables suggested by Grimshaw (appendix in [3])

$$B(\sigma,\xi) = f(\sigma)[Q(s,x) - m_1\xi], \quad x = f(\sigma)\xi, \quad s = g(\sigma), \tag{4.7}$$

where m_1 is some constant, $g(\sigma)$ is to be determined, and $Q(s,x)$ satisfies the standard KdV equation

$$Q_s + QQ_x + Q_{xxx} = 0. \tag{4.8}$$

Substituting (4.7) into (4.5) yields

$$\begin{aligned}(f' - m_1 f^3)Q + (\xi f f' - m_1\xi f^4)Q_x + (m_1^2\xi f^3 - m_1\xi f') \\ + fg'Q_s + f^4QQ_x + f^4Q_{xxx} = 0.\end{aligned} \tag{4.9}$$

Therefore $Q(s,x)$ will satisfy (4.8) if each bracketed term in (4.9) vanishes and $fg' = f^4$. This means that when $m_1 \neq 0$, f and g have to satisfy

$$f(\sigma) = (-2m_1\sigma + m_2)^{-1/2}, \quad g(\sigma) = \frac{1}{m_1}(-2m_1\sigma + m_2)^{-1/2}. \tag{4.10}$$

where m_2 is an integrating constant. This is consistent with the integrability condition (2.28) with $c_2 = 0$. According to (4.6), (4.10) and definition of σ, the bottom function has to satisfy

$$h(r) = (-72m_1)^{1/4}(m_3 r^{-8/9} + 1)^{1/4}. \tag{4.11}$$

When h is chosen as in (4.11), we may combine (4.4), (4.7) and any exact solution to the standard KdV to obtain a new solution to the vc-cy-KdV equation. Nevertheless, the presence of the term "$m_1\xi$" in transformation (4.7) changes the condition of the solution at infinity both in temporal and spacial variables. Consequently, solutions to the standard KdV with physical meanings become unbounded solutions to vc-cy-KdV. As an illustration to it, let us take a soliton solution to (4.8)

$$Q(s,x) = 3c\,\mathrm{sech}^2[\frac{\sqrt{c}}{2}(x - cs - b)], \tag{4.12}$$

where b is any constant and c encodes information of amplitude, wavenumber and frequency of the soliton. Then

$$A(r,t) = \kappa_1 r^{-1}\,\mathrm{sech}^2[\kappa_2 r^{-\frac{1}{2}}(r - \kappa_3 t) + \kappa_4 r^{-\frac{1}{2}} + \kappa_5] + \kappa_6 r^{-\frac{1}{2}}(r - \kappa_3 t), \tag{4.13}$$

for some constants $\kappa_i, (i = 1, ..., 6)$. Solution (4.13) is plotted in figure 4.

We notice that when $f(\sigma)$ is a constant ($m_1 = 0$), there is no need of applying another transformation to (4.5); thereupon any bounded solution to it remains bounded to vc-cy-KdV. In particular, solitary waves will keep their soliton-like shape except that the amplitude may vary due to combined effects of geometrical spreading and shoaling. Suppose $f = 6\lambda_1$ for some constant λ_1, then one soliton solution to (4.5) is

$$B = \frac{c}{2\lambda_1}\,\mathrm{sech}^2[\frac{\sqrt{c}}{2}(\xi - c\sigma - b)], \quad \text{for a constant } b. \tag{4.14}$$

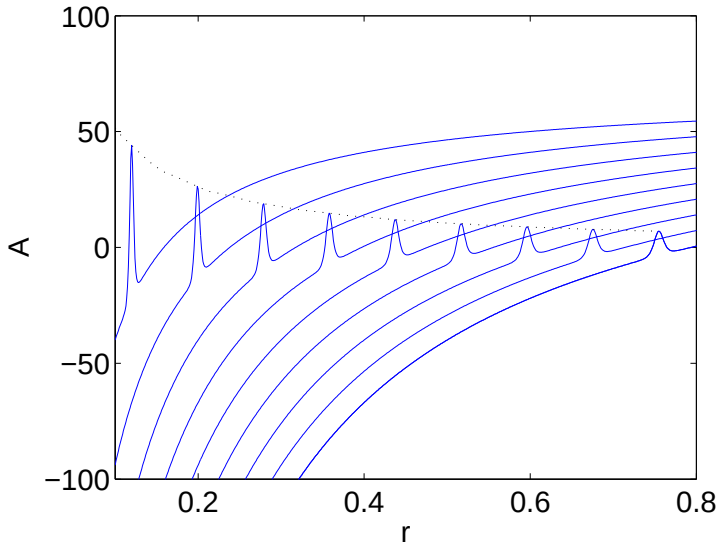


Figure 4: An unbounded solution (4.13) at different times. The dotted curve is the envelope of local solitons.

Now

$$\sigma = 2^{-\frac{35}{9}} 3^{\frac{10}{9}} \lambda_1^{-\frac{2}{9}} r^{\frac{8}{9}} - \lambda_2, \tag{4.15a}$$

$$h(\sigma) = [\frac{8\lambda_1^2}{9}(\sigma + \lambda_2)]^{-\frac{1}{4}} = (\frac{36}{\lambda_1^4})^{1/9} r^{-\frac{2}{9}}. \tag{4.15b}$$

where λ_2 is an arbitrary constant. Combining (4.4),(4.14),(4.15) and definition of ξ we obtain a bounded soliton solution to vc-cy-KdV (4.3):

$$A(r, \xi) = 2^{\frac{2}{9}} 3^{-\frac{7}{9}} r^{-\frac{2}{9}} \lambda_1^{-1} \operatorname{sech}^2[\frac{\sqrt{c}}{2}(\epsilon^{-1}\xi - c\sigma - b)], \tag{4.16}$$

which is plotted in figure 5 for $r \geq r_0$. An examination of solution (4.16) shows that for large r the amplitude decays to 0 as $r^{-2/9}$. Moreover, the dominant term inside the hyperbolic secant function is $C_1 r^{\frac{10}{9}} - C_2 t$ for some constants C_1 and C_2, indicating that the traveling speed of the soliton decays to 0 as $r^{-\frac{1}{9}}$.

4.2 Approximate solutions – Cnoidal waves.

As we can see in previous section, the choice of the topography is fairly restrictive in order for us to obtain explicit solutions. Even in the most special case when the bottom is flat, we are not able to obtain a solution which is physically meaningful. Therefore developing an approach to find approximate solutions is of necessity. We look for an approximate solution to vc-KdV (4.5) which is periodic (cnoidal waves) in this section, and non-periodic (solitary waves) in next section.

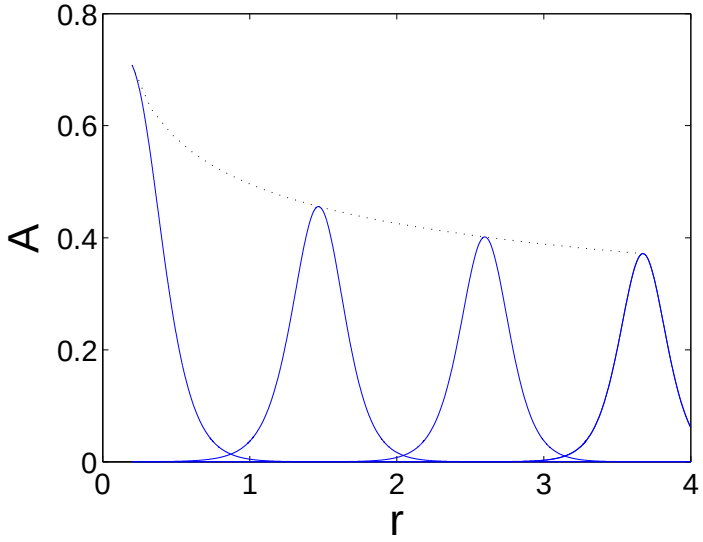


Figure 5: An exact solution $A(r,\xi)$ given by (4.16), plotted at time $t = 0, 1, 2, 3$ with parameters $c = 0.01, b = 1, \lambda_1 = 1, \lambda_2 = 0, r_0 = 0.2$ and $\epsilon = 0.01$. The dotted curve is the envelope of the moving soliton.

Assume that $f(\sigma)$ slowly varies in σ, then we can introduce a slow-time variable $T = \epsilon\sigma$ ($\epsilon << 1$) and write $f(\sigma)$ as

$$f = f(T). \tag{4.17}$$

Denote

$$\zeta = k(\xi - \epsilon^{-1}\int^T V(T)dT), \tag{4.18}$$

where k is some constant wavenumber. We seek for a periodic solution B which is periodic in ζ with a period 2π and has the asymptotic expansion

$$B = B_0(\zeta, T) + \epsilon B_1(\zeta, T) + O(\epsilon^2). \tag{4.19}$$

Plugging it into (4.5) we obtain the leading order term as

$$-VB_{0\zeta} + fB_0B_{0\zeta} + k^2B_{0\zeta\zeta\zeta} = 0. \tag{4.20}$$

Equation (4.20) is an ordinary differential equation in ζ, with T as a parameter; furthermore, it has a cnoidal wave solution

$$B_0(\zeta, T) = a[b + \mathrm{cn}^2(\gamma\zeta; m)] + d, \tag{4.21}$$

where

$$b = \frac{1-m}{m} - \frac{E(m)}{mK(m)}, \tag{4.22a}$$

$$a = \frac{12mK^2(m)k^2}{\pi^2 f}, \tag{4.22b}$$

$$\text{and } V = [d + a(\frac{2-m}{3m} - \frac{E(m)}{mK(m)})]f. \tag{4.22c}$$

Here notations are: $\mathrm{cn}(x;m)$ is the Jacobian elliptic function with parameter $m \in (0,1)$; $K(m)$ and $E(m)$ (or K and E for necessary abbreviations) are complete elliptic integrals of the first and second kind respectively. Since $\mathrm{cn}^2(x;m)$ is known to have a period of $2K(m)$, $\gamma = \frac{K(m)}{\pi}$. The constant b is so chosen that the integration of the sum in the bracket of (4.21) over one period is zero.

There are three free parameters in the cnoidal wave solution, which are chosen by us to be the amplitude a, mean level d and elliptic parameter m. They are functions of T (or r) describing how waves are modulated. To determine a, d and m, we apply the so-called Whitham averaging method ([11]), which considers the following two conservation laws directly deduced from (4.5)

$$\frac{\partial}{\partial t}\int_0^{2\pi} B d\zeta = 0, \tag{4.23}$$

$$\frac{\partial}{\partial t}\int_0^{2\pi} B^2 d\zeta = 0. \tag{4.24}$$

These conservation laws hold for each term in the expansion of B. It follows from (4.23) that $d = const$. Plugging (4.21) into (4.24) and using (4.22b) we get

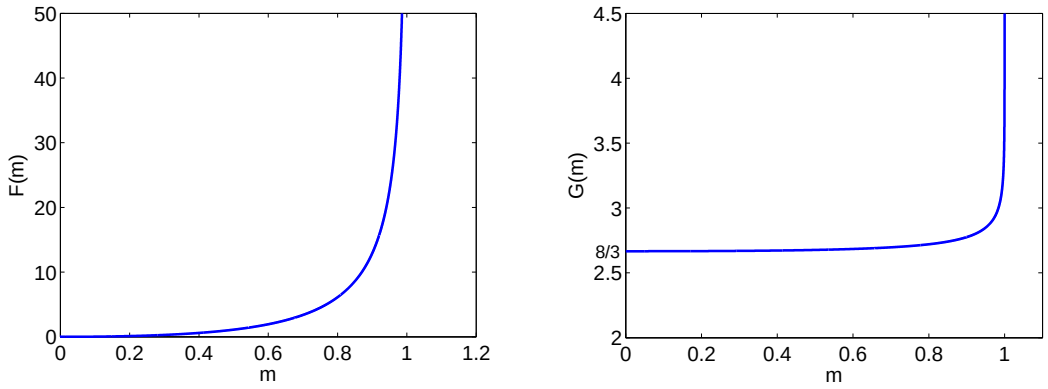
$$a^2 = \frac{m^2 K(m)^4 I_0}{F(m)} \equiv G(m) I_0, \tag{4.25}$$

$$\text{and} \quad F(m) = \frac{\pi^4 I_0}{144 k^4} f^2, \tag{4.26}$$

where

$$F(m) \equiv K(m)^2[(4-2m)E(m)K(m) - 3E(m)^2 - (1-m)K(m)^2], \tag{4.27}$$

and I_0 is a sum of a constant depending only on d and another integrating constant. $F(m)$ and $G(m)$ are plotted in figure 6. The wavenumber k and V can be determined from (4.22b) and (4.22c), respectively.



(a) (b)

Figure 6: Plots of $F(m)$ and $G(m)$ versus m.

To study the remote behavior of the amplitude of a cnoidal wave solution (4.21), us choose a representative example of bottom function

$$h(r) = (r + r_1)^{-\beta} \qquad r_1 \geq 0,\ \beta > 0. \tag{4.28}$$

Then it follows from (4.6) that

$$f(r) = 6\sqrt{6} r^{-1/2} (r + r_1)^{9\beta/4}. \tag{4.29}$$

We remark that in a more generalized case the bottom can be chosen to be $h = a_1(r + r_1)^{-\beta} + a_2$ which takes into account the slope of the basin at origin and the place where water depth vanishes.

At large r, from (4.29) we have $f \sim r^{\frac{9\beta-2}{4}}$. Thus when $\beta < \frac{2}{9}$, $f(r) \to 0$ as $r \to \infty$. From (4.26) and monotonicity of $F(m)$(figure 6a), we have $\lim_{r\to\infty} m(r) = 0$. Since $\lim_{m\to 0} G(m) = \frac{8}{3}$(figure 6b), it follows from (4.25) that the amplitude a decreases to a finite value as $r \to \infty$. Then the physical amplitude

$$\widehat{a} = h^{-5/4} r^{-1/2} a \sim r^{(5\beta-2)/4} \tag{4.30}$$

decreases to zero as well. When $\beta > \frac{2}{9}$, $\lim_{r\to\infty} f(r) = \infty$, and $\lim_{r\to\infty} m(r) = 1$, so $a \to \infty$. However the physical amplitude

$$\widehat{a} = h^{-\frac{5}{4}} r^{-\frac{1}{2}} a \sim F^{4(\beta-\frac{1}{3})/(9\beta-2)}, \quad \text{as } r \to \infty. \tag{4.31}$$

Therefore if $\frac{2}{9} < \beta < \frac{1}{3}$, the amplitude of cnoidal wave $\widehat{a}$ still decreases to 0. But if $\beta > \frac{1}{3}$, (4.31) still holds by the same argument, and $\widehat{a} \to \infty$ as $r \to \infty$. At the critical slope $\beta = \frac{1}{3}$, the wave amplitude remains finite and converges to a non-zero constant. It has the physical implication that the effects of geometrical spreading and shoaling are offset.

4.3 Approximate solutions – Solitary waves.

Heuristically, cnoidal waves will become solitary waves when the elliptic parameter equals unity. However, techniques discussed in section 4.2 can not be directly applied to find solitary wave solution by merely taking the limit m approaching 1, because the two processes $m \to 1$ and $\epsilon \to 0$ do not commute. A new concept of slowly-varying is required and we adopt the one given by Grimshaw[4]. Equation (4.17) is to be used as before but (4.18) is now replaced by

$$\delta = \xi - \epsilon^{-1} \int^T V(T) dT, \tag{4.32}$$

and (4.19) replaced by

$$B = B_0(\delta, T) + \epsilon B_1(\delta, T) + O(\epsilon^2). \tag{4.33}$$

The solution B is no longer required to be periodic in δ. Substitution of (4.33) into (4.5) gives the leading order term as

$$-V B_{0\delta} + f B_0 B_{0\delta} + B_{0\delta\delta\delta} = 0, \tag{4.34}$$

which has a soliton solution

$$B_0 = a\,\mathrm{sech}^2(q\delta), \tag{4.35}$$

where $V = \frac{af}{3} = 4q^2$, for some constant q. Substituting (4.35) into the conservation law

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} B_0^2\; d\delta = 0 \tag{4.36}$$

leads to

$$a^3 = \text{const.} f. \tag{4.37}$$

Combining (4.29) and (4.37) we can see that for large r,

$$a \sim r^{(9\beta-2)/4}.$$

The physical amplitude is then

$$\widehat{a} \sim h^{-5/4} r^{-1/2} a \sim r^{2(\beta-\frac{1}{3})}, \quad \text{as} \quad r \to \infty. \tag{4.38}$$

Therefore $\beta = \frac{1}{3}$ is again a critical value for the behavior of the amplitude.

Let us now examine when the approximate solutions of cnoidal waves and solitary waves are valid. We treat f and a, m etc. as functions of r instead of T(or σ) since the variable r represents the radial distance from the origin. Approximate solutions are only true for a slowly varying function f, which is determined by the bottom function $h(r)$ as in (4.6). Moreover, generation of circular waves (such as dropping a stone into a pond) usually involves very complicated mechanism near the origin where current theory fails. These two difficulties can be circumvented by considering the flow domain to be the whole plane excluding vicinity of the origin.

As a summary to this section, we characterize as follows the amplitude $\widehat{a}$ to vc-cy-KdV equation (4.3) in terms of β while the bottom function is in the form of $h(r) = (r + r_1)^{-\beta}$:

(i) when $0 < \beta < \frac{1}{3}$, $\widehat{a} \to \infty$ as $r \to \infty$;

(ii) when $\beta = \frac{1}{3}$, $\widehat{a} \to$ const. as $r \to \infty$;

(iii) when $\frac{1}{3} < \beta < 2$, $\widehat{a} \to 0$ as $r \to \infty$.

It is perhaps an interesting but not a surprising fact that both cnoidal waves and solitary waves share the same critical value $\beta = \frac{1}{3}$, even though the analysis leading to the result are not identical.

5 Discussion.

Equation (2.27) was originally derived by Shen&Keller[8] for a compressible fluid rotating at a constant angular velocity about the $z-$axis. In that case an adiabatic equation must be added to the governing equations and the equation of motion must be modified to include the effect of rotation. Therefore our equation is a special case of the KdV equation

obtained by them. The features of nonlinearity and dispersion of waves traveling in shallow water are captured in this variable coefficient KdV equation.

The critical value $\beta = \frac{1}{3}$ for which wave amplitude approaches to zero/finity/infinity may not be verified in practice, since the occurrence of reflected waves from the rising topography is neglected. As remarked in section 3.3, waves are certainly reflected if $\beta < 2$. This raises the possibility that returning waves interact with advancing waves and modify the wave amplitude in a non-trivial way. Mathematically, a possible remedy is to allow the sign in front of variable t in the phase function (eqn.(4.2)) to be reversed, giving rise to another vc-cy-KdV equation. The new KdV equation needs to be solved together with the existing one.

Acknowledgement

I am grateful to Joseph Keller for his advice during the whole summer. His insightful ideas always enlightened me to view and solve difficult problems from broader perspectives. My acknowledgement is also to Harvey Segur and Roger Grimshaw for their principle lectures as well as many helpful discussions afterwards. Finally I would like to extend my thanks to all GFD staff and fellows for making this summer program so exciting and memorable.

References

[1] P. Carbonaro, *Cylindrical Solitons in Shallow Water of Variable Depth*, IL Nuovo Cimento, **6**, 133-144, (1983).

[2] E. Fan, *Auto-Bäklund transformation and similarity reductions for general variable coefficient KdV equations*, Phys. Lett. A, **294**, 26-30, (2002).

[3] R. Grimshaw, *Slowly varying solitary waves. I. Korteweg-de Vries equation*, Proc. R. Soc. London, A **368**, 359-375, (1979).

[4] R. Grimshaw, *Internal solitary waves in a variable medium*, Gesellschaft Angew. Math., 30, 96-109, (2007).

[5] N. Joshi, *Painlev property of general variable-coefficient versions of the Korteweg-de Vries and non-linear Schrödinger equations*, Physics Letters A. **125**, 456-460, (1987).

[6] D. Peregrine, *Long waves on a beach*, J. Fluid Mech., 27, 815-827, (1967).

[7] B. Saha & R. Roychoudhury, *Exact solutions of cylindrical and spherical dust ion acoustic waves*, Phys. Plasma, **10**, 4162-4165, (2003).

[8] M. Shen & J. Keller, *Ray method for nonlinear wave propagation in a rotating fluid of variable depth*, Phys. Fluids **16**, 1565-1572, (1973).

[9] S. Shen & M. Shen, *On a Korteweg-de Vries equation with variable coefficients in cylindrical coordinates*, Phys. Fluids **29**, 1759-1760, (1986).

[10] S. Voit, *Tsunamis*, Ann. Rev. Fluid Mech. **19**, 217-236, (1987).

[11] G. Whitham, *Linear and Nonlinear Waves*, Wiley, New York, (1974).

[12] D. Zhang *Conservation laws and Lax pair of the variable coefficient KdV equation*, Chin. Phys. Lett., 24, No.11, 3021, (2007).

The derivation and application of convective pattern equations

Yiping Ma

Advised by Edward A Spiegel

Abstract

In this report, we review an elegant technique, known as the Bogoliubov method, for deriving amplitude equations in pattern forming systems, through detailed solution to the classical problem of Rayleigh-Bénard convection, with both free and rigid upper and lower boundaries. The computation is facilitated by the use of a newly proposed diagrammatic technique. The resulting equation is a nonlocal pattern equation, which reduces to the 1D Swift-Hohenberg equation for 2D convection. We show that the nonlocal pattern equation is variational by finding a Lyapunov functional.

Subsequently, we formulate a few properties of steady state solutions of general variational PDE, and test their utility in pattern prediction for models including the Swift-Hohenberg equation.

Finally, we describe normal form theory for Hopf bifurcation in the context of pattern equations, and point out possible extensions of the diagrammatic technique.

1 Introduction

In the principal lectures, we learned much about water waves, mostly described by the Korteweg-de Vries (KdV) equation, the nonlinear Schrödinger (NLS) equation, and their generalizations. Both the KdV equation and the NLS equation are Hamiltonian, and thus admit travelling wave solutions with particle-like behavior ("solitons" in the usual sense).

However, if we need to take dissipation (e.g. viscosity, heat diffusivity) and/or forcing (e.g. mechanical vibrations, heat flux) into account, the Hamiltonian structure will likely be destroyed. In hydrodynamics, this happens in the subject of instability [4], within which the Rayleigh-Bénard convection is a prototypical problem.

Near the onset of the Rayleigh-Bénard instability, one can derive an equation for the amplitude of the nearly marginal modes [7]. A systematic procedure (referred to as the Bogoliubov method) for deriving such amplitude equations for a class of instability problems is formulated in [5]. We find that this procedure possesses a mode interaction structure that makes it feasible for expression in terms of Feynman diagrams. Therefore, throughout the derivation we draw these diagrams wherever applicable, and in the end we sketch how this diagrammatic technique can be generalized to other situations, e.g. Hopf bifurcation.

Insofar as the Rayleigh-Bénard convection is concerned, we get a nonlocal equation for the amplitude. We show that this equation possesses a Lyapunov functional $\mathcal{F}$, and thus exhibits features of a gradient dynamical system, namely that the system always evolves towards a local minimum of $\mathcal{F}$. We discover a set of conditions that any solution at a local minimum must satisfy, as long as it is embedded in a zero free-energy background. We

also review the implications of the variational principle for behavior of steady solution near the boundary, as well as a conservation law that follows from Noether's theorem (cf. [12] §9.2.1). We point out that the latter has a natural interpretation in terms of the former.

2 Rayleigh-Bénard Convection

In this section, we formulate in detail the Rayleigh-Bénard convection problem with fixed temperature boundaries.

2.1 Boussinesq Equations

The convection problem is described by the following standard set of non-dimensionalized hydrodynamic equations (cf. [4] II.§7)

$$\begin{aligned}
\partial_t \mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} + w\partial_z \mathbf{v} &= -\nabla\Pi + \varpi(\Delta + \partial_z^2)\mathbf{v}, && (1)\\
\partial_t w + (\mathbf{v}\cdot\nabla)w + w\partial_z w &= -\partial_z\Pi + \varpi(\Delta + \partial_z^2)w + R\varpi\theta, && (2)\\
\partial_t \theta + (\mathbf{v}\cdot\nabla)\theta + w\partial_z \theta &= (\Delta + \partial_z^2)\theta + w, && (3)\\
\nabla\cdot\mathbf{v} + \partial_z w &= 0, && (4)
\end{aligned}$$

where $\nabla = \hat{x}\partial_x + \hat{y}\partial_y$ ($\hat{x}$ and $\hat{y}$ are the two horizontal unit vectors), $\Delta = \nabla^2$, $\mathbf{v}$ is the horizontal velocity, $w\hat{z}$ is the vertical velocity ($\hat{z}$ is the vertical unit vector), and θ and Π are the deviations of the temperature and pressure from their static values. The Prandtl number ϖ and the Rayleigh number R are defined as usual. Typically, we can impose either free or rigid boundary conditions, defined as

$$w = 0, \quad \theta = 0; \quad \partial_z^2 w = 0\,(\text{free}) \quad \text{or} \quad \partial_z w = 0\,(\text{rigid}) \quad \text{on} \quad z = \pm\frac{1}{2}. \tag{5}$$

Let us denote the full velocity as $\mathbf{u} = (\mathbf{v}, w)$, and the full vorticity as $\boldsymbol{\omega} = \boldsymbol{\nabla}\times\mathbf{u}$, where $\boldsymbol{\nabla} = \hat{x}\partial_x + \hat{y}\partial_y + \hat{z}\partial_z$. The full vorticity equation is then (cf. [13] §5.5)

$$\frac{D\boldsymbol{\omega}}{Dt} = (\boldsymbol{\omega}\cdot\boldsymbol{\nabla})\mathbf{u} + \varpi\boldsymbol{\nabla}^2\boldsymbol{\omega} + \boldsymbol{\nabla}\times(R\varpi\theta\hat{z}). \tag{6}$$

From (5), we get the following boundary conditions on the vertical vorticity $\zeta \equiv (\nabla\times\mathbf{v})\cdot\hat{z}$

$$\partial_z\zeta = 0\,(\text{free}) \quad \text{or} \quad \zeta = 0\,(\text{rigid}) \quad \text{on} \quad z = \pm\frac{1}{2}. \tag{7}$$

The significance of ζ is that, given w and ζ, we can uniquely determine $\mathbf{v}$ up to a gauge using the relations $\nabla\cdot\mathbf{v} = -\partial_z w$, $\nabla\times\mathbf{v} = \zeta\hat{z}$. Therefore, (ζ, w, θ) provides a complete description of the state of the fluid. Now we formulate linear theory based on this set of variables.

2.2 Linear Theory

Taking the $\hat{z}$-component of (6) and linearizing, we get an uncoupled equation for ζ

$$\partial_t\zeta = \varpi\boldsymbol{\nabla}^2\zeta.$$

In view of (7), if we take the boundaries at $z = \pm 1/2$ both to be rigid surfaces, then all the vertical vorticity modes are strongly damped. However, for two free boundaries, there is a solution with ζ constant in space and time corresponding to a rigid, undamped rotation of the whole fluid. In a finite system, such modes of motion play no role, and the only motions consisting of purely vertical vorticity are viscously damped. These will be slave modes, in the jargon of dynamical systems theory. Hence we shall not allow for their nonlinear excitation here and so make the approximation

$$\zeta = 0 \quad \text{for all} \quad t. \tag{8}$$

The case of one rigid and one free boundary is mathematically similar to two rigid boundaries. The case of nonzero ζ is studied in [17] and [21].

Now the fields w and θ completely specify the state of the fluid. To write the Boussinesq equations only in terms of w and θ, we shall keep (3) but for (1) and (2), we need to eliminate Π by $-\partial_z \nabla\cdot(1)+\Delta(2)$. The linear parts of these two equations are

$$\partial_t(\partial_z^2 + \Delta)w = \varpi(\partial_z^2 + \Delta)^2 w + R\varpi\Delta\theta, \quad \partial_t\theta = w + (\partial_z^2 + \Delta)\theta. \tag{9}$$

It is thus convenient to work in the Fourier space with

$$\Diamond(\mathbf{x}, z, t) = \int \Diamond_{\mathbf{k}}(z, t) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k},$$

where $\Diamond = w$, θ or $\mathbf{v}$. We may define $\mathbf{U}_{\mathbf{k}} \equiv (w_{\mathbf{k}}, \theta_{\mathbf{k}})^T$ and write (9) succinctly as

$$\partial_t \mathbb{M}_{\mathbf{k}} \mathbf{U}_{\mathbf{k}} = \mathbb{L}_{\mathbf{k}} \mathbf{U}_{\mathbf{k}}, \tag{10}$$

where

$$\mathbb{M}_{\mathbf{k}} = \begin{pmatrix} \partial_z^2 - \mathbf{k}^2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbb{L}_{\mathbf{k}} = \begin{pmatrix} \varpi\left(\partial_z^2 - \mathbf{k}^2\right)^2 & -\mathbf{k}^2 R\varpi \\ 1 & \partial_z^2 - \mathbf{k}^2 \end{pmatrix}, \tag{11}$$

and the boundary conditions follow from (5)

$$\mathbf{U}_{\mathbf{k}} = \mathbf{0}, \quad (1,0)\partial_z^2 \mathbf{U}_{\mathbf{k}} = 0 \,(\text{free}) \quad \text{or} \quad (1,0)\partial_z \mathbf{U}_{\mathbf{k}} = 0 \,(\text{rigid}) \quad \text{on} \quad z = \pm\frac{1}{2}. \tag{12}$$

This linear problem is separable, and we may seek solutions of the form $\Phi_{\mathbf{k}}(z)e^{st}$. For either free or rigid boundaries, there is a pair of $\Phi_{\mathbf{k}}(z)$ with l nodes for each $l = 0, 1, 2, \cdots$. One lowest vertical mode, denoted by $\phi_{\mathbf{k}}(z)$, can go unstable with growth rate denoted by $\sigma_{\mathbf{k}}$. The other lowest vertical mode, denoted by $\varphi_{\mathbf{k}}(z)$, and the higher vertical modes, denoted by $\chi^{\pm}_{\mathbf{k},l}(z)$ $(l = 1, 2, \cdots)$, are always stable with growth rate denoted by $\tau_{\mathbf{k}}$ and $\omega^{\pm}_{\mathbf{k},l}$. The functions $\phi_{\mathbf{k}}$, $\varphi_{\mathbf{k}}$ and $\chi^{\pm}_{\mathbf{k},l}$ form an orthogonal basis with the inner product

$$\langle \mathbf{U}_{\mathbf{k}}, \tilde{\mathbf{U}}_{\mathbf{k}} \rangle \equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(w^*_{\mathbf{k}}(\mathbf{k}^2 - \partial_z^2)\tilde{w}_{\mathbf{k}} + R\varpi\theta^*_{\mathbf{k}}\mathbf{k}^2\tilde{\theta}_{\mathbf{k}} \right) dz,$$

where * denotes complex conjugate.

It follows from (10) that the linear problems for $\phi_{\mathbf{k}}$, and its adjoint $\phi^{\dagger}_{\mathbf{k}}$, are

$$\mathbb{L}_{\mathbf{k}}\phi_{\mathbf{k}} = \sigma_{\mathbf{k}}\mathbb{M}_{\mathbf{k}}\phi_{\mathbf{k}}, \quad \mathbb{L}^{\dagger}_{\mathbf{k}}\phi^{\dagger}_{\mathbf{k}} = \sigma_{\mathbf{k}}\mathbb{M}^{\dagger}_{\mathbf{k}}\phi^{\dagger}_{\mathbf{k}},$$

where the adjoint operators are the transposes of the original ones, and the adjoint boundary conditions coincide with the original ones ($\sigma_{\mathbf{k}}$ can be shown to be real for our range of R).

The critical point is $(R_c, k_c) = (27\pi^4/4, \pi/\sqrt{2})$ for free boundaries, and $(R_c, k_c) = (1708, 3.116)$ for rigid boundaries (cf. [4] II.§15(b)). For (R, k) near (R_c, k_c), we have the expansion [7] for free boundaries

$$\sigma_k = \frac{k_c^2 \varpi}{\varpi + 1} \left(3(\frac{R}{R_c} - 1) - \frac{(k^2 - k_c^2)^2}{k_c^4} \right), \tag{13}$$

and for rigid boundaries

$$\sigma_k = \frac{19.65\varpi}{\varpi + 0.5117} \left((\frac{R}{R_c} - 1) - 0.3593 \frac{(k^2 - k_c^2)^2}{k_c^4} \right). \tag{14}$$

The solution at (R_c, k_c) and its adjoint are, for free boundaries

$$\phi_{k_c}(z) = \hat{\phi}_{k_c} \cos(\pi z), \quad \hat{\phi}_{k_c} = \begin{pmatrix} 1 \\ 1/(3{k_c}^2) \end{pmatrix};$$

$$\phi^\dagger_{k_c}(z) = \hat{\phi}^\dagger_{k_c} \cos(\pi z), \quad \hat{\phi}^\dagger_{k_c} = \begin{pmatrix} 1 \\ -9k_c^4 \varpi \end{pmatrix}.$$

For rigid boundaries

$$\phi_{k_c}(z) = \begin{pmatrix} W_{k_c}(z) \\ \left(k_c^2 R\right)^{-1/3} \Theta_{k_c}(z) \end{pmatrix}, \quad \phi^\dagger_{k_c}(z) = \begin{pmatrix} W_{k_c}(z) \\ -\varpi \left(k_c^2 R\right)^{2/3} \Theta_{k_c}(z) \end{pmatrix},$$

where the functions W_{k_c} and Θ_{k_c} are defined by

$$\begin{aligned} W_{k_c}(z) &= \cos q_0 z + (A_1 + iA_2) \cos i(q_1 + iq_2)z + (A_1 - iA_2) \cos i(q_1 - iq_2)z, \\ \Theta_{k_c}(z) &= \cos q_0 z + (B_1 + iB_2) \cos i(q_1 + iq_2)z + (B_1 - iB_2) \cos i(q_1 - iq_2)z, \end{aligned}$$

with $q_0 = 3.974$, $q_1 = 5.194$, $q_2 = 2.126$, $A_1 = -0.03076$, $A_2 = -0.05196$, $B_1 = 0.06038$ and $B_2 = -0.0006647$.

2.3 Nonlinear Terms

In the Fourier space, $\mathbf{v_k}$ can be expressed in terms of $w_{\mathbf{k}}$, by (4) and (8), as

$$\mathbf{v_k} = \frac{i\mathbf{k}}{\mathbf{k}^2} \partial_z w_{\mathbf{k}}. \tag{15}$$

With this, we can work out the nonlinear terms in (1-4). First, we transform (3) into the form

$$\partial_t \theta_{\mathbf{k}} = w_{\mathbf{k}} + (\partial_z^2 - \mathbf{k}^2)\theta_{\mathbf{k}} + \iint \mathcal{N}_\theta \delta(\mathbf{k} - \mathbf{p} - \mathbf{q}) d\mathbf{p} d\mathbf{q}. \tag{16}$$

To determine the nonlinear part $\mathcal{N}_\theta$, we note that $-w\partial_z\theta$ yields $-w_{\mathbf{q}}\partial_z\theta_{\mathbf{p}}$, and $-(\mathbf{v}\cdot\nabla)\theta$ yields

$$-(\mathbf{v}\cdot\nabla)\theta = -\int (i\mathbf{v}\cdot\mathbf{p})\theta_{\mathbf{p}} e^{i\mathbf{p}\cdot\mathbf{x}} d\mathbf{p},$$

where, by (15)

$$-(i\mathbf{v}\cdot\mathbf{p}) = \int \frac{\mathbf{p}\cdot\mathbf{q}}{\mathbf{q}^2}\partial_z w_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{x}} d\mathbf{q}.$$

We then obtain

$$\mathcal{N}_\theta = \frac{\mathbf{p}\cdot\mathbf{q}}{\mathbf{q}^2}\theta_{\mathbf{p}}\partial_z w_{\mathbf{q}} - w_{\mathbf{q}}\partial_z\theta_{\mathbf{p}}. \tag{17}$$

As for (1) and (2), let us recall that we need to first eliminate Π by $-\partial_z\nabla\cdot(1)+\Delta(2)$. The resulting equation has four nonlinear terms on the right

$$(A)\partial_z\nabla\cdot((\mathbf{v}\cdot\nabla)\mathbf{v}); (B)\partial_z\nabla\cdot(w\partial_z\mathbf{v}); (C)-\Delta((\mathbf{v}\cdot\nabla)w); (D)-\Delta(w\partial_z w). \tag{18}$$

The incarnation of this equation in the Fourier space is

$$\partial_t(\partial_z^2-\mathbf{k}^2)w_{\mathbf{k}} = \varpi(\partial_z^2-\mathbf{k}^2)^2 w_{\mathbf{k}} - \mathbf{k}^2 R\varpi\theta_{\mathbf{k}} + \iint \mathcal{N}_w\delta(\mathbf{k}-\mathbf{p}-\mathbf{q})d\mathbf{p}d\mathbf{q}. \tag{19}$$

The contributions of (18) to $\mathcal{N}_w$ can be worked out by switching to index notation, doing the Fourier transform, and finally returning to vector notation. The results are

$$(A)(\frac{(\mathbf{p}\cdot\mathbf{q})^2}{\mathbf{p}^2\mathbf{q}^2}+\frac{\mathbf{p}\cdot\mathbf{q}}{\mathbf{p}^2})\partial_z(\partial_z w_{\mathbf{p}}\partial_z w_{\mathbf{q}}); \qquad (B)-(\frac{\mathbf{p}\cdot\mathbf{q}}{\mathbf{q}^2}+1)\partial_z(w_{\mathbf{p}}\partial_z^2 w_{\mathbf{q}});$$
$$(C)-(\mathbf{p}\cdot\mathbf{q}+2\frac{(\mathbf{p}\cdot\mathbf{q})^2}{\mathbf{p}^2}+\frac{\mathbf{q}^2}{\mathbf{p}^2}\mathbf{p}\cdot\mathbf{q})w_{\mathbf{q}}\partial_z w_{\mathbf{p}}; \qquad (D)(\mathbf{p}^2+2\mathbf{p}\cdot\mathbf{q}+\mathbf{q}^2)w_{\mathbf{p}}\partial_z w_{\mathbf{q}}.$$

To simplify the expression, we can exchange $\mathbf{p}$ and $\mathbf{q}$ in any term. The overall sum is

$$\begin{aligned}\mathcal{N}_w &= (2\frac{(\mathbf{p}\cdot\mathbf{q})^2}{\mathbf{p}^2\mathbf{q}^2}+\frac{\mathbf{p}\cdot\mathbf{q}}{\mathbf{p}^2}-1)\partial_z w_{\mathbf{p}}\partial_z^2 w_{\mathbf{q}} - (\frac{\mathbf{p}\cdot\mathbf{q}}{\mathbf{q}^2}+1)w_{\mathbf{p}}\partial_z^3 w_{\mathbf{q}} \\ &\quad -(2\frac{(\mathbf{p}\cdot\mathbf{q})^2}{\mathbf{q}^2}+(\frac{\mathbf{p}^2}{\mathbf{q}^2}-1)\mathbf{p}\cdot\mathbf{q}-\mathbf{p}^2-\mathbf{q}^2)w_{\mathbf{p}}\partial_z w_{\mathbf{q}}.\end{aligned} \tag{20}$$

In terms of $\mathbf{U}_{\mathbf{k}} \equiv (w_{\mathbf{k}}, \theta_{\mathbf{k}})^T$, we may write (16) and (19) succinctly as

$$\partial_t\mathbb{M}_{\mathbf{k}}\mathbf{U}_{\mathbf{k}} = \mathbb{L}_{\mathbf{k}}\mathbf{U}_{\mathbf{k}} + \iint \mathcal{N}\delta(\mathbf{k}-\mathbf{p}-\mathbf{q})d\mathbf{p}d\mathbf{q}, \tag{21}$$

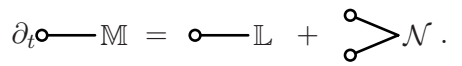
where $\mathcal{N} \equiv (\mathcal{N}_w, \mathcal{N}_\theta)^T$, and the boundary conditions are given in (12).

3 The Expansion Procedure

In this section, we follow [5] §6 to employ the Bogoliubov approach to derive the pattern equation for the convection problem. Note that the results differ from [5] in some details.

3.1 The Diagrammatic Notation

We can express (21) diagrammatically, with the replacement $\mathbf{U} \to \circ$, as



This type of diagram, regarded for now simply as a device to keep track of terms in perturbation series, bears some resemblance to Feynman diagrams in quantum field theory. The lines carry momenta, or wave vectors, although these are typically not labelled. The vertices conserve momenta, as reflected by the δ-function of wave vectors. Each diagram corresponds to a unique term in the perturbation series. At any given order, one first draws all possible (Feynman) diagrams consistent with certain physical laws, and then translates these diagrams back to algebraic expressions with a prescribed set of Feynman rules. As Feynman diagrams are frequently used in particle physics [16], condensed matter physics [15] as well as the theories of turbulence in fluids [25] and wave propagation in random media [10], we hope that their use here will simplify the calculation of modal interactions in convection.

The diagrams are largely self-explanatory except for a few conventions. All the symbols in the diagrams are vertices except for the special operator ∂_t. For any vertex, the lines immediately to its left (right) are called incoming (outgoing) lines, and the number of incoming (outgoing) lines is called the indegree (outdegree) of that vertex. We observe that all the vertices have outdegree one, but the indegree can be one (e.g. $\mathbb{M}$, $\mathbb{L}$) or more (e.g. $\mathcal{N}$). If only such vertices are present, the diagram necessarily takes the form of a tree, consisting of the *root* connected to the *leaves* via the *stems*. These diagrams differ from those in quantum field theory, where particles can be either created or annihilated, thereby making loops possible. Note that the leftmost incoming line(s) entering the leaves and the rightmost outgoing line exiting the root are not shown explicitly. The total number of leaves is the order of the diagram, consistent with perturbation theory practice.

How to translate a diagram back to an algebraic expression? The case of a first order diagram is obvious. For an n-th order diagram ($n \geq 2$), we first label the wave vectors entering the leaves as, say, $\mathbf{p}_1, \cdots, \mathbf{p}_n$, and the wave vector exiting the root as, say, $\mathbf{k}$. Then we label the wave vectors on the stems in terms of $\mathbf{p}_1, \cdots, \mathbf{p}_n$, using conservation of wave wave vectors at the vertices. Any vertex corresponds to a function the wave vectors on its incoming lines and the sub-diagrams connected to it through these lines ($\mathcal{N}$ is a somehow complicated example). With these guidelines, for any tree diagram, we can start from its root, traverse its stems and finally reach its leaves. The expression that we get, multiplied by $\delta(\mathbf{k} - \mathbf{p}_1 - \cdots - \mathbf{p}_n)$ and integrated over $\mathbf{p}_1, \cdots, \mathbf{p}_n$, is the final result. We remark that the wave wave vectors carried by the intermediate stems are automatically integrated out, so only the δ-function describing the overall wave vector conservation remains.

Now things are ready for deriving the amplitude equation. For this purpose, we need to eliminate the fast modes and keep the slow modes, much like the process of center manifold reduction in bifurcation theory for ODE. However, in generalizing the procedure to PDE, we encounter an inherent difficulty. Before dealing with the convection problem, we explain the reduction procedure with an illustrative example, where this difficulty clearly figures.

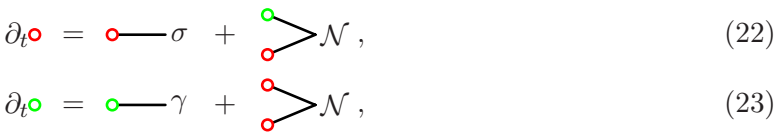
3.2 Irremovable Resonances

To illustrate the procedure we consider the following equations [24] whose structures are reminiscent of (21)

$$\partial_t f_{\mathbf{k}} = \sigma_{\mathbf{k}} f_{\mathbf{k}} + \iint f_{\mathbf{p}} g_{\mathbf{q}} \delta(\mathbf{k}-\mathbf{p}-\mathbf{q}) d\mathbf{p} d\mathbf{q},$$

$$\partial_t g_{\mathbf{k}} = \gamma_{\mathbf{k}} g_{\mathbf{k}} + \iint f_{\mathbf{p}} f_{\mathbf{q}} \delta(\mathbf{k}-\mathbf{p}-\mathbf{q}) d\mathbf{p} d\mathbf{q},$$

which, with the replacements $f \to \circ$, $g \to \circ$, can be represented diagrammatically as



(22)

(23)

where $\sigma_{\mathbf{k}} = \sigma - (k^2 + k_1^2)^2$ and $\gamma_{\mathbf{k}} = \gamma - (k^2 + k_2^2)^2$ with $\gamma < 0$ and $|\sigma| \ll |\gamma|$. The vertex $\mathcal{N}$ here simply takes the product of the two sub-diagrams connected to it.

We seek a near-identity coordinate change (here $F_{\mathbf{k}}$ is the coordinate on the "center manifold")

$$f_{\mathbf{k}} = F_{\mathbf{k}} + \iint \mathcal{I}_{\mathbf{k}}(\mathbf{p},\mathbf{q}) F_{\mathbf{p}} F_{\mathbf{q}} d\mathbf{p} d\mathbf{q} + \cdots,$$

$$g_{\mathbf{k}} = G_{\mathbf{k}} + \iint \mathcal{J}_{\mathbf{k}}(\mathbf{p},\mathbf{q}) F_{\mathbf{p}} F_{\mathbf{q}} d\mathbf{p} d\mathbf{q} + \cdots,$$

represented diagrammatically, with $F \to \bullet$, $G \to \bullet$, as

(24)

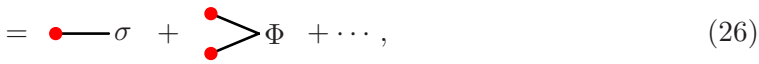
(25)

which turns (22-23) into the standard form

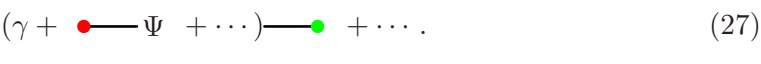
$$\partial_t F_{\mathbf{k}} = \sigma_{\mathbf{k}} F_{\mathbf{k}} + \iint \Phi_{\mathbf{k}}(\mathbf{p},\mathbf{q}) F_{\mathbf{p}} F_{\mathbf{q}} d\mathbf{p} d\mathbf{q} + \cdots,$$

$$\partial_t G_{\mathbf{k}} = (\gamma_{\mathbf{k}} + \int \Psi_{\mathbf{k}}(\mathbf{p}) F_{\mathbf{p}} d\mathbf{p} + \cdots) G_{\mathbf{k}} + \cdots,$$

represented diagrammatically as

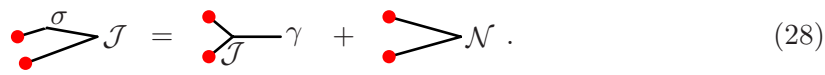


(26)



(27)

In order that (24-25) produce (27) from (23) (or equivalently, a functional analogue of a center manifold exists in a loose sense), we must satisfy, at second order



(28)

Two points should be noted in the interpretation of the first diagram. First, the factor of 2, known as the symmetry factor in quantum field theory, refers to the two distinct

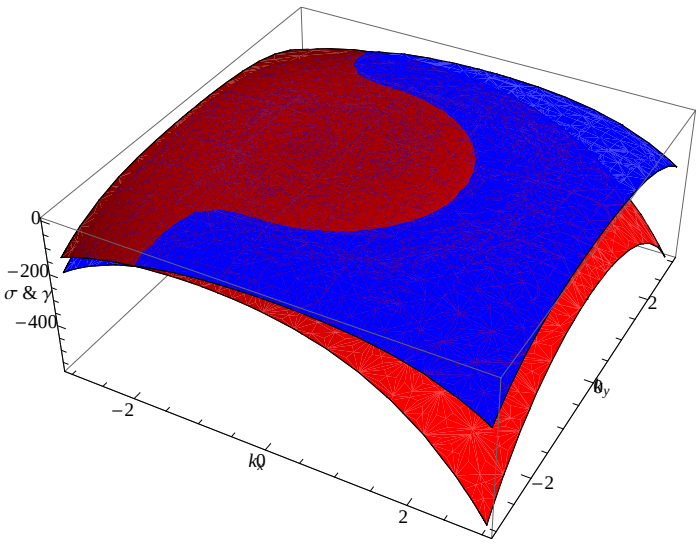


Figure 1: The curve of resonances. We plot $\sigma + \sigma_{\mathbf{k}+k_1\hat{x}}$ in red and $\gamma_{\mathbf{k}}$ in blue, with their intersection implying resonances. Parameters: $k_1 = 1$, $k_2 = 2$, $\sigma = 0.1$, $\gamma = -5$.

ways to permutate the incoming edges of $\mathcal{J}$. Graphically these two permutations can be combined, but algebraically they represent two different terms involving the two arguments of $\mathcal{J}$. Second, the vertex σ has indegree and outdegree both one, so it simply propagates the wave vector through itself. At any such vertex, the line is deliberately bent to signify the presence of the vertex. In contrast, for a vertex with indegree two (or more) and outdegree one (e.g. $\mathcal{J}$), its presence is already apparent from the topology of the diagram.

It follows that (28) translates to the condition

$$\iint \left(\mathcal{D}_{\mathbf{k}}(\mathbf{p},\mathbf{q})\mathcal{J}_{\mathbf{k}}(\mathbf{p},\mathbf{q}) - \delta(\mathbf{p}+\mathbf{q}-\mathbf{k})\right) F_{\mathbf{p}}F_{\mathbf{q}}d\mathbf{p}d\mathbf{q} = 0, \tag{29}$$

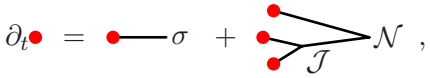
where $\mathcal{D}_{\mathbf{k}}(\mathbf{p},\mathbf{q}) = \sigma_{\mathbf{p}} + \sigma_{\mathbf{q}} - \gamma_{\mathbf{k}}$. There are resonances when

$$\mathbf{p}+\mathbf{q}-\mathbf{k} = 0, \quad \mathcal{D}_{\mathbf{k}}(\mathbf{p},\mathbf{q}) = 0.$$

These resonances cannot occur if both $|\mathbf{p}|$ and $|\mathbf{q}|$ are near k_1, but they do occur when we take, say, only $\mathbf{p} = k_1\hat{x}$. We can find the curve of resonances (Fig. 1) with the graphical procedure (cf. [2], [27]) mentioned in Lecture 12 – Triad Resonances.

It is an open question whether these resonances are artifacts of the procedure, although there is evidence that they indeed are [24].

To get the pattern equation, we introduce (25) into (22). On the center manifold, by definition, $\bullet = 0$. In addition, we have the freedom to identify $\circ$ with $\bullet$. Therefore, the pattern equation truncated to third order is

∂_t  , (30)

which translates to

$$\partial_t F_{\mathbf{k}} = \sigma_{\mathbf{k}}F_{\mathbf{k}} + \iiint \mathcal{J}(\mathbf{q},\mathbf{r})\delta(\mathbf{k}-\mathbf{p}-\mathbf{q}-\mathbf{r})F_{\mathbf{p}}F_{\mathbf{q}}F_{\mathbf{r}}d\mathbf{p}d\mathbf{q}d\mathbf{r},$$

where $\mathcal{J}$ is determined from (29) by the method of steepest descent [24].

3.3 Nonlinear Expansions

Now we attempt to apply center manifold reduction to the Boussinesq convection problem. A crucial observation is that once we admit that an effective (or functional) center manifold exists, all physical quantities depend only on the amplitudes of the slow modes. This is the spirit of the Bogoliubov method that we use.

Since the growth rate $\sigma_{\mathbf{k}}$ depends continuously on $\mathbf{k}$, one needs to introduce a cutoff in $\mathbf{k}$ to define which modes are slow. We will choose the cutoff for computational convenience, but point out that no choice is truly justified due to the lack of spectral gap.

3.3.1 Setup

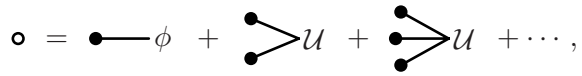
The key idea is the expressibility of $\mathbf{U}_{\mathbf{k}}$ as the functional power series

$$\mathbf{U}_{\mathbf{k}} = A_{\mathbf{k}}\phi_{\mathbf{k}} + \iint d\mathbf{p}d\mathbf{q}A_{\mathbf{p}}A_{\mathbf{q}}\mathcal{U}^{(2)}_{\mathbf{k}}(\mathbf{p},\mathbf{q};z) + \cdots , \tag{31}$$

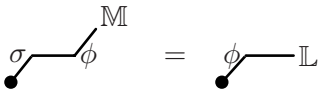
where $A_{\mathbf{k}}$ satisfies

$$\partial_t A_{\mathbf{k}} = \sigma_k A_{\mathbf{k}} + \iint d\mathbf{p}d\mathbf{q}A_{\mathbf{p}}A_{\mathbf{q}}\Gamma^{(2)}_{\mathbf{k}}(\mathbf{p},\mathbf{q}) + \cdots . \tag{32}$$

We can express (31) and (32) diagrammatically with $A \to \bullet$ as

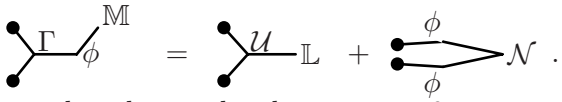


We substitute into (21) and gather terms of the same order in $A_{\mathbf{k}}$. The linear condition



is satisfied for a suitable choice of $\phi_{\mathbf{k}}$. At each higher order, the kernel must cancel out because $A_{\mathbf{k}}$ is arbitrary. Let $\mathcal{N}^{(m)}$ denote the coefficient of the δ-function in the kernel of $\mathcal{N}$ at the m-th order. To approximate, we take $A_{\mathbf{k}}$ to have *infinitesimal support* around the critical circle $|\mathbf{k}| = k_c$. Therefore, we can set the corresponding growth rate $\sigma_{\mathbf{k}} = 0$.

3.3.2 Second order



Setting the kernel of the gathered second order terms to 0, we get

$$-\mathbb{L}_{\mathbf{k}}\mathcal{U}^{(2)}_{\mathbf{k}} = \mathcal{N}^{(2)}(\mathbf{p},\mathbf{q})\delta(\mathbf{k}-\mathbf{p}-\mathbf{q}) - \Gamma^{(2)}_{\mathbf{k}}(\mathbf{p},\mathbf{q})\mathbb{M}_{\mathbf{k}}\phi_{\mathbf{k}}.$$

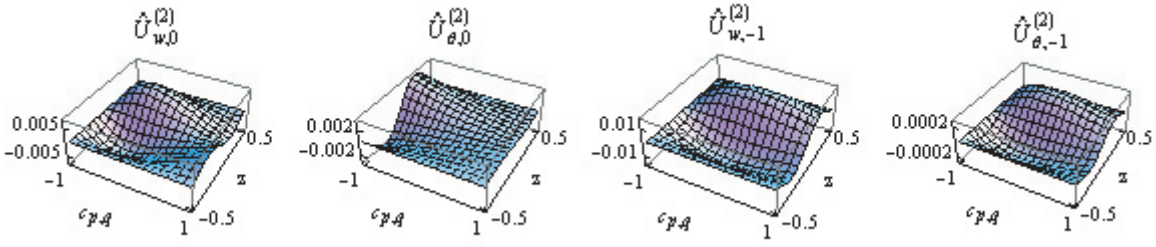


Figure 2: Plot of $\hat{\mathcal{U}}^{(2)}_{w,0}$, $\hat{\mathcal{U}}^{(2)}_{\theta,0}$, $\hat{\mathcal{U}}^{(2)}_{w,-1}$ and $\hat{\mathcal{U}}^{(2)}_{\theta,-1}$ as functions of z and $c_{\mathbf{p},\mathbf{q}}$.

The solvability condition ($\mathcal{N}^{(2)}$ being odd does not contain the lowest vertical mode)

$$0 = -\Gamma^{(2)}_{\mathbf{k}}(\mathbf{p},\mathbf{q})\langle(\phi^{\dagger}_{\mathbf{k}})^T|\mathbb{M}_{\mathbf{k}}\phi_{\mathbf{k}}\rangle$$

determines that $\Gamma^{(2)}_{\mathbf{k}}(\mathbf{p},\mathbf{q}) = 0$. Here $\langle\cdot|\cdot\rangle$ denotes the usual inner product. We are left with an always solvable equation

$$-\mathbb{L}_{\mathbf{k}}\mathcal{U}^{(2)}_{\mathbf{k}} = \mathcal{N}^{(2)}(\mathbf{p},\mathbf{q})\delta(\mathbf{k}-\mathbf{p}-\mathbf{q}).$$

Here, $\mathbf{p}$ and $\mathbf{q}$ are constrained to lie on the critical circle but $\mathbf{k}$ is not. If we introduce the shorthand $c_{\mathbf{p},\mathbf{q}} \equiv \mathbf{p}\cdot\mathbf{q}/\sqrt{\mathbf{p}^2\mathbf{q}^2}$, then

$$\mathcal{N}_{\theta} = c_{\mathbf{p},\mathbf{q}}\theta_{\mathbf{p}}\partial_z w_{\mathbf{q}} - w_{\mathbf{q}}\partial_z\theta_{\mathbf{p}},$$
$$\mathcal{N}_{w} = (2c^2_{\mathbf{p},\mathbf{q}} + c_{\mathbf{p},\mathbf{q}} - 1)\partial_z w_{\mathbf{p}}\partial^2_z w_{\mathbf{q}} - (c_{\mathbf{p},\mathbf{q}}+1)w_{\mathbf{p}}\partial^3_z w_{\mathbf{q}} - 2k_c^2(c^2_{\mathbf{p},\mathbf{q}} - 1)w_{\mathbf{p}}\partial_z w_{\mathbf{q}}.$$

The matrix $\mathbb{L}_{\mathbf{k}} = \mathbb{L}_{\mathbf{p}+\mathbf{q}}$ also depends on $c_{\mathbf{p},\mathbf{q}}$. These lead to

$$\mathcal{U}^{(2)}_{\mathbf{k}}(\mathbf{p},\mathbf{q};z) = \hat{\mathcal{U}}^{(2)}(\varpi,z,c_{\mathbf{p},\mathbf{q}})\delta(\mathbf{k}-\mathbf{p}-\mathbf{q}).$$

For free boundaries, we have the relatively simple expression

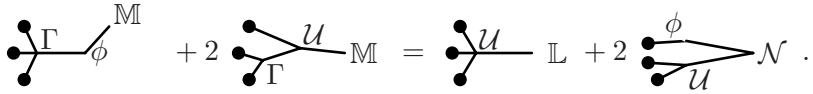
$$\hat{\mathcal{U}}^{(2)}(\varpi,z,c_{\mathbf{p},\mathbf{q}}) = \frac{1-c_{\mathbf{p},\mathbf{q}}}{(5+c_{\mathbf{p},\mathbf{q}})^3 - \frac{27}{4}(1+c_{\mathbf{p},\mathbf{q}})}\begin{pmatrix} \frac{3(1+c_{\mathbf{p},\mathbf{q}})(3+2(5+c_{\mathbf{p},\mathbf{q}})\varpi^{-1})}{4\pi} \\ \frac{2(5+c_{\mathbf{p},\mathbf{q}})^2+9(1+c_{\mathbf{p},\mathbf{q}})\varpi^{-1}}{6\pi^3} \end{pmatrix}\sin(2\pi z).$$

For rigid boundaries, it turns out easier to evaluate $\hat{\mathcal{U}}^{(2)}$ numerically rather than analytically. The result can be written as

$$\hat{\mathcal{U}}^{(2)}(\varpi,z,c_{\mathbf{p},\mathbf{q}}) = \begin{pmatrix} \hat{\mathcal{U}}^{(2)}_{w,0}(z,c_{\mathbf{p},\mathbf{q}}) \\ \hat{\mathcal{U}}^{(2)}_{\theta,0}(z,c_{\mathbf{p},\mathbf{q}}) \end{pmatrix} + \begin{pmatrix} \hat{\mathcal{U}}^{(2)}_{w,-1}(z,c_{\mathbf{p},\mathbf{q}}) \\ \hat{\mathcal{U}}^{(2)}_{\theta,-1}(z,c_{\mathbf{p},\mathbf{q}}) \end{pmatrix}\varpi^{-1},$$

where the functions $\hat{\mathcal{U}}^{(2)}_{w,0}$, $\hat{\mathcal{U}}^{(2)}_{\theta,0}$, $\hat{\mathcal{U}}^{(2)}_{w,-1}$ and $\hat{\mathcal{U}}^{(2)}_{\theta,-1}$ are plotted in Fig. 2.

3.3.3 Third order



Setting the kernel of the gathered third order terms to 0, we get

$$-\mathbb{L}_{\mathbf{k}}\mathcal{U}_{\mathbf{k}}^{(3)} = \mathcal{N}^{(3)}(\mathbf{p},\mathbf{q},\mathbf{r})\delta(\mathbf{k}-\mathbf{p}-\mathbf{q}-\mathbf{r}) - \Gamma_{\mathbf{k}}^{(3)}(\mathbf{p},\mathbf{q},\mathbf{r})\mathbb{M}_{\mathbf{k}}\phi_{\mathbf{k}}.$$

Because $\mathbf{k}$, $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ are all restricted to the critical circle, we have only three possibilities

$$\text{(I)}\,\mathbf{k}=\mathbf{p},\,\mathbf{q}=-\mathbf{r};\quad \text{(II)}\,\mathbf{k}=\mathbf{q},\,\mathbf{r}=-\mathbf{p};\quad \text{(III)}\,\mathbf{k}=\mathbf{r},\,\mathbf{p}=-\mathbf{q}. \tag{33}$$

Now we let $\mathbf{p}$ (resp. $\mathbf{q}$ and $\mathbf{r}$) denote the wave vector(s) in the first (resp. second) order. If we suitably permutate $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$, we can transform all three cases to the geometry of case (II). Thus the contribution from case (III) is identical to case (II), but differs from case (I). We find, with some algebra, that

$$\mathcal{N}^{(3)} = 2\begin{pmatrix} \hat{\mathcal{U}}_w\left(k_c^2\left(2+c_{q,r}\right)\phi_w' + c_{q,r}{\phi_w}^{(3)}\right) + \frac{1}{2}\cdot \\ \cdot\left(\left(3k_c^2\phi_w + \left(-1+2c_{q,r}\right)\phi_w''\right)\hat{\mathcal{U}}_w{}^{(0,1,0)} - 2\phi_w'\hat{\mathcal{U}}_w{}^{(0,2,0)} - \phi_w\hat{\mathcal{U}}_w{}^{(0,3,0)}\right) \\ \frac{1}{2}\left(-2\hat{\mathcal{U}}_\theta\left(1+c_{q,r}\right)\phi_w' - 2\hat{\mathcal{U}}_w\phi_\theta' - \phi_\theta\hat{\mathcal{U}}_w{}^{(0,1,0)} - 2\phi_w\hat{\mathcal{U}}_\theta{}^{(0,1,0)}\right) \end{pmatrix} + \mathcal{N}_{(\mathrm{I})}^{(3)},$$

where subscript (I) denotes case (I). The following solvability condition must be satisfied

$$\langle(\phi_{\mathbf{k}}^\dagger)^T|\mathcal{N}^{(3)}(\mathbf{p},\mathbf{q},\mathbf{r})\delta(\mathbf{k}-\mathbf{p}-\mathbf{q}-\mathbf{r}) - \Gamma_{\mathbf{k}}^{(3)}(\mathbf{p},\mathbf{q},\mathbf{r})\mathbb{M}_{\mathbf{k}}\phi_{\mathbf{k}}\rangle = 0.$$

For either free or rigid boundaries, we have the general expression

$$\Gamma_{\mathbf{k}}^{(3)}(\mathbf{p},\mathbf{q},\mathbf{r}) = \hat{\Gamma}^{(3)}(\varpi, c_{\mathbf{q},\mathbf{r}})\delta(\mathbf{k}-\mathbf{p}-\mathbf{q}-\mathbf{r}). \tag{34}$$

In the free case, we have the relatively simple expression

$$\hat{\Gamma}^{(3)}(\varpi, c_{\mathbf{q},\mathbf{r}}) = 2\hat{\Gamma}_{(\mathrm{II})}^{(3)} + \hat{\Gamma}_{(\mathrm{I})}^{(3)} \tag{35}$$

where

$$\hat{\Gamma}_{(\mathrm{II})}^{(3)} = -\frac{\varpi}{4(1+\varpi)}\frac{(1-c_{\mathbf{q},\mathbf{r}})^2((5+c_{\mathbf{q},\mathbf{r}})^2 + 9(1+c_{\mathbf{q},\mathbf{r}})\varpi^{-1} + 3(1+c_{\mathbf{q},\mathbf{r}})(5+c_{\mathbf{q},\mathbf{r}})\varpi^{-2})}{(5+c_{\mathbf{q},\mathbf{r}})^3 - \frac{27}{4}(1+c_{\mathbf{q},\mathbf{r}})}$$

and

$$\hat{\Gamma}_{(\mathrm{I})}^{(3)} = -\frac{\varpi}{4(1+\varpi)}. \tag{36}$$

The expression for $\hat{\Gamma}_{(\mathrm{II})}^{(3)}$ agrees with [7] (which is based on the calculations done in [20]) up to a constant factor due to normalization convention, and the expression for $\hat{\Gamma}_{(\mathrm{I})}^{(3)}$ agrees with $\hat{\Gamma}_{(\mathrm{II})}^{(3)}$ when $c_{\mathbf{q},\mathbf{r}} = -1$. In the rigid case, these expressions are again found to agree with [7] up to normalization, though they are not explicitly shown here.

3.4 The Evolution Equation

The evolution equation truncated to leading order is therefore

$$\partial_t A_{\mathbf{k}} = (\sigma_0 + \sigma_2(k^2 - k_c^2)^2)A_{\mathbf{k}} + \iiint_{\text{(II)}} \Gamma^{(3)}_{\mathbf{k}}(\mathbf{p},\mathbf{q},\mathbf{r})A_{\mathbf{p}}A_{\mathbf{q}}A_{\mathbf{r}}d\mathbf{p}d\mathbf{q}d\mathbf{r},$$

where (13)-(14) have been used, the kernel is given in (34), and the integration is done in case (II) of (33). If the deviation from the marginal Rayleigh number scales as $\sigma_0 \sim \epsilon^2$, consistent scalings near marginality are $\partial_t \sim \epsilon^2$, $k^2 - k_c^2 \sim \epsilon$, $A \sim \epsilon$. In the physical space, with additional re-scaling of t and x, we finally have the nonlocal pattern equation

$$\partial_t u = ru - (\Delta + k_0^2)^2 u + \iiint_{\text{(II)}} \hat{\Gamma}^{(3)}(\varpi, c_{\mathbf{q},\mathbf{r}})e^{i(\mathbf{p}+\mathbf{q}+\mathbf{r})\cdot\mathbf{x}}u_{\mathbf{p}}u_{\mathbf{q}}u_{\mathbf{r}}d\mathbf{p}d\mathbf{q}d\mathbf{r}, \tag{37}$$

where all the coefficients are $O(1)$.

For 2D convection, $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ are constrained to be collinear, so $\hat{\Gamma}^{(3)}$ reduces to a constant, and (37) reduces to the 1D Swift-Hohenberg equation [23]

$$\partial_t u = ru - (\Delta + k_0^2)^2 u + f_3 u^3, \tag{38}$$

where f_3 is a constant depending on ϖ. The usual Swift-Hohenberg model for the standard Rayleigh-Bénard convection problem has had a remarkable qualitative success in reproducing the gamut of observed patterns in the Boussinesq context. However, to make quantitative comparison between theory and experiment a nonlocal pattern equations such as the one found here and in earlier discussions (e.g. [19]) may be called for. At any rate, our analysis indicates that the constant part of the kernel $\hat{\Gamma}^{(3)}_{\text{(I)}}$ does yield a local cubic term in the evolution equation, but the nonconstant part leads to nonlocal interactions.

4 The Variational Structure and Its Applications

An evolution equation $\partial_t u = M[u]$ is called variational if it can be written

$$\partial_t u = -\delta\mathcal{F}/\delta u,$$

where $\delta\mathcal{F}/\delta u$ is the functional (or Fréchet) derivative. It may then be shown that $d\mathcal{F}/dt \leq 0$. If, in addition, $\mathcal{F}[u]$ is bounded from below, it is called a Lyapunov functional and functions u that minimize it are stable. It is known that (38) has a Lyapunov functional (cf. §4.2.1) and so does (37), under suitable restrictions, as we next see.

4.1 Variational Structure of the Nonlocal Pattern Equation

To find the Lyapunov functional for (37), we seek

$$\mathcal{G}[u] = \iiiint_{\text{(II)}} \Lambda(\varpi, c_{\mathbf{q},\mathbf{r}})\delta(\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s})u_{\mathbf{p}}u_{\mathbf{q}}u_{\mathbf{r}}u_{\mathbf{s}}d\mathbf{p}d\mathbf{q}d\mathbf{r}d\mathbf{s}$$

such that

$$\frac{\delta\mathcal{G}}{\delta u} = \iiint_{\text{(II)}} \hat{\Gamma}^{(3)}(\varpi, c_{\mathbf{q},\mathbf{r}})e^{i(\mathbf{p}+\mathbf{q}+\mathbf{r})\cdot\mathbf{x}}u_{\mathbf{p}}u_{\mathbf{q}}u_{\mathbf{r}}d\mathbf{p}d\mathbf{q}d\mathbf{r},$$

where out of the three equivalent scenarios

$$(\mathrm{I})\,\mathbf{s}=-\mathbf{p},\,\mathbf{q}=-\mathbf{r};\quad (\mathrm{II})\,\mathbf{s}=-\mathbf{q},\,\mathbf{r}=-\mathbf{p};\quad (\mathrm{III})\,\mathbf{s}=-\mathbf{r},\,\mathbf{p}=-\mathbf{q},$$

we assume that the integration for $\mathcal{G}$ is done in case (II).

After some manipulations with functional derivatives and Fourier transforms, we get

$$\frac{\delta\mathcal{G}}{\delta u}=\iiint_{(\mathrm{II})}\frac{1}{\pi^2}\Lambda(\varpi,c_{\mathbf{q},\mathbf{r}})e^{i(\mathbf{p}+\mathbf{q}+\mathbf{r})\cdot\mathbf{x}}u_{\mathbf{p}}u_{\mathbf{q}}u_{\mathbf{r}}d\mathbf{p}d\mathbf{q}d\mathbf{r},$$

in light of $c_{\mathbf{q},\mathbf{r}}=c_{\mathbf{r},\mathbf{q}}=c_{\mathbf{p},\mathbf{s}}=c_{\mathbf{s},\mathbf{p}}$. Therefore we should pick

$$\Lambda(\varpi,c_{\mathbf{q},\mathbf{r}})=\pi^2\hat{\Gamma}^{(3)}(\varpi,c_{\mathbf{q},\mathbf{r}}). \tag{39}$$

The Lyapunov functional for (37) is

$$\mathcal{F}[u]=\int\left(-\frac{1}{2}ru^2+\frac{1}{2}((\nabla^2+k_0^2)u)^2\right)d\mathbf{x}-\mathcal{G}[u]. \tag{40}$$

To show that $\mathcal{F}[u]$ is bounded from below, note that $u_{-\mathbf{k}}=u_{\mathbf{k}}^*$ leads to

$$u_{\mathbf{p}}u_{\mathbf{q}}u_{\mathbf{r}}u_{\mathbf{s}}=|u_{\mathbf{p}}|^2|u_{\mathbf{q}}|^2\geq 0.$$

In addition, (35,36,39) imply that (note that $\hat{\Gamma}^{(3)}_{(\mathrm{II})}$ is always non-positive)

$$-\Lambda(\varpi,c_{\mathbf{q},\mathbf{r}})\geq\pi^2\frac{\varpi}{4(1+\varpi)}\equiv\pi^2M>0.$$

Therefore, we can establish the estimate (now the integration is done over all three cases)

$$-\mathcal{G}[u]\geq\frac{\pi^2M}{3}\iiiint\delta(\mathbf{p}+\mathbf{q}+\mathbf{r}+\mathbf{s})u_{\mathbf{p}}u_{\mathbf{q}}u_{\mathbf{r}}u_{\mathbf{s}}d\mathbf{p}d\mathbf{q}d\mathbf{r}d\mathbf{s}=\frac{M}{12}\int u^4d\mathbf{x},$$

where the identity $u_{\mathbf{s}}=\frac{1}{(2\pi)^2}\int ue^{-i\mathbf{s}\cdot\mathbf{x}}d\mathbf{x}$ is used. It then follows from (40) that

$$\mathcal{F}[u]\geq\int\left(-\frac{1}{2}ru^2+\frac{1}{2}((\nabla^2+k_0^2)u)^2+\frac{M}{12}u^4\right)d\mathbf{x}\equiv\tilde{\mathcal{F}}[u],$$

where $\tilde{\mathcal{F}}[u]$ is nothing but the Lyapunov functional for (38) with $f_3=-M/3$. It is known (cf. [11] §7.3) that $\tilde{\mathcal{F}}[u]$ is bounded from below for $f_3<0$, so our claim is proved.

We expect the nonlocal pattern equation (37) to accurately capture the physics, but PDE models like (38) enjoy much greater popularity because of their mathematical simplicity. Hence in the following, we derive a few lesser known properties of steady states in dissipative PDE with Lyapunov functionals, and provide a few examples. In terms of notation, repeated Latin indices are summed over, but Greek indices are not.

In general, consider any system in d spatial dimensions with a Lyapunov functional

$$\mathcal{F}[u]=\int\mathcal{L}(u,\partial_iu,\Delta u)d\mathbf{x}\equiv\int\mathcal{L}(q,\mathbf{p}_i,r)d\mathbf{x},\quad\mathbf{x}\in\mathbb{R}^d.$$

We will focus on $u(\mathbf{x})$ that locally minimizes $\mathcal{F}[u]$, although the equalities that appear below also apply to other stationary points of $\mathcal{F}[u]$ (i.e. saddles and maxima).

4.2 Generalized Virial Theorem

We generalize upon [9] §4 to derive an analog of virial theorem. If we introduce a parameter λ and define $\mathcal{F}(\lambda) \equiv \mathcal{F}[u(\mathbf{x} + \lambda \mathbf{y}(\mathbf{x}))]$, then stationarity and minimality require

$$\mathcal{F}'(\lambda = 0) = 0, \tag{41}$$

$$\mathcal{F}''(\lambda = 0) \geq 0. \tag{42}$$

Consider a linear function $\mathbf{y}(\mathbf{x})$, for which the Jacobian matrix $\mathbf{J}$ (defined by $\mathbf{J}_{ij} \equiv \partial \mathbf{y}_j / \partial \mathbf{x}_i$) is a constant matrix independent of $\mathbf{x}$. For any λ, we can transform the coordinate by $\mathbf{x} + \lambda \mathbf{y}(\mathbf{x}) \to \mathbf{x}$ to obtain the exact expression

$$\mathcal{F}(\lambda) = \int \frac{\mathcal{L}(u, \partial_i u + \lambda \mathbf{J}_{ij} \partial_j u, \Delta u + \lambda (\mathbf{J}_{jk} + \mathbf{J}_{kj}) \partial_{jk} u + \lambda^2 \mathbf{J}_{ij} \mathbf{J}_{ik} \partial_{jk} u)}{\det(\mathbf{I} + \lambda \mathbf{J})} d\mathbf{x}, \tag{43}$$

where $\mathbf{I}$ is the identity matrix. Here the integration domain is left unspecified, but certain conditions near the boundary must be satisfied. As a simple example from calculus, for an arbitrary function $\varphi(x)$

$$\int_0^L \varphi(\lambda x) d(\lambda x) = \int_0^{L/\lambda} \varphi(x) dx \neq \int_0^L \varphi(x) dx$$

unless $\varphi(x) \to 0$ at $x \to L$. Therefore, we may either take a finite domain and require $\mathcal{L} = 0$ near the boundary, or take the domain to be $\mathbb{R}^d$ and require $\mathcal{L} \in L^1(\mathbb{R}^d)$. In other words, what follows should be applied to steady localized states with bounded total free energy.

We see that in (43), λ and $\mathbf{J}$ always appear as the combination $\lambda \mathbf{J}$, so $\mathcal{F}^{(n)}(\lambda = 0)$ must be n-linear in $\mathbf{J}$. For $|\lambda| \ll 1$, the denominator can be expanded as

$$\det(\mathbf{I} + \lambda \mathbf{J}) = \exp(\mathrm{tr}(\log(\mathbf{I} + \lambda \mathbf{J}))) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(- \sum_{j=1}^{\infty} \frac{(-\lambda)^j}{j} \mathrm{tr}(\mathbf{J}^j) \right)^k,$$

from which follows the useful expression

$$\frac{1}{\det(\mathbf{I} + \lambda \mathbf{J})} = 1 - \lambda \mathrm{tr} \mathbf{J} + \frac{\lambda^2}{2} (\mathrm{tr}^2 \mathbf{J} + \mathrm{tr} \mathbf{J}^2) + O(\lambda^3).$$

Since $\mathcal{F}'(\lambda = 0)$ is linear in $\mathbf{J}$, we only need to impose (41) on the following two classes of transformations. The first is the class of scaling transforms $\mathbf{J}_{ij} = \delta_{i\mu} \delta_{j\mu}$, which leads to

$$\int \left(\mathcal{L}_{\mathbf{p}_\mu} \partial_\mu u + 2 \mathcal{L}_r \partial_{\mu\mu} u - \mathcal{L} \right) d\mathbf{x} = 0, \tag{44}$$

Summation over μ yields the "global" virial theorem

$$\int \left(\mathcal{L}_{\mathbf{p}} \cdot \mathbf{p} + 2 \mathcal{L}_r r - d\mathcal{L} \right) d\mathbf{x} = 0,$$

or equivalently (in fact, more generally)

$$\sum_i (i - d) \int \mathcal{L}_i d\mathbf{x} = 0, \tag{45}$$

where $\mathcal{L}_i$ is the term in $\mathcal{L}$ with i spatial derivatives in total. The second is the class of shear transforms $\mathbf{J}_{ij} = \delta_{i\mu}\delta_{j\nu}$ $(\mu \neq \nu)$, which leads to

$$\int \left(\mathcal{L}_{\mathbf{p}_\mu}\partial_\nu u + 2\mathcal{L}_r\partial_{\mu\nu}u\right) d\mathbf{x} = 0, \tag{46}$$

We may summarize (44) and (46) as an orthogonality condition (μ and ν are arbitrary)

$$\boxed{\int(\mathcal{L}_{\mathbf{p}_\mu} + 2\mathcal{L}_r\partial_\mu)\partial_\nu u d\mathbf{x} = \delta_{\mu\nu}\int \mathcal{L} d\mathbf{x}.} \tag{47}$$

Since $\mathcal{F}''(\lambda = 0)$ is a quadratic form in the matrix elements $\mathbf{J}_{ij}$, (42) is equivalent to the requirement that the $d^2 \times d^2$ matrix

$$\begin{aligned}\mathbf{H}_{\kappa\lambda\mu\nu} = \int(&\mathcal{L}\left(\delta_{\kappa\lambda}\delta_{\mu\nu} + \delta_{\kappa\nu}\delta_{\lambda\mu}\right) + 2\mathcal{L}_r(\delta_{\kappa\mu}\partial_{\lambda\nu}u - 2\delta_{\kappa\lambda}\partial_{\mu\nu}u) - 2\mathcal{L}_{\mathbf{p}_\mu}\delta_{\kappa\lambda}\partial_\nu u \\ &+4\mathcal{L}_{rr}\partial_{\kappa\lambda}u\partial_{\mu\nu}u + 4\mathcal{L}_{r\mathbf{p}_\mu}\partial_\nu u\partial_{\kappa\lambda}u + \mathcal{L}_{\mathbf{p}_\mu\mathbf{p}_\kappa}\partial_\nu u\partial_\lambda u)d\mathbf{x},\end{aligned}$$

defined such that $\mathcal{F}''(\lambda = 0) = \mathbf{J}_{kl}\mathbf{H}_{klmn}\mathbf{J}_{mn}$, is nonnegative definite. However, we cannot see any immediate application of this result, except in the special case $\mathbf{J}_{ij} = \delta_{ij}$. Then

$$\mathcal{F}''(\lambda = 0) = \int \left(d(d+1)\mathcal{L} + 2(1-2d)\mathcal{L}_r r - 2d\mathcal{L}_{\mathbf{p}}\cdot\mathbf{p} + 4\mathcal{L}_{rr}r^2 + 4\mathcal{L}_{r\mathbf{p}}\cdot\mathbf{p}r + \mathcal{L}_{\mathbf{p}_i\mathbf{p}_j}\mathbf{p}_i\mathbf{p}_j\right) d\mathbf{x} \geq 0,$$

or equivalently (in fact, more generally)

$$\mathcal{F}''(\lambda = 0) = \sum_i (i-d)(i-d-1)\int \mathcal{L}_i d\mathbf{x} \geq 0. \tag{48}$$

4.2.1 Applications

Let us introduce the shorthand $I_i \equiv \int \mathcal{L}_i d\mathbf{x}$. Sometimes, (45) and (48) together with the positivity (or negativity) of I_i can be restated as a linear programming problem. The nonexistence of solutions for I_i then implies the nonexistence of stable localized states.

To take an example, the generalized Swift-Hohenberg equation

$$\frac{\partial u}{\partial t} = ru - (\nabla^2 + k_0^2)^2 u + f(u), \tag{49}$$

where the nonlinear function f may be either quadratic-cubic or cubic-quintic in u, has a Lyapunov functional with free energy density

$$\mathcal{L}^{SH} = -\frac{1}{2}ru^2 + \frac{1}{2}((\nabla^2 + k_0^2)u)^2 - F(u), \quad F' = f.$$

For another example, the steady state Proctor equation [18]

$$0 = \mu^2\nabla^2 u + \nabla^4 u - \nabla\cdot(|\nabla u|^2\nabla u) + \beta u, \quad \beta > 0 \tag{50}$$

has a Lyapunov functional with free energy density

$$\mathcal{L}^P = \frac{1}{4}|\nabla u|^4 + \frac{1}{2}|\nabla^2 u|^2 - \frac{\mu^2}{2}|\nabla u|^2 + \frac{\beta}{2}u^2.$$

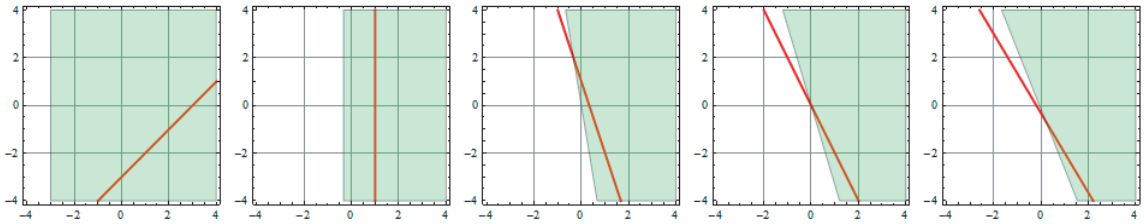


Figure 3: The R_0(horizontal)-R_2(vertical) plane. (45)/(48) defines the solid line/shaded region. Their intersection forms the segment of possible stable localized states. From left to right, d varies from 1 to 5, whereas $d \geq 6$ is qualitatively the same as $d = 5$.

The free energies for both (49) and (50) consist of I_0, I_2 and $I_4 > 0$. Hence we may introduce $R_i = I_i/I_4$ $(i = 0, 2)$, and rewrite (45) and (48) as

$$(4-d)+(2-d)R_2+(-d)R_0 = 0, \quad (4-d)(3-d)+(2-d)(1-d)R_2+(-d)(-1-d)R_0 \geq 0.$$

The segment that they define on the R_0-R_2 plane depends on d (Fig. 3). In addition

$$I_2^{SH} = k_0^2 \int \left(u\nabla^2 u\right) d\mathbf{x} = -k_0^2 \int |\nabla u|^2 d\mathbf{x} < 0, \quad I_2^P = -\frac{\mu^2}{2} \int |\nabla u|^2 d\mathbf{x} < 0.$$

As for I_0, we have

$$I_0^{SH} \geq C \int u^2 d\mathbf{x}, \quad I_0^P = \frac{\beta}{2} \int u^2 d\mathbf{x} > 0,$$

where $C > 0$ for some choices of parameters. These inequalities imply that both (49) and (50) can possibly have stable localized solutions for any d. One can make a better estimate for R_0 from the Cauchy-Schwarz inequality

$$\left| \int \left(u\nabla^2 u\right) d\mathbf{x} \right| \leq \left(\int u^2 d\mathbf{x} \right)^{1/2} \left(\int |\nabla^2 u|^2 d\mathbf{x} \right)^{1/2}.$$

This further shortens the segment on the R_0-R_2 plane, but does not change the conclusion.

It is necessary that the PDE takes the precise form $u_t = -\delta\mathcal{F}/\delta u$. As an example to the contrary, the Cahn-Hilliard equation [3]

$$\partial_t u = \Delta(u^3 - u - \gamma\Delta u), \quad \gamma > 0 \tag{51}$$

has a Lyapunov functional with free energy density

$$\mathcal{L}^{CH} = \frac{1}{4}(u^2-1)^2 + \frac{\gamma}{2}|\nabla u|^2,$$

such that $u_t = \Delta(\delta\mathcal{F}/\delta u)$ where $\mathcal{F} = \int \mathcal{L}^{CH} d\mathbf{x}$. Stable localized solutions to (51) must satisfy $\delta\mathcal{F}/\delta u = 0$, but are not necessarily local minima of $\mathcal{F}[u]$. For example, a spherical *bubble* of $u = -1$ with any finite radius, embedded in a $u = 1$ background, is stable. We could decrease $\mathcal{F}$ by shrinking the bubble radially, but this is forbidden since it violates

$$\frac{d}{dt} \int u d\mathbf{x} = 0,$$

a conservation law that follows from (51) rewritten as

$$\partial_t u = \nabla \cdot \mathbf{j} \quad \text{with} \quad \mathbf{j} = \nabla(u^3 - u - \gamma \Delta u).$$

We refer the reader to [8] for such reasoning applied to nonlinear wave equations.

4.3 Least Action Principle

Consider a finite spatial domain $\mathcal{D}$. The change in free energy under perturbation δu is

$$\delta \mathcal{F} = \mathcal{F}[u + \delta u] - \mathcal{F}[u] = \int_{\partial \mathcal{D}} (\mathcal{L}_r \nabla \delta u + \delta u (\mathcal{L}_{\mathbf{p}} - \nabla \mathcal{L}_r)) \cdot \hat{n} dA - \int_{\mathcal{D}} H \delta u d\mathbf{x}, \tag{52}$$

where $H = -\mathcal{L}_q + \partial_i \mathcal{L}_{\mathbf{p}_i} - \Delta \mathcal{L}_r$, $\partial \mathcal{D}$ is the boundary of $\mathcal{D}$, $\hat{n}$ is the unit outward normal to $\partial \mathcal{D}$ and dA is the surface element on $\partial \mathcal{D}$. If (52) is to vanish for arbitrary δu, $H = 0$ must be satisfied on $\mathcal{D}$, and in addition the following boundary conditions must be fulfilled

$$\boxed{\mathcal{L}_r = 0, \quad (\mathcal{L}_{\mathbf{p}} - \nabla \mathcal{L}_r) \cdot \hat{n} = 0 \quad \text{on} \quad \partial \mathcal{D}.}$$

4.3.1 Applications

For the Swift-Hohenberg equation, the boundary conditions are (here $\mathcal{L}_r = (\nabla^2 + k_0^2)u$)

$$\mathcal{L}_r = 0, \quad \nabla \mathcal{L}_r \cdot \hat{n} = 0 \quad \text{on} \quad \partial \mathcal{D}.$$

These conditions require that at the boundary, both u and $\nabla u \cdot \hat{n}$ must have planform wavelengths k_0. However, if the convection rolls merge perpendicularly with the boundary, then $\nabla u \cdot \hat{n} = 0$ and only the constraint on u remains. This is a plausible reason why such behavior is most often observed in convection experiments near the onset of instability.

For a more quantitative approach to the roll merging problem, we refer the reader to [26].

4.4 A Conservation Law

According to [14], for any $\mathcal{L}$ not explicitly dependent on $\mathbf{x}$, we have the conservation law

$$\partial_{x_1} \mathcal{L} - \nabla \cdot (u_{x_1} \mathcal{L}_{\mathbf{p}} + \mathcal{L}_r \nabla u_{x_1} - u_{x_1} \nabla \mathcal{L}_r) = 0 \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^d, \tag{53}$$

where ∂_{x_1} can be replaced by any other directional derivative. To interpret (53) physically, we make a perturbation δu proportional to u_{x_1}. In view of (52), the second term of (53) is the local contribution to the bilinear concomitant, or equivalently the free energy flux. On the other hand, the perturbation essentially translates u in the $-x_1$ direction, so the free energy flux can also be expressed as the first term of (53). Thus, (53) is a consequence of the special choice of perturbation by a translational mode.

4.4.1 Applications

For $d = 2$ in the Swift-Hohenberg equation, (53) is directly responsible for selecting the wavelength of hexagonal pattern that connects through a front to the trivial state [14]. For the PDE studied in [9], however, (53) simply reduces to the original PDE.

5 Normal Forms of Pattern Equations

In bifurcation theory for ODE, (24) and (25) in §3.2 represent two steps of reduction [22]. In the first step (center manifold reduction), the power series corresponding to (25) is uniquely determined but in terms of the *original* center manifold coordinate (denoted by $\boldsymbol{\alpha}$). This results in an evolution equation for $\boldsymbol{\alpha}$

$$\frac{d\boldsymbol{\alpha}}{dt} = \mathbb{M}\boldsymbol{\alpha} + \boldsymbol{\Gamma}(\boldsymbol{\alpha}), \tag{54}$$

where $\boldsymbol{\Gamma}(\boldsymbol{\alpha})$ is purely nonlinear in $\boldsymbol{\alpha}$. The matrix $\mathbb{M}$ depends only on the control parameters, with its eigenvalues having real parts all equal to 0 at criticality. In the second step (normal form theory), we choose a near-identity coordinate change corresponding to (24)

$$\boldsymbol{\alpha} = \mathbf{A} + \boldsymbol{\Psi}(\mathbf{A}),$$

where $\mathbf{A}$ is the *new* center manifold coordinate, and $\boldsymbol{\Psi}(\mathbf{A})$ is purely nonlinear in $\mathbf{A}$. The purpose is to simplify the nonlinear term $\boldsymbol{\Gamma}(\boldsymbol{\alpha})$ in (54) as much as possible. In the end we obtain an evolution equation for $\mathbf{A}$

$$\frac{d\mathbf{A}}{dt} = \mathbb{M}\mathbf{A} + \mathbf{g}(\mathbf{A}), \tag{55}$$

where $\mathbf{g}(\mathbf{A})$ represents the simplified nonlinear terms. It can be shown that (55), the (not necessarily unique) normal form, depends only on $\mathbb{M}$ (cf. [6]), and is therefore universal for all bifurcations with the same linear part.

In §3.2, the PDE model actually undergoes a pitchfork bifurcation at finite wavenumber. It is known that normal form theory cannot simplify the nonlinear terms for steady state bifurcations, which is why we can identify the new center manifold coordinate F with the original coordinate f. In this section, however, we derive the normal form of pattern equation for Hopf bifurcation, the simplest situation where normal form theory is necessary.

5.1 Hopf Bifurcation

Consider a physical system with an active control parameter λ, like the Rayleigh-Bénard convection, but whose critical modes are a pair of complex conjugate normal modes $(\phi_{\mathbf{k}}, \phi_{\mathbf{k}}^*)$ with growth rates $(\mu_{\mathbf{k}} + i\omega_{\mathbf{k}}, \mu_{\mathbf{k}} - i\omega_{\mathbf{k}})$, where $\mathbf{k}$ is the horizontal wave vector. This pair of critical modes loses stability, that is, $\mu_{\mathbf{k}}$ crosses through 0, at a critical wavenumber $|\mathbf{k}| = k_c$ as λ crosses through a critical value λ_c. We assume that all the other normal modes remain linearly stable when λ is near λ_c.

We expect that the dynamics is captured by the amplitudes of the pair of critical modes $(\phi_{\mathbf{k}}, \phi_{\mathbf{k}}^*)$, which are denoted by $(\alpha_{\mathbf{k}}, \alpha_{\mathbf{k}}^*)$. Hence we may utilize the Bogoliubov method to derive the following amplitude equation for λ near λ_c and $|\mathbf{k}|$ near k_c

$$\partial_t \alpha_{\mathbf{k}} = (\mu_{\mathbf{k}} + i\omega_{\mathbf{k}})\alpha_{\mathbf{k}} + \Gamma_{\mathbf{k}}[\alpha_{\mathbf{p}}, \alpha_{\mathbf{q}}^*], \tag{56}$$

where $\Gamma_{\mathbf{k}}$ is a purely nonlinear functional power series of $\alpha_{\mathbf{p}}$ and $\alpha_{\mathbf{q}}^*$, which possibly contains nonlocal interactions. The amplitude $\alpha_{\mathbf{k}}^*$ satisfies the complex conjugate of (56). In normal form theory, we try to simplify (56) by the near-identity coordinate change

$$\alpha_{\mathbf{k}} = A_{\mathbf{k}} + \Psi_{\mathbf{k}}[A_{\mathbf{p}}, A_{\mathbf{q}}^*],$$

so that $A_{\mathbf{k}}$ satisfies

$$\partial_t A_{\mathbf{k}} = (\mu_{\mathbf{k}} + i\omega_{\mathbf{k}})A_{\mathbf{k}} + g_{\mathbf{k}}[A_{\mathbf{p}}, A^*_{\mathbf{q}}].$$

The equation for $\Psi_{\mathbf{k}}$ is found by substitution into (56) to be

$$\mathcal{L}_{\mathbf{k}}\Psi_{\mathbf{k}} = T_{\mathbf{k}} - g_{\mathbf{k}} \tag{57}$$

where the linear operator is

$$\mathcal{L}_{\mathbf{k}}\Psi_{\mathbf{k}} = \int \frac{\delta\Psi_{\mathbf{k}}}{\delta A_{\mathbf{p}}}(\mu_{\mathbf{p}} + i\omega_{\mathbf{p}})A_{\mathbf{p}}d\mathbf{p} + \int \frac{\delta\Psi_{\mathbf{k}}}{\delta A^*_{\mathbf{q}}}(\mu_{\mathbf{q}} - i\omega_{\mathbf{q}})A^*_{\mathbf{q}}d\mathbf{q} - (\mu_{\mathbf{k}} + i\omega_{\mathbf{k}})\Psi_{\mathbf{k}}$$

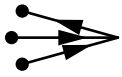
and the nonlinear term has been abbreviated by

$$T_{\mathbf{k}} = \Gamma_{\mathbf{k}}[A_{\mathbf{p}} + \Psi_{\mathbf{p}}, A^*_{\mathbf{q}} + \Psi^*_{\mathbf{q}}] - \int \frac{\delta\Psi_{\mathbf{k}}}{\delta A_{\mathbf{p}}}g_{\mathbf{p}}d\mathbf{p} - \int \frac{\delta\Psi_{\mathbf{k}}}{\delta A^*_{\mathbf{q}}}g^*_{\mathbf{q}}d\mathbf{q}. \tag{58}$$

The functional power series above are formed by monomials conveniently denoted as

$$\begin{aligned} |DK\rangle_{\mathcal{Q}} &= \left(\prod_{i=1}^{D-K}\int d\mathbf{p}_i\right)\left(\prod_{j=1}^{K}\int d\mathbf{q}_j\right)\mathcal{Q}_{DK}(\mathbf{p}_1,\cdots,\mathbf{p}_{D-K},\mathbf{q}_1,\cdots,\mathbf{q}_K) \\ &\times \delta(\sum_{i=1}^{D-K}\mathbf{p}_i + \sum_{j=1}^{K}\mathbf{q}_j - \mathbf{k})\prod_{i=1}^{D-K}A_{\mathbf{p}_i}\prod_{j=1}^{K}A^*_{\mathbf{q}_j}, \end{aligned} \tag{59}$$

where $\mathcal{Q} = \Psi$, g or T. When the kernel $\mathcal{Q}_{DK}$ is constant (denoted by $\bar{\mathcal{Q}}_{DK}$), the monomial $|DK\rangle_{\mathcal{Q}}$ reduces to the local expression $\bar{\mathcal{Q}}_{DK}(A^{D-K}(A^*)^K)_{\mathbf{k}}$. It may seem that in (59) we need to deal with two fields $A_{\mathbf{k}}$ and $A^*_{\mathbf{k}}$, but the relation $A^*_{\mathbf{k}} = A_{-\mathbf{k}}$ suggests that A^* can be transformed to A by reversal of wave vector. Therefore, we may choose to represent $|DK\rangle_{\mathcal{Q}}$ by a diagram with $D-K$ lines with right arrows and K lines with left arrows, both connected to a vertex labelled $\mathcal{Q}$. We can use these two notations interchangeably, e.g.

$|31\rangle_\Gamma \quad \Leftrightarrow$  .

Regardless of the kernel $\mathcal{Q}_{DK}$, any monomial is an eigenvector of $\mathcal{L}_{\mathbf{k}}$

$$\mathcal{L}_{\mathbf{k}}|DK\rangle_{\mathcal{Q}} = \Lambda_{DK}|DK\rangle_{\mathcal{Q}}$$

if we take $A_{\mathbf{k}}$ to have infinitesimal support around $|\mathbf{k}| = k_c$. The eigenvalue is

$$\Lambda_{DK} = i\omega_{k_c}(D - 2K - 1) + \mu_{k_c}(D-1). \tag{60}$$

Then (57) becomes

$$\Lambda_{DK}|DK\rangle_\Psi = |DK\rangle_T - |DK\rangle_g \tag{61}$$

for $D \geq 2$. From (58) we know $|2K\rangle_T$, so we can solve (61) when $D = 2$, once we choose $|2K\rangle_g$. Thereafter, at each new D, $|DK\rangle_T$ is known if we solve sequentially.

As long as $\Lambda_{DK} \neq 0$, we can always choose $g_{DK} = 0$. However, (60) suggests that when $\mu_{k_c} = 0$ (i.e. $\lambda = \lambda_c$), Λ_{DK} vanishes when $K = (D-1)/2$, or equivalently $\Lambda_{2L+1,L} = 0$ for

$L = 1, 2, \cdots$. To avoid the small denominator $\Lambda_{2L+1,L}$ near λ_c, we require that $g_{2L+1,L} = T_{2L+1,L}$. All the other g_{DK} may be set equal to 0. Therefore the normal form is

$$\partial_t A_{\mathbf{k}} = (\mu_{\mathbf{k}} + i\omega_{\mathbf{k}})A_{\mathbf{k}} + |31\rangle_T + |52\rangle_T + \cdots . \tag{62}$$

The diagrammatic notation is not needed to get (62), but it can be useful for calculating the normal form "coefficients" T_{31}, T_{52}, etc. As an example, to expand the convolution $\int(\delta\Psi_{\mathbf{k}}/\delta A_{\mathbf{p}})g_{\mathbf{p}}d\mathbf{p}$ in (58), we first cross out any of the $D-K$ instances of A that occur in $|DK\rangle_\Psi$, which results in $D-K$ diagrams with an empty slot each. Then we fill each slot with the functional power series $g_{\mathbf{p}}$, and sum up these $D-K$ diagrams in the end.

Overall, the above normal form theory for Hopf bifurcation in nonlocal pattern equation parallels the ODE case (cf. [22] §4). We will not study the properties of (62), but mention that when T_{31} is constant and $k_c = 0$, (62) truncated to third order is the complex Ginzburg-Landau equation (CGLE), whose solutions have been extensively documented [1]. The CGLE, or more generally (62), is invariant under the phase-shift $A \to Ae^{i\phi}$ where ϕ is a constant. Interestingly, in terms of diagrams, we may assign A (A^*) spin 1/2 ($-1/2$), and state this invariance as the conservation of total spin among all diagrams.

6 Conclusion

In this report, we derived a nonlocal pattern equation for Rayleigh-Bénard convection with fixed temperature boundaries, found a Lyapunov functional for this equation, and then formulated a few properties of steady state solutions of variational PDE. The diagrammatic technique proved instrumental in the derivation of convective pattern equations.

There exist unresolved issues and open questions. First, for Rayleigh-Bénard convection, it remains to find a suitable way of removing the singularities that come from resonances among fast and slow modes (cf. §3.2). Second, we need to interpret the generalized virial theorem physically, and work out more applications. It is also interesting to study how these properties of variational PDE generalize to nonlocal pattern equations. Finally, we plan to extend [22] §5 on multiple instabilities to nonlocal pattern equation. However, for higher codimension bifurcations, even if center manifold reduction works as before, it is not clear how normal form theory can be formulated. In this case, the linear term does not uniquely determine the nonlinear terms that cannot be discarded in the normal form.

Acknowledgement

First and foremost, I would like to thank Prof. Ed Spiegel for his constant guidance and encouragement, as well as jokes and anecdotes, throughout the summer. I am also grateful to Prof. Predrag Cvitanovic for helpful discussions near the end of the program.

The unique combination of faculty and fellows at Walsh offered an experience almost reminiscent of the golden high school days. To all these people I owe my sincere thanks, especially to my roommate Andong He for many hilarious moments.

References

[1] **The world of the complex Ginzburg-Landau equation**, Aranson, I. S. and Kramer, L. 2002, *Rev. Mod. Phys.*, **Vol. 74 No. 1**, 99-143.

[2] **Energy transfer between external and internal gravity waves**, Ball, F. K. 1964, *J. Fluid Mech.*, **19**, 465-478.

[3] **Free energy of a nonuniform system. I. Interfacial free energy**, Cahn, J. W. and Hilliard, J. E. 1958, *J. Chem. Phys.*, **28**, 258-267.

[4] **Hydrodynamic and hydromagnetic stability**, Chandrasekhar, S. 1961, *Oxford: Clarendon Press.*

[5] **Evolution equations for extended systems**, Coullet, P. H. and Spiegel, E. A. 1987, in *Energy stability and convection*, Pitman Research Notes 168, edited by G. P. Galdi and B. Straughan (Wiley, New York).

[6] **Introduction to bifurcation theory**, Crawford, J. D. 1991, *Rev. Mod. Phys.*, **Vol. 63 No. 4**, 991-1037.

[7] **Derivation of the amplitude equation at the Rayleigh-Bénard instability**, Cross, M. C. 1980, *Phys. Fluids*, **23**(9), 1727-1731.

[8] **Comments on nonlinear wave equations as models for elementary particles**, Derrick, G. H. 1964, *J. Math. Phys.*, **Vol. 5 No. 9**, 1252-1254.

[9] **Complexity from thermal instability**, Elphick, C., Regev, O. and Spiegel, E. A. 1991, *Mon. Not. R. Astr. Soc.*, **250**, 617-628.

[10] **Wave propagation in random media**, Frisch, U. 1968, in *Probabilistic methods in applied mathematics*, edited by A. T. Bharucha-Reid (Academic, New York).

[11] **Pattern Formation**, Hoyle, R. 2006, *Cambridge University Press.*

[12] **Classical Dynamics: A Contemporary Approach**, Jose, J. V. and Saletan, E. J. 1998, *Cambridge University Press.*

[13] **Fluid Mechanics**, Kundu, P. K. and Cohen, I. M. 2007, *Academic Press.*

[14] **Localized hexagon patterns of the planar Swift-Hohenberg equation**, Lloyd, D. J. B., Sandstede, B., Avitabile, D. and Champneys, A. R. 2008, *SIAM J. Appl. Dyn. Syst.*, **Vol. 7 No. 3**, 1049-1100.

[15] **A Guide to Feynman Diagrams in the Many-Body Problem**, Mattuck, R. D. 1992, *Dover Publications.*

[16] **An Introduction to Quantum Field Theory**, Peskin, M. E. and Schroeder, D. V. 1995, *Westview Press.*

[17] **Inertial effects in long-scale thermal convection**, Pismen, L. M. 1986, *Phys. Lett. A*, **Vol. 116 No. 5**, 241-244.

[18] **Planform selection by finite-amplitude thermal convection between poorly conducting slabs**, Proctor, M. R. E. 1981, *J. Fluid Mech.*, **113**, 469-485.

[19] **The Swift-Hohenberg equation requires nonlocal modifications to model spatial pattern evolution of physical problems**, Roberts, A. J. 1994, *Preprint*, arXiv:patt-sol/9412002.

[20] **On the stability of steady finite amplitude convection**, Schlüter, A., Lortz, D. and Busse, F. 1965, *J. Fluid Mech.*, **Vol. 23 Part 1**, 129-144.

[21] **Pattern selection in Rayleigh-Bénard convection near threshold**, Siggia, E. D. and Zippelius, A. 1981, *Phys. Rev. Lett.*, **Vol. 47 No. 12**, 835-838.

[22] **Cosmic arrhythmias**, Spiegel, E. A. 1985, in *Chaos in astrophysics; Proceedings of the Advanced Research Workshop*, Dordrecht, D. Reidel Publishing Co., 91-135.

[23] **Hydrodynamic fluctuations at the convective instability**, Swift, J. and Hohenberg, P. C. 1977, *Phys. Rev. A*, **Vol. 15 No. 1**, 319-328.

[24] **Instability in extended systems**, Tao, L. and Spiegel, E. A. 1993, in *Proceedings of the Summer Study Program in G.F.D.*, Woods Hole Oceanographic Institution, pp. 177-184.

[25] **Formulation of the theory of turbulence in an incompressible fluid**, Wyld, H. W. 1961, *Annals of Physics*, **14**, 143-165.

[26] **Optimal merging of rolls near a plane boundary**, Zaleski, S., Pomeau, Y. and Pumir, A. 1984, *Phys. Rev. A*, **Vol. 29 No. 1**, 366-370.

[27] **Electrons and Phonons**, Ziman, J. M. 1960, *Oxford*.

Variation of the Eddy Diffusivity Across Jets in the Southern Ocean

by:

Ali Mashayek

Supervisors:

Oliver Bühler & Raffaele Ferrari

August 2009

1 Introduction

Large scale ocean circulation is one of the main components of the climate system. The enormous heat capacity of the oceans makes general oceanic circulation a great means of transport of considerable amount tracers such as heat across the globe. Thus, a growing number of research activities have been focused on understanding the oceans as a part of the climate system. The main goal is to develop tools that will enable scientists to understand the past history of the climate system and more importantly, to make predictions about the future. The complexities of the components of the climate system such as the oceans and the atmosphere demand numerical models to simulate their behaviors in conditions close to reality. In particular, ocean general circulation models (GCM) have been developed with the aim of simulating the general circulation and more local characteristics of the ocean. The current ocean models used for climate studies cannot resolve the fine scale turbulence and eddy characteristics due to computational cost limitations. Thus, resorting to parameterizations for fine scale behaviors is inevitable. One of the key parameters used in the GCMs to capture the turbulent characteristics is the eddy diffusivity, K, which relates the flux of tracers to their mean gradients. This parametrizaion however requires proper knowledge of the magnitude and variations of K as a function of latitude-longitude and depth as well as its temporal variations. Great effort has been devoted to enhancing our knowledge of the diffusivity over the past years. As one particular case, estimating the tracer transport across a permanent jet is of great importance. The transport of heat across the Antarctic Circumpolar Current (ACC) for example is of great importance specially in the context of global warming and melting of the land ice sheets located in the Antarctica. Ferrari and Nikurashin (2009) [1] (hereafter referred to as FN09) studied the variations in the eddy diffusivity across the jets in the southern ocean. They compared different methods of measurement of the the diffusivity and investigated the effect of the eddy-mean flow interaction on the diffusivity across the jet. Several studies (such as those listed in FN09) have used the mixing length theory to relate the diffusivity to the r.m.s velocity of the eddies and the mixing length l. The mixing length was often set proportional to the observed eddy sizes. This method led to a peak in the values of K in the core of the ACC due to larger eddy velocities. A different approach taken by Marshal *et al.*(2006) [3] provided estimates of the diffusivity by feeding flow simulations by velocities obtained from observation and studying the diffusion of an artificially (numerically) injected passive tracer field into the domain. They suggested that the diffusivity is rather suppressed in the core of the ACC and enhanced in its flanks as opposed to the first group of studies. FN09 explained this discrepancy by pointing out that the mixing length used in the first class of studies is not necessarily of the same order of the eddy sizes. Using a dynamic model, FN09 showed that the mixing length is modulated by the mean flow and the relative velocity between the main flow and the speed of travelling of the eddies reduces the diffusivity. Thus, the large values of K found by the earlier studies was mainly due to not taking the reduction of the mixing length into account. It should also be mentioned that Marshal *et al.*(2006) speculated that the large values of K in the flank of the ACC can be due to the presence of critical layers.

Since our motivation for the present work is based on the model proposed by FN09, a brief introduction to their model is necessary before the motivation of our work can be

explained. Starting from a uniform zonal flow of constant velocity U_0, FN09 applied the surface Quasi-Geostrophic (QG) formulation to describe the dynamics and used it to derive an expression for the eddy diffusivity. The governing equations they started from are

$$b_t + J(\psi - U_0 y, b - \Gamma y) = 0, \tag{1}$$

$$\partial_x^2 \psi + \partial_y^2 \psi + \frac{f^2}{N^2} \partial_z^2 \psi = 0, \tag{2}$$

where ψ is the perturbation (from the mean) geostrophic stream function, b is the perturbation surface buoyancy, J is the Jacobian operator, Γ is the constant lateral gradient of the background buoyancy $\partial_y B = -\Gamma$, and f and N are the inertial and stratification frequencies respectively. They expanded (1) to get

$$\partial_t b + U_0 \partial_x b - \Gamma \psi_x = fN\sqrt{\gamma} r(t) e^{i(kx+ly)} - \gamma b, \tag{3}$$

where the nonlinear term $J(\psi, b)$ in (1) is expressed in the form of a fluctuation-dissipation stochastic model which is a crude representation of excitation of waves by baroclinic instability at the most unstable wave number (k, l). The model is in terms of a stochastic variable $r(t)$ and a linear damping rate γ. The variable $r(t)$ is a white-noise function with the property $< r(t) r(t')^* >= \delta(t - t')$ where '$< >$' denotes the "expected value". The stochastic forcing "mimics the nonlinear damping of each wave through interaction with other waves"[1]. The constant f sets the forcing amplitude. The forcing is kept monochromatic to keep the problem linear. FN09 showed that this model generates a velocity field with a correlation function decaying exponentially at the rate γ. The solution of the stream function was obtained from (1,2) in the form of

$$\psi = \frac{f}{\kappa}\sqrt{\gamma} \int_0^\infty r(t-\tau) e^{i(kx+ly-kc_w\tau)-\gamma\tau+\frac{N\kappa}{f}z} d\tau, \tag{4}$$

where $\kappa^2 = k^2 + l^2$, $\kappa_d^{-1} = \frac{NH}{f}$ is the deformation radius, and $c_w = (1 - \frac{\kappa_d}{\kappa})U_0$ is the phase speed of the surface waves embedded in the current U. Next they considered a tracer S with a constant gradient $\Gamma_s = \frac{dS}{dy}$ embedded in the eddy field (4) given by

$$S_t + J(\psi - U_0 y, S) = -\Gamma_s \partial_x \psi. \tag{5}$$

After solving for S from the above equation, and obtaining the jet cross-stream velocity v from the stream function in (4), they calculated the tracer flux across the jet to be

$$< vS >= -[\frac{1}{2} f^2 \frac{k^2}{\kappa^2} \frac{\gamma}{\gamma^2 + k^2(c_w - U_0)^2} e^{2\frac{N\kappa}{f}z}]\Gamma_s. \tag{6}$$

The eddy diffusivity $K_\perp$ can be obtained by dividing (6) by Γ_s. The simple expression (6) has an important physical interpretation. It shows that for an eddy phase speed equal to the jet mean velocity U_0, the eddy diffusivity has a maximum value and indeed yields a value similar to those obtained by the group of studies which used the mixing length to relate the observations to the diffusivity. However, as the relative velocity of the eddies to the main current increases, the denominator of (6) becomes larger leading to a suppression in the diffusivity. In practice, this relative velocity is large (about 13 cm/s for the ACC)

and comparable to the main jet speed U_0 (which is about 15 cm/s in the core of the ACC). Thus, the suppression effect is pronounced in the core of the ACC as shown by Marshal *et al.*(2006) [3]. Neglecting the $(c_w - U)^2$ in the denominator of (6) leads to overestimation of the mixing length and thus, over prediction of the diffusivity. It should also be noted that the suppression is strong if the decorrelation timescale of the eddy is much slower than the advection of the tracer out of the eddy due to the relative velocity (i.e. if $k^2(c_w - U)^2 \gg \gamma^2$).

The FN09 expression for tracer flux mentioned above is for an ideal environment of infinite extent and with a constant velocity and constant stochastic forcing over the whole domain. Left panel of figure 1 shows the zonally averaged eddy kinetic energy and mean jet velocity curves for a patch of the southern ocean between 125^oW-150^oW and 66^oS-30^oS. The right panel of the figure shows a snapshot of the observed sea surface height anomalies for the same patch. As the figures show, there is considerable variations in both the forcing (eddy field) and the jet velocity from the main core of the ACC to its flanks. Thus, one can curiously wonder what happens if continuous or piece-wise continuous variations in the forcing and jet velocity are considered in the context of the model developed by FN09. To be more clear, the question is whether (and how) the expression (6) would change if for example we considered two semi-infinite velocity zones of adjacent flows with different forcing or different velocities? The main motivation of this study is to take the model explained above few steps further by allowing for these variations and investigating their effects on the tracer fluxes. In what follows in the remainder of this article, we start from a simple case of two semi-infinite adjacent zones and allow for variations in the forcing and the velocities between the two zones. The approach is to start from the simplest case and add complexity to the problem step by step to gain insight into the basic physics of the problem.

2 Barotropic Quasi-Geostrophic Formulation

As mentioned by FN09, their final results on eddy mixing are independent of their particular choice of the surface Quasi Geostrophic (QG) model. We opt to use two-dimensional barotropic QG formulations in this work and will recover the FN09 results for the surface of the ocean ($z = 0$). Modifying the formulation and results of this work to the surface QG formulation is an easy and straight forward task and only adds an exponential z-dependence to the stream functions and the fluxes obtained. Starting from 2D barotropic QG equations we have

$$q_t + J(\psi, q) = 0, \tag{7}$$

$$\nabla^2\psi - \kappa_D^2\psi = q - \beta y, \tag{8}$$

where q is the potential vorticity, $\kappa_D^{-1} = \sqrt{gH_0}/f_0$ is the deformation radius, and ψ is the velocity stream function. In the QG formulation, $\psi = \frac{g\eta}{f_0}$ where η is the surface elevation (or depression) from the reference surface level (located at H_0 above the flat bottom boundary). Since η is easily related to the pressure, the stream function can be interpreted as the surface pressure or the surface height and thus, can be estimated from altimetry data. In the two above formulae and all that follows beyond this point, a subscript denotes a differentiation. Next we make the assumption that the potential vorticity, q, is a function of the meridional

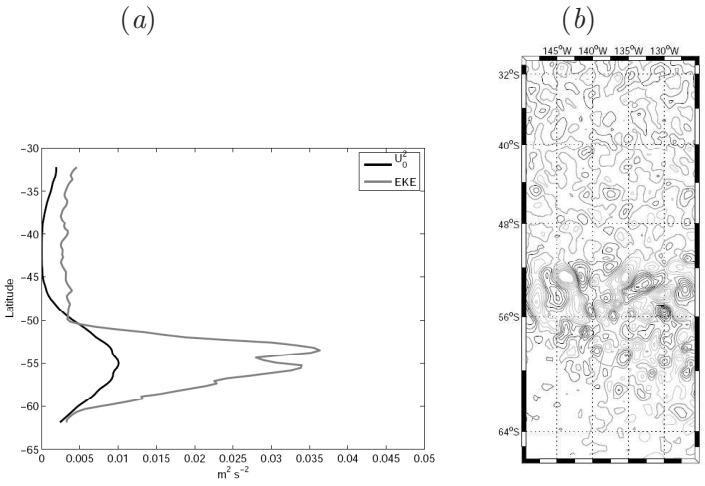


Figure 1: (*a*)Zonally averaged mean kinetic energy (black) and eddy kinetic energy (EKE) (gray) at the surface for a patch in the Pacific sector of the Southern Ocean between 125W-150W and 66S-30S; (*b*) Snapshot of sea surface height anomaly for the same patch: (black) positive anomaly, (gray) negative anomaly. (Taken from Ferrari and Nikurashin 2009 by permission).

coordinate y alone and so $q = Q(y)$. This assumption along with (8) mean that ψ is also only a function of y and so $u = -\psi_y = U(y)$ and $v = \psi_x = 0$. The equation (8) takes the simpler form

$$\frac{d^2\psi}{dy^2} - \kappa_D^2\psi = Q(y) - \beta y. \tag{9}$$

Next we perturb equations (7,8) around a mean flow $u = U(y)$ by introducing

$$\begin{aligned} q &= Q + \epsilon q', \\ \psi &= \Psi + \epsilon\psi', \\ u &= U + \epsilon u', \\ v &= \epsilon v'. \end{aligned} \tag{10}$$

Substituting (10) into (7,8) we get

$$q_t + \frac{1}{\epsilon}Q_t + J(\Psi, q') + J(\psi', Q) + \frac{1}{\epsilon}J(\Psi, Q) + \epsilon J(\psi', q') = 0, \tag{11}$$

$$(\nabla^2 - \kappa_D^2)\psi' = q'. \tag{12}$$

The second and the fifth term in (11) form the background field and sum up to zero. The $J(\Psi, q')$ term simplifies to $U(y)a'_x$ and the $J(\psi', Q)$ simplifies to $Q(y)\psi'_x$ leading to

$$(\partial_t + U\partial_x)q' + Q_y\psi'_x = \mathcal{F} \tag{13}$$

where $\mathcal{F}$ is the nonlinear term coming from $J(\psi', q')$. Dropping the primes we reach the linearized QG equations for the perturbation quantities q and ψ:

$$(\nabla^2 - \kappa_D^2)\psi = q, \tag{14}$$

$$(\partial_t + U(y)\partial_x)q + Q_y\psi_x = \mathcal{F}. \tag{15}$$

The Q_y term can be obtained by taking the y-derivative of (9)

$$Q_y = \beta + \kappa_D^2 U(y) - U(y)_{yy}. \tag{16}$$

The set of equations (14,15) will be the starting point of our analysis in the consequent sections. As we will see, the background velocity $U(y)$ is considered to be constant (or piecewise constant) throughout this article and thus the third term in (16) drops out and the second term simplifies to a constant value.

3 Basic Case: Non-stochastic forcing

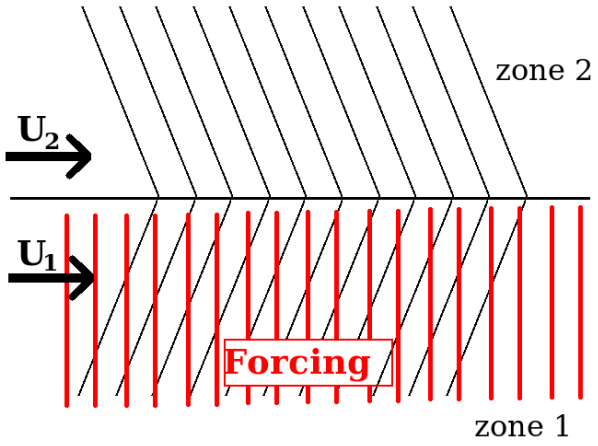


Figure 2: Schematic view of the basic case flow configuration.

As the first basic case, we consider two semi-infinite adjacent zones with velocities U_1 and U_2. Figure 2 shows the schematic of the problem. The top zone (zone 2) is an unforced region while the bottom zone is forced with a monochromatic forcing in the form of:

$$\mathcal{F} = fe^{ik(x-ct)}, \tag{17}$$

where f is the forcing amplitude, k is its x-direction wavenumber and c is the corresponding phase speed. The forcing is chosen to be monochromatic to keep the problem linear. Although a y-dependence can also be included in the expression for forcing, it does not add new physics into the problem and (17) can be used without loss of generality. The forcing is demonstrated by the red vertical lines in the figure. The linearized form of the governing equations in the top layer takes the form

$$(\nabla^2 - \kappa_d)\psi = q_2 \tag{18}$$

$$(\partial_t + U\partial_x)q_2 + \beta\psi_{2,x} + \gamma q_2 = 0, \tag{19}$$

and in the bottom layer we have

$$(\nabla^2 - \kappa_d)\psi = q_1 \tag{20}$$

$$(\partial_t + U\partial_x)q_1 + \beta\psi_{1,x} + \gamma q_1 = \mathcal{F}. \tag{21}$$

It is easy to see that the solution of (19) and the homogeneous solution of(21) have the general form:

$$\psi = ae^{ily} + be^{-ily}, \tag{22}$$

where a and b are constants and l is

$$l = \sqrt{\frac{\Gamma}{U-c} - (k^2 + \kappa_D^2)}, \tag{23}$$

where $\kappa_D^{-1} = \sqrt{gH_0}/f_0$ is the deformation radius for the barotropic case and $\Gamma = \beta + \kappa_D^2 U$ is a constant. We consider homogeneous solutions that radiate out from the interface and pick the proper term of (22) for each zone. Next, we set to calculate the particular solution of (21). To do that, we notice that the following expressions for ψ and q satisfy (21):

$$\psi = -A\frac{f}{k^2 + \kappa_D^2}\, e^{ik(x-ct)}, \tag{24}$$

$$q = Af\, e^{ik(x-ct)}. \tag{25}$$

Replacing these into (21), the coefficient A becomes

$$A = \frac{1}{\gamma + ik(U_1 - c - \frac{\Gamma}{k^2+\kappa_D^2})}. \tag{26}$$

It should be noted that A is an imaginary number and would be pure imaginary in the absence of the damping coefficient, γ. So, the total solution in the two zones become

$$\psi_1 = [a_1 e^{-il_1 y} - A\frac{f}{k^2 + \kappa_D^2}]e^{ik(x-ct)}, \tag{27}$$

$$\psi_2 = a_2 e^{il_2 y} e^{ik(x-ct)}. \tag{28}$$

Coefficients a_1 and a_2 can be found by matching the interfaces using two jump conditions which are originally due to Rayleigh (1880) [4]. The first jump condition ensure that the stream function (which can be interpreted as surface height in the QG formulation) is continuous across the interface. The second jump condition is that the displacement of the interface between the two regions of the flow be the same on both sides of the interface to ensure no cavitation occurs. A general derivation of these two conditions will be presented in the next section. For this case however, these conditions simply take the form

$$\Delta(\frac{\psi}{U-c}) = 0, \tag{29}$$

$$\Delta(\psi_y(U-c)) = 0, \tag{30}$$

were Δ stands for jump across the interface and both conditions are enforced at $y = 0$. Applying these conditions to the two stream functions we get

$$a_1 = \frac{ABl_2}{R^2 l_1 + l_2}, \tag{31}$$

$$a_2 = -\frac{ABRl_1}{R^2 l_1 + l_2}, \tag{32}$$

where $R = \frac{U_1 - c}{U_2 - c}$ and $B = \frac{f}{k^2 + \kappa^2}$. Next we set to calculate the cross-jets diffusivities in the two zones. Following FN09, we introduce the equation for the transport of a passive scalar in the fields described by stream functions (27, 28):

$$S_t = -J(\psi_i - U_i y, S - \Gamma_s y), \tag{33}$$

where Γ_s is a constant background tracer gradient equal to $\frac{d_B}{dy}$ and $i = 1, 2$ corresponds to the two zones. Rearranging we get

$$S_t + J(\psi, S) + U_i S_x = -\Gamma_s (\psi_x)_i = -\Gamma_s v_i. \tag{34}$$

We will shortly see that the second term in (34) is zero and the scalar equation reduces to

$$S_t + U_i S_x = -\Gamma_s v_i. \tag{35}$$

Solving (35) for the two zones using (27,28), the solutions of the scalar fields becomes:

$$S_1 = \frac{-\Gamma_s}{U_1 - c}(a_1 e^{-il_1 y} - AB)e^{ik(x-ct)}, \tag{36}$$

$$S_2 = \frac{-\Gamma_s}{U_2 - c}(a_2 e^{-il_2 y} - AB)e^{ik(x-ct)}. \tag{37}$$

One can substitute (36, 37) back into the $J(\psi, S)$ in (34) to confirm that the Jacobian vanishes. With the solutions of the scalar field obtained in both zones, we can calculate the diffusivity

$$K_\perp = \frac{1}{\Gamma_s} \Re < Sv > . \tag{38}$$

According to (36) for the first zone we have

$$S_1 = \frac{\Gamma_s}{ik(U_1 - c)}, \tag{39}$$

and thus the cross-jet flux becomes

$$\Re < Sv > = \frac{k(U_1 - c)}{\Gamma_s} \Re < iS_1 S_1 > = 0. \tag{40}$$

This simply means that for a velocity field of the general form $v = \cos(\alpha) + i \sin(\alpha)$, the solution to the scalar equation takes the form $S = c - \sin(\alpha) + i \cos(\alpha)$ and so the flux $\Re < Sv >$ takes the form $< \sin(\alpha) \cos(\alpha) >$ which is identically zero. Repeating the analysis

for the second zone yields the same result. This outcome simply indicates that a linear monochromatic forcing (which yields to north and south monochromatic radiations) cannot mix the fluid. To introduce diffusion (i.e., a nonzero flux) one has to either take into effect the nonlinear wave dynamics and wave interactions or introduce a forcing (and the consequent radiations) that decorelate from itself in a finite time. The second approach was that taken by FN09. In the next section we introduce a random forcing to investigate its effects. But before that, an interesting point can be observed by a closer look at equations (36, 37, 40). If the velocity of the main flow in one of the zones (lets say U_1) is equal to the zonal phase speed of the wave propagation, then the conclusion obtained above is not strictly valid. To better understand this fact, one can assume a single one-dimensional wave propagating eastward in the absence of any mean flow. The streamlines of the flow inside the fluid induced by the wave propagation form closed circulating regions and particle paths form closed elliptic curves. So, a particle (or a tracer filament) departing from some arbitrary initial location will return to its initial location due to the linear wave propagation. An introduction of any nonlinear effect (such as the Stoke's drift) will lead to the particle not returning to its initial position and thus mixing. This is also true for the case were a uniform background velocity U is superposed as one can go to the reference frame moving with U and observe the wave with the relative phase speed $U-c$. However, if the wave phase speed and the background flow velocities are the same (i.e., $U-c=0$), then an observer sitting on the reference frame which moves with U will observe a frozen wave pattern with fixed recirculating regions (a critical layer). Particles or tracer filaments left in these regions will mix with the ambient fluid at a high rate and considerable mixing will occur. That is why the value of the flux in equation (40) can not be clearly estimated in this limit as the denominator tends to zero. This behavior will be further examined in the next section.

4 Case I: Semi-infinite white-noise forcing

As the first case with random forcing, we consider two semi-infinite adjacent zones with similar velocities U. Figure 3 shows the schematic of the problem. The top zone (zone 2) is an unforced region while the bottom zone is forced similar to the previous section but with a stochastic forcing in the form of:

$$\mathcal{F} = f\kappa\sqrt{\gamma}\, r(t) e^{ikx} - \gamma q, \tag{41}$$

where f is the forcing amplitude, and $r(t)$ is a stochastic stationary variable satisfying $< r(t)r^*(t') >= \delta(t-t')$. The forcing is chosen to be monochromatic to keep the problem linear and is a crude representation of the wave excitation by the baroclinic instability and thus, k can be interpreted as the wave number of the most unstable mode of the instability.

The linearized form of the governing equations in the top layer takes the form

$$(\nabla^2 - \kappa_d)\psi = q_2 \tag{42}$$

$$(\partial_t + U\partial_x)q_2 + \beta\psi_{2,x} + \gamma q_2 = 0, \tag{43}$$

and in the bottom layer we have

$$(\nabla^2 - \kappa_d)\psi = q_1 \tag{44}$$

$$(\partial_t + U\partial_x)q_1 + \beta\psi_{1,x} + \gamma q_1 = \mathcal{F}. \tag{45}$$

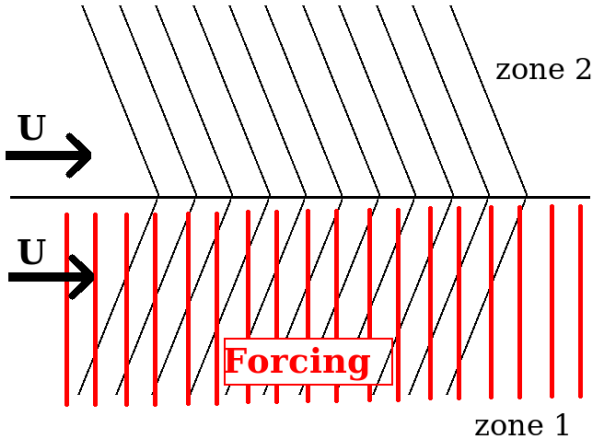


Figure 3: Schematic view of the case I flow configuration.

From previous section we know that the homogeneous solutions of equations (43, 45) can be written in the form of

$$\psi_h = e^{ikx} \int A(\omega)\, e^{il(\omega)y}\, e^{i\omega t} d\omega, \tag{46}$$

Substitution of (46) in (43) and in the homogeneous part of (45) leads to the dispersion relation

$$\omega = Uk + i\gamma - \frac{\beta k}{l^2 + k^2 + \kappa_d^2}. \tag{47}$$

Using 47 one can calculate the vertical wavenumber l for any particular frequency ω. However, the choice of l in zone 2 should be made in the way that all wave components travel away from the interface (and thus $\Re(l_2) > 0$) and all components decay as $y \to \infty$ (and thus $\Im(l_2) > 0$). A single component of such form is shown in zone 2 in figure 3 as the representative of the homogeneous solution in that zone. The components of the homogeneous solution in zone 1 should also travel downward and decay as $y \to -\infty$. This translates to $l_1(\omega) = -l_2(\omega)$. A single component of such form is also shown in zone 1 in the figure in black. We next set to derive the particular solution of the forced equation (45). To do that, we note that

$$q_{p1} = -a(t)\kappa f e^{ikx}, \qquad \psi_{p1} = a(t)\frac{f}{\kappa} e^{ikx} \tag{48}$$

satisfy the equations (44, 45). Substitution of (48) into (45) gives

$$a_t + [\gamma + ik(U - \frac{\beta}{\kappa^2})]a = \sqrt{\gamma} r(t), \tag{49}$$

which can be solved to give

$$a(t) = \sqrt{\gamma} \int_0^\infty e^{-(\gamma + ikc_w)\tau}\, r(t-\tau) d\tau, \tag{50}$$

where $c_w = U - \frac{\beta}{\kappa^2}$ is the phase speed of the eddies embedded in the current U. Thus the particular solution in zone 1 takes the form

$$\psi_{p1} = \frac{f}{\kappa}\sqrt{\gamma}\; e^{ikx} \int_0^\infty e^{-(\gamma + ikc_w)\tau}\; r(t-\tau)d\tau, \tag{51}$$

and the total solutions in the two zones become

$$\psi_2 = \psi_{h2} = e^{ikx} \int A_2(\omega)\; e^{il(\omega)y}\; e^{i\omega t} d\omega, \tag{52}$$

$$\psi_1 = \psi_{h1} + \psi_{p1} = e^{ikx} \int A_1(\omega)\; e^{-il(\omega)y}\; e^{i\omega t} d\omega + \frac{f}{\kappa}\sqrt{\gamma}\; e^{ikx} \int_0^\infty e^{-(\gamma + ikc_w)\tau}\; r(t-\tau)d\tau. \tag{53}$$

The stream functions (52 and 53) have to satisfy the jump conditions across the interface. The fist jump condition can be obtained by integrating the stream functions across the interface:

$$\int_{-\epsilon}^{0} \psi_1\; dy = \int_0^{+\epsilon} \psi_2\; dy. \tag{54}$$

Calculating the limit of (54) as $\epsilon \to 0$ gives the fist jump condition at the interface

$$A_1(\omega) = -A_2(\omega). \tag{55}$$

The second jump condition can be obtained by requiring the solution for the interface perturbation η (shown schematically in figure 11) obtained in the two zones to be equal to each other. Using (101) from appendix **A** and evaluating the jump condition at $y = 0$ we get

$$\hat{A_1} = \frac{1}{2}\hat{\psi_{p1}}, \tag{56}$$

where ' $\hat{}$ ' denotes the Fourier transform. Using (55, 56) and taking the Fourier transform of ψ_{p1} in (51) we get

$$\hat{A_1} = \frac{f}{2\kappa}\frac{\sqrt{\gamma}\; e^{ikx}}{\gamma + i(\omega - kc_w)}\hat{r}(\omega), \tag{57}$$

$$\hat{A_2} = -\frac{f}{2\kappa}\frac{\sqrt{\gamma}\; e^{ikx}}{\gamma + i(\omega - kc_w)}\hat{r}(\omega). \tag{58}$$

Next we set to calculate the $< vS >$ flux for the unforced region (zone 2). As before, the flux equation is

$$(\partial_t + ikU)S = -ik\Gamma\psi \tag{59}$$

multiplying both sides by e^{ikUt} and rearranging we get

$$\tilde{S_t} = -ik\Gamma e^{ikUt}\psi = H(t), \tag{60}$$

where $\tilde{S} = e^{iKUt}S$. As before, the flux can be written in the form

$$< Sv >= \frac{1}{2} < S^*v + Sv^* > . \tag{61}$$

We can write

$$< S^*v >=< S^*(-ik\psi) >=< e^{-iUkt}S^*(-ik\psi)e^{iUkt} >=< \tilde{S}^*\frac{H(t)}{\Gamma} > \tag{62}$$

and according to (60) we get

$$< S^*v >= \frac{< \tilde{S}^*H(t) >}{\Gamma} = \frac{1}{2\Gamma}\frac{d}{dt} < \tilde{S}_t^* > . \tag{63}$$

It is shown in the appendix B that for any two functions $V(t)$ and $S(t)$ if we have $V(t) = bA(t) * r(t)$ and $S_t = V(t)$ (where b is a constant), then $< SV >= \frac{1}{2}\frac{d}{dt} < S^2 >= \frac{1}{2}|bb^*||\hat{A}(0)\hat{A}^*(0)|$.

Thus if we can show that $H(t)$ can be written in the form of $H(t) = bA(t) * r(t)$ then from (63) and (60) we get

$$< S^*v >= \frac{1}{2\Gamma}|\hat{H}(0)|^2 = \frac{1}{2\Gamma}|\hat{H}(0)||\hat{H}^*(0)|. \tag{64}$$

Its easy to show that $< Sv^* >$ has also the same value as $< S^*v >$ and so $< Sv >=< S^*v >$ according to (61). In order to calculate the flux we have to calculate $\hat{H}(0)$ and also show that $H(t)$ has the $bA(t) * r(t)$ form. From (60) we have

$$\hat{H}(\omega) = -ik\Gamma\hat{\psi}(\omega - Uk). \tag{65}$$

Taking the Fourier transform of (52,53) and using (56,57,58) we get

$$\hat{\psi_1} = [\hat{A}_1(e^{-il(\omega)y} + 2)]e^{ikx} = \frac{f}{2\kappa}\frac{\sqrt{\gamma}\, e^{ikx}}{\gamma + i(\omega - kc_w)}\hat{r}(\omega)(e^{-il(\omega)y} + 2)e^{ikx}, \tag{66}$$

$$\hat{\psi_2} = \hat{A}_2 e^{il(\omega)y}e^{ikx} = \frac{f}{2\kappa}\frac{\sqrt{\gamma}\, e^{ikx}}{\gamma + i(\omega - kc_w)}\hat{r}(\omega)e^{il(\omega)y}e^{ikx}, \tag{67}$$

and so both ψ_1 and ψ_2 (and thus H) have the desired $bA(t) * r(t)$ (or $b\hat{A}\hat{r}$) form. It should be noted that in calculating the Fourier transform of ψ_p, we have assumed a non-zero value for the damping rate γ. Thus, in interpreting the results that follow, $\gamma = 0$ is not permitted event hough some equations might suggest it has a meaning. So, according to appendix B and (64), the diffusivity in the forced zone becomes

$$K_1 = \frac{1}{2}|\hat{A}_1\hat{A}_1^*||1 - \frac{e^{-il(\omega)y}}{2}|^2 = \frac{1}{8}\frac{f^2k^2}{\kappa^2}\frac{\gamma}{\gamma^2 + k^2(U - c_w)^2}\,(4 + e^{2\Im(l)y} - 4e^{\Im(l)y}\cos(\Im(l)y)), \tag{68}$$

and in the unforced zone we get

$$K_2 = \frac{1}{8}\frac{f^2k^2}{\kappa^2}\frac{\gamma}{\gamma^2 + k^2(U - c_w)^2}\, e^{-2\Im(l)y} \tag{69}$$

Two interesting points immediately emerge by looking at the expression for K_2. First, the magnitude of the diffusivity at the interface is a quarter of that of the diffusivity of the one infinite zone problem studied by FN09. This could already be known by the second jump condition (56) and noting that $K \propto |\hat{A(0)}|^2$. The second point is the y-dependence of the result in (69). As we move away from the interface in the unforced region, the diffusivity decays exponentially to zero. So the mere introduction of a jump in the forcing leads to differences in magnitude and behavior of the diffusivity in the unforced region.

Expression (68) shows that similar to K_2, K_1 also takes the value of one quarter of that of the infinite-domain case at the interface ($y = 0$) and so the solutions in the two zones match at the interface as expected. However, as $y \to -\infty$, the diffusivity tends to the result obtained in the infinite-domain case. So in the forced region, there is a transition from the value of one quarter of the constant infinite-domain flux at the interface to the full value in the far distances from the interface. This transition however, does not have an exponential decay pattern such as zone 2 and shows oscillatory behavior as determined by the terms in the parenthesis on the right hand side of (68). To demonstrate these behaviors better, figure 4 shows a series of plots of the wave patterns and diffusivity variations in the neighborhood of the interface for a set of parameters. The left column shows the wave pattern for the intrinsic ($\omega - Uk = 0$) wave pattern in the two zones and the right column shows the corresponding diffusivity curves. The relative magnitude of the ratio of the imaginary part of the vertical wave number l (which determines the decay rate in both zones) to the real part of l (which determines the oscillations in the forced zone) changes from 0 in case (a) to 1 in case (e). The green lines in the right column represent the constant diffusivity corresponding to the one-zone problem for the same parameters of each case. Starting from panel (a) where $\Im(l) = 0$, we can clearly see pure oscillatory behavior in both zones. The left panel shows the north-east propagation of waves coming from the homogeneous solution in the upper zone and superposition of a south-east propagation wave(coming from the homogeneous solution in zone 1) and a east propagating wave (coming from the particular solution in zone 1) in the lower zone. These results were expected from the $\omega - Uk = 0$ mode of the solutions (52, 53) in the Fourier space. The right panel shows the corresponding effective diffusivity. Pure oscillatory motion due to the absence of an imaginary part to l is clear as the diffusivity oscillates around the infinite-zone value (green line). It should be noted that as mentioned earlier, the value of the diffusivity at the interface (the point of meeting of the red and blue curves) is one fourth of that of the green line. This value retains a constant value in the unforced zone indicating a constant diffusivity. It is important to note that the $\Im(l) = 0$ condition can be achieved only when $\gamma = 0$ which means in the absence of any damping. Since the assumption of $\gamma \neq 0$ was used in the derivation of (69, 68), case (a) should be treated as a limiting case where $\gamma \to 0$.

As $\Im(l)$ grows larger than zero, the decay in the vertical direction leads to exponential decay of this one-fourth value to zero away from the interface in the upper zone and an oscillatory decay to the green line value in the forced zone. This can be seen as $\Im(l)$ grows to 0.1 in the case (b). The left panel in case (b) also show the decay of the waves to zero as we move away from the interface in the north zone and the decay of the radiative waves in the lower zone until only the forcing pattern remains. As the $\Im(l)/\Re(l)$ ratio increases in cases (c-e), the decaying layer in each zone becomes compressed about the interface. For the final case with $\Im(l)/\Re(l) = 1$, we almost have two regions of constant diffusivity with

a value of zero in the upper zone and that of the one-zone problem in the lower region. Although K still has the one-fourth value (of the green line) at the interface, the transition to the two limiting constant values in the two regions is very fast and limited to the close vicinity of the interface.

It would now be interesting to see how the real physical case corresponding to the ACC would compare to the cases of figure 4. To do that, we substitute for parameters in the obtained expressions for the diffusivity in the two zones by typical values of the southern ocean. The results for the diffusivity and the corresponding eddy kinetic energy are shown in figure 5. As the figure shows, the decay rates of the information away from the interface are very large meaning that $\Im(l)/\Re(l) > 1$ according to figure 4. In fact, $\Im(l)/\Re(l) \sim 40$ for figure 5,.

To investigate if $\Im(l)/\Re(l)$ can be smaller, we set to calculate real and imaginary parts of l. From the dispersion relation we have:

$$l^2 = \frac{\beta k}{(Uk - w) + i\gamma} - k^2 - \kappa_d^2 \tag{70}$$

for $(Uk - w)$ and $k \sim \kappa_d$ we get

$$l^2 = -\frac{i\beta k}{\gamma} - 2k^2 \tag{71}$$

it then can be shown that

$$\Re(l) = k\sqrt{\sqrt{(\frac{\beta}{2k\gamma})^2 + 1} - 1} \tag{72}$$

$$\Im(l) = k\sqrt{\sqrt{(\frac{\beta}{2k\gamma})^2 + 1} + 1} \tag{73}$$

and so

$$\frac{\Im(l)}{\Re(l)} = (\frac{\sqrt{(\frac{\beta}{2k\gamma})^2 + 1} + 1}{\sqrt{(\frac{\beta}{2k\gamma})^2 + 1} - 1})^{\frac{1}{2}}. \tag{74}$$

The fraction $\frac{\Im(l)}{\Re(l)}$ tends to 1 for very large values of $\frac{\beta}{2k\gamma}$ and tends to ∞ for very small values of $\frac{\beta}{2k\gamma}$. So, even for the smallest value of this ratio the interface effects are very limited to the interface neighborhood as shown in case (e) of figure 4. Substituting typical values of the southern ocean in (74) gives a ratio of 40.

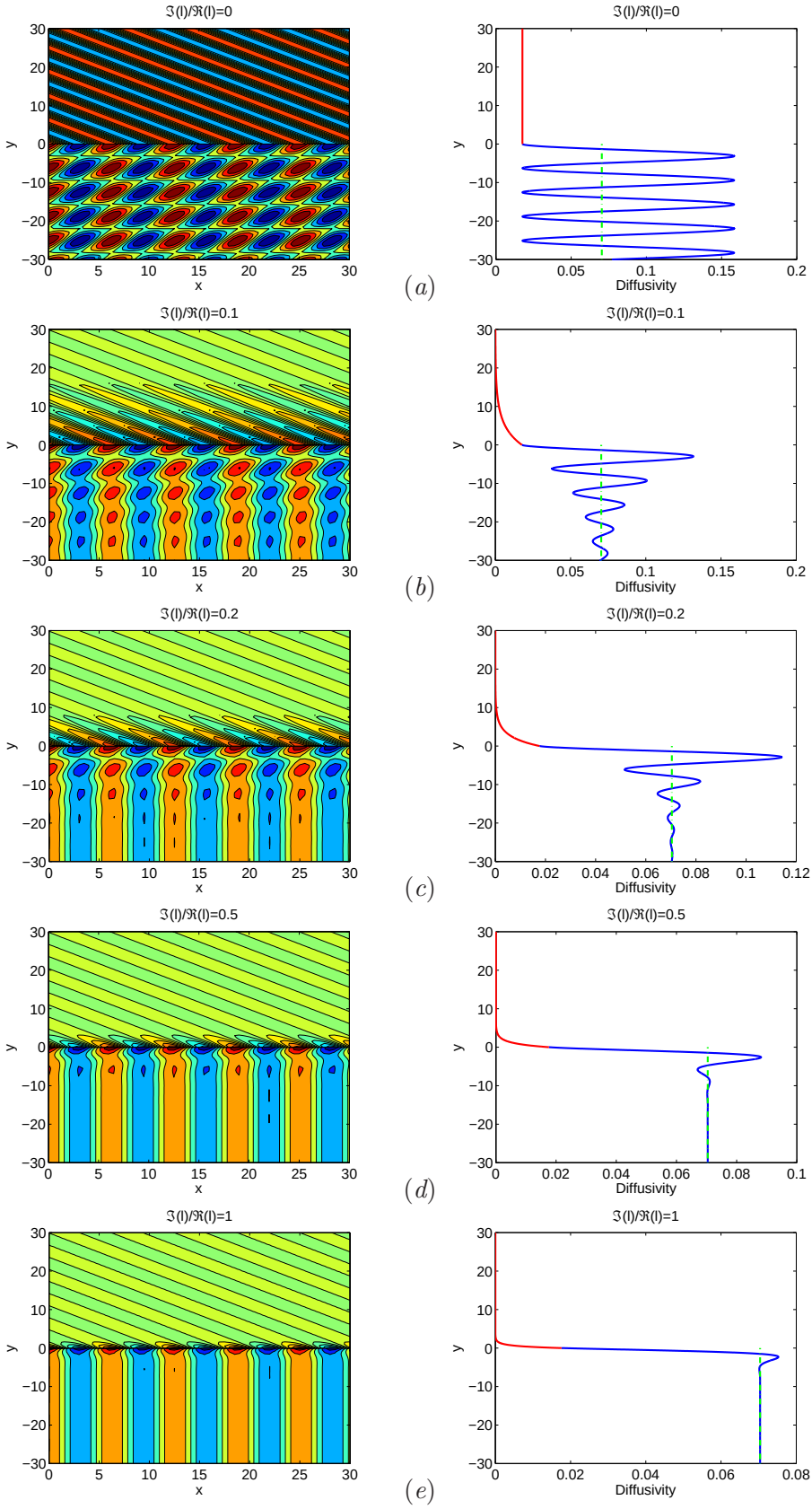


Figure 4: (a) Wave patterns for the intrinsic zero wave velocity in the two zones (left column) and the corresponding effective diffusivity (right column) for $U = 1$, $c_w = U/7$, $k = 1$ and $\Re(l) = 1$. (a) $\Im(l)/\Re(l) = 0$; (b) $\Im(l)/\Re(l) = 0.1$; (c) $\Im(l)/\Re(l) = 0.2$; (d) $\Im(l)/\Re(l) = 0.5$; (e) $\Im(l)/\Re(l) = 1$. Green lines in the right column represent the cons tan diffusivity corresponding to the one-zone problem for the same parameters of each case.

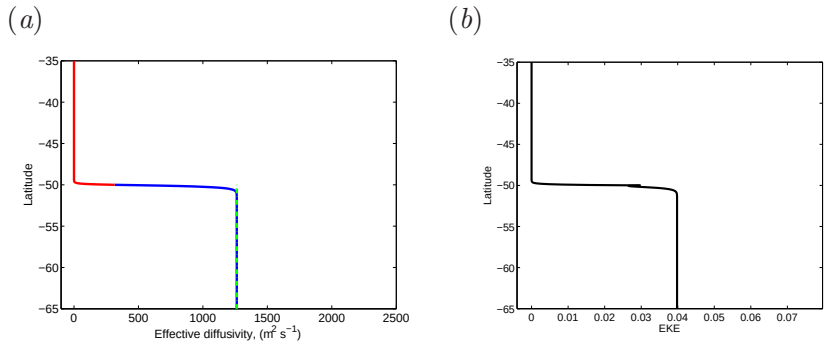


Figure 5: The diffusivity (left) and the corresponding eddy kinetic energy (right) for a case with parameters chosen to be in the range of those of the ACC.

5 Case II: Confined white-noise forcing

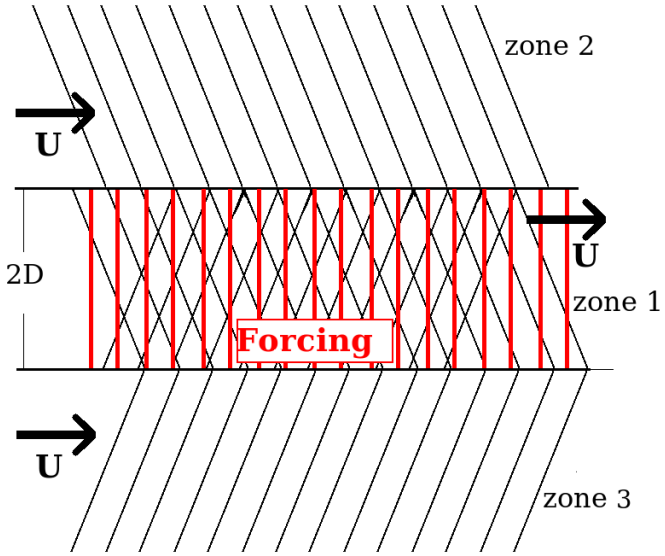


Figure 6: Schematic view of the case II flow configuration.

For the second case with white-noise forcing, we extend the previous case by confining the forcing region into a zone with width $2D$. The flow configuration is shown in figure 6. The velocities are uniform and equal in all three zones. The forcing is only applied to the middle zone (zone 1) and has a form similar to the previous case. The $y = 0$ point is chosen to be half way between the two interfaces in zone 1. As before, the stream functions take

the form

$$\hat{\psi}_1 = (\hat{A}_{1,1}e^{ily} + \hat{A}_{1,2}e^{-ily}) + \hat{\psi}_p, \tag{75}$$

$$\hat{\psi}_2 = \hat{A}_2 e^{ily}, \tag{76}$$

$$\hat{\psi}_3 = \hat{A}_3 e^{-ily}, \tag{77}$$

where ψ_p in zone 1 is similar to the previous case. It should be noted that due to symmetry, both north propagating and south propagating waves are allowed to exist in the middle zone. Applying the first of the jump conditions to the two interfaces gives

$$\begin{aligned} \hat{A}_{1,1}e^{ilD} - \hat{A}_{1,2}e^{-ilD} &= \hat{A}_2 e^{ilD}, \\ \hat{A}_{1,1}e^{-ilD} - \hat{A}_{1,2}e^{ilD} &= -\hat{A}_3 e^{ilD}. \end{aligned} \tag{78}$$

And applying the second jump condition at the two interfaces gives

$$\begin{aligned} \hat{A}_{1,1}e^{ilD} + \hat{A}_{1,2}e^{-ilD} + \hat{\psi}_p &= \hat{A}_2 e^{ilD}, \\ \hat{A}_{1,1}e^{-ilD} + \hat{A}_{1,2}e^{ilD} + \hat{\psi}_p &= \hat{A}_3 e^{ilD}. \end{aligned} \tag{79}$$

And so we get

$$\hat{A}_2 = \hat{A}_3 = -\frac{1}{2}\hat{\psi}_p(e^{ilD} - e^{-ilD}) \tag{80}$$

Considering a tracer equation like before and following the procedure explained in the last case for calculating the flux $< Sv >$ closely, the diffusivities in the three zones become

$$K_1 = \frac{1}{8}\frac{f^2k^2}{\kappa^2}\frac{\gamma}{\gamma^2 + k^2(U - c_w)^2}\,|2 - e^{ilD}(e^{ily} + e^{-ily})|^2, \tag{81}$$

$$K_2 = \frac{1}{8}\frac{f^2k^2}{\kappa^2}\frac{\gamma}{\gamma^2 + k^2(U - c_w)^2}\,|e^{ily}(e^{ilD} - e^{-ilD})|^2, \tag{82}$$

$$K_3 = \frac{1}{8}\frac{f^2k^2}{\kappa^2}\frac{\gamma}{\gamma^2 + k^2(U - c_w)^2}\,|e^{-ily}(e^{ilD} - e^{-ilD})|^2, \tag{83}$$

To demonstrate the variations in the diffusivity expressions obtained for the three zones, figure 7 shows a series of plots similar to those of figure 4. The third zone is not plotted in the figure since it is similar to the second zone but just mirrored with respect to the $y = 0$ line. The green lines in the right column represent the constant diffusivity corresponding to the one-zone problem for the same parameters of each panel. The right panel of the figure shows pure oscillatory motion due to the absence of an imaginary part to l for panel (a). Similar to the previous case, the values of the diffusivity at the interfaces are one fourth of that of the green line. This value retains a constant value in the unforced zone indicating a constant diffusivity. It is important to note that the $\Im(l) = 0$ condition can be achieved only when $\gamma = 0$ which is not permitted as explained before.

As $\Im(l)$ grows larger than zero, the decay in the vertical direction leads to exponential decay of this one-fourth value to zero away from the interface in the upper zone and an oscillatory symmetric pattern in the middle zone forced zone. This can be seen as $\Im(l)$ grows to 0.1 in the case (b). With further increase in $\Im(l)/\Re l$, the decaying layer in each zone

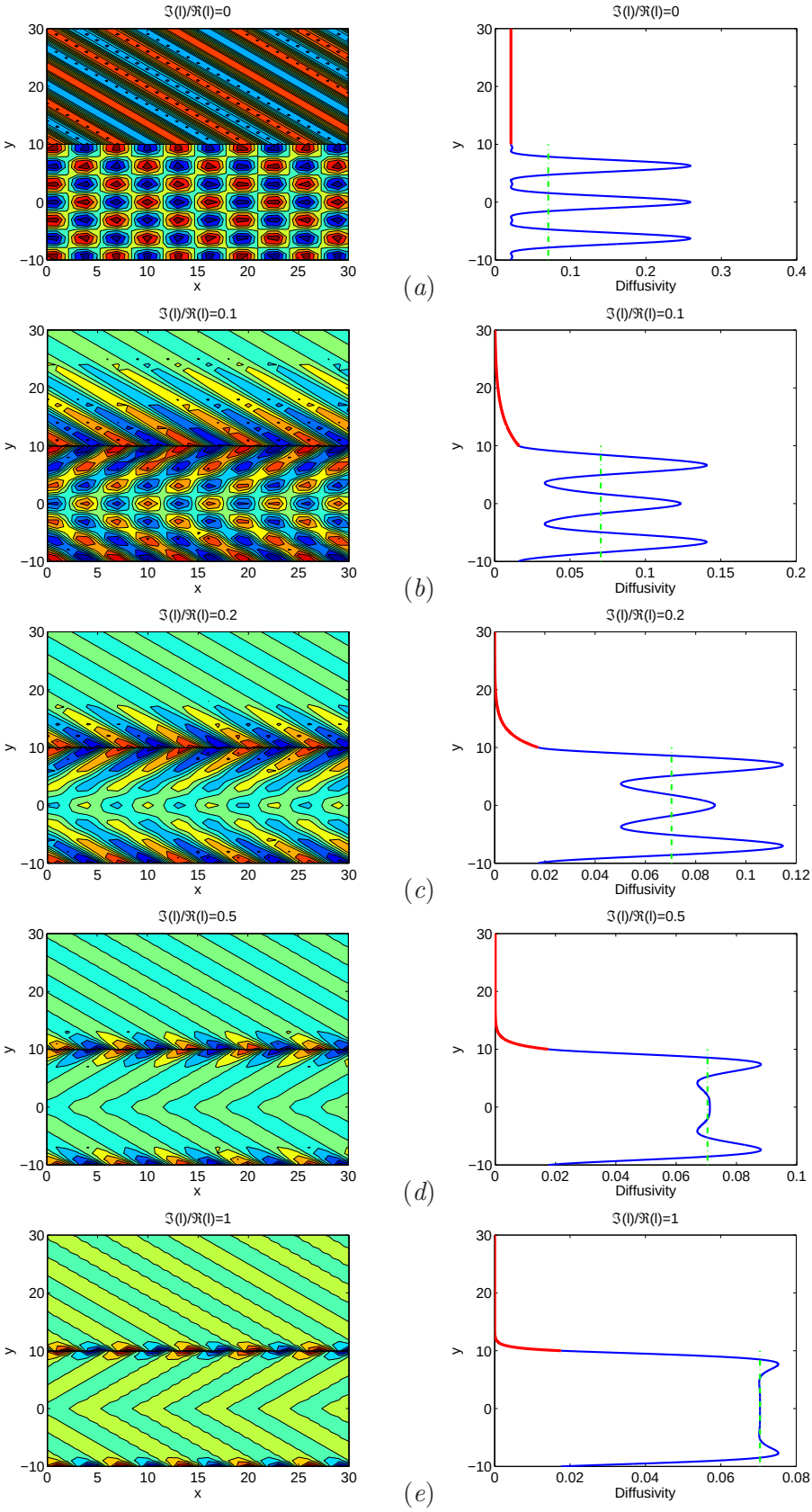


Figure 7: (a) Wave patterns for the intrinsic zero wave velocity in the two zones (left column) and the corresponding effective diffusivity (right column) for $U = 1$, $c_w = U/7$, $k = 1$ and $\Re(l) = 1$. (a) $\Im(l)/\Re(l) = 0$; (b) $\Im(l)/\Re(l) = 0.1$; (c) $\Im(l)/\Re(l) = 0.2$; (d) $\Im(l)/\Re(l) = 0.5$; (e) $\Im(l)/\Re(l) = 1$. Green lines in the right column represent the constant diffusivity corresponding to the one-zone problem for the same parameters of each case.

becomes more compressed about the interface. If we calculate the diffusivity curves for this case using values typical to the southern ocean (similar to figure 5 for the previous case), we find again that the variations in the diffusivities are limited to a very close proximity of the interfaces. And so, The two interfaces will not even feel the effect of each other as if they are each located between two semi-infinite zones.

Comparison of the right column of the figure for larger values of $\Im(l)/\Re l$ (panels (b, c, d)) with figure 7 of FN09 (reproduced in figure 8d in this article) poses some similarities. This may suggest that confinement of the forcing to a zone of a a limited width (2D in our case) can affect the general pattern of the diffusivity in the core of the ACC where the mean velocity is larger than the flanks of the ACC. A close look at figure 2 of FN09 (reproduced in figure 8c in this article) justifies the choice of a confined forcing for this particular patch of the pacific in which we are making the comparisons. As the figure shows, the eddy kinetic energy is dominant in the core of the ACC in the band between 50^oS and 60^oS.

As both figures 4 and 7 showed, the curve for the diffusivity in the unforced zone decay to a small value with a rapid rate, whereas the diffusivity curve in the forced region shows oscillatory behavior. This is mainly due to the absence of the forcing in the top zone. The forcing in the governing equations gives rise to a particular solution from which the oscillatory terms in the diffusivity expressions are originated. So a smoother transition in forcing from the main current to its flanks (zone 1 to zone 2) is needed to capture more variability in the second zone. Also, a change in the velocity from zone 1 to zone 2 might have consequences which can imply more variations in the diffusivity in the unforced zone. The effect of a piece-wise discontinuity in the velocities of the two zones will be briefly visited in the next section.

Regardless of the above arguments however, we have already found out that within the ranges of the parameters associated with the southern ocean all the variations in the diffusivity are compressed to the interface neighborhood. This simply implies that one has to consider smooth variations in the velocity and forcing fields rather than discontinuities considered here. In short, all the interesting physics are compressed into the region of discontinuity by the flow configurations used in our study. However, one can still use the expressions obtained to obtain realistic estimates of the diffusivity from observational data. FN09 described the procedure of translating the observational data into the parameters needed by the expressions we have developed for the diffusivity. Partly (and relatively crudely) following their steps, we can relate the damping rate γ to the eddy kinetic energy through

$$\gamma = d_0^{-1}\sqrt{\kappa^2 EKE}, \tag{84}$$

were d_0 is a constant defined in FN09. We also relate the amplitude of the forcing, f, to the eddy kinetic energy by

$$EKE = \frac{1}{2}f^2, \tag{85}$$

which is not exact for our case but provides a good approximation. We apply these assumptions to the same patch of the pacific ocean as discussed before. The observed EKE for that patch is shown in figure 8(c). To facilitate the calculation, an analytical form if (shown in panel a) is fitted to the EKE curve in panel c. There are few other parameters

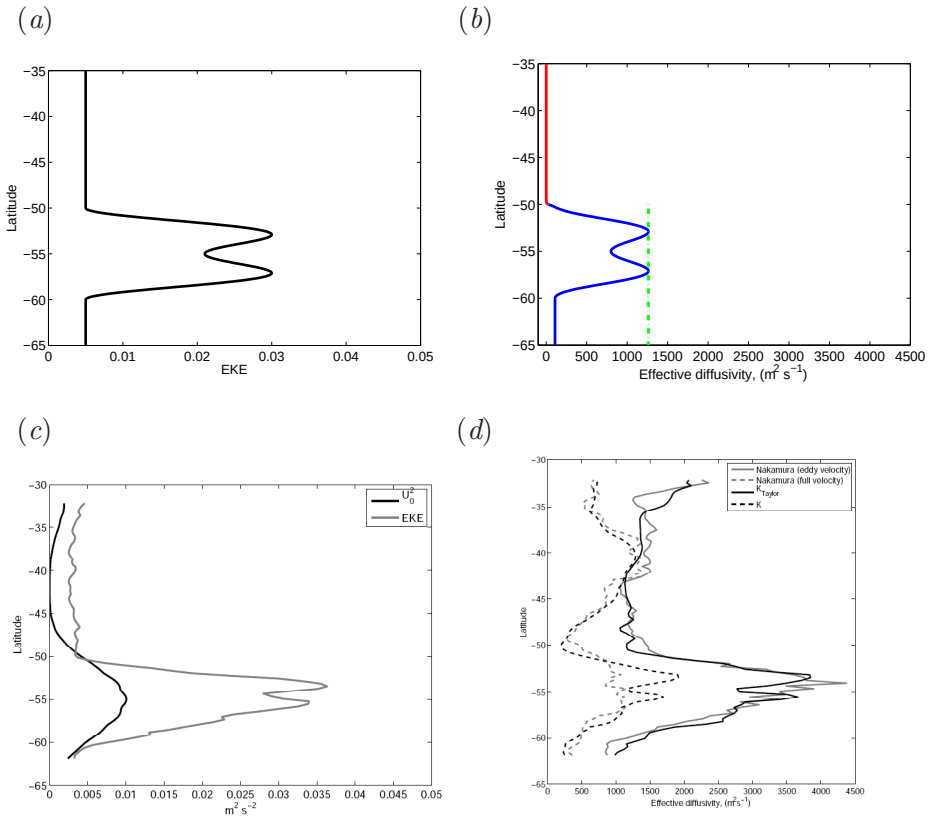


Figure 8: (a) An analytical fit to the observational EKE (shown in panel (c)) used in the present model; (b) Calculated diffusivity corresponding to the EKE of panel (a); (c) Zonally averaged eddy kinetic energy EKE in gray for a patch in the pacific sector of the Southern Ocean; (d) Diffusivities calculated by using various methods. (c,d) are from Ferrari and Nikurashin(2009)

.

needed to be determined for calculating the diffusivity from (81,82). The y-wavenumber, l, can be obtained from the dispersion relation (47) by setting $\omega = U_i k$ for each case. The x-wavenumber k is set to be that of the most unstable mode of the baroclinic instability and a typical value for the southern ocean is used for the deformation radius. Plugging these information into (81,82) we obtain the curves for diffusivities in the two regions. The curves are shown in the panel (b) of figure 8. The green line is defined similar to figures 4 and 7. Comparing the result with those predicted by FN09 (dashed lines in panel (d) of our figure 8), it seems that our calculation captures the size and pattern reasonably well. Indeed, all that we have achieved is to map the EKE map of panel (a) into the diffusivity curve of panel (b) through the parameters γ and f. And hence, there is not much variation in our diffusivity curve in the unforced zone because of the very small constant value of EKE considered in that zone. One can wonder how come the EKE curve of panel (c) leads to the diffusivity curves of panel (d) whereas that is not the case in our calculations

from (a) to (b). Probably the answer is that some physical phenomena accruing in the flanks are being missed out of our modeling due to our flow configurations. This will be further discussed in the discussion section.

6 Case III: Semi-infinite forcing with discontinuous velocity

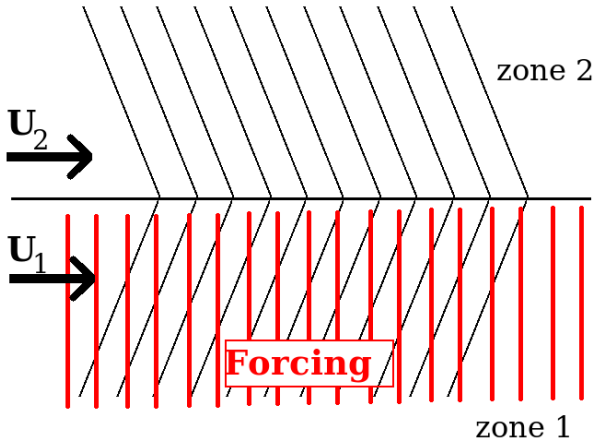


Figure 9: Schematic view of the case III flow configuration.

As the last case in this study, we also allow for a piecewise discontinuity in the velocity field (unlike the previous two cases) with white-noise forcing. Although previous cases have taught us that we should expect the consequences of the velocity discontinuity to be confined to the close neighborhood of the interface, this case might give some insight into possible outcomes of a change in the jet velocity. The flow configuration is schematically shown in figure 9. As previous cases, we start by writing the general form of the stream functions in the two regions

$$\hat{\psi}_1 = \hat{A}_1 e^{il_1 y} + \hat{\psi}_p, \tag{86}$$

$$\hat{\psi}_2 = \hat{A}_2 e^{il_2 y}. \tag{87}$$

It should be noted that the y-wavenumber components $l_1(\omega), l_2(\omega)$ are not the same anymore due to the velocity discontinuity. The jump conditions for this case become

$$l_2(\omega + U_1 k)(k^2 + l_1^2)\hat{A}_1 = l_1(\omega + U_2 k)(k^2 + l_2^2)\hat{A}_2, \tag{88}$$

$$(\omega + U_2 k)(\hat{A}_1 + \hat{\psi}) = (\omega + U_1 k)\hat{A}_2. \tag{89}$$

Solving for $\hat{A}_1$ and $\hat{A}_2$ we get

$$\hat{A}_1(\omega) = \frac{l_1(\omega + U_2 k)^2(k^2 + l_2^2)}{l_2(\omega + U_1 k)^2(k^2 + l_1^2) - l_1(\omega + U_2 k)^2(k^2 + l_2^2)}\hat{\psi}_p \tag{90}$$

$$\hat{A}_2(\omega) = \frac{l_2(\omega + U_1 k)(\omega + U_2 k)(k^2 + l_1^2)}{l_2(\omega + U_1 k)^2(k^2 + l_1^2) - l_1(\omega + U_2 k)^2(k^2 + l_2^2)}\hat{\psi}_p \tag{91}$$

In the case of $U_1 = U_2$, we have $l_1 = -l_2$ and these jump conditions reduce to (55, 56) obtained for Case I. The stream functions become

$$\hat{\psi}_1 = (\tilde{A}_1 e^{il_1 y} + 1)\hat{\psi}_p, \tag{92}$$

$$\hat{\psi}_2 = \tilde{A}_2 e^{il_2 y}\hat{\psi}_p, \tag{93}$$

where $\tilde{A}_1$ and $\tilde{A}_2$ are the fractions on the right hand sides of (90, 91) respectively. Finally, the expressions for the diffusivity in the two regions take the form

$$K_1 = \frac{1}{2}\frac{f^2k^2}{\kappa^2}\frac{\gamma}{\gamma^2 + k^2(U_1 - c_w)^2}\ |\tilde{A}_1 e^{il_1 y} + 1|^2_{(at\ U_1 k)}, \tag{94}$$

$$K_2 = \frac{1}{2}\frac{f^2k^2}{\kappa^2}\frac{\gamma}{\gamma^2 + k^2(U_2 - c_w)^2}|\tilde{A}_2 k e^{il_2 y}|^2_{(at\ U_2 k)}, \tag{95}$$

where the subscripts $(at\ U_i k)$ imply that the expressions should be evaluated at $\omega = U_i k$ for each case. Two points can be pointed out by looking at (94,95). First, similar to previous cases the expression for K_2 decays exponentially in zone 2 while the $|\tilde{A}_1 e^{il_1 y} + 1|^2$ term in (94) allows for oscillatory behavior in the forced zone. Second point is that one might argue that only the difference between the two velocities U_1 and U_2 matters as one can move with the reference frame attached to one of the zones and hence set its velocity to zero. However, that is not possible due to the dependence of the diffusivities in the two zones on the $(U_1 - c_w)$ and $(U_2 - c_w)$ terms. So it is the difference of the velocity of the main flow and the eddy propagation velocity that really matters for mixing in each zone.

Figure 10 shows the diffusivity curves in the two zones for different U_2/U_1 ratios. The plots are made for typical values for the southern ocean similar to figure 5. It should be noted that the figures (5, 10) differ from figure 8 in that fixed values for EKE (and subsequently fixed values of γ and f) are used in obtaining them. The velocity of zone 1 is chosen to be 15 cm/s and c_w is taken to be $U1/7$. The first panel simply recovers the results of case I as the velocity ratio is unity. However, the value of the diffusivity at the interface is changed as the velocity ratio is varied. This is shown by the overlapping of the blue and red curves in panel (b) and the discontinuity between them in panel (c). So, the unphysical sharp discontinuity in the velocity field leads to an expected unphysical discontinuity in the diffusivity. It should also be noted that the value of the diffusivity on each side of the interface does not show a monotonic behavior with the change in the velocity ratio. This can be clearly seen by comparing the tips of the red and blue lines at the interface and comparing them for the three cases in the figure. As before, the fast decay of the information away from the interface emphasizes the necessity of considering smoother changes in the velocity and forcing fields.

7 Conclusion

The aim of this project was to investigate the effects of variations in the forcing and the velocity field employed in the FN09 model on the expression for the diffusivity. Several test cases where considered with piecewise discontinuities in both the forcing and the velocity fields. It was shown mathematically that among the numerous radiative waves produced

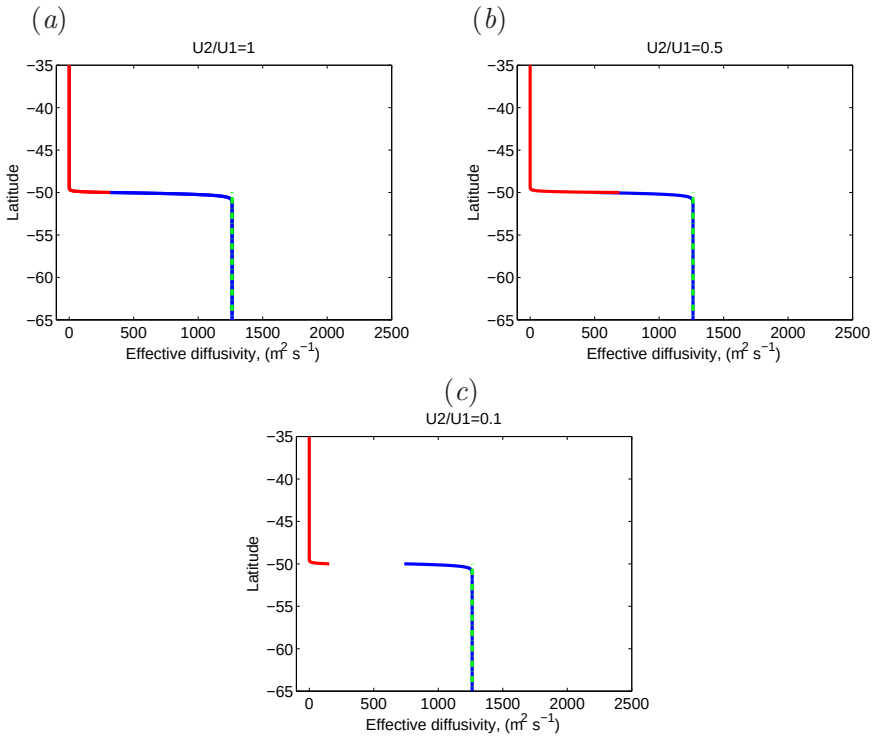


Figure 10: The diffusivity for three cases with different velocity rations. (a) $U_2/U_1 = 1$; (b) $U_2/U_1 = 0.5$; (c) $U_2/U_1 = 0.1$. The parameters are chosen in a similar way to figure 5.

through the discontinuity at the interface, only the one with a phase speed equal to the ambient jet velocity is the one contributing to the mixing process. This finding is of a great physical importance and also provides a very efficient tool for calculating the diffusivities. The mathematical derivations would be more rigorous without this finding. The results of this study suggest that one needs to consider a smoother variation in the velocity field to obtain more information about the mixing in the flanks of the jet. An estimate for the meridional variations in the velocity field can be obtained from the observational data such as those presented in FN09 both for the pacific patch and for the zonally averaged case. Considering a smoothly varying jet velocity enables investigating the presence of the critical layers in the flanks of the ACC and their implications for the mixing. As an instructional example, one can simply apply the expression derived for the eddy diffusivity in FN09 (equation (6) in this article) to a jet with a zonal velocity which uniformly decreases with y. By keeping the other parameters constant (for the sake of physical intuition), the denominator of the expression for the flux becomes smaller as the jet velocity is reduced as the $(U - c_w)$ becomes smaller. At a critical layer where the jet velocity reaches the speed of propagation of the eddies, the diffusivity obtains a maximum value. Although this argument is only true as long as the other parameters are kept constant, it shows that one can expect a rise in the eddy diffusivity in the regions where the jet velocity has reduced to a value close to c_w. The enhancing effect of the critical layers was suggested by Marshal *et al.*(2006) [3]

and seems to be in agreement with the results of FN09. The introduction of a continuously varying jet velocity into the governing equations may pose analytical difficulties but one might be able to choose a profile which makes the mathematical derivations easier (such as the work of Talley (1983) [5]).

Another important feature of all the results of our study was the fast decay of the effects of the interface away from it. By revisiting the procedure we followed in the stability analysis, we note that the solutions were sought in the form of $e^{i(kx-ct)}$, where c was assumed real. This assumption was made to exclude the temporally growing modes. However, the final results indicated a fast spacial decay (and so "trapped" waves). One can wonder if there might be modes that can grow in time but decay in space? Indeed stability analysis done with allowance for a complex c (and thus the possibility of growing modes) shows that those modes can exist. They might grow in time and decay in space (in the y direction) in a way that they form a wave-packet travelling north (or south) with a group velocity. Or, they might still be trapped but decay to zero much slower than our results (due to their temporal growth). Either way, these waves can propagate information much further from the interface compared to our results. A nice discussion on these groups of waves can be found in Kamenkovich and Pedlosky (1995) [2].

So, it seems that the first step in extending this study is to revisit the cases considered in this work by allowing for modes with temporal growth rates which were not included. This can allow for the information to propagate further from the interface which can have interesting implications for the mixing in the flanks of the jet. Allowing for continuous variations in the velocity would also be a natural step towards building a more realistic model.

In the end, I would like to thank

Both my advisers for all the time they spent with me on this project,
Joe Pedlosky for the instructive discussions we had,
The GFD program for accepting me as a fellow,
George Veronis for the softball experience,
Marco for being such a nice roommate and helping me out in the world of vegetable,
The 2009 fellows for the fun experience,
and
All the GFD staff for helping such a wonderful program to exist.

References

[1] R. Ferrari and M. Nikurashin, *Suppression of eddy diffusivity across jets in the southern ocean*, submitted to J. Phys. Oceanogr., (2009).

[2] I. Kamenkovich and J. Pedlosky, *Radiating instability of nonzonal ocean currents*, J. Phys. Oceanogr., (1995).

[3] J. Marshall, E. Shuckburgh, H. Jones, and C. Hill, *Estimates and implications of surface eddy diffusivity in the southern ocean derived from tracer transport*, J. Phys. Oceanogr., (2006).

[4] L. Rayleigh, *On the stability or instability of certain fluid motions.*, Scientific Papers, Vol.1, 474-487, (1980).

[5] L. Talley, *Radiating barotropic instability*, J. Phys. Oceanogr., (1983).

8 Appendix A

The second jump condition can be obtained by setting the solution for the interface perturbation η obtained in the two zones to be equal to each other. The interface perturbation η is shown schematically in figure 11. The equations for η in the two zones have the form

$$\frac{D\,\eta}{D\,t} = \partial_t \eta + U_1 \partial_x \eta = v_1 = -\partial_x \psi_1 \tag{96}$$

$$\frac{D\,\eta}{D\,t} = \partial_t \eta + U_2 \partial_x \eta = v_2 = -\partial_x \psi_2 \tag{97}$$

Taking the Fourier transform of the two equations and expecting η to have the same x-dependence as the stream functions (i.e., e^{ikx}), we get

$$\partial_t \hat{\eta} + ikU_1 \hat{\eta} = ik\hat{\psi}_1 \tag{98}$$

$$\partial_t \hat{\eta} + ikU_2 \hat{\eta} = ik\hat{\psi}_1 \tag{99}$$

Subtracting (99) from (98) we get

$$\hat{\eta} = \frac{\hat{\psi}_1 - \hat{\psi}_2}{U_1 - U2}. \tag{100}$$

substituting (100) into (98) gives the jump condition

$$(\partial_t + ikU_1)(\hat{\psi}_1 - \hat{\psi}_2) = ik\hat{\psi}_1(U_1 - U_2). \tag{101}$$

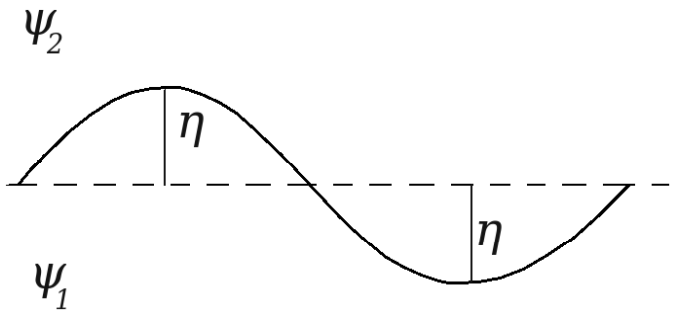


Figure 11: Schematic view of the perturbation of the interface between zones 1 and 2.

9 Appendix B

The goal of this appendix is to to show that **if** two complex functions $S(t)$ and $V(t)$ satisfy the relations

$$S_t = V(t), \tag{102}$$
$$V(t) = A(t) * r(t), \tag{103}$$

where $A(t)$ is a general function and $r(t)$ satisfies

$$C(t-t') =< r(t)r(t')^* >= a\ stationary\ function, \tag{104}$$

then the following is true:

$$< Sv >= \frac{1}{2}|\hat{A}(0)|^2 C(0). \tag{105}$$

C is the covariance function and '*' denotes the convolution of the two functions. The fact that $C(t-t')$ is stationary means that it only depends on $\Delta t = t - t'$ and not on t or t'. To start the proof we have

$$< Sv >= \frac{1}{2}\frac{d}{dt} < S^2 >= \frac{1}{2}\int_0^t < V^*(t)V(t') > dt' = \frac{1}{2}\int_{-\infty}^{\infty} < V^*(t)V(t-t') > dt'. \tag{106}$$

Its easy to show that the second and term expressions are equal in the above equality. The integrand in the last expression in (106) can be simplified to

$$\begin{aligned} &< V^*(t)V(t-t') >=< (A*r)^*|_t (A*r)|_{t-t'} >= \\ &< \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A^*(t-t_1)r^*(t_1)A(t-t'-t_2)r(t_2)dt_1dt_2 >= \\ &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A^*(t-t_1)r^*(t_1)A(t-t'-t_2)C(t_1-t_2)dt_1dt_2, \end{aligned} \tag{107}$$

where C is defined in (104). Replacing the last expression back into (106) and integrating with respect to t' we get

$$< Sv >= \frac{1}{2}\hat{A}(0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A^*(t-t_1)C(t_1-t_2)dt_1dt_2.$$

Next integrating with respect to t_2 we get

$$< Sv >= \frac{1}{2}\hat{A}(0)\hat{C}(0)\int_{-\infty}^{\infty} A^*(t-t_1)dt_1.$$

And finally integrating with respect to t_1 we get

$$< Sv >= \frac{1}{2}\hat{A}(0)\hat{A}^*(0)\hat{C}(0) = \frac{1}{2}|\hat{A}(0)|^2\hat{C}(0). \tag{108}$$

This completes the proof. Had we included a constant b in the definition of $V(t)$ at the beginning (i.e., $V(t) = bA(t) * r(t)$), the final result (108) would become

$$< Sv >= \frac{1}{2}|bb^*||\hat{A}(0)|^2\hat{C}(0). \tag{109}$$

For a white noise function $r(t)$ (used throughout this article) we get

$$C(t-t') =< r(t)r(t')^* >= \delta(t-t'),$$

and so

$$\hat{C}(0) = 1. \tag{110}$$

Flow-destabilized seiche modes with a movable dam

Hélène Scolan

August 2009

1 Introduction

Volcanic tremors are seismic signals emanating from fluid channels encased in rock. The understanding of the mechanism of this phenomenon is of great importance to improve evaluation of volcanic eruptions. Many possibilities have been advanced to understand the mechanism responsible for these low-frequency signals.

It has been suggested recently [5] that the seismicity could be the result of flow-destabilized oscillations whose frequency is set by an adjacent reservoir in the rock. Thus, the channel in the rock would act like a clarinet reed exciting and interacting with standing waves in an adjacent reservoir (Fig 1).

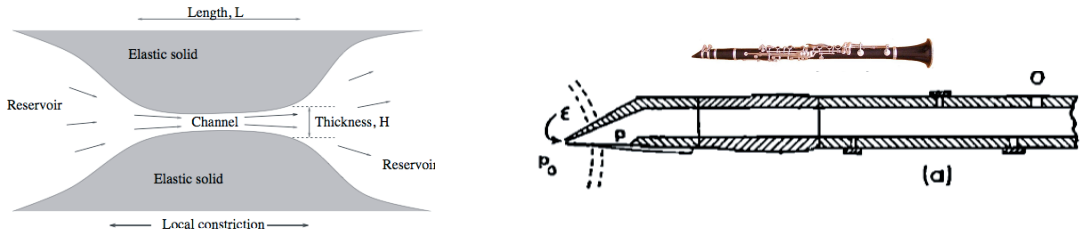


Figure 1: Oscillating reed in a clarinet (Backus 1963 [1]) or rock excitating and interacting with standing waves in the adjacent reservoir [5].

In the musical instrument, the sound is produced thanks to an instability arising from the coupling between the flow in the reed and the feedback from the resonating cavity. In the same way, crack or 'sloshing' modes (instead of acoustic modes) in a reservoir adjacent to a channel in the rock could be the relevant mechanism of generation of the low-frequency seismicity in volcanic tremor.[5]

So far, the theory of destabilized sloshing modes has not been deeply studied. The purpose of this project is to explore and illustrate the mechanism using a simple analogous laboratory configuration involving water instead of magma and looking for flow-destabilized seiche modes (Fig 2).

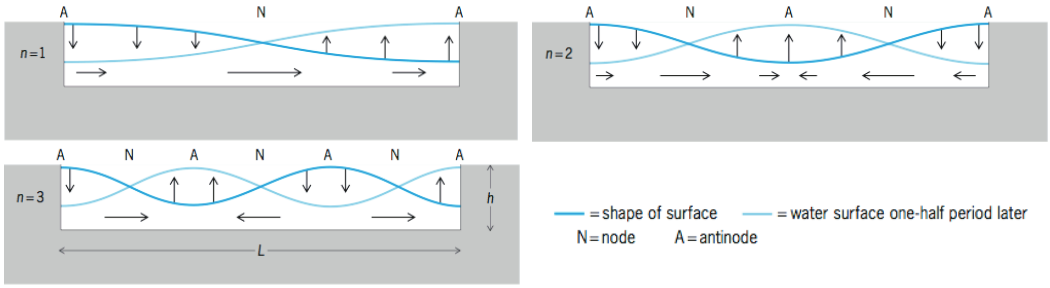


Figure 2: Seiches modes in a box: standing waves with $T = \frac{2L}{n\sqrt{gh}}$

2 Experiment in water : a reservoir and a movable dam

Figure 3 illustrates the experimental setup chosen to investigate the phenomenon. It looks like a clarinet in that a movable plexiglas plate (the "paddle") plays the role of the reed in the instrument and there is an adjacent reservoir. Water flows at constant flow rate into the reservoir. The plexiglas plate attached to a pivoted rod and masses are first added at the other extremity in order to counterbalance the weight of the plate. A addtional mass m_{added} is then put at varying positions x on the rod to make the plate go down.

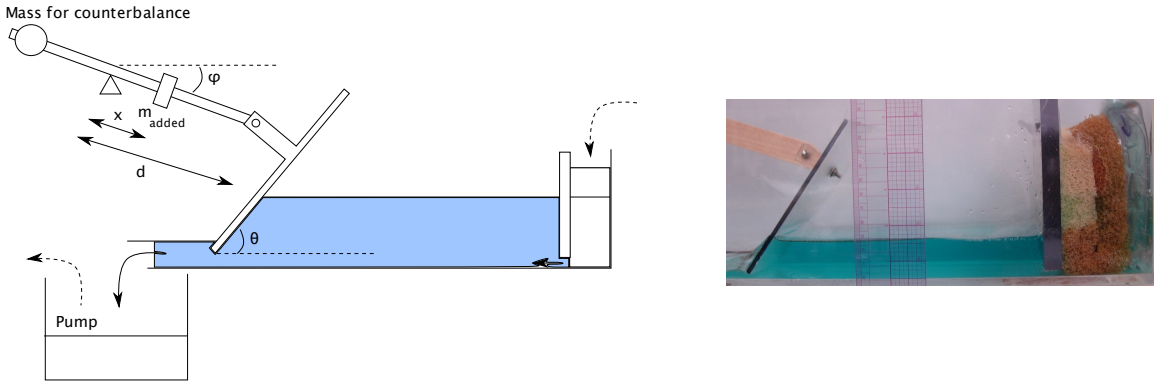


Figure 3: Experimental setup: the reservoir length is around 20 cm, d=50cm. m_{added} varies from 3 to 100g and the height of water can vary from around 2 cm to 5 cm.

Blue dye is added in the water to enhance the contrast in the images taken at 15 fps with a Pulnix camera. Turbulence near the inflow to the reservoir is prevented by inserting a sponge or polystyrene foam. However, as that dampening structure also absorbs waves, a plexiglas barrier is placed in front to promote wave reflections.

First, the goal will be to understand theoretically under what conditions the instability appears and then to observe the phenomenon experimentally.

3 A shallow water model

3.1 Governing equations

3.1.1 Equation for the paddle

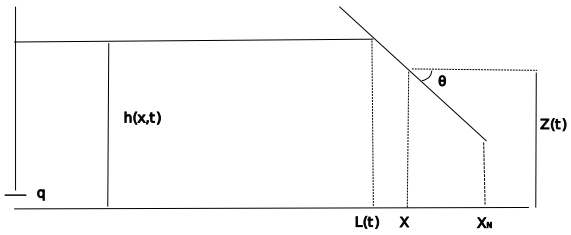


Figure 4: Diagram of the system and notations.

The motion of the paddle is described by the conservation of angular momentum applied on the pivot point (triangle in figure 3):

$$J\ddot{\phi} = -mgd\cos(\phi) + F\cos(\phi) \tag{1}$$

where J is the moment of inertia, that can be written as $J = (M+m)d^2$, with a global effective mass M and an additional equivalent mass m placed on the end on the paddle. (i.e. m is linked to the excess mass added on the paddle by $m_{\text{added}}x^2 = md^2$).

We consider here only the force due to the gravity and F, the vertical force of the water on the paddle. Friction (for example in the hinge) is neglected. The force F is essentially the vertical component of the normal pressure force (dominating drag force) exerted on the paddle by the fluid. Introducing W, the transversal width of the reservoir, we have therefore:

$$J\ddot{\phi} = -mgd\cos(\phi) + Wd\cos(\phi)\cos\theta\int_L^{X_N} p dx \tag{2}$$

If ϕ is a small angle, $\cos(\phi) \approx 1$ and the position $Z(t)$ of the paddle satisfies: $\dot{Z}(t) \approx d\dot{\phi}(t)$ then the governing equation of the paddle is finally:

$$(m+M)\ddot{Z} = -mg + W\cos\theta\int_L^{X_N} p dx \tag{3}$$

3.1.2 Equations in the fluid

We will work here with an inviscid shallow water model (assuming the height remains smaller than the horizontal scale). The problem can be divided up into two regions:

• For $x > L$, the water flows underneath the paddle, and the height and velocity satisfy the following equations:

$$h = Z(t) + (X - x)\tan\theta \tag{4}$$

$$h_t = \dot{Z} = -(hu)_x \tag{5}$$

$$u_t + uu_x = -\frac{p_x}{\rho} \tag{6}$$

where u_i is the derivative of u with respect to i. X is the position of the middle of the paddle and $Z(t)$ its vertical position. The boundary conditions specify the continuity of the flow velocity and height at $x = L$:

$$u(x \to L^-) = u_L \tag{7}$$
$$h(x \to L^-) = h_L = Z + (X - L)\tan\theta \tag{8}$$

Equation (5) can be integrated between L and x to give an expression for the velocity:

$$u = \frac{h_L u_L + (L - x)\dot{Z}}{Z + (X - x)\tan\theta}$$

• For $0 < x < L$, the water surface is free, implying a hydrostatic pressure, and we have:

$$h_t + (hu)_x = 0 \tag{9}$$
$$u_t + uu_x = -gh_x \tag{10}$$
$$\tag{11}$$

The imposition of a constant flow rate q at $x = 0$, and the continuity at the right are written as:

$$[hu]_{x=0} = q \tag{12}$$
$$h(x = L) = h_L = Z + (X - L)\tan\theta \tag{13}$$

The problem contains three unknowns, $Z(t)$, $L(t)$ and $u_L(t)$, that are determined using the equation of motion of the paddle, the equations of the fluid for $0 < x < L$, and equation (6) integrated between $x = L$ and the end of the paddle, $x = X_N$, where we assume that the pressure becomes hydrostatic again: $-\int_L^{X_N}(u_t + uu_x)dx = \rho g(h_N - h_L)$.

3.2 Adimensionalisation

3.2.1 Scaling

In order to simplify the problem and uncover the important non-dimensional parameters, we consider different scalings for x in the two regions:

• For the reservoir, the natural scale for distance is X, the position of the paddle. So we use $\hat{x} = \dfrac{x}{X}$ for $0 < x < L$.

• Under the paddle, the relevant distance is $X_N - X$, so we introduce another variable ξ such that:

$$x = (X_N - X)\xi + X \tag{14}$$

with $l < \xi < 1$ where $l = \frac{L-X}{X_N-X}$. i.e.

$$\hat{x} = R\xi + 1$$

with the ratio, $R = \dfrac{X_N - X}{X}$

The vertical scale is set by $(X_N - X)\tan\theta$ so all the other non-dimensional quantities are written:

$$\hat{h} = \frac{h}{(X_N - X)\tan\theta}$$

$$\hat{u} = \frac{u}{\sqrt{g(X_N - X)\tan\theta}}$$

$$\hat{t} = \frac{t}{\frac{X}{\sqrt{g(X_N - X)\tan\theta}}}$$

$$\hat{p} = \frac{p}{\rho g(X_N - X)\tan\theta}$$

3.2.2 Adimensionalised equations

We then derive the adimensionalised form of the equations of the fluid:

• For $x > L$:

$$\hat{h} = \hat{Z} - \xi \tag{15}$$

$$\hat{h_t} = \dot{\hat{Z}} = -(\hat{h}\hat{u})_{\hat{x}} \tag{16}$$

$$\hat{p}_\xi = -\hat{u}\hat{u}_\xi - R\hat{u}_{\hat{t}} \tag{17}$$

with the boundary conditions:

$$\hat{u}_{\hat{L}} = u(\hat{x} = 1 + Rl) \tag{18}$$

$$\hat{h}_{\hat{L}} = h(\hat{x} = 1 + Rl) \tag{19}$$

$$\hat{u} = \frac{\hat{q_L}}{\hat{Z} - \xi} - R\frac{\xi - l}{\hat{Z} - \xi} \tag{20}$$

• For $0 < x < L$, it is convenient to write everything in terms of the flow rate $q = uh$ and the height h (instead of u and h):

$$\hat{h}_{\hat{t}} + \hat{q}_{\hat{x}} = 0 \tag{21}$$

$$\hat{q}_{\hat{t}} + (\frac{\hat{q}^2}{\hat{h}})_{\hat{x}} = -\hat{h}\hat{h}_{\hat{x}} \tag{22}$$

with the boundary conditions:

$$[\hat{h}\hat{u}]_{\hat{x}=0} = \hat{q} \tag{23}$$

$$\hat{h}_{\hat{L}} = h(\hat{x} = 1 + Rl) \tag{24}$$

The equation 3 for the paddle becomes:

$$I\ddot{\hat{Z}} = -1 + \mu \int_l^1 p d\xi \tag{25}$$

where I and μ are two dimensionless parameters.

3.2.3 Control parameters

We distinguish 5 parameters in the system:

- $I = \dfrac{m+M}{m} R^2 \tan^2\theta$ the inertia term,
- $\mu = \dfrac{\rho W (X_N - X)^2 \sin\theta}{m}$ corresponding to the ratio $\dfrac{\text{mass of water under the paddle}}{\text{mass on the paddle}}$
- $R = \dfrac{X_N - X}{X} = \hat{X}_N - 1$ the aspect ratio horizontally between the distance under the paddle and the length of the reservoir.
- $\hat{Q} = \dfrac{q}{\sqrt{g}((X_N - X)\tan\theta)^{3/2}}$ the dimensionless imposed flow rate.

3.3 Approximation for small R

To simplify the problem, we add the hypothesis that R is small, implying that the distance under the paddle is small compared to the length of the reservoir. Then, if we recall the equations for $x > L$:

$$\hat{u} = \frac{\hat{q_L}}{\hat{Z} - \xi} - R\frac{\xi - l}{\hat{Z} - \xi}$$
$$\hat{p}_\xi = -\hat{u}\hat{u}_\xi - R\hat{u}_{\hat{t}}$$
$$I\ddot{\hat{Z}} = -1 + \mu \int_l^1 p d\xi$$

We obtain $u \approx \dfrac{q_L}{Z - \xi} = \dfrac{q_L}{h}$ and $p_\xi \approx -u u_\xi$.
(Now we work with the non-dimensional variables and omit the hat in the formulas).

If we integrate the latter equation, we can see that this additional hypothesis means that Bernoulli's law holds underneath the paddle:

$$p + 1/2u^2 = B = \text{constant} = h_N + 1/2u_N^2 = h_L + 1/2u_L^2 \tag{26}$$

because the pressure is hydrostatic at X_N and L.

Physically, it also amounts to neglecting the temporal dependence of the fluid variables under the paddle. i.e. the flow is in a quasi-steady state under the paddle. That is to say that the time to pass under the paddle is smaller that the time of change of velocity.

Note that the equality (26) takes the compact form:

$$\frac{q_L^2}{2h_N^3} = \frac{Fr_N^2}{2} = \frac{\alpha^2}{\alpha + 1} > 1 \tag{27}$$

where $\alpha = \dfrac{h_L}{h_N}$ is the ratio of depths at the two extremities of the paddle. Fr_N is the Froude number at the end of the paddle and the flow is supercritical.

From the Bernoulli equation (26), the pressure can be written:

$$p = \left(h_N + \frac{{q_L}^2}{2h_N}\right) - \frac{{q_L}^2}{2h} \tag{28}$$

and by integrating to get the total pressure force on the paddle, the equation of motion (25) becomes:

$$I\ddot{\hat{Z}} = -1 + \frac{q_L^{4/3}\mu}{2^{2/3}}F(\alpha) \tag{29}$$

where

$$F(\alpha) = \frac{(\alpha-1)(\alpha^2+1)}{\alpha^{4/3}(\alpha+1)^{1/3}} \tag{30}$$

3.4 Steady state

The steady state (Q_L, H_L, H_N) is calculated easily:

$$F(\alpha_0) = \frac{2^{2/3}}{Q_L^{4/3}\mu} \tag{31}$$

3.5 Linear stability analysis

We now investigate the linear stability around the steady state. From the equations in the reservoir $(0 < x < L)$:

$$h_t + q_x = 0 \tag{32}$$

$$q_t + (\frac{q^2}{h})_x = -hh_x \tag{33}$$

$$\tag{34}$$

We introduce the perturbations about equilibrium:

$$h = H + h' \quad h_L = H_L + h'_L \text{ and } h_N = H_N + Z'$$
$$q = Q + q'$$
$$Z = Z_0 + Z'$$

where $H(= H_L)$ and $Q(= Q_L)$ denote the stationary state.

The linearized equations are then:

$$h'_t + q'_x = 0 \tag{35}$$

$$q'_t + \frac{2Q}{H}q'_x - \frac{Q^2}{H}h'_x = -Hh'_x \tag{36}$$

which can be combined to give:

$$\left(\partial_t + \frac{Q}{H}\partial_x\right)^2 h' = Hh'_{xx} \tag{37}$$

We look for normal modes with $q' = \tilde{q}e^{-i\omega t}$ and $h' = \tilde{h}e^{-i\omega t}$, where ω is the dimensionless frequency:

$$\left(-i\omega + \frac{Q}{H}\partial_x\right)^2 \tilde{h} = H\tilde{h}_{xx} \tag{38}$$

The solutions take the spatial form, $e^{\lambda x}$, where

$$(-i\omega + \frac{Q}{H}\lambda)^2 - H\lambda^2 = 0 \tag{39}$$

$$\lambda_1 = \frac{i\omega}{\frac{Q}{H} - \sqrt{H}} \quad \text{and} \quad \lambda_2 = \frac{i\omega}{\frac{Q}{H} + \sqrt{H}} \tag{40}$$

Thus $\tilde{q} = A_1 e^{\lambda_1 x} + A_2 e^{\lambda_2 x}$ and $\tilde{h} = A_3 e^{\lambda_1 x} + A_4 e^{\lambda_2 x}$. Because the linear problem can be normalized arbitrarily, we set $A_1 = 1$. The boundary conditions q'=0 at x=0 (no perturbed flow rate) then demand $A_2 = -1$. From the continuity equation (35), we have $i\omega\tilde{h} = \tilde{q}_x$, which gives the coefficients,

$$A_3 = \frac{1}{\frac{Q}{H} + \sqrt{H}} \quad \text{and} \quad A_4 = -\frac{1}{\frac{Q}{H} - \sqrt{H}} \tag{41}$$

By introducing $\beta = \dfrac{Q}{H^{3/2}}$ and $\tilde{\Omega} = \dfrac{\Omega}{(1-\beta^2)}$, where $\Omega = \dfrac{\omega}{\sqrt{H}}$, $q'$and h' are finally written:

$$q' = e^{-i\tilde{\Omega}\beta x} 2i \sin(\tilde{\Omega}x) e^{-i\omega t} \tag{42}$$

$$h' = e^{-i\tilde{\Omega}\beta x} \frac{2\tilde{\Omega}}{\omega} \left[\cos(\tilde{\Omega}x) - \beta i \sin(\tilde{\Omega}x)\right] e^{-i\omega \hat{t}} \tag{43}$$

In the same way, we can linearize the equation of the paddle and the Bernoulli equation. Given $h_L = H_L + h'_L$ and $h_N = H_N + Z'$, we have $\alpha = \alpha_0 + \alpha'$ with $\alpha' = \dfrac{h'_L - \alpha_0 Z'}{H_N}$. Then the two linearized equations are:

$$-\omega^2 I Z' = \frac{4}{3}\frac{\mu}{2^{2/3}} Q^{1/3} q'_L F(\alpha_0) + \frac{\mu}{H_N}\frac{F'(\alpha_0)}{F(\alpha_0)}(h'_L - \alpha_0 Z') \tag{44}$$

$$Q q'_L - h'_L H_N^2 \frac{\alpha_0(\alpha_0+2)}{(\alpha_0+1)^2} = \frac{H_N^2 \alpha_0^2 Z'(2\alpha_0+1)}{(\alpha_0+1)^2} \tag{45}$$

where $F'(\alpha_0) = \dfrac{dF}{d\alpha}(\alpha_0)$.

At $x = L$ (with $\xi = l$), the flow rate and the height in the reservoir must be continuous to those in the paddle area (q_L, h_L), so the matching conditions are

$$q_L = q(x = 1 + R\ l(t), t) \tag{46}$$

Using the approximation that R is small, we find

$$q_L \approx q(x = 1, t) \tag{47}$$

In the same way, we have for the perturbed variables:

$$q'_L = q'(x = 1 + R\ l(t), t) \approx q'(x = 1, t) \tag{48}$$

and also $h'_L \approx h'(x = 1)$

Finally, the Bernoulli equation in combination with the equation of the paddle, plus the matching conditions, furnishes an equation of the form, $D(\omega; \alpha_0, I, \mu, Q, H) = 0$, which allows us to determine ω, and where

$$\begin{aligned} D(\omega) = \cos\left(\frac{\Omega}{1-\beta^2}\right)\left[3(1+\alpha_0)F'(\alpha_0) - (2+\alpha_0)\frac{\Omega^2 I}{\mu^2}\left(\frac{\alpha_0+1}{\alpha_0^2}\right)^{2/3}\right] \\ - i\sin\left(\frac{\Omega}{1-\beta^2}\right)\left[\sqrt{2\alpha_0}(1+\alpha_0)^{3/2}(1-\beta^2)\left(F'(\alpha_0) - \frac{2}{3}\frac{2\alpha_0+1}{\alpha_0+1}\frac{F'(\alpha_0)}{\alpha_0\mu} - \frac{\Omega^2 I}{\mu^2}\left(\frac{\alpha_0+1}{\alpha_0^2}\right)^{2/3}\right)\right. \\ \left. + 3\left(\frac{2(1+\alpha_0)}{\alpha_0}\right)^{1/2} F'(\alpha_0) - (\alpha_0+2)\frac{\Omega^2 I}{\mu^2}2^{1/2}\frac{(\alpha_0+1)^{1/6}}{\alpha_0^{11/6}}\right] \end{aligned} \tag{49}$$

The solutions $\omega = \omega_{\rm r} + i\ \omega_{\rm i}$ can be found numerically. In our convention, modes grow when $\omega_{\rm i} > 0$.

3.6 Approximation of small flow rate: $Q \ll 1$ or $\alpha_0 \gg 1$

If the flow rate is small, i.e. $\alpha_0 \gg 1$, then $F(\alpha_0) \sim \alpha_0^{4/3}$ and by keeping the leading order in α_0 in $D(\omega) = 0$, we find $\Omega \approx n\pi$. i.e. in dimensional variables:

$$\omega = n\pi\frac{\sqrt{gH_L}}{X} \tag{50}$$

These are the frequencies of the seiche modes of the reservoir in the absence of the fluid flow.

Proceeding further, and setting $\Omega = n\pi + \gamma$, we find

$$\gamma = \frac{i(1 - \frac{n^2\pi^2 I}{4\alpha_0\mu})}{\frac{1}{3}\alpha_0\sqrt{2}(-1 + \frac{1}{\mu} + \frac{3n^2\pi^2 I}{4\mu^2\alpha_0})} \tag{51}$$

Thus, since the right-hand side of this relation can become positive, the seiche modes can be destablized by the flow. Figure 5 shows the numerical results calculated with experimental parameters and highlights the growth of the instability for the mode $n = 1$.

Thus, the model predicts the existence of flow-destabilized seiches modes. We now follow on and present experimental observations of this phenomenon.

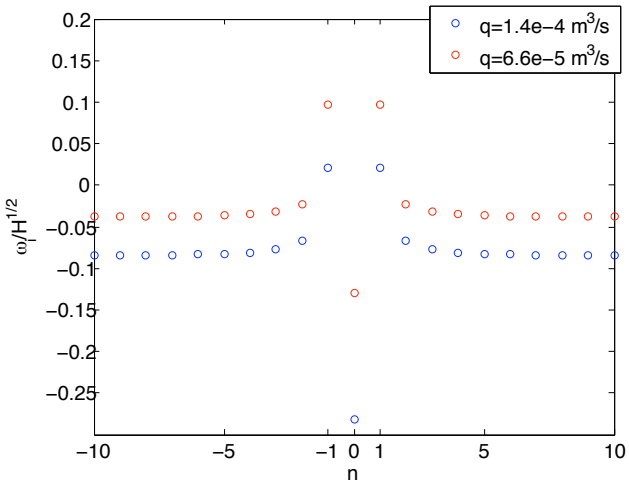
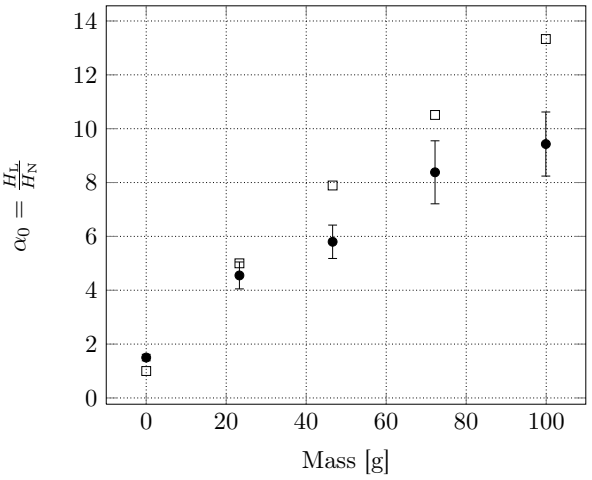


Figure 5: Numerical results: $m_{\text{added}} = 3.1$ g, $x = 5$ cm, $X = 18$ cm, $\theta = 60°$, $H_L \approx 2$ cm.

4 Experimental results

4.1 Steady state

To begin, we explored steady flow states and made measurements of the water depth under the paddle for different masses added on the paddle. These measurements could then be compared with the theoretical ratio α_0 given by the inversion of equation (31). The results, shown in figure 6 (left), reveal agreement between experiment and theory for small masses, but an increasing discrepancy for masses bigger than around 40 g.



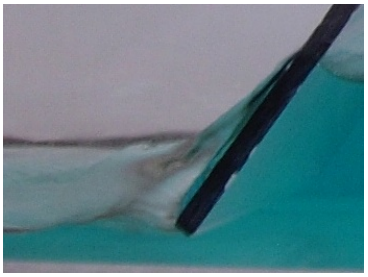


Figure 6: Left: Comparison between experimental points (dots) and numerical calculation using the inversion of $F(\alpha_0) = \frac{2^{2/3}}{Q_L^{4/3}\mu}$ (squares). The angle was fixed at $\theta = 30°$ and $q = 1.6 \times 10^{-4}$ m^3/s. Right: Flow around the paddle instead of underneath in cases of large mass added.

We explain this discrepancy with an observed change of structure in the flow depending

on the value of the mass added on the paddle: for small mass, the fluid flows largely underneath the paddle (as supposed in our model), whereas for large mass, much fluid is also diverted around the sides of the paddle, as illustrated in figure 6 (right).

4.2 Experimental unstable modes

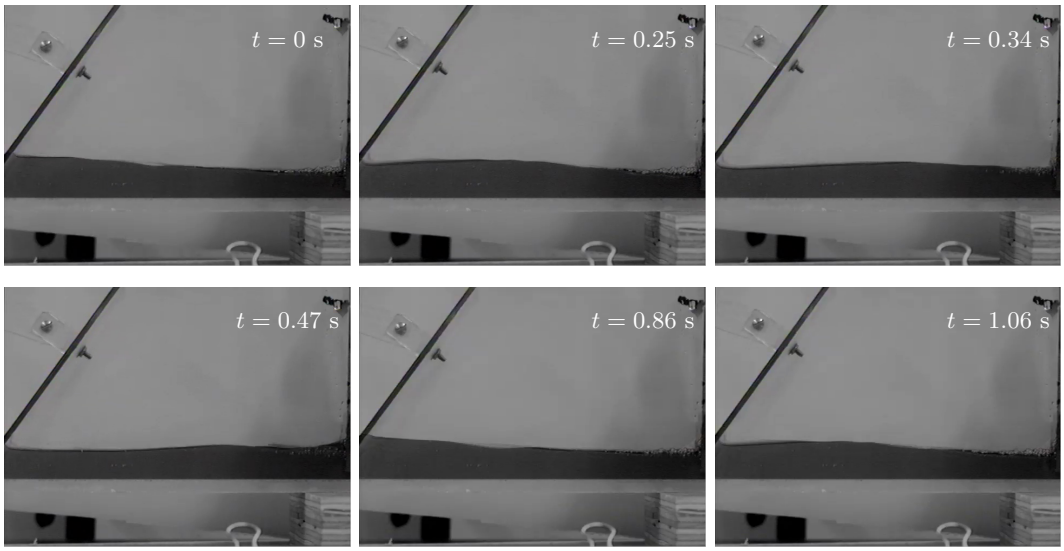


Figure 7: Series of images showing the presence of a flow-destabilized seiche mode n=1 during a period. Experiment with $m_{\text{added}} = 3.1$ g at 5 cm of the pivot.

Figure 7 shows images from an experiment in which a flow-destabilized sloshing mode appeared. From such images, the water level could be extracted by applying a threshold to the intensity of the picture. A time series of the water surface measured in this way is shown in figure 8.

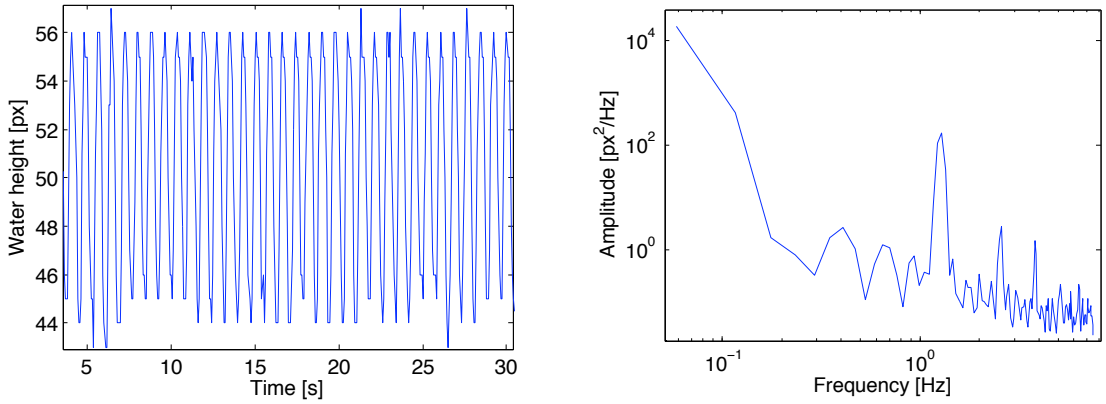


Figure 8: Time series and spectrum for the same experiment as in figure 7.

A comparison between the experiment and the linear theory was made by superimposing the interface observed experimentally with theoretical results using (52):

$$h'(x,t) = e^{-i(\omega t - \tilde{\Omega}\beta x)} \ \frac{2\tilde{\Omega}}{\omega} cos(\tilde{\Omega}x) \quad \text{for } n = 1 \tag{52}$$

On figure 9, the red line corresponds to the theory, with the variables Ω, β and ω calculated from the experimental parameters, except the amplitude and the phase, which were fit using a least squares algorithm.

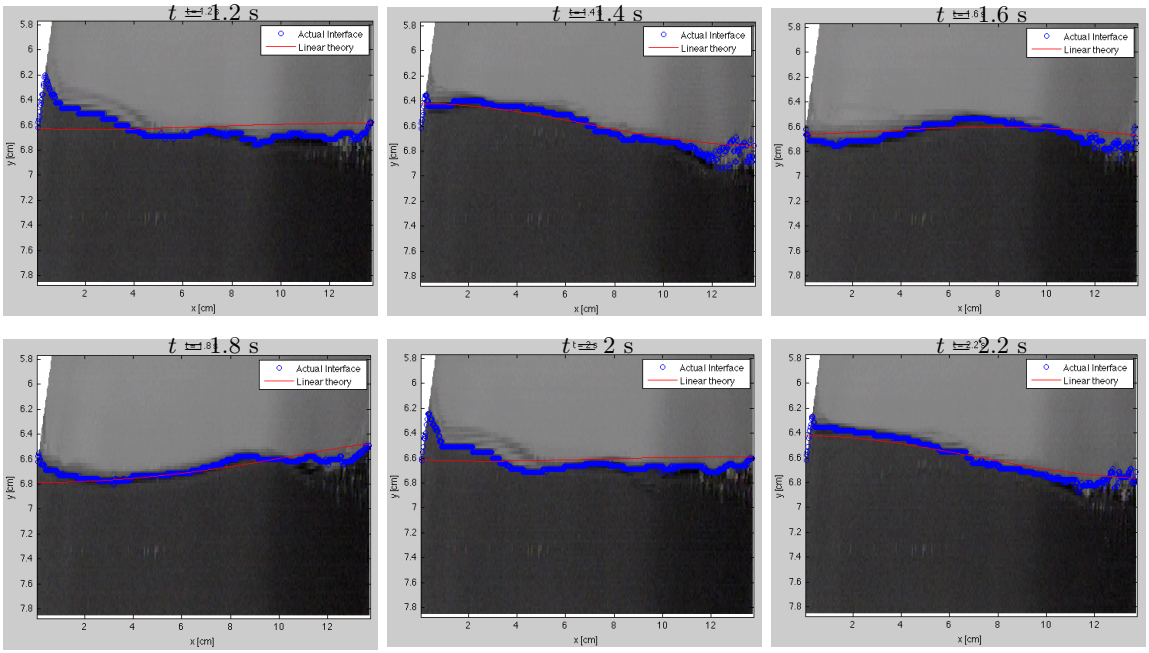


Figure 9: Comparison of the experiment and the linear theory.

Even though the theoretical predictions do not follow exactly the actual interface, especially near the paddle and the wall at the right (where some new elements of physics, like surface tension, should probably be added for a complete description), there is a fair agreement overall, since the main features of the sloshing mode $n = 1$ are well captured.

Higher-order modes with $n = 2$ and 3 were also observed in the experiments. However, it appears that the $n = 1$ mode was usually dominant, probably because the other modes are more strongly damped due to fluid viscosity and friction in the pivot.

4.3 Variations of parameters

After demonstrating the existence of unstable modes, we attempted to further investigate the characteristics of the system by changing control parameters. Figure 10 presents the frequency of the paddle oscillations for masses placed at different positions on the rod (left) and varying flow rate (right).

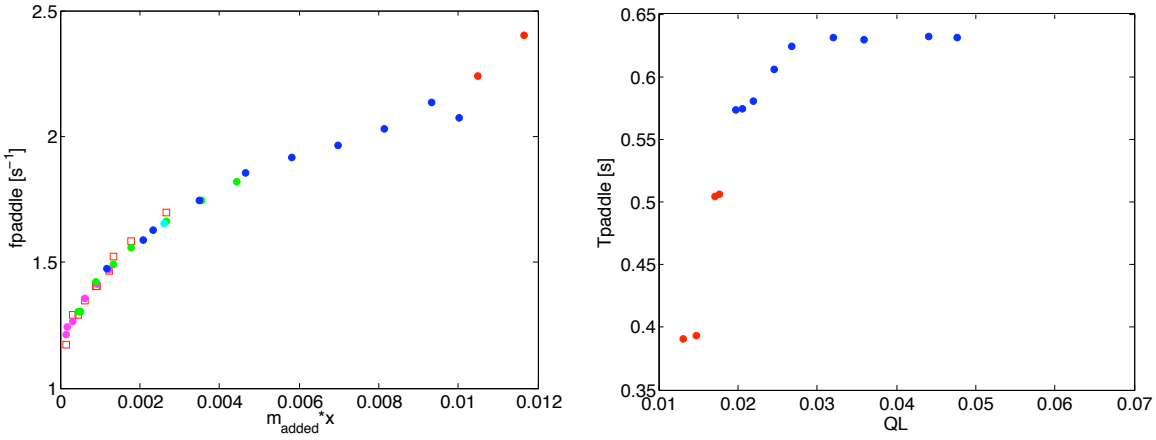


Figure 10:
Left: Variation of m_{added} at q fixed. ($q = 1.42 \times 10^{-4}$ m^3/s). Each colour represents one series of experiments realized with one particular m_{added} translated at different positions on the rod. The frequency of the paddle motion has been measured with a stopwatch (looking at 50 oscillations). For checking, the square points are the frequency of the paddle determined using the spectrum of the time series of the water level (like in figure 8) for some experiments.
Right: Variation of q at m_{added} fixed. ($m_{added} = 8.9$ g and $x = 25$ cm). See also section 4.4.2.

4.4 Evidence of non-linearities

4.4.1 Beating

In certains experiments, the water level was characterized by a beating motion, as illustrated in figure 11. Though beating is not itself a non-linear phenomenon, the spectrum reveals a possible resonant triad with the superposition of two frequencies (the main frequency and another one smaller) which creates the two additional peaks around the main peak and its harmonics.

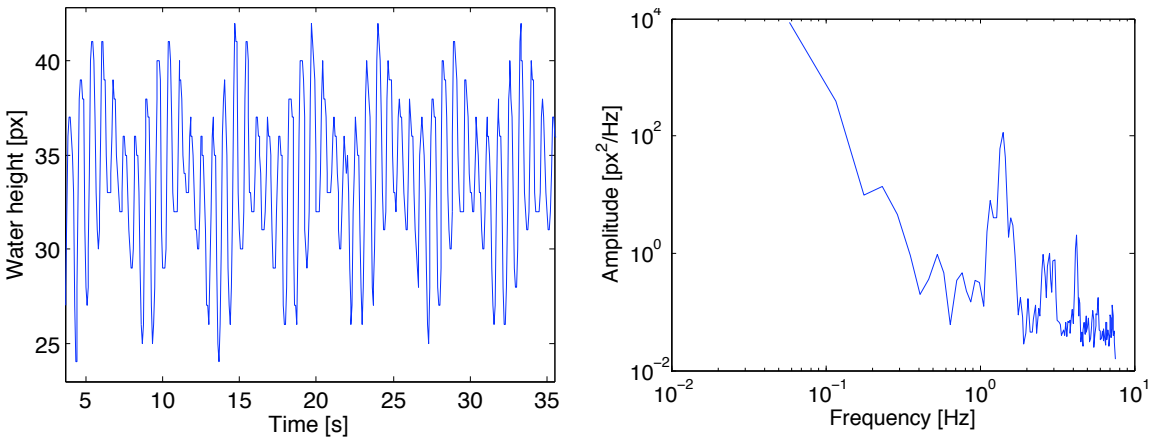


Figure 11: Time series and spectrum for $m_{\text{added}} = 3.1$ g at 30 cm of the pivot.

4.4.2 Bistability

Another observed feature typical of a non-linear behaviour is bistability: in some experiments, the unstable behaviour could be stopped by gently bringing the paddle to rest with one's hand; after being released the paddle remained stationary and no instability grew. At the same time, by giving a little kick to the paddle, the flow-destabilized sloshing could be resumed. Thus, for the same set of parameters, the system could be both in a steady or sloshing regime. On the other hand for other ranges of parameters, the system was only unstable (see figure 12).

The existence of bistability suggests that the system could follow a subcritical bifurcation, as sketched in figure 12 (right). The limit between the bistable regime and unstable regime can then be considered as the linear threshold of the instability.

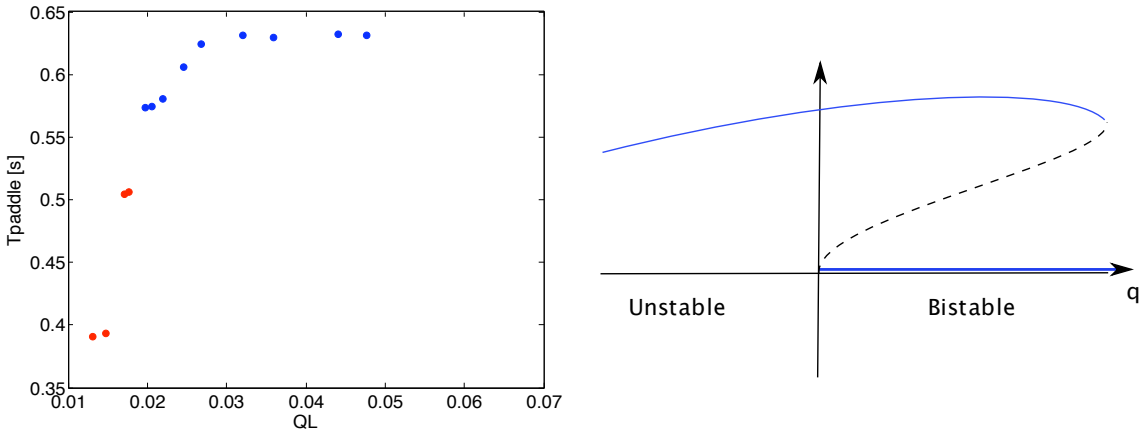


Figure 12: Presence of bistability: ●: unstable points. ●: bistable points. Right: Subcritical bifurcation.

5 Conclusion

The phenomenon of flow-destabilized sloshing modes has been demonstrated experimentally with a setup analogous to a water clarinet'. The linear theory shows some agreement with the experimental unstable modes.

However, several limits of the model can be pointed out. It does not take into account water flow around the sides of the paddle. Friction in the hinge has been neglected, together with fluid viscosity and surface tension. The experiment also sometimes suggests a possible collision of the paddle with the bottom, an interaction not included in the model. We can also question the validity of the shallow water hypothesis and the Bernoulli approximation under the paddle for the experimental conditions.

Finally, experiments reveal that work must be done to go beyond linear theory in order to capture the observed non-linear dynamics like resonant triad and bistability.

6 Acknowlegments

I would like to thank my supervisor Neil Balmforth for all his help and ideas, together with Bill Young for his help in the first calculations and Andrew Belmonte for his advice in the experimental setup. I am also very grateful to Roger Grimshaw and Harvey Segur for their very good lectures on non-linear waves. Finally, I thank all the fellows and the staff for the pleasant atmosphere during the whole summer at Walsh Cottage.

References

[1] J. Backus, *Small-vibration theory of the clarinet*, The Journal of the Acoustical Society of America, 35 (1963), pp. 305–313.

[2] B. J. Binder and J.-M. Vanden-Broeck, *Free surface flows past surfboards and sluice gates*, European Journal of Applied Mathematics, 16 (2005), pp. 601–619.

[3] J.-P. Dalmont, J. Gilbert, J. Kergomard, and S. Ollivier, *An analytical prediction of the oscillation and extinction thresholds of a clarinet*, The Journal of the Acoustical Society of America, 118 (2005), pp. 3294–3305.

[4] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, *On the oscillations of musical instruments*, The Journal of the Acoustical Society of America, 74 (1983), pp. 1325–1345.

[5] A. C. Rust, N. Balmforth, and S.Mandre, *The feasibility of generating low-frequency volcano seismicity by flow through a deformable channel*, Geological Society, London, Special Publications, 307 (2008), pp. 45–56.

Nonlinear Shelf Waves in a Rotating Annulus

Andrew Stewart

October 13, 2009

1 Introduction

Shelf breaks are a ubiquitous and dynamically important feature in the topography of the coastal oceans. A region of particular interest is the southeastern coast of Africa, where the Agulhas current flows southwest at speeds of up to 1 ms^{-1} over a shelf break [2]. This has motivated a series of dynamical investigations. Gill and Schumann [6] studied the influence of topography on a coastal jet with specific application to the Agulhas current, whilst Mysak and others have considered the phenomenon of coastally-trapped waves [19, 20, 21].

Where there is a very sharp drop from the coastal shelf to the ocean floor, the topography may be approximated by a discontinuity in depth, an approach that is appealing due to its analytical simplicity. This approach was first used to study coastally-trapped waves by Longuet-Higgins [17], and was subsequently applied to coastal currents by Johnson [10, 13].

The present work concerns a model for nonlinear Rossby shelf waves propagating along a discontinuity in depth that was first developed by Haynes et al [7], who derived a fully nonlinear wave equation that is valid in the limit of infinitely long waves. This model, which will be discussed further in Section 3, was subsequently extended to include higher-order dispersive terms [3, 11]. In the weakly nonlinear limit, the model reduces to a Korteweg-de Vries (KdV)-type equation [4], which will be discussed in Section 4. These models, and the broader subject of Rossby wave hydraulics, were reviewed by Johnson & Clarke [12].

The theory of nonlinear Rossby shelf waves hinges upon the assumption of columnar motion in the fluid, even when it crosses the step, which has motivated several studies [9, 14, 15] of the interaction of vortices with step topography. In a recent experiment [8], dipoles fired over a step were shown largely to maintain their columnar structure, albeit with some additional diffusion of vorticity. This suggests that fluid columns in a shallow rotating flow will be preserved across a discontinuity in depth.

The goal of this project is to generate nonlinear Rossby shelf waves experimentally and compare their properties with theoretical predictions. Whilst the previous theory is based on an infinitely long straight channel, we have constructed our experiment in an annulus, and we will show that analogous theoretical results hold in this geometry. We will first introduce the set-up of the laboratory experiment and reformulate the finite-amplitude and weakly nonlinear wave equations for the annular domain. We will then discuss the inclusion of bottom friction in the model and its effect on the waves, and analyse some instabilities that were found to arise in the course of the experiment. Finally, we will present some experimental results for breaking lee waves and compare them with numerical simulations based on the finite-amplitude wave equation.

2 Experimental Set-Up

The theoretical analysis presented in this report has been motivated by the laboratory experiment that served as the focus of this project. In this section we will only provide an overview of the experiment, postponing a discussion of the results to Sections 6 and 7.

The experiment was designed to replicate the theoretical investigations of [7], who studied the behaviour of waves propagating about a discontinuity in depth in a straight channel with a rigid lid in a rotating frame. The most practical way to achieve this in the laboratory was to create an annular channel in a cylindrical tank on a rotating table. A shelf of width 10.5 cm and height 5.0 cm was built around the outer edge of the tank, which measured approximately 2.13 m in diameter, to create a discontinuity in depth, and a weighted sheet of flexible plastic was used to create the inner wall of the annulus. A photograph of this apparatus is shown in Figure 1.

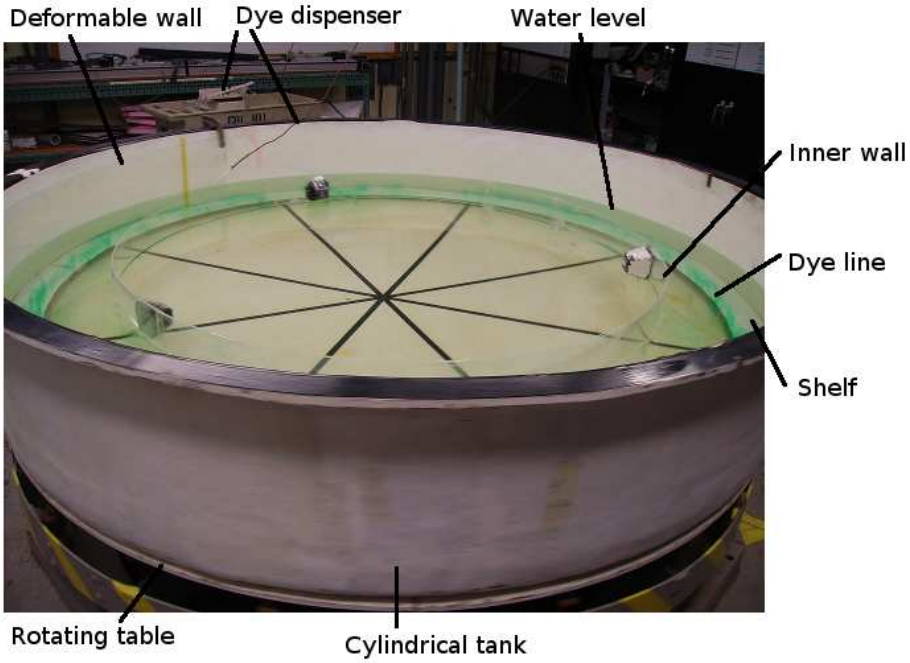


Figure 1: A photograph of the experimental set-up.

To perform the experiments, we filled the tank with water to a prescribed depth, typically 10 – 25 cm above the base of the tank, and then accelerated it to a constant angular velocity. We then left the tank to rotate until the fluid inside had reached a stationary state in the rotating frame of reference. We perturbed the fluid from this state either by the moving a deformable wall attached to the edge of the tank, or by creating a bump in the outer wall and rapidly changing the rate of rotation of the tank. To visualise the flow, we dispensed a line of dye into the fluid directly above the shelf line (along the potential vorticity interface) and recorded the results with a digital camera that positioned directly above the centre of the tank and made to rotate at the same speed as the tank.

This experiment does not perfectly replicate the theoretical conditions of [7], most obviously because the channel is curved rather than straight. We have therefore modified and extended the previous theory accordingly, and this will be covered in the following sections. Note also that the inner wall is not sealed against the bottom of the tank, thereby permitting a small flux of fluid between the outer annulus and the central part of the tank, and that the fluid has a free surface that deforms under the influence of the centrifugal force, rather than a rigid lid. However, both of these effects were found to make negligible contributions to the behaviour of the fluid.

Another theoretical inconsistency arises from the fact that the ratio of the depth of the fluid to the width of the channel is typically close to 1, casting doubt upon the validity of the shallow water theory. In practice the rate of rotation of the tank was sufficiently high as to ensure approximately columnar motion anyway, so the shallow water theory may be expected to hold quite well. In fact, the most important deviation from the theory was the absolute discontinuity in the depth at the shelf line, which in theory is an approximation of a steep slope. In later experiments a slope of width 2.5 cm was added to the end of the shelf, substantially altering the behaviour of the experiment.

3 Topographic Rossby Waves in a Rotating Annulus

We shall now reformulate the theory of [3] for application to the annular domain of the laboratory experiment. The fluid lies in the region $R_w(\theta) \leq r \leq R_c(\theta)$, $0 \leq \theta \leq 2\pi$, $h(r) \leq z \leq H$, where h is the height of the bottom topography,

$$h = \begin{cases} 0, & R_w < r < R_h \\ H_s, & R_h < r < R_c \end{cases}, \tag{1}$$

and $r = R_h(\theta)$ is the equation of the shelf line. Here we have applied the rigid lid approximation, assuming that perturbations to the upper surface of the fluid are much smaller than the average depth H. A diagram of this set-up is presented in Figure 2.

We assume that the flow is sufficiently shallow ($H \ll R_c - R_w$), and has sufficiently small topography ($H_s \ll H$), that it is governed approximately by the shallow water quasigeostrophic equation for fluid under a rigid lid,

$$\frac{\mathrm{D}q}{\mathrm{D}t} = 0, \quad q = \nabla^2\Psi + \frac{fh}{H}, \tag{2}$$

where f is the Coriolis parameter and q is the quasigeostrophic potential vorticity. Here $\Psi = p_s/f\rho$ is the quasigeostrophic streamfunction, where p_s is the pressure at the rigid lid and ρ is the density of the fluid. With this definition, Ψ satisfies

$$u = -\frac{1}{r}\frac{\partial\Psi}{\partial\theta}, \quad v = \frac{\partial\Psi}{\partial r}, \tag{3}$$

and the velocity field is $\mathbf{u} = u\mathbf{e_r} + v\mathbf{e}_\theta$, such that $\omega = (\nabla \times \mathbf{u}) \cdot \hat{\mathbf{z}} = \nabla^2\Psi$.

We prescribe a mean flow $\mathbf{u}_m = -\nabla \times \psi_m(r,\theta)\hat{\mathbf{z}}$ in the channel, modelling the presence of an along-shore current in the real ocean. We require that the flow should be unidirectional,

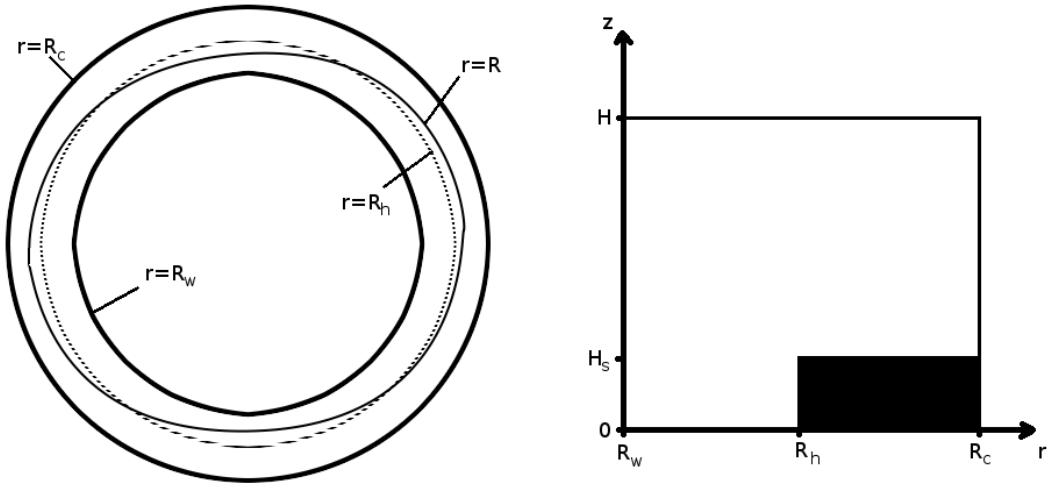


Figure 2: Top-down (left) and side-on (right) views of the physical set-up for the model.

that the walls $r = R_w$ and $r = R_c$ should be streamlines, and that the volume flux should be constant around the annulus,

$$\frac{\partial}{\partial\theta}\int_{R_w(\theta)}^{R_c(\theta)} v_m \mathrm{d}r = 0. \tag{4}$$

Here we neglect the contribution to the volume from the variation of the bottom topography in order to be consistent with the quasigeostrophic approximation. We also require that the vorticity of the mean flow should be materially conserved. These requirements may be met in a radially symmetric annulus, but when R_w and R_c vary with θ we must allow for bidirectional flow to find a steady solution. We must therefore prescribe a mean flow streamfunction that approximately satisfies these conditions,

$$\psi_m = \tfrac{1}{2}\overline{\Omega_m}\frac{\left(\overline{R_c}^2 - \overline{R_w}^2\right)\left(r^2 - R_w^2\right)}{R_c^2 - R_w^2}, \tag{5}$$

where an overbar $\overline{\ }$ denotes the value of a function at some fixed $\theta = \theta_m$. This mean flow satisfies all requirements except that its vorticity should be materially conserved, which is only approximately satisfied under the assumption that the walls of the annulus vary very slowly with θ,

$$\left|\frac{\partial R_w}{\partial\theta}\right|, \left|\frac{\partial R_c}{\partial\theta}\right| \ll 1 \implies \frac{\mathrm{D}\omega}{\mathrm{D}t} \approx 0. \tag{6}$$

In the case that R_w and R_c do not vary with θ, (5) reduces to a flow of constant angular velocity and exactly satisfies the above requirements.

It is now convenient to write the overall streamfunction as

$$\Psi = \psi_m + \psi(r, \theta, t), \tag{7}$$

where the first term corresponds to the mean flow and the second to any deviation. On applying (6), the potential vorticity equation (2) then becomes

$$\frac{\mathrm{D}q'}{\mathrm{D}t} \approx 0, \quad q' = \nabla^2\psi + \frac{fh}{H}, \tag{8}$$

Suppose that the fluid is perturbed from a basic state in which the only motion is that of the mean flow, such that $\nabla^2\psi = 0$ and

$$q' = \begin{cases} 0, & R_w < r < R_h \\ Q, & R_h < r < R_c \end{cases}. \tag{9}$$

We denote the position of the interface between these regions of differing potential vorticity as $r = R(\theta, t)$, assuming that this interface remains single-valued in θ. If in some region $R > R_h$ then, by conservation of potential vorticity, we have

$$\begin{aligned} q' = \nabla^2\psi + 0 = 0 \quad &\implies \nabla^2\psi = 0 \quad && \text{in } R_w < r < R_h, \\ q' = \nabla^2\psi + Q = 0 \quad &\implies \nabla^2\psi = -Q \quad && \text{in } R_h < r < R, \\ q' = \nabla^2\psi + Q = Q \quad &\implies \nabla^2\psi = 0 \quad && \text{in } R < r < R_c, \end{aligned}$$

and if $R < R_h$ then

$$\begin{aligned} q' = \nabla^2\psi + 0 = 0 \quad &\implies \nabla^2\psi = 0 \quad && \text{in } R_w < r < R, \\ q' = \nabla^2\psi + 0 = Q \quad &\implies \nabla^2\psi = Q \quad && \text{in } R < r < R_h, \\ q' = \nabla^2\psi + Q = Q \quad &\implies \nabla^2\psi = 0 \quad && \text{in } R_h < r < R_c, \end{aligned}$$

which may be expressed succinctly as the following Poisson equation for the streamfunction,

$$\nabla^2\psi = Q\left(\mathcal{H}(r - R) - \mathcal{H}(r - R_h)\right), \tag{10}$$

where $\mathcal{H}$ denotes the Heaviside step function. We also require that the streamfunction vanishes on the boundaries of the annulus,

$$\psi = 0 \text{ on } r = R_w \text{ and } r = R_c. \tag{11}$$

Finally, we require that particles on the interface $r = R$ remain on the interface, i.e. $D/Dt(r - R) = 0$, which leads to the following condition,

$$\frac{\partial R}{\partial t} = -\frac{1}{R}\frac{\partial}{\partial\theta}\Psi(R(\theta, t), \theta, t). \tag{12}$$

We now define a length scale $L = \overline{R_c} - \overline{R_w}$ and nondimensionalise $r = L\hat{r}$, $t = Q^{-1}\hat{t}$ and $\psi = QL^2\hat{\psi}$, where a hat $\hat{}$ denotes a dimensionless variable. This yields the following system of dimensionless equations,

$$\frac{\partial^2\hat{\psi}}{\partial\hat{r}^2} + \frac{1}{\hat{r}}\frac{\partial\hat{\psi}}{\partial\hat{r}} + \frac{1}{\hat{r}^2}\frac{\partial^2\hat{\psi}}{\partial\theta^2} = \mathcal{H}\left(\hat{r} - \hat{R}\right) - \mathcal{H}\left(\hat{r} - R_H\right), \tag{13a}$$

$$\hat{\psi} = 0 \text{ on } \hat{r} = R_I \text{ and } \hat{r} = R_O, \tag{13b}$$

$$\frac{\partial\hat{R}}{\partial\hat{t}} = -\frac{1}{\hat{R}}\frac{\partial}{\partial\theta}\hat{\Psi}\left(\hat{R}\left(\theta, \hat{t}\right), \theta, \hat{t}\right), \tag{13c}$$

where $R_H = R_h/L$, $R_I = R_w/L$ and $R_O = R_c/L$ are the dimensionless radii of the annulus. We will henceforth drop the hat notation for dimensionless variables. We seek long wave solutions by prescribing that the width of the channel should be small compared to the circumference of the shelf line and assuming that the solution varies slowly with θ and t,

$$\theta = \mu^{-1/2}\phi, \quad t = \mu^{-1/2}\tau, \tag{14}$$

where $\mu = (L/2\pi\overline{R_h})^2 \ll 1$. Equations (13a)–(13c) then take the form,

$$\frac{\partial^2\psi}{\partial r^2} + \frac{1}{r}\frac{\partial\psi}{\partial r} + \frac{\mu}{r^2}\frac{\partial^2\psi}{\partial\phi^2} = \mathcal{H}(r-R) - \mathcal{H}(r-R_H), \tag{15a}$$

$$\psi = 0 \text{ on } r = R_I \text{ and } r = R_O, \tag{15b}$$

$$\frac{\partial R}{\partial\tau} = -\frac{1}{R}\frac{\partial}{\partial\phi}\Psi(R(\phi,\tau),\phi,\tau). \tag{15c}$$

We now expand ψ asymptotically in the small parameter μ,

$$\psi = \psi^{(0)} + \mu\psi^{(1)} + \mu^2\psi^{(2)} + \dots. \tag{16}$$

Substituting this in to (15a) we find that the leading order equation is

$$\frac{\partial^2\psi^{(0)}}{\partial r^2} + \frac{1}{r}\frac{\partial\psi^{(0)}}{\partial r} = \mathcal{H}(r-R) - \mathcal{H}(r-R_H), \tag{17}$$

which may be solved subject to the boundary conditions (15b) to give

$$\begin{aligned}\psi^{(0)} = {} & \tfrac{1}{4}\left(r^2 - R^2\right)\mathcal{H}(r-R) - \tfrac{1}{4}\left(r^2 - R_H^2\right)\mathcal{H}(r-R_H) \\ & + \tfrac{1}{2}R_H^2\mathcal{H}(r-R_H)\ln(r/R_H) - \tfrac{1}{2}R^2\mathcal{H}(r-R)\ln(r/R) \\ & + \frac{\ln(r/R_I)}{\ln(R_O/R_I)}\left[\tfrac{1}{2}R^2\ln(R_O/R) - \tfrac{1}{2}R_H^2\ln(R_O/R_H) + \tfrac{1}{4}\left(R^2 - R_H^2\right)\right].\end{aligned} \tag{18}$$

We may substitute this leading-order streamfunction in to (15c), but we then obtain a non-dispersive evolution equation for R. This may be remedied by continuing the expansion of (15a) to $\mathrm{O}(\mu)$,

$$\frac{\partial^2\psi^{(1)}}{\partial r^2} + \frac{1}{r}\frac{\partial\psi^{(1)}}{\partial r} + \frac{1}{r^2}\frac{\partial^2\psi^{(0)}}{\partial\phi^2} = 0, \tag{19}$$

which we solve to find the following expression for $\psi^{(1)}$,

$$\begin{aligned}\psi^{(1)} = {} & \tfrac{1}{6}\frac{\partial}{\partial\phi}\left\{RR_\phi\ln^3(r/R)\mathcal{H}(r-R) - R_HR_{H\phi}\ln^3(r/R_H)\mathcal{H}(r-R_H)\right\} \\ & + \frac{\ln(r/R_I)}{\ln(R_O/R_I)}\left\{\tfrac{1}{6}\left(R_HR_{H\phi}\right)_\phi\ln^3(R_O/R_H) - \tfrac{1}{2}R_{H\phi}^2\ln^2(R_O/R_H)\right. \\ & \qquad \left. - \tfrac{1}{6}\left(RR_\phi\right)_\phi\ln^3(R_O/R) + \tfrac{1}{2}R_\phi^2\ln^2(R_O/R)\right\} \\ & + \ln(R_O/r)\ln(r/R_I)\left\{\tfrac{1}{6}\ln(R_Or/R_I^2)\frac{\partial^2}{\partial\phi^2} - \frac{R_{I\phi}}{R_I}\frac{\partial}{\partial\phi} + \frac{R_{I\phi}^2 - R_IR_{I\phi\phi}}{2R_I^2}\right\}G,\end{aligned} \tag{20}$$

where

$$G(\phi,\tau)=\frac{\frac{1}{2}R^2\ln(R_O/R)-\frac{1}{2}R_H^2\ln(R_O/R_H)+\frac{1}{4}(R^2-R_H^2)}{\ln(R_O/R_I)}. \tag{21}$$

Substituting (18) and (20) into (15c) yields the following evolution equation for finite-amplitude waves on the interface,

$$\begin{aligned}\frac{\partial R}{\partial\tau}=-\frac{1}{R}\frac{\partial}{\partial\phi}\Bigg\{&\tfrac{1}{2}\alpha_m\frac{\left(\overline{R_O}^2-\overline{R_I}^2\right)\left(R^2-R_I^2\right)}{R_O^2-R_I^2}\\
&+\ \mathcal{H}(R-R_H)\left[\tfrac{1}{2}R_H^2\ln(R/R_H)+\tfrac{1}{4}(R_H^2-R^2)\right]+\ln(R/R_I)\,G\\
&+\ \mu\Bigg\{\mathcal{H}(R-R_H)\left[\tfrac{1}{2}R_{H\phi}^2\ln^2(R/R_H)-\tfrac{1}{6}\left(R_HR_{H\phi}\right)_\phi\ln^3(R/R_H)\right]\\
&+\ \frac{\ln(R/R_I)}{\ln(R_O/R_I)}\Big[\tfrac{1}{6}\left(R_HR_{H\phi}\right)_\phi\ln^3(R_O/R_H)-\tfrac{1}{2}R_{H\phi}^2\ln^2(R_O/R_H)\\
&\qquad\qquad-\tfrac{1}{6}\left(RR_\phi\right)_\phi\ln^3(R_O/R)+\tfrac{1}{2}R_\phi^2\ln^2(R_O/R)\Big]\\
&+\ \ln(R_O/R)\ln(R/R_I)\left[\tfrac{1}{6}\ln(R_OR/R_I^2)\frac{\partial^2}{\partial\phi^2}-\frac{R_{I\phi}}{R_I}\frac{\partial}{\partial\phi}+\frac{R_{I\phi}^2-R_IR_{I\phi\phi}}{2R_I^2}\right]G\Bigg\}\Bigg\},\end{aligned} \tag{22}$$

where $\alpha_m=\overline{\Omega_m}/Q$. We may obtain higher-order approximations to the streamfunction, and thereby to the evolution equation (22), by solving further equations with the same form as (19), but the expressions involved quickly become unmanagable.

It is important to note that whilst we can account for azimuthally varying R_w, R_h and R_c, we are unable to apply this theory in the limits as (a) $R_w\to 0$ or (b) $R_c\to\infty$ uniformly. We can still apply a long-wave scaling in these cases by redefining $L=\overline{R_c}-\overline{R_h}$ or $L=\overline{R_h}-\overline{R_w}$ respectively. In case (a) the asymptotic expansion breaks down when $r=\mathrm{O}\left(\mu^{1/2}\right)$ because the derivative with respect to ϕ in (15a) is then $\mathrm{O}\,(1)$. In case (b) we require that $|\mathbf{u}|\to 0$ as $r\to\infty$ so that the total energy remains finite, but this condition permits an infinite family of functions $\psi^{(0)}$ that satisfy the leading order Poisson equation (17), and so we require a stronger condition on ψ as $r\to\infty$ in order to determine ψ uniquely. We are therefore restricted to cases where R_I is $\mathrm{O}\,(1)$ and R_O is finite.

4 Weakly Nonlinear Waves

The evolution equation (22) descibes the fully nonlinear behaviour of our system, but it is somewhat too complicated to analyse directly. We therefore first examine the dynamics of weakly nonlinear waves by assuming that the wave amplitude is $\mathrm{O}\,(\mu)$ and that the variations in R_I, R_H and R_O are $\mathrm{O}\left(\mu^2\right)$. These scalings are chosen such that nonlinear and dispersive terms are of the same order, and such that only the leading-order effects of a radially asymmetric annulus are included. We set $R=\overline{R_H}+\mu\hat{A}$, $R_I=\overline{R_I}+\mu^2\hat{w}(\phi)$, $R_H=\overline{R_H}+\mu^2\hat{h}(\phi)$ and $R_O=\overline{R_O}+\mu^2\hat{c}(\phi)$ in (22), and retain terms up to $\mathrm{O}\left(\mu^2\right)$. Note

that the contributions from $\psi^{(2)}$ and higher-order terms will be at least $\mathrm{O}\left(\mu^3\right)$ due to the ϕ derivative on the right hand side of (22). We obtain the following weakly nonlinear wave equation,

$$\hat{A}_\tau + (\alpha+\alpha_m)\hat{A}_\phi + \mu\beta\hat{A}_{\phi\phi\phi} - \frac{\mu}{2\overline{R_H}}\left(\mathrm{sign}\left(\hat{A}\right)+\gamma\right)\hat{A}\hat{A}_\phi = -\mu\alpha\hat{h}_\phi + \mu\alpha_m\left(\delta_c\hat{c}_\phi + \delta_w\hat{w}_\phi\right), \tag{23}$$

where

$$\alpha = \frac{\ln\left(\overline{R_H}/\overline{R_I}\right)\ln\left(\overline{R_O}/\overline{R_H}\right)}{\ln\left(\overline{R_O}/\overline{R_I}\right)},\ \beta = \frac{\ln^2\left(\overline{R_H}/\overline{R_I}\right)\ln^2\left(\overline{R_O}/\overline{R_H}\right)}{3\ln\left(\overline{R_O}/\overline{R_I}\right)},\ \gamma = 3\frac{\ln\left(\overline{R_H}^2/\overline{R_O R_I}\right)}{\ln\left(\overline{R_O}/\overline{R_I}\right)},$$
$$\delta_c = \frac{\overline{R_O}\left(\overline{R_H}^2-\overline{R_I}^2\right)}{\overline{R_H}\left(\overline{R_O}^2-\overline{R_I}^2\right)},\ \delta_w = \frac{\overline{R_I}\left(\overline{R_O}^2-\overline{R_H}^2\right)}{\overline{R_H}\left(\overline{R_O}^2-\overline{R_I}^2\right)}. \tag{24}$$

The terms on the left hand side of (23) resemble those of the KdV equation, whilst those on the right are all contributions from variations of the annulus' walls and shelf line. The most significant contribution comes from variations of the shelf line $\hat{h}_\phi$, whilst variation of the walls, $\hat{c}_\phi$ and $\hat{w}_\phi$, only makes a contribution in the case that there is a mean flow ($\alpha_m \neq 0$).

For convenience, we rewrite (23) in terms of unscaled dimensionless variables by reversing the transformation (14) and setting $A = \mu\hat{A}$, $h = \mu^2\hat{h}$, $c = \mu^2\hat{c}$ and $w = \mu^2\hat{w}$. This yields the following alternative form of the weakly nonlinear wave equation,

$$A_t + (\alpha+\alpha_m)A_\theta + \beta A_{\theta\theta\theta} - \frac{1}{2\overline{R_H}}\left(\mathrm{sign}(A)+\gamma\right)AA_\theta = -\alpha h_\theta + \alpha_m\left(\delta_c c_\theta + \delta_w w_\theta\right). \tag{25}$$

The difference between (25) and KdV lies in the coefficient of the nonlinear term (AA_ϕ), which may be positive or negative depending on the sign of A and the positions of the channel walls and shelf line. It is exactly zero if

$$A>0 \text{ and } \overline{R_H} = \left(\overline{R_I}^2\overline{R_O}\right)^{1/3} = \overline{R_{H1}}, \text{ or } A<0 \text{ and } \overline{R_H} = \left(\overline{R_I R_O}^2\right)^{1/3} = \overline{R_{H2}}. \tag{26}$$

If we let $\overline{R_I}\to\infty$ then $\overline{R_{H1}}\to(2\overline{R_I}+\overline{R_O})/3$ and $\overline{R_{H2}}\to(\overline{R_I}+2\overline{R_O})/3$, which are exactly the conditions found for the analogous straight-channel case in [3]. These values of $\overline{R_H}$ define different regimes for the coefficient of the nonlinear term, and therefore for the direction in which nonlinear steepening occurs. Specifically, the direction of nonlinear steepening is always opposite to the direction of propagation, except in the cases $\overline{R_H}<\overline{R_{H1}}$ and $A>0$ or $\overline{R_H}>\overline{R_{H2}}$ and $A<0$, when the waves steepen towards the direction of propagation.

4.1 Cnoidal Waves

When A is single-signed and the annulus is radially symmetric ($R_O \equiv \overline{R_O}$, $R_H \equiv \overline{R_H}$, $R_I \equiv \overline{R_I}$), equation (25) reduces to the KdV equation, which is known [25] to possess cnoidal wave solutions. We shall restrict our attention to the case $A \leq 0$, with the understanding that the $A \geq 0$ case has analogous results. Equation (25) then becomes

$$A_t + \alpha A_\theta + \beta A_{\theta\theta\theta} + \gamma_1 AA_\theta = 0, \qquad \gamma_1 = \frac{\ln(R_O^2 R_I/R_H^3)}{R_H\ln(R_O/R_I)}. \tag{27}$$

We seek travelling wave solutions of the form $A(\theta,t)=F(\xi)$, where $\xi=\theta-(\alpha+\Delta)t$, and then apply the transformation $F(\xi)=6B(\xi)/\gamma_1$, $\xi=\sqrt{\beta}\zeta$, to obtain the standard form of the KdV equation

$$-\Delta B_\zeta+6BB_\zeta+B_{\zeta\zeta\zeta}=0. \tag{28}$$

This equation has well-documented [25] 'cnoidal' wave solutions of the form,

$$B=2p^2\kappa^2\mathrm{cn}^2(p(\zeta+\zeta_0);\kappa)+B_0, \qquad \Delta=8p^2\kappa^2-4p^2+\gamma_1 A_0, \tag{29}$$

where $0\le\kappa\le1$ is the modulus of the Jacobi elliptic cn function [1] and ζ_0, p and B_0 are constants. Thus, the cnoidal wave solutions to (27) are

$$A(\theta,t)=A_m\,\mathrm{cn}^2\left\{\frac{1}{2\kappa}\sqrt{\frac{A_m\gamma_1}{3\beta}}\,(\theta-(\alpha+\Delta)t+\theta_0)\,;\kappa\right\}+A_0, \tag{30}$$

with

$$\Delta=\gamma_1\left(A_0+\left(2-\frac{1}{\kappa^2}\right)\frac{A_m}{3}\right), \tag{31}$$

where A_0 is the reference amplitude and A_m is the maximum displacement from A_0. The argument of the cn function must be real for A to be bounded, so we must have $A_m\gamma_1\ge0$ because $\beta>0$ always. The sign of A_m is therefore determined by the position of the shelf line relative to the inner and outer walls of the annulus. Specifically, if $R_H\ge R_{H2}$ then $\gamma_1\le0$ and so $A_m\le0$. Thus $A\le0$ as long as $A_0\le0$. If $R_H\le R_{H2}$ then $\gamma_1\ge0$ and so $A_m\ge0$. Thus $A\le0$ as long as $A_0\le-A_m$.

A further restriction is imposed by the periodicity of the annular domain. The angular wave length of the cnoidal waves is

$$\lambda_\theta=4\kappa K(\kappa)\sqrt{\frac{3\beta}{A_m\gamma_1}}, \tag{32}$$

where K is the complete elliptic integral of the first kind [1]. We require that $\lambda_\theta=2\pi/n$ for any positive integer n, which yields the following expression for the wave amplitude,

$$A_m=\frac{12\beta n^2\kappa^2K^2(\kappa)}{\pi^2\gamma_1}. \tag{33}$$

Thus, for a given mode n, the amplitude of the wave A_m determines the modulus κ, and thereby the shape of the wave. Note also that because the amplitude is limited by the width of the channel, and because $\kappa K(\kappa)$ is a strictly increasing function of κ, higher-frequency modes must have smaller moduli κ, and so will be more similar to cosine waves. Meanwhile lower-frequency modes can have larger moduli and so are closer to the soliton solution ($\kappa=1$). The complete range of solutions may be written as

$$A(\theta,t)=\frac{12\beta n^2\kappa^2K^2(\kappa)}{\pi^2\gamma_1}\mathrm{cn}^2\left\{\frac{nK(\kappa)}{\pi}(\theta-(\alpha+\Delta)t+\theta_0)\,;\kappa\right\}+A_0. \tag{34}$$

Whilst these solutions describe only a small subset of the possible travelling wave solutions of (27), we may expect to see the same broad characteristics in all such solutions. That is, that the longest waves should exhibit the most nonlinear behaviour, whilst shorter waves should be well-approximated by linear theory.

4.2 Phase Plane Analysis

We will now expand our analysis of the weakly nonlinear wave equation to include travelling wave solutions that are not single-signed. Following the method used in Section 4.1, we seek travelling wave solutions of the form $A(\theta, t) = F(\xi)$, $\xi = \theta - (\alpha + \Delta)t$, which leads to the following travelling wave equation,

$$-\Delta A_\xi + \beta A_{\xi\xi\xi} - \frac{1}{2R_H}\left(\text{sign}(A) + \gamma\right) A A_\xi = 0. \tag{35}$$

Integrating with respect to ξ yields

$$-\Delta A + \beta A_{\xi\xi} - \frac{1}{4R_H}\left(\text{sign}(A) + \gamma\right) A^2 = -\Delta A_0 - \frac{1}{4R_H}\left(\text{sign}(A_0) + \gamma\right) A_0^2, \tag{36}$$

where A_0 is a reference amplitude at which $A_{\xi\xi} = 0$. Multiplying through by A_ξ allows us to integrate again to obtain

$$\tfrac{1}{2}\beta A_\xi^2 + V(A) = E, \tag{37}$$

where E is a constant and V is given by

$$V(A) = A\left[\Delta\left(A_0 - \tfrac{1}{2}A\right) + \frac{\gamma}{4R_H}\left(A_0^2 - \tfrac{1}{3}A^2\right) + \frac{1}{4R_H}\left(|A_0|A_0 - \tfrac{1}{3}|A|A\right)\right] \tag{38}$$

The critical points of (37) are defined as solutions of the equation $V'(A_c) = 0$, and their nature is determined by $V''(A_c)$: if $V''(A_c) > 0$ then the critical point is a centre, and if $V''(A_c) < 0$ then the critical point is a saddle.

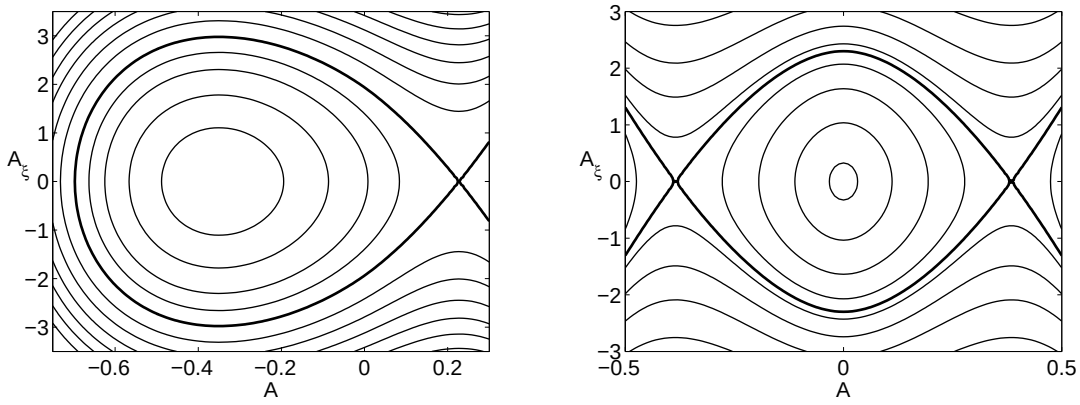


Figure 3: Phase plane for weakly nonlinear waves in the range $R_I \leq A + R_H \leq R_O$ when $R_c = 1.065\,\text{m}$, $R_h = 0.960\,\text{m}$, and (a) $R_w = 0.665\,\text{m}$, $\Delta = -0.01$, $A_0 = -0.35$, (b) $R_w = 0.8654\,\text{m}$, $\Delta = -0.02$, $A_0 = 0$. The thicker lines highlight the orbits of the critical points.

We determine the forms of the solutions by analysing the phase plane in A and A_ξ, where each contour corresponds to a different travelling wave solution. We are restricted to those

solutions whose amplitudes lie within the boundaries of the annulus, i.e. $R_I \leq R_H + A \leq R_O$. In order to fit in the annular domain, solutions must also be periodic with a period that exactly divides 2π, so only closed orbits correspond to viable solutions. The reference amplitude $A = A_0$ is always a critical point, and is always a centre as long as

$$\Delta + \frac{1}{2R_H}\left(\gamma + \operatorname{sign}(A_0)\right) A_0 < 0. \tag{39}$$

If this point is not a centre then the phase plane has no closed orbits, so this serves as a sufficient condition for there to be physical solutions.

Additionally, because we assumed in our derivation that the solution has been perturbed from a basic state in which $R = R_H$, we must ensure that mass has been in conserved in the transition to this steady solution,

$$\int_0^{2\pi} \int_{R_H}^{R} r \, \mathrm{d}r\mathrm{d}\theta = 0. \tag{40}$$

Thus, only a small subset of the curves in the phase plane are permissible as travelling wave solutions in the annulus. This excludes the possibility of the soliton solutions, which correspond to the homoclinic orbit of the critical point in Figure 3(a). In Figure 3(b) we present the case where $R_h = \sqrt{R_c R_w}$, such that $\gamma = 0$ and the nonlinear term in (35) is symmetric about $A = 0$. In this case there is a heteroclinic orbit between the saddle critical points that corresponds to the 'kink' solitons studied in [3], but again the periodicity of the domain means that such solutions are not permissible.

5 Frictional Effects

The theory of Ekman layers [22, 23, 24] suggests that the fluid in the annulus will return to solid-body rotation due to frictional effects at the base of the layer of fluid. It may be shown [18] that the leading-order effects of bottom friction can be incorporated in the rotating shallow water equations as follows,

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla\pi + f\hat{\mathbf{z}}\times\mathbf{u} + \frac{\tilde{k}}{\eta}\mathbf{u} = 0, \tag{41a}$$

$$\nabla\cdot(\eta\mathbf{u}) = 0, \tag{41b}$$

where $\mathbf{u} = (u, v)$ is the vertically-averaged horizontal fluid velocity, $\eta = H - h$ is the vertical thickness of the fluid layer and π is the fluid pressure at the rigid lid. The dissipation constant $\tilde{k}$ is typically inversely proportional to the Ekman spin-down time [23], $\tilde{k}/H \propto 1/\tau_E = \sqrt{f\nu/H^2} \implies \tilde{k} \propto \sqrt{f\nu}$, where τ_E is the Ekman spin-down time. The constant ν is the kinematic viscosity of the fluid in the experiment or the vertical eddy viscosity in the ocean. We will now consider the influence of friction on the system studied in Sections 3 and 4.

We first apply quasigeostrophic scaling [22, 23, 24] to (41a) and (41b). We assume that the Rossby number is very small, $\mathrm{R}o = U/fL \ll 1$, where U is a velocity scale for the flow and $L = \overline{R_c} - \overline{R_w}$ as before, and that h/H and $\tilde{k}/Hf$ are $\mathrm{O}\,(\mathrm{Ro})$. Expanding $\mathbf{u}$ and

π asymptotically in Ro then yields geostrophic balance at leading order, with contributions from friction, bottom topography and advection included at first order in a modified quasigeostrophic potential vorticity equation,

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\omega + Q\hat{h}\right) = -k\omega, \tag{42}$$

where $k = \tilde{k}/H$, and $\hat{h} = h/H_s$ is the dimensionless height of the bottom topography.

5.1 Vorticity Evolution of Fluid Columns

As in Section 3, we seek to determine the vorticity distribution and thereby formulate a nonlinear wave equation for the interface $r = R(\theta, t)$. Let us consider the vorticity $\omega_p(t)$ of an infinitesimal fluid column that lies in the region $R_w < r < R_h$ at $t = 0$, such that $\omega_p(0) = q_p(0) = 0$. Equation (42) allows us to write down an evolution equation for ω_p,

$$\frac{\mathrm{d}\omega_p}{\mathrm{d}t} + k\omega_p + Q\frac{\mathrm{d}\hat{h}_p}{\mathrm{d}t} = 0, \tag{43}$$

where $Q = fH_s/H$ as before and $\hat{h}_p(t)$ is the dimensionless height of the topography directly below the fluid column at time t. If the fluid column crosses the shelf line at times $t = t_1, t_2, \ldots, t_N$, then we may write $\hat{h}_p$ explicitly as

$$\hat{h}_p = \begin{cases} 0, & t_i < t < t_{i+1} \quad i = 0, 2, 4, \ldots, \\ 1, & t_i < t < t_{i+1} \quad i = 1, 3, 5, \ldots, \end{cases} \tag{44}$$

where we define $t_0 = 0$. Thus in $t_i < t < t_{i+1}$ for all i, $\partial\hat{h}_p/\partial t = 0$, and integrating equation (43) yields

$$\omega_p = \omega_p|_{t=t_i}\, e^{-k(t-t_i)}, \qquad t_i < t < t_{i+1}. \tag{45}$$

In order to determine the evolution of ω_p at $t = t_i$, $i = 1, \ldots, N$, we integrate (43) over $[t_i - \delta t, t_i + \delta t]$ and take the limit $\delta t \to 0$,

$$\lim_{\delta t \to 0} \int_{t_i - \delta t}^{t_i + \delta t} \left\{ \frac{\mathrm{d}\omega_p}{\mathrm{d}t} + k\omega_p + Q\frac{\mathrm{d}\hat{h}_p}{\mathrm{d}t} \right\} \mathrm{d}t = \lim_{\delta t \to 0} \left[\omega_p + Q\hat{h}_p\right]_{t_i - \delta t}^{t_i + \delta t} = 0, \tag{46}$$

where the first equality follows from the requirement that ω_p must remain finite. This condition states that potential vorticity is exactly conserved following the fluid column at the instant when it crosses the shelf line. Using (44) we may then write the jump condition for the vorticity of the fluid column as

$$[\omega_p]_{t=t_i} = (-1)^i Q, \tag{47}$$

where the square brackets [ ] denote the change of ω_p over an infinitesimally short period around $t = t_i$. Combining (45) and (47) yields the complete time-evolution of ω_p,

$$\omega_p = \begin{cases} 0, & 0 \le t < t_1, \\ Q\displaystyle\sum_{i=1}^{n} (-1)^i e^{-k(t-t_i)}, & t_n < t < t_{n+1}, \quad n < N, \\ Q\displaystyle\sum_{i=1}^{N} (-1)^i e^{-k(t-t_i)}, & t_N < t. \end{cases} \tag{48}$$

The vorticity of an infinitesimal fluid element that is initially on the shelf and that crosses the discontinuity N times may be obtained by multiplying (48) by -1.

This analysis highlights an important difference between the frictionless and frictional potential vorticity equations: in the frictionless case, fluid columns can only have non-zero vorticity if they lie between the shelf line $r = R_h$ and the interface $r = R$, whereas now every fluid column that has at any point crossed the shelf line will have nonzero vorticity. This complication means that it is no longer possible to perform a derivation similar to that described in Section 3. However, the fact that the evolution of the vorticity of individual fluid columns can be expressed in a simple form makes this system an appropriate candidate for a vortex element method [16], a Lagrangian numerical method that advects individual vortices to calculate the flow field.

5.2 Approximate Conservation Law Form

In order to include some dissipative effects in our nonlinear model, we require a conservation law that approximates (42). It is not possible to rewrite this equation in an exact conservation law form, so instead we use the following approximation,

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\omega e^{kt} + Q\hat{h}\right) = 0. \tag{49}$$

We shall now motivate this choice by considering the evolution of the total vorticity in the annulus, rather than the local vorticity field.

We obtain an evolution equation for the total vorticity by integrating (42) over the entire area of the annulus, which we denote by A,

$$\iint_A \left\{ \frac{\partial \omega}{\partial t} + k\omega + \mathbf{u} \cdot \nabla \left(\omega + Q\hat{h}\right) \right\} \mathrm{d}A = 0. \tag{50}$$

In the quasigeostrophic limit $\nabla \cdot \mathbf{u} = 0$, so we may write the final term in the integrand as

$$\iint_A \mathbf{u} \cdot \nabla \left(\omega + Q\hat{h}\right) \mathrm{d}A = \iint_A \nabla \cdot \left(\left(\omega + Q\hat{h}\right)\mathbf{u}\right) \mathrm{d}A = \int_{\partial A} \left(\omega + Q\hat{h}\right) \mathbf{u} \cdot \hat{\mathbf{n}} \, \mathrm{d}s = 0, \tag{51}$$

where ∂A denotes the horizontal boundaries of the domain, $\hat{\mathbf{n}}$ denotes a unit vector normal to the boundary, and $\mathrm{d}s$ denotes an infinitesimal unit of distance along the boundary. The second equality follows from an application of the divergence theorem, and the final equality holds because $\mathbf{u}$ must be tangential to the boundaries. Applying (51) to (50) yields the following evolution equation for the total vorticity,

$$\frac{\partial}{\partial t}\left(\langle\omega\rangle e^{kt}\right) = 0, \qquad \text{where} \qquad \langle\omega\rangle = \iint_A \omega \, \mathrm{d}A. \tag{52}$$

Via a similar procedure, it may be shown that exactly the same evolution equation for $\langle\omega\rangle$ holds in (49), which suggests that the global behaviours of (42) and (49) should be broadly similar. However, if the interface line $r = R(\theta, t)$ has been perturbed from a state where $R = R_h$ and $\omega = 0$ everywhere then this simply states that $\langle\omega\rangle = 0$ for all time, as in the frictionless case. We can therefore make a stronger evaluation of our parametrisation by considering the total absoulute vorticity in the annulus.

Recall from (48) that when $Q > 0$, the vorticity is always positive in $R_w < r < R_h$ and always negative in $R_h < r < R_c$. We shall henceforth restrict our attention to this case, with the understanding that analogous results hold when $Q < 0$. This motivates the following integration of (42),

$$\left(\iint_{A_1} - \iint_{A_2}\right)\left\{\frac{\partial\omega}{\partial t} + k\omega + \nabla\cdot\left(\left(\omega + Q\hat{h}\right)\mathbf{u}\right)\right\}\mathrm{d}A = 0 \tag{53}$$

where we have applied $\nabla\cdot\mathbf{u} = 0$ in the final term of the integrand. We define the regions $A_1 = \{(r,\theta) : R_w < r < R_h, 0 < \theta < 2\pi\}$ and $A_2 = \{(r,\theta) : R_h < r < R_c, 0 < \theta < 2\pi\}$, such that $A = A_1 \cup A_2$. Thus $\omega \geq 0$ everywhere in A_1 and $\omega \leq 0$ everywhere in A_2. The first two terms in the integrand in (53) may then be rewritten as

$$\begin{aligned}\left(\iint_{A_1} - \iint_{A_2}\right)\left\{\frac{\partial\omega}{\partial t} + k\omega\right\}\mathrm{d}A &= \left(\frac{\partial}{\partial t} + k\right)\left(\iint_{A_1} - \iint_{A_2}\right)\omega\,\mathrm{d}A \\ &= \left(\frac{\partial}{\partial t} + k\right)\left(\iint_{A_1} + \iint_{A_2}\right)|\omega|\,\mathrm{d}A = \left(\frac{\partial}{\partial t} + k\right)\langle|\omega|\rangle,\end{aligned} \tag{54}$$

where we use the shortened notation $(\iint_{A_1} - \iint_{A_2})F\,\mathrm{d}A = \iint_{A_1} F\,\mathrm{d}A - \iint_{A_2} F\,\mathrm{d}A$, and where the angular brackets $\langle\ \rangle$ denote integration over A, as in (52). The second equality in (54) follows from the fact that ω is always positive in A_1 and always negative in A_2. The final term of the integrand in (53) may be simplified using the divergence theorem,

$$\begin{aligned}I = \left(\iint_{A_1} - \iint_{A_2}\right)\nabla\cdot\left(\left(\omega + Q\hat{h}\right)\mathbf{u}\right)\mathrm{d}A &= \left(\int_{\partial A_1} - \int_{\partial A_2}\right)\left(\omega + Q\hat{h}\right)\mathbf{u}\cdot\hat{\mathbf{n}}\,\mathrm{d}s \\ &= \int_{r=R_h}\left\{\left(\omega_1 + Q\hat{h}_1\right)\mathbf{u}_1\cdot\hat{\mathbf{n}}_1 - \left(\omega_2 + Q\hat{h}_2\right)\mathbf{u}_2\cdot\hat{\mathbf{n}}_2\right\}\mathrm{d}s.\end{aligned} \tag{55}$$

Here again ∂A_1 and ∂A_2 denote the boundaries of A_1 and A_2, $\hat{\mathbf{n}}$ denotes a unit vector pointing normally outwards from the boundary, and $\mathrm{d}s$ denotes an infinitesimal distance along the boundary. The second equality in (55) follows from the requirement that $\mathbf{u}\cdot\hat{\mathbf{n}} = 0$ on $r = R_w$ and $r = R_c$, so we need only integrate over the intersection of ∂A_1 and ∂A_2, which is the shelf line $r = R_h$. The unit vectors $\hat{\mathbf{n}}_i$ are directed normally to the boundary $r = R_h$ and outwards from region A_i, so in fact $\hat{\mathbf{n}}_1 = -\hat{\mathbf{n}}_2$. Otherwise the subscripts i, where $i = 1, 2$, denote that a quantity should be evaluated infinitesimally close to the boundary in the region A_i.

Now by definition $\hat{h}_1 = 0$, and $\hat{h}_2 = 1$, and as we have taken the quasigeostrophic limit, $\mathbf{u}$ must be continuous at $r = R_h$. This allows us to further simplify our expression for I to

$$I = \int_{r=R_h}\left\{\omega_1 + \omega_2 + Q\left(\hat{h}_1 + \hat{h}_2\right)\right\}\mathbf{u}\cdot\hat{\mathbf{n}}\,\mathrm{d}s = \int_{r=R_h}\left(|\omega_1| - |\omega_2|\right)\mathbf{u}\cdot\hat{\mathbf{n}}\,\mathrm{d}s. \tag{56}$$

The second equality follows from the fact that the total mass flux across $r = R_h$ must be zero, $\int_{r=R_h}\mathbf{u}\cdot\hat{\mathbf{n}} = 0$, by conservation of mass. Thus I is the average around the shelf line of the difference between the vorticity fluxes across the shelf line. Combining (54) and (56), we obtain the following evolution equation for the total absolute vorticity in the annulus,

$$\frac{\partial}{\partial t}\left(\langle|\omega|\rangle e^{kt}\right) + Ie^{kt} = 0. \tag{57}$$

Applying the same integration to (49) yields exactly the same evolution equation for $\langle|\omega|\rangle$. The dependence of $\langle|\omega|\rangle$ on I, and thereby on the average of the vorticity field around the shelf line, means that the total absolute vorticity will evolve differently under (42) and (49).

Our parametrisation assumes, as before, that the fluid has been perturbed from a rest state with $r = R_h$ and $\omega = 0$. All vorticity then decays exponentially at a rate k, regardless of when the vorticity is generated by fluid crossing the shelf. As in the frictionless case, only fluid within the envelope of the wave (between $r = R_h$ and $r = R$) can have nonzero vorticity. Thus, this equation exactly replicates the behaviour of fluid columns that are perturbed across the shelf at $t = 0$ and remain there, or that do not cross the shelf at all. It least accurately captures the behaviour of fluid columns that cross the shelf, remain there for a period of time that is $\mathrm{O}\,(1/k)$, and then return, because these fluid columns end up outside the envelope of the wave with non-negligible vorticity. One might expect to see such behaviour, for example, in travelling wave solutions.

Nevertheless, we shall see in Section 7 that this parametrisation yields good agreement with experimental results. This is because fluid columns tend to experience a loss of vorticity when they cross the shelf line in the experiment, and so tend not to retain large vorticities outside the envelope of the waves. Equation (49) is also useful because it allows a straightforward derivation of an evolution equation for the interface $r = R$. Recalling that in the quasigeostrophic limit we have $\omega = \nabla^2\Psi$, as in Section 3, we let $\Psi_f = \Psi e^{kt}$, such that (49) becomes

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\nabla^2\Psi_f + \frac{fh}{H}\right) = 0. \tag{58}$$

We may then perform exactly the same analysis as was performed in Section 3, but replacing Ψ and ψ by Ψ_f and ψ_f throughout. The equation for the evolution of the interface is then

$$\frac{\partial R}{\partial t} = -\frac{1}{R}\frac{\partial}{\partial\theta}\Psi(R(\theta,t),\theta,t) = -\frac{e^{-kt}}{R}\frac{\partial}{\partial\theta}\Psi_f(R(\theta,t),\theta,t). \tag{59}$$

Thus, we obtain exactly the same solution for the evolution of the interface $\partial R/\partial t$, but multiplied by $\exp(-kt/Q)$, where t is now the dimensionless time. This means that the solution evolves in a way that is identical to the frictionless case, but with the motion of the interface decaying exponentially to zero.

6 Stability

The laboratory experiments with a rotating annulus have highlighed some instabilities of the fluid under certain conditions. We now examine two specific types of instability in an attempt to explain these results.

6.1 Stability of the Mean Flow

In Figure 4 we present snapshots of a laboratory experiment in which a mean flow is induced by rapidly changing the rotation rate of the annulus once the fluid is in solid body rotation. The flow becomes unstable and forms large eddies along the shelf line. It is desirable to include a mean flow in our experiments to represent the presence of a current along the shelf break, so it is of interest to determine the cause of this instability.

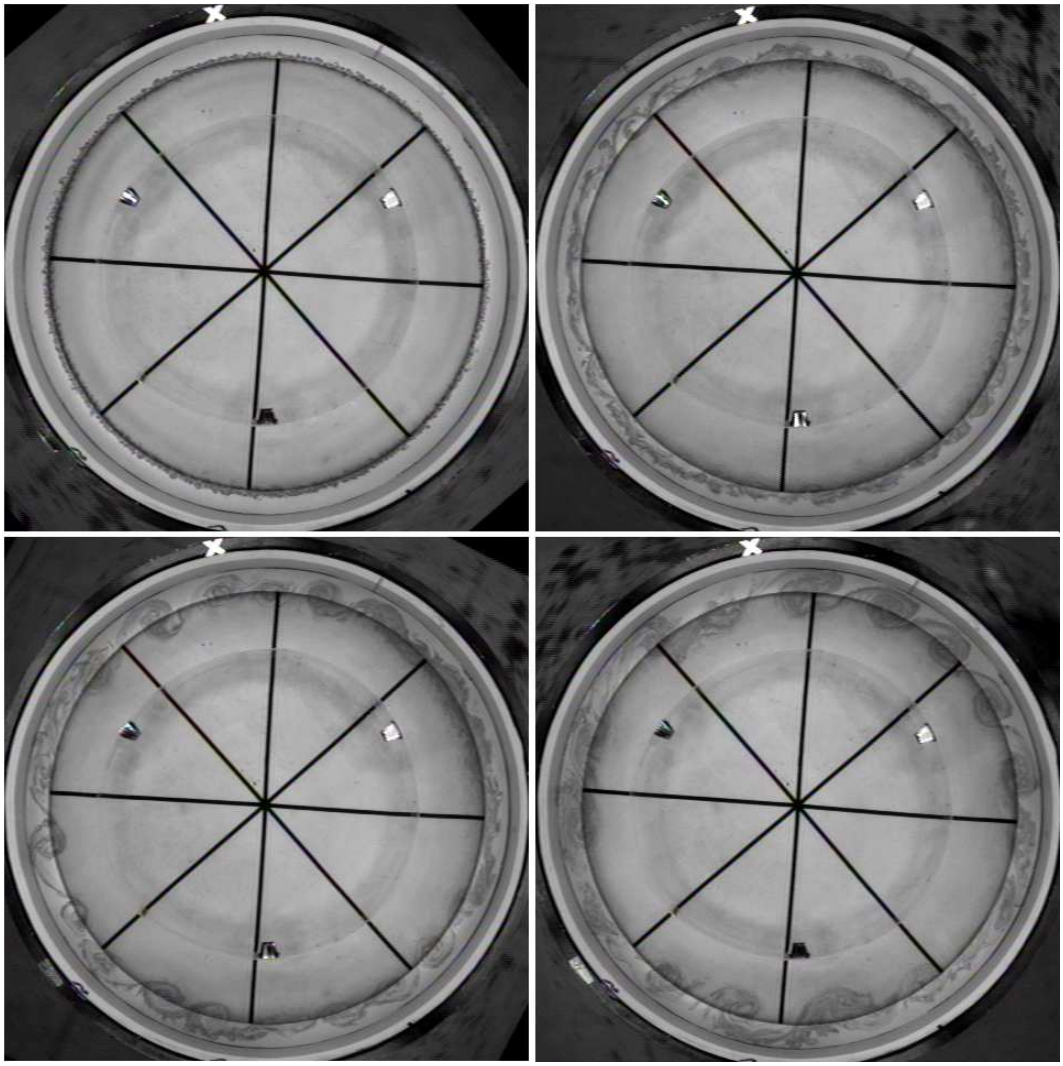


Figure 4: Snapshots of an experiment exhibiting instability to a mean flow. Once the fluid in the tank had achieved solid-body rotation, the rate of rotation of the tank was rapidly reduced from $f = 2.00\,\mathrm{rad\,s^{-1}}$ to $f = 1.85\,\mathrm{rad}\,s^{-1}$. The fluid thus continues to move anticlockwise around the tank, and develops large instabilities resembling Kelvin-Helmholz rolls as it does so. These snapshots were taken 0 (top left), 45 (top right), 59 (bottom left) and 77 (bottom right) seconds after the change in the tank's rotation rate. The maximum depth of the fluid is 25cm, and the radius of the inner wall is approximately 66cm.

Consider a mean flow $\mathbf{u}_m = v_m(r,t)\hat{\mathbf{e}}_\theta$ in a rotating annulus whose walls R_c and R_w and shelf line R_h are independent of θ, and where $\mathbf{u}_m$ is initially a flow of constant angular velocity α_0, $v_m(r,0) = \alpha_0 r$. We seek a radially symmetric solution of the dissipative shallow water equations (41a), ensuring that mass and surface pressure are conserved across the shelf line at $r = R_h$. The solution, which must be time-dependent due to the dissipative terms, is

$$v_m = \alpha_0 r e^{-\hat{k}t/\eta}, \quad \pi_m = \pi_0 + \tfrac{1}{2}\alpha_0 \left(r^2 - R_h^2\right) e^{-\hat{k}t/\eta} \left(f + \alpha_0 e^{-\hat{k}t/\eta}\right), \tag{60}$$

The fluid depth is $\eta = H - h$, where h is defined in Section 3. For $t > 0$ there is a discontinuity in v_m at $r = R_h$, and we expect that this velocity shear should be unstable to small perturbations.

To demonstrate that this configuration is unstable we consider the simple problem of a steady shear flow in the presence of a discontinuity in depth. We prescribe a mean flow of the form

$$v_m = \begin{cases} \alpha_1 r, & R_w < r < R_h, \\ \alpha_2 r, & R_h < r < R_c. \end{cases} \tag{61}$$

We might now naively seek a linear wave solution by following a method analogous to that of [17], linearising the governing equations about the mean flow solution and requiring that the mass flux and pressure across $r = R_h$ should be continuous. However, this yields a dispersion relation that has purely real modes, indicating that there is no instability in the mean flow, and that does not reduce to the expected dispersion relation when there is no discontinuity in depth. We must therefore approach the problem more carefully.

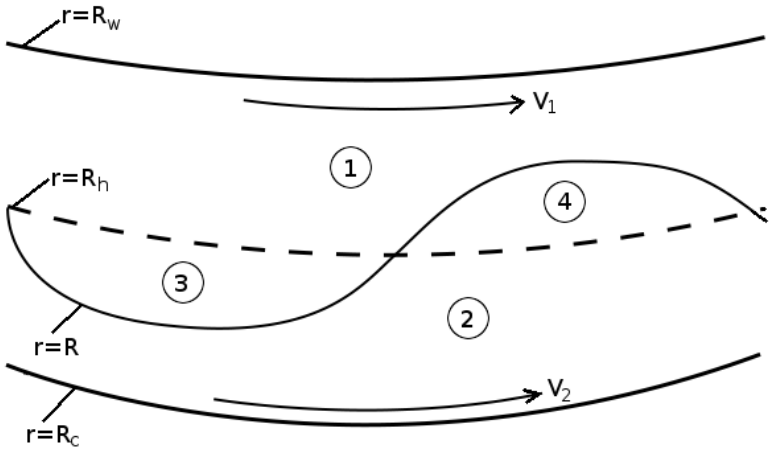


Figure 5: Schematic of a linear wave propagating in a small region of the annulus in the presence of a sheared mean flow.

Let us consider the propagation of a linear wave along the line of the shelf, presented diagrammatically in Figure 5. We split the domain into four regions, similar to the approach taken in the derivation of the nonlinear wave equation in Section 3. In order to properly

incorporate the influence of the shear, we must recognise that the mean flow in region 3 is the same as that in region 1, and similarly that the mean flow in region 4 is the same as that in region 2. The complete mean flow is then

$$\mathbf{u}_{mi} = \alpha_j r\,\hat{\mathbf{e}}_\theta, \quad \pi_{mi} = \pi_0 + \tfrac{1}{2}\alpha_j\left(r^2 - R^2\right), \tag{62}$$

where $j = 1$ for $i = 1, 3$ and $j = 2$ for $i = 2, 4$, which also ensures continuity of pressure across $r = R$ and $r = R_h$. We linearise (41a) with $\hat{k} = 0$ about the mean flow solution, setting $\mathbf{u}_i = \mathbf{u}_{m,i} + \mathbf{u}_i'$, $\pi_i = \pi_{m,i} + \pi_i'$ and $R = R_h + R'$, where $\mathbf{u}_i' \ll u_{m,i}$, $\pi_i' \ll \pi_{m,i}$ and $R' \ll L$. This yields

$$\frac{\partial \mathbf{u}_i'}{\partial t} + \alpha_j \frac{\partial \mathbf{u}_i'}{\partial \theta} + (f + \alpha_j)\hat{\mathbf{z}} \times \mathbf{u}_i' + \nabla \pi_i' = 0, \qquad \nabla \cdot \mathbf{u}_i' = 0, \tag{63}$$

which is valid everywhere except at $r = R_h$. However, in the linear limit, the fluid in regions 3 and 4 becomes infinitesimally close to the shelf line $r = R_h$, and so there we must apply the linearised form of the complete mass conservation equation (41b), rather than the two-dimensional incompressibility condition. Thus (63) holds only for $i = 1, 2$.

In order to avoid a detailed consideration of the dynamics in regions 3 and 4, we instead make use of the fact that the linearised kinematic boundary conditions and pressure continuity condition for the interface $r = R_h + R'$ are imposed on the shelf line $r = R_h$,

$$\frac{\partial R'}{\partial t} + \alpha_1 \frac{\partial R'}{\partial \theta} = u_3', \quad \frac{\partial R'}{\partial t} + \alpha_2 \frac{\partial R'}{\partial \theta} = u_2', \quad \pi_2' = \pi_3', \quad \text{on} \quad r = R_h, \quad R' > 0, \tag{64a}$$

$$\frac{\partial R'}{\partial t} + \alpha_1 \frac{\partial R'}{\partial \theta} = u_1', \quad \frac{\partial R'}{\partial t} + \alpha_2 \frac{\partial R'}{\partial \theta} = u_4', \quad \pi_1' = \pi_4', \quad \text{on} \quad r = R_h, \quad R' < 0. \tag{64b}$$

We combine these with the jump conditions for the shelf line, which state that mass flux and surface pressure must be continuous across the discontinuity in depth,

$$H u_1' = (H - H_s) u_3', \quad \pi_1' = \pi_3', \quad \text{on} \quad r = R_h, \quad R' > 0, \tag{65a}$$

$$H u_4' = (H - H_s) u_2', \quad \pi_4' = \pi_2', \quad \text{on} \quad r = R_h, \quad R' < 0. \tag{65b}$$

We may now eliminate u_3', u_4', π_3', and π_4' between (65a), (65b), (64a) and (64b). We thereby avoid any consideration of the structure of regions 3 and 4, instead connecting regions 1 and 2 directly via the following jump conditions,

$$\frac{\partial R'}{\partial t} + \alpha_1 \frac{\partial R'}{\partial \theta} = \frac{1}{\Delta_H} u_1', \quad \frac{\partial R'}{\partial t} + \alpha_2 \frac{\partial R'}{\partial \theta} = u_2', \quad \pi_1' = \pi_2', \quad \text{on} \quad r = R_h, \quad R' > 0, \tag{66a}$$

$$\frac{\partial R'}{\partial t} + \alpha_2 \frac{\partial R'}{\partial \theta} = \Delta_H u_2', \quad \frac{\partial R'}{\partial t} + \alpha_1 \frac{\partial R'}{\partial \theta} = u_1', \quad \pi_1' = \pi_2', \quad \text{on} \quad r = R_h, \quad R' < 0, \tag{66b}$$

where $\Delta_H = (H - H_s)/H \leq 1$ is the depth ratio. Note that in the absence of a mean flow, $\alpha_1 = \alpha_2 = 0$, we may reduce the boundary conditions to $(H - H_s)u_2' = H u_1'$ and $\pi_1' = \pi_2'$ at $r = R_h$ for all θ, which corresponds to the approach of [17]. In fact, this reduction can be performed for any continuous mean flow profile, and it is only in the case of a discontinuous profile that this special consideration is required.

We now make use of the fact that the velocity is divergence-free in regions 1 and 2 to write $\mathbf{u}_j' = -\nabla \times \psi_j' \, \hat{\mathbf{z}}$, and seek plane wave solutions of the form $\psi_j' = \mathrm{Re}\{\hat{\psi}_j(r) \exp[i(l\theta - \sigma t)]\}$. Taking the curl of the linearised momentum equations in (63) yields a vorticity equation that states that the vorticity must be uniformly equal to zero at all times, so $\omega_1' = \nabla^2 \psi_1' = 0$ and $\omega_2' = \nabla^2 \psi_2' = 0$. We may then solve for $\hat{\psi}_i$ subject to the boundary conditions $\psi_1' = 0$ at $r = R_w$ and $\psi_2' = 0$ at $r = R_c$. Substituting the result into (66a) and (66b) yields the following dispersion relation,

$$\sigma = \frac{1}{c_1 \Delta_H + c_2} \left\{ \tfrac{1}{2}(f_2 - \Delta_H f_1) + (\Delta_H c_1 \alpha_1 + c_2 \alpha_2)\, l \right.$$

$$\left. \pm \sqrt{\tfrac{1}{4}(f_2 - \Delta_H f_1)^2 + \Delta_H (f_2 c_1 + f_1 c_2)(\alpha_1 - \alpha_2) l - \Delta_H c_1 c_2 (\alpha_1 - \alpha_2)^2 l^2} \right\}, \qquad (67)$$

where $f_1 = f + 2\alpha_1$, $f_2 = f + 2\alpha_2$, and

$$c_1(l) = \frac{(R_h/R_w)^l + (R_h/R_w)^{-l}}{(R_h/R_w)^l - (R_h/R_w)^{-l}}, \qquad c_2(l) = \frac{(R_c/R_h)^l + (R_c/R_h)^{-l}}{(R_c/R_h)^l - (R_c/R_h)^{-l}}. \qquad (68)$$

When there is no mean flow ($\alpha_1 = \alpha_2 = 0$) this reduces to the dispersion relation for linear shelf waves in an annulus. In the case of a sheared mean flow with no topography and no rotation ($\Delta_H = 1$, $f = 0$) we recover the expected instability of all wave numbers to small perturbations.

In the short-wave limit, $c_1, c_2 \to \pm 1$ as $l \to \pm\infty$, so the term proportional to l^2 in the square root of (67) causes σ to have a positive complex component proportional to l for sufficiently large l. Thus very short waves are always unstable. However, for small l the frequency may become real, and so long waves may be stable. It may be shown that in the limit of infinitely long waves, $l \to 0$, σ is real if the velocity shear, $\alpha_1 - \alpha_2$, satisfies

$$\left| \frac{\alpha_1 - \alpha_2}{\frac{1}{2}(f_1 \ln(R_h/R_w) + f_2 \ln(R_c/R_h))} - 1 \right| < \sqrt{1 + \frac{(f_2 - \Delta_H f_1)^2 \ln(R_c/R_h) \ln(R_h/R_w)}{\Delta_H (f_1 \ln(R_h/R_w) + f_2 \ln(R_c/R_h))^2}}. \qquad (69)$$

This is not an exact condition for the velocity shear because α_1 and α_2 appear within f_1 and f_2. However, typically in both the laboratory experiments and in the ocean, $\alpha_1, \alpha_2 \ll f$, and so (69) still provides us with some insight. Long waves are stable as long as the shear is sufficiently small. For $f > 0$ a comparatively small negative shear ($\alpha_1 < \alpha_2$) will make all wave numbers unstable, whereas a relatively large positive shear ($\alpha_1 > \alpha_2$) is required to achieve the same effect. In Figure 6 we present plots of the real and imaginary parts of (67) for typical experimental parameter values with a positive shear.

Long waves can be stabilised because the generation of vorticity within the envelope of the wave creates a velocity shear about the shelf line. This shear, which is most pronounced in long waves, may exceed the size of the mean shear and thereby stabilise the disturbances. Despite this, short waves are almost completely unaffected by the presence of rotation and a discontinuity in depth. Although we still expect the shortest waves to be the most unstable, the largest instabilities in the experiment of Figure 4 have wavelengths on the order of 30cm, so in fact the longer perturbations grow faster in practice. Thus, whilst our idealised theory can qualitatively predict this instability, we require a more sophisticated treatment to make realistic quantitative predictions.

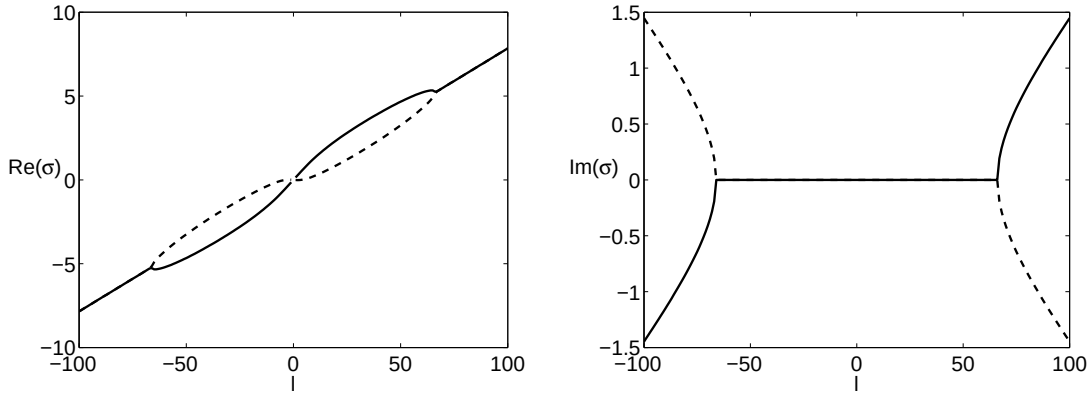


Figure 6: Real (left) and imaginary (right) parts of the dispersion relation for linear waves propagating in the presence of a sheared mean flow. Here $f = 1.5\,\mathrm{rad\,s^{-1}}$, $\alpha_1 = 0.1\,\mathrm{rad\,s^{-1}}$, $U_2 = 0.05\,\mathrm{rad\,s^{-1}}$, $\Delta_H = 0.8$, $R_w = 66\,\mathrm{cm}$, $R_h = 96\,\mathrm{cm}$ and $R_c = 106.5\,\mathrm{cm}$, as in the experiment shown in Figure 4.

6.2 Stability of Long Waves

We now turn our attention to another laboratory experiment, illustrated in Figure 7. Here a long wave is generated by reducing the rotation rate of the tank for a few seconds, creating a temporary mean flow that generates a perturbation past a long bump in the outer wall. The perturbation quickly breaks up into short waves, with the rest of the dyed interface following suit shortly afterwards. This raises the possibility that the nonlinear topographic Rossby waves themselves may be intrinsically unstable.

We perform a simple stability analysis of the waves by using the fact that they are long and slowly-varying to approximate them as a radially symmetric strip of constant vorticity ω_0, as shown in Figure 8. We retain the shelf line at $r = R_h$, but for now we let ω_0 be arbitrary, and do not require it to be equal to Q as we did in Section 3. We shall also only consider the case where $R > R_h$, as in the experiment shown in Figure 7, with the understanding that similar results hold in the case that $R < R_h$. In the analogous problem for a straight channel with a symmetric strip of constant vorticity [5], the flow is found to be unstable to long waves if the width of the strip is less than half the width of the channel.

We write the basic-state vorticity as

$$\overline{\omega} = \omega_0 \left[\mathcal{H}(r - R) - \mathcal{H}(r - R_h)\right], \tag{70}$$

as in (10), where the bar notation $\overline{\ }$ will be used henceforth to denote the basic flow. This means that for $\omega_0 > 0$, $\overline{\omega} < 0$ in $R_h < r < R$, reflecting the $f > 0$ case in Section 3. We seek a radially symmetric basic flow $\overline{\mathbf{u}} = \overline{v}(r)\hat{\mathbf{e}}_\theta$, which by definition is incompressible, $\nabla \cdot \overline{\mathbf{u}} = 0$. The corresponding streamfunction $\overline{\psi}$ matches the leading-order streamfunction (18) from Section 3 exactly, subject to the requirement that it must be uniformly equal to zero on the

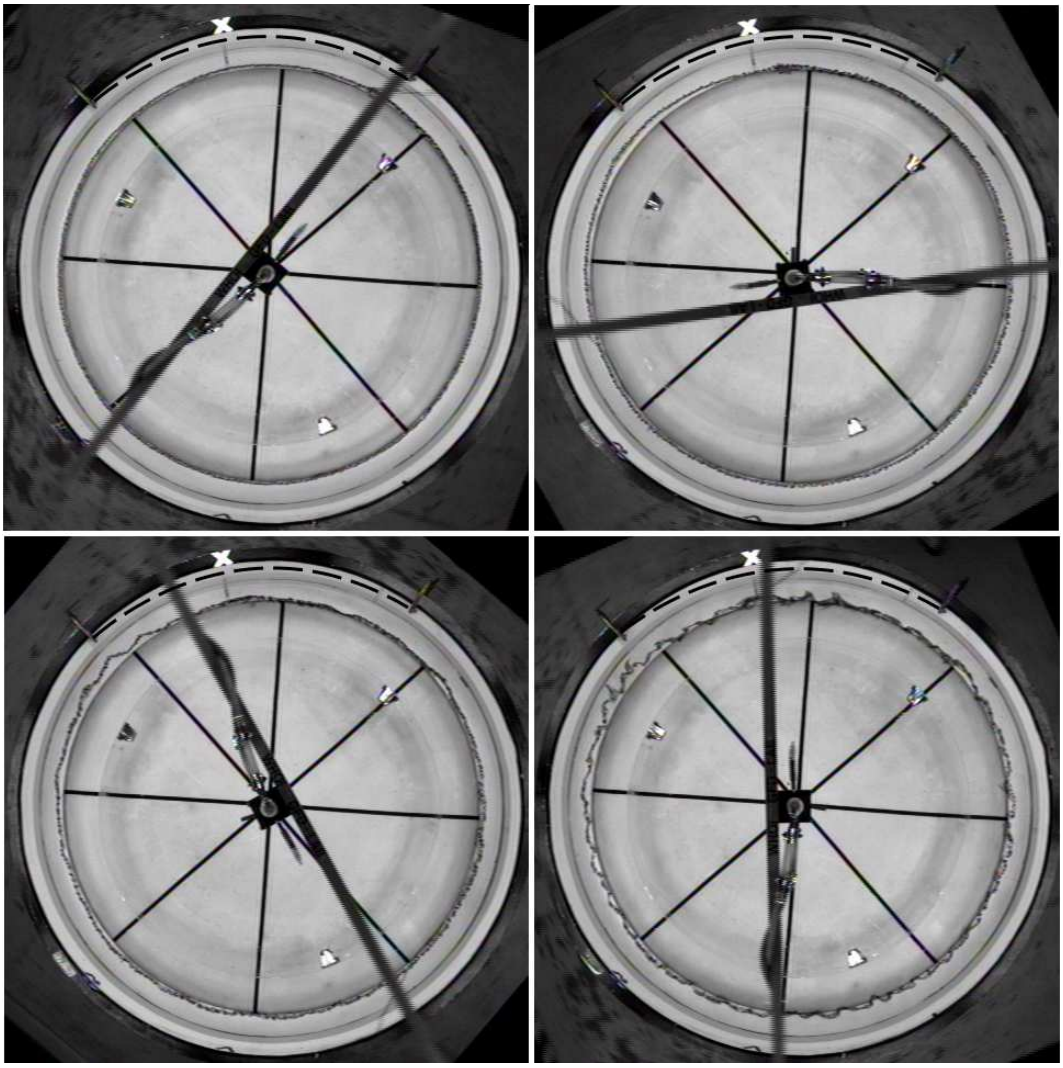


Figure 7: Snapshots of an experiment in which an instability arises in a long perturbation. Once the fluid in the tank had achieved solid-body rotation (top left), the rate of rotation of the tank was rapidly decelerated from $f = 1.50\,\mathrm{rad\,s^{-1}}$ to $f = 1.20\,\mathrm{rad\,s^{-1}}$, and then rapidly accelerated again to $f = 1.45\,\mathrm{rad\,s^{-1}}$. The effect of this is to cause the fluid to move anticlockwise around the tank for 6 seconds and then return to approximate solid-body rotation. A bump in the side wall of length 1.5m and maximum amplitude 2cm, highlighted in the diagram by the superposed dashed lines, leads to the formation of a perturbation ahead of the bump (top right). After 12 seconds (bottom left), a growing short wave instability appears in the perturbation, and after 17 seconds this instability is affecting the entire annulus. The maximum depth of the fluid in this experiment is 20cm, and the radius of the inner wall is approximately 75cm.

walls $r = R_w$ and $r = R_c$. Thus the basic-state velocity is

$$\overline{v}(r) = \frac{\omega_0}{2r}\left[\mathcal{H}(r-R)(r^2-R^2) - \mathcal{H}(r-R_h)(r^2-R_h^2) + \frac{\frac{1}{2}(R^2-R_h^2) + R^2\ln(R_c/R) - R_h\ln(R_c/R_h)}{\ln(R_c/R_w)}\right]. \tag{71}$$

Note that although the vorticity is discontinuous at $r = R$ and $r = R_h$, the velocity is continuous at both radii. The corresponding basic-state surface pressure satisfies $\mathrm{d}\overline{\pi}/\mathrm{d}r = f\overline{v} + \overline{v}^2/r$, but its exact form has no bearing on the stability of the flow.

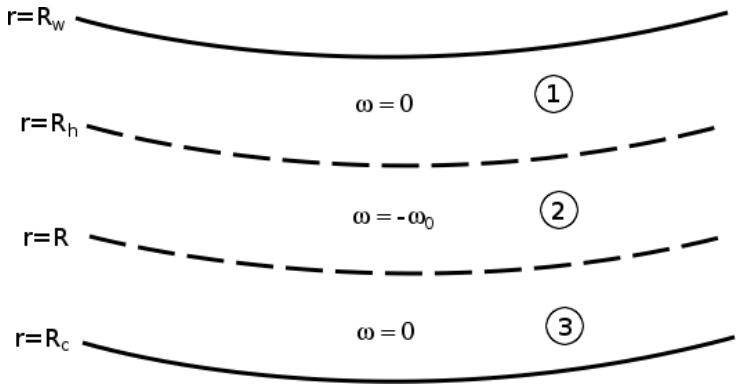


Figure 8: A strip of constant vorticity in a section of the annulus, approximating a long, slowly-varying topographic Rossby shelf wave.

We now consider infinitesimal perturbations to the basic state, setting $\mathbf{u} = \overline{\mathbf{u}} + \mathbf{u}'$ and $\pi = \overline{\pi} + \pi'$, where $\mathbf{u}' \ll \overline{\mathbf{u}}$ and $\pi' \ll \overline{\pi}$. Substituting these into the non-dissipative shallow water equations, (41a) and (41b), and linearising yields

$$\frac{\partial u_j'}{\partial t} + \frac{\overline{v}}{r}\frac{\partial u_j'}{\partial \theta} - \left(f + \frac{2\overline{v}}{r}\right)v_j' + \frac{\partial \pi_j'}{\partial r} = 0, \tag{72a}$$

$$\frac{\partial v_j'}{\partial t} + \frac{\overline{v}}{r}\frac{\partial v_j'}{\partial \theta} + \left(f + \frac{\overline{v}}{r} + \frac{\partial \overline{v}}{\partial r}\right)u_j' + \frac{1}{r}\frac{\partial \pi_j'}{\partial \theta} = 0, \tag{72b}$$

$$\nabla \cdot \mathbf{u}_j' = 0, \tag{72c}$$

for $j = 1, 2, 3$, which is valid everywhere except at $r = R_h$. We connect regions 1 and 2 across $r = R_h$ and regions 2 and 3 across $r = R$ by requiring that an interface propagating about either of them should satisfy kinematic boundary conditions and continuity of surface pressure. At $r = R$, these reduce to

$$u_2' = u_3', \qquad \pi_2' = \pi_3', \qquad \text{at} \quad r = R. \tag{73}$$

At $r = R_h$ we must also ensure that mass flux and surface pressure are continuous across the discontinuity in depth. An analysis similar to that performed in Section 6.1 shows that,

because the basic velocity $\overline{v}$ is continuous, these are exactly the jump conditions required to connect regions 1 and 2,

$$Hu_1' = (H - H_s)u_2', \qquad \pi_1' = \pi_2', \qquad \text{at} \quad r = R_h. \tag{74}$$

Taking the curl of (72a) and (72b) yields the vorticity equation

$$\frac{\partial \omega_j'}{\partial t} + \frac{\overline{v}}{r}\frac{\partial \omega_j'}{\partial \theta} + u_j'\frac{\partial \overline{\omega}}{\partial r} = 0. \tag{75}$$

The last term on the left hand side of this equation is equal to zero because $\overline{\omega}$ is piecewise-constant. Thus, if we seek plane wave solutions of the form $\psi_j' = \mathrm{Re}\{\hat{\psi}_j \exp[i(l\theta - \sigma t)]\}$, then (75) states that $\omega_j' = 0$ holds everywhere except at $r = R_h$.

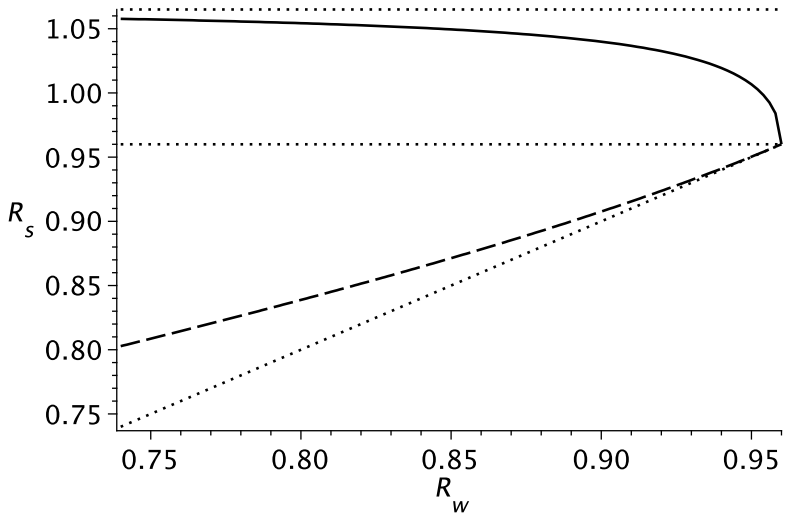


Figure 9: Plot of the stability radius R_s (in metres) for the case $R > R_h$ (solid line), defined such that if $R_h < R < R_s$ then the flow is unstable, but for $R_s < R < R_c$ the flow is stable. We also show the stability radius for the case $R < R_h$ (dashed line), defined such that if $R_s < R < R_h$ then the flow is unstable, but for $R_w < R < R_s$ the flow is stable. The dotted lines mark the positions of the outer wall R_c, the shelf line R_h and the inner wall R_w.

We may now obtain the complete solution using $\omega_j = \nabla^2 \psi_j' = 0$ in region j for $j = 1, 2, 3$, which has the solution

$$\psi_j' = \mathrm{Re}\left\{\left(A_j r^l + B_j r^{-l}\right) e^{i(l\theta - \sigma t)}\right\}, \tag{76}$$

where A_j and B_j are constants. We then apply the jump conditions (73) and (74), and the boundary conditions $\psi_1' = 0$ on $r = R_w$, $\psi_3' = 0$ on $r = R_c$. This determines all of the constants A_j, B_j in terms a single constant representing the amplitude of the wave, which remains arbitrary, and yields a dispersion relation for the wave frequency $\sigma(l)$. Unfortunately the calculation quickly becomes too complicated to obtain meaningful analytical results, and is best handled using symbolic computation packages and numerical calculation. We find that if there is no shelf ($\Delta_H = 1$) then, as in the case of a straight channel, the flow

is unstable to long waves ($l \to 0$) unless one of the discontinuities lies sufficiently close to the edge of the channel. The stability here is also independent of the vorticity ω_0 and the Coriolis parameter f. This is illustrated in Figure 9.

If we include a shelf ($\Delta_H < 1$) then, as we found in Section 6.1, long waves may be stabilised. In Figure 10 we compare $\sigma(l)$ between cases where $\Delta_H = 0.75$ and $\Delta_H = 1$, with all other parameters chosen to match the experiment presented in Figure 7. We see that the flow is stable to all wave numbers in the presence of the shelf, but is unstable to long waves when the shelf is absent. The most unstable wave in the latter case has a length of approximately $\pi/16$ radians, which almost exactly matches the the length of the instabilities observed in the bottom-left frame in Figure 7. This raises the possibility that the stability of the long nonlinear wave in this experiment, which utilised an absolute discontinuity in the bottom topography, is not affected by the shelf.

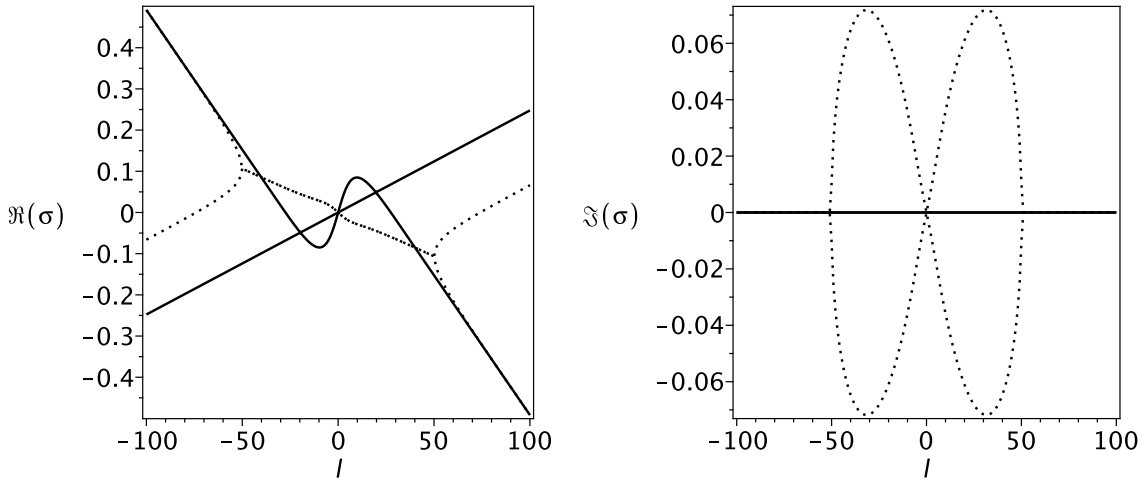


Figure 10: Plot of the wave frequency σ in rad s^{-1} as a function of the wave number l in rad^{-1} for linear perturbations to an annular strip of vorticity, where $R_w = 0.75\,\mathrm{m}$, $R_h = 0.96\,\mathrm{m}$, $R = 0.985\,\mathrm{m}$ and $R_c = 1.065\,\mathrm{m}$. The solid line corresponds to the wave frequency when there is a discontinuity in depth ($\Delta_H = 0.75$) at $r = R_h$, whilst the dotted line corresponds to the case of no discontinuity in depth ($\Delta_H = 1$). In both cases $f = 1.45\,\mathrm{rad\,s^{-1}}$ and $\omega_0 = 0.3625\,\mathrm{rad\,s^{-1}}$, corresponding the theoretical vorticity of the nonlinear topographic Rossby wave in Figure 7.

The analysis performed here is based on infinitesimal horizontal perturbations to the fluid directly above the discontinuity. To comply with this theory, fluid columns are required to cross the discontinuity rapidly and repeatedly, which is physically counter-intuitive, so it is not surprising to see that this analysis breaks down. If indeed the stability of long topographic Rossby shelf waves propagating about an absolute discontinuity in depth is determined by the theory for the case of a flat channel, then we should expect almost all such waves to be unstable, according to Figure 9. However, in both the real ocean and in the later laboratory experiments, the discontinuity in the theory is actually an approximation of a slope, whose width is assumed to be negligible compared to the amplitude of the

waves. This casts doubt upon the validity of the linear theory, which assumes waves of infinitesimally small amplitude. In order to accurately examine the stability of these waves, we would require a considerably more complex treatment of the problem that accounted for the slope connecting the regions of different depths across the shelf line.

7 Breaking Lee Waves

We shall now turn our attention to waves generated in the lee of a bump in the outer wall of the annulus. Of all of the experiments performed in the course of this project, these were the only ones that led to consistent and measurable results, as most were subject to rapidly growing instabilities similar to those described in Section 6.

In Figure 11 we present a specimen experiment in which a bump has been created on the inside edge of the tank using the deformable wall. This is compared with a numerical solution to the nonlinear wave equation (22), computed using a pseudo-spectral code. Once the fluid had reached solid-body rotation, we rapidly increased the rate of rotation of the tank from $f = 1.50\,\mathrm{rad\,s^{-1}}$ to $f = 1.60\,\mathrm{rad\,s^{-1}}$ for 6 seconds, and then returned the rotation rate to $f = 1.54\,\mathrm{rad\,s^{-1}}$. This resulted in the generation of a wave behind the bump and a mean flow of $\alpha_m \approx -0.02\,\mathrm{rad\,s^{-1}}$. This state is used as the starting point ($t = 0\,\mathrm{s}$) of the experiment for the purpose of numerical calculation, and corresponds to the top frame in each of the series of images in Figure 11.

In the laboratory experiment, both with and without a sloping shelf, the long lee wave breaks at the end furthest from the shelf after approximately 8 seconds. This is not possible in the numerical solution of the full nonlinear wave equation (22), as dispersion will always "smooth out" any large gradients. We therefore also present the numerical solution when the dispersive $\mathrm{O}(\mu)$ terms are excluded, visible only in the second frame of the numerical solution. In Figure 12 we compare the length of the lee wave at the point of breaking between the experiments and the non-dispersive computations for a range of mean flow speeds. We also plot the dispersive wavelengths at times correspoinding to the point of wave breaking in the experiments. These results suggest that the dispersive theory accurately predicts the wavelength for mean flow speeds less than -0.01 rad $\mathrm{s^{-1}}$, whilst the non-dispersive theory is more accurate for larger mean flow speeds. However, there is some uncertainty in the wavelength and exact time of breaking in the experiment. For $\alpha_m > 0.01\,\mathrm{rad\,s^{-1}}$ the wave is drawn under the bump in the wall before breaking can occur, whilst for $\alpha_m < -0.02\,\mathrm{rad\,s^{-1}}$ the wave tends to be swept away from the bump by the strong mean flow.

At later times the behaviour of the experiments and the numerical solution diverge from one another. In the presence of a steep slope, the wave grows in length and amplitude, building up behind the bump in the wall. Meanwhile, instabilities arise at the end of the wave furthest from the bump, and subsequently grow into short breaking waves. Thus the wave gradually sheds vorticity and its amplitude slowly decays. By contrast, in the presence of a discontinuous shelf the wave quickly loses energy and decays back down to the shelf line. In the numerical solution, the wave also build up behind the bump in the wall, but remains there because it can not shed vorticity. The dissipation introduced via the parameterisation of section 5 gradually slows all motion of the interface, such that the wave eventually comes to rest behind the bump. The behaviour described here was found to be similar for any negative mean flow, $\alpha_m < 0$. For $\alpha_m \geq 0$ the wave moves under the

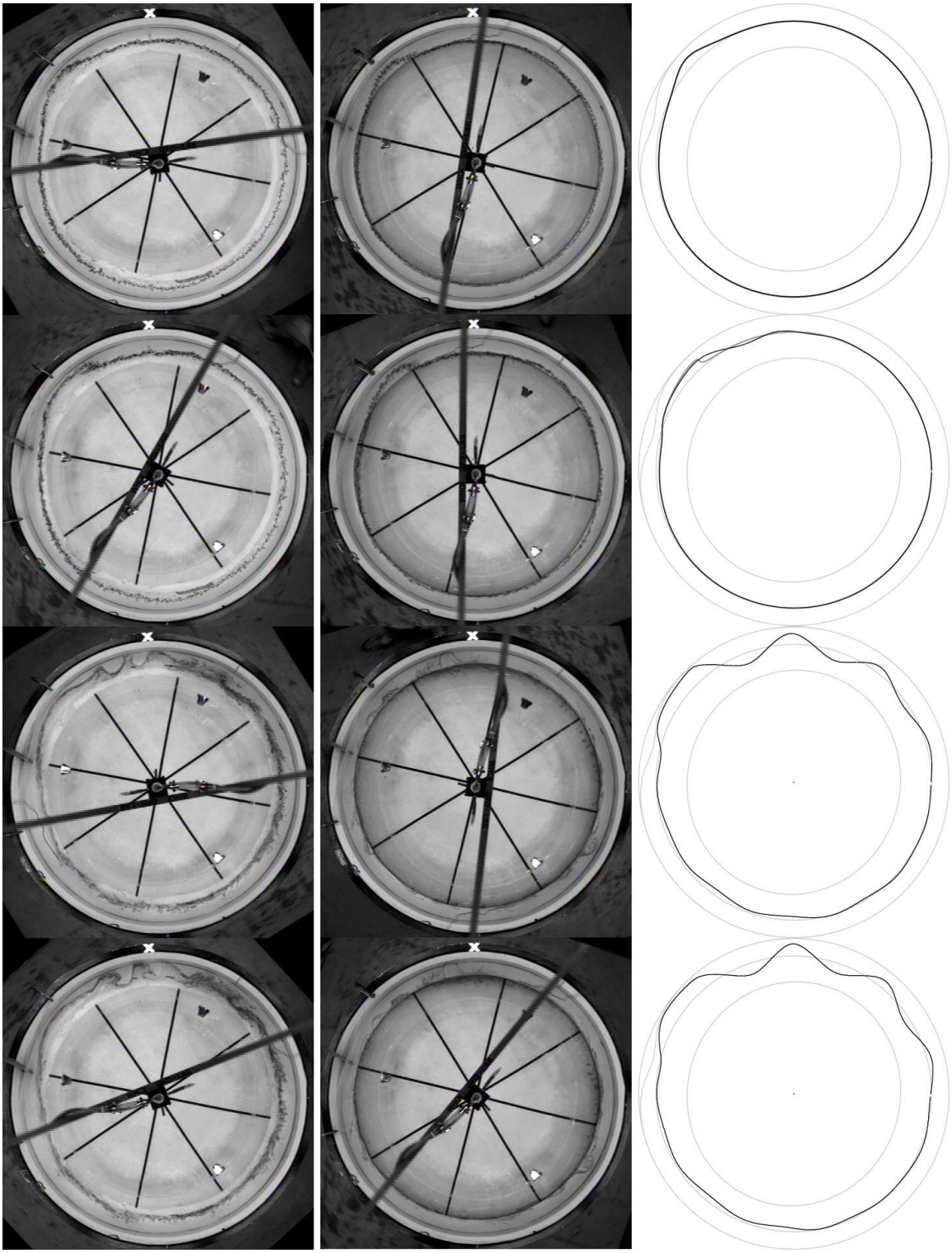


Figure 11: Snapshots of an experiment with a sloping shelf (left), and a discontinuous shelf (middle), and in a numerical solution (right) of the nonlinear wave equation with (black line) and without (dark grey line) its dispersive terms. The pictures have been taken 0, 8, 94 and 113 seconds after the generation of the lee wave.

bump, but friction around the shelf line causes it to decay rapidly as it does so.

As was mentioned in Section 2, the most serious deviation of the experiment from the theory was the use of an actual discontinuity in the depth. This experiment suggests that the fluid loses substantially more energy in crossing the discontinuous shelf than it does in crossing the steeply sloping shelf. Intuitively, this is because a Stewartson boundary layer will form at vertical wall of the shelf, whereas the sloping shelf will only have a bottom Ekman layer (see [22], p219). We therefore expect fluid columns to encounter more resistance as they cross the discontinuity than they do when they cross the slope.

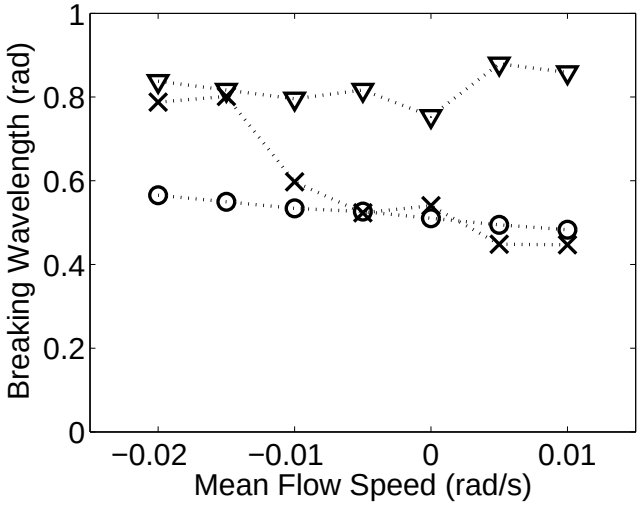


Figure 12: A comparison of the lengths of breaking lee waves in laboratory experiments (crosses) and nondispersive computations (circles), defined as the angular distance from the point of breaking to the front of the wave. In the dispersive computations (triangles), the length of the region of positive amplitude is calculated at the time of breaking in the corresponding laboratory experiment.

Between the dispersive and non-dispersive nonlinear wave equations, the main features of the flow in the sloping shelf experiment are captured quite well. The instability that arises in the lee waves may be due to a mechanism similar to that discussed in Section 6.2, but there is no reason to expect that linear analysis to apply to the case of a sloping shelf. Either way, it seems that short-wave processes invariably become important within a finite time.

8 Conclusion

This project was initially conceived as a means of testing the fundamental theory of nonlinear long Rossby shelf waves in a channel, but it has broadened to cover a much wider range of topics. The traditional theory for a straight channel has been redeveloped for application to an annular domain, and extended to include the effects of irregular walls, yielding a similar nonlinear wave equation (22). In the weakly nonlinear limit, the equation

is exactly the same as that found in the straight-channel case, but with slightly different coefficients, so the amplitude of weakly nonlinear waves is too small for them to be affected by the geometry of the annulus. The additional requirements that the waves must fit into the annular domain, and conserve mass about the shelf line, severely constrain the types of solution that are possible, and suggest that longer waves should exhibit the most nonlinear behaviour.

We have also considered the way that bottom friction might act to dissipate energy from the system. Even a simple representation of this dissipation in the quasigeostrophic shallow water equations is impossible to include exactly in our nonlinear wave model, so we constructed an alternative parametrisation on the basis of conservation laws for the global vorticity.

In the experiment itself, the waves generated were found to deviate from the long-wave theory in two important ways: they were able to break, and they were unstable to shorter waves. We explored possible mechanisms for the observed instabilities in Section 6, where a linear analysis showed that a mean flow about a discontinuity should be unstable, and that long waves of sufficiently small amplitude should themselves be unstable to long perturbations. A comparison with the experiment depicted in Figure 7 suggests that in fact the stability may be unaffected by the presence of the shelf line, explaining the observed instability at somewhat larger amplitudes.

The experiments discussed in Section 7 exhibited both breaking of long waves and instability to shorter waves that themselves subsequently went on to break. The fact that the instability arises at the rear of the lee waves, where the amplitude approaches zero, suggests an agreement with the linear stability theory. Meanwhile, the breaking of long waves suggests that in reality, dispersive effects are not strong enough to counteract the nonlinear steepening of the waves. Thus, although the nonlinear long-wave theory provides a good qualitative description of the behaviour, in practice short-wave phenomena tend to become just as important in a short time.

Acknowledgements

I would like to thank Ted Johnson, with whom I developed the ideas for the laboratory experiment and this project, for his supervision over the summer. I would also like to thank Karl Helfrich and Harvey Segur for useful meetings and discussions throughout the project. I am particularly grateful to Anders Jensen for putting much time and effort into helping me with the construction of this experiment. Finally, thanks to all of the fellows and staff for an inspiring and highly enjoyable summer at Woods Hole.

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, ninth Dover printing, tenth GPO printing ed., 1964.

[2] H. Bryden, L. Beal, and L. Duncan, *Structure and transport of the Agulhas Current and its temporal variability*, Journal of Oceanography, 61 (2005), pp. 479–492.

[3] S. Clarke and E. Johnson, *Finite-amplitude topographic Rossby waves in a channel*, Physics of Fluids, 11 (1999), pp. 107–120.

[4] ——, *The weakly nonlinear limit of forced Rossby waves in a stepped channel*, Proceedings: Mathematics, Physical and Engineering Sciences, (2001), pp. 2361–2378.

[5] P. Drazin and W. Reid, *Hydrodynamic Stability*, Cambridge University Press, Cambridge, 1981.

[6] A. Gill and E. Schumann, *Topographically induced changes in the structure of an inertial coastal jet: Application to the Agulhas Current*, Journal of Physical Oceanography, 9 (1979), pp. 975–991.

[7] P. Haynes, E. Johnson, and R. Hurst, *A simple model of Rossby-wave hydraulic behaviour*, Journal of Fluid Mechanics, 253 (1993), pp. 359–384.

[8] A. Hinds, I. Eames, E. Johnson, and N. McDonald, *Laboratory study of vortex dipoles interacting with step topography*, Journal of Geophysical Research-Oceans, 114 (2009), pp. C06006–+.

[9] A. Hinds, E. Johnson, and N. McDonald, *Vortex scattering by step topography*, Journal of Fluid Mechanics, 571 (2007), pp. 495–505.

[10] E. Johnson, *Topographic waves and the evolution of coastal currents*, Journal of Fluid Mechanics, 160 (1985), pp. 499–509.

[11] E. Johnson and S. Clarke, *Dispersive effects in Rossby-wave hydraulics*, Journal of Fluid Mechanics, 401 (1999), pp. 27–54.

[12] ——, *Rossby Wave Hydraulics*, Annual Review of Fluid Mechanics, 33 (2001), pp. 207–230.

[13] E. Johnson and M. Davey, *Free-surface adjustment and topographic waves in coastal currents*, Journal of Fluid Mechanics, 219 (1990), pp. 273–289.

[14] E. Johnson, A. Hinds, and N. McDonald, *Steadily translating vortices near step topography*, Physics of Fluids, 17 (2005), pp. 056601–+.

[15] E. Johnson and N. McDonald, *Surf-zone vortices over stepped topography*, Journal of Fluid Mechanics, 511 (2004), pp. 265–283.

[16] A. Leonard, *Vortex methods for flow simulation*, Journal of Computational Physics, 37 (1980), pp. 289–335.

[17] M. Longuet-Higgins, *On the trapping of waves along a discontinuity of depth in a rotating ocean*, Journal of Fluid Mechanics, 31 (1968), pp. 417–434.

[18] F. Marche, *Derivation of a new two-dimensional viscous shallow water model with varying topography, bottom friction and capillary effects*, European Journal of Mechanics/B Fluids, 26 (2007), pp. 49–63.

[19] L. Mysak, *Recent advances in shelf wave dynamics*, Reviews of Geophysics, 18 (1980), pp. 211–241.

[20] ——, *Topographically trapped waves*, Annual Review of Fluid Mechanics, 12 (1980), pp. 45–76.

[21] L. Mysak, P. Leblond, and W. Emery, *Trench waves*, Journal of Physical Oceanography, 9 (1979), pp. 1001–1013.

[22] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer, 1987.

[23] R. Salmon, *Lectures on geophysical fluid dynamics*, Oxford University Press, USA, 1998.

[24] G. Vallis, *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*, Cambridge University Press, 2006.

[25] G. Whitham, *Linear and nonlinear waves*, Wiley-Interscience, 1974.

Nonlinear peristaltic waves: a bitter pill to swallow

Daisuke Takagi

September 28, 2009

Abstract

Nonlinear waves of fluid are driven in an elastic tube by imposing a radial force of sinusoidal form. The governing equations of the deformation of the tube and the flow rate inside the tube are derived using linear elasticity theory and lubrication theory. Steady and periodic solutions in the reference frame of a steadily propagating wave are obtained by either asymptotic theory in the two limits of small and large forcing amplitudes or numerical techniques for moderate forcing amplitudes. A strongly deformed tube of Newtonian fluid is shown to feature an occluded region and a peak region, which depends importantly on the elastic properties of the tube and weakly on the large forcing amplitude. The flow rate inside the tube reduces significantly when the fluid has a yield stress, as investigated using a Bingham plastic model. The flow of Newtonian fluid containing a rigid rod in the tube shows that a maximal speed of the rod is attained by imposing a radial force of moderate amplitude. The rod generally declines in speed with increasing radius, suggesting that the python, which must take in food by peristalsis without grinding into smaller pieces, has a bitter pill to swallow.

1 Introduction

Fluid inside a deformable tube can be driven by the mechanism of peristaltic action. Many biological ducts convey contents, including a bolus in the gastronomical duct and urine in the ureter, by propagating waves of muscular contraction and relaxation. A python is able to swallow prey of considerable size this way. A related type of flow in a deformed tube occurs in peristaltic pumps, which are used for the infusion of medication into the circulatory system and the treatment of wastewater in the environment, amongst many other applications. Pushing toothpaste out of its tube is another example of relevance.

Mathematical models of peristaltic motion can be developed using lubrication theory, provided that effects due to fluid inertia are negligible. The low-Reynolds-number flow of Newtonian fluid was described in an axisymmetric tube with either a sinusoidal (Shapiro et al., 1969) or non-sinusoidal (Lykoudis and Roos, 1970) deformation in its radius. The flow of non-Newtonian fluid

was studied in a similar fashion, given a small deformation in the tube radius (Frigaard and Ryan, 2004; Vajravelu et al., 2005). Another variation is to introduce a peripheral layer of Newtonian fluid adjacent to the wall, which has a different viscosity from that of the inner fluid (Brasseur et al., 1987). These models prescribe the tube deformation without taking elastic properties of the tube into consideration.

Of interest is a situation where the shape of the tube is unknown a priori and must be solved as part of the problem, given some coupling between the hydrodynamics and the mechanics of the elastic tube. Related free-boundary problems of flow near an elastic material arise in a range of contexts, including the swimming of a microorganism near a rigid wall (Argentina et al., 2007; Balmforth et al., 2009) and premelting of ice in a deformable capillary (Wettlaufer and Worster, 1995; Wettlaufer et al., 1996). In the context of peristaltic motion with a prescribed activation wave of muscular contraction, where elastic tubes result in finite-amplitude deformations, tubes containing Newtonian fluid (Carew and Pedley, 1997) and a rigid bolus (Bertuzzi et al., 1983) have been solved numerically. Analytic solutions are desirable for gaining a deeper understanding of peristaltic flow inside a strongly deformed tube, where the response in tube radius is a nonlinear function of the forces driving the flow.

Here, theoretical models of peristaltic flow are developed, given a sinusoidal wave of radial force of arbitrary amplitude that translates along an elastic tube. In the reference frame of the wave, steady and periodic solutions are obtained to describe the motion of three different materials inside the tube. In section 2, the flow of Newtonian fluid is investigated inside a linearly elastic tube and separately inside a tube of finite bending stiffness. In section 3, the flow governed by a Bingham plastic model is considered, which exhibits the dual behaviour of a fluid and a solid. Familiar examples of non-Newtonian fluid described by the Bingham model include mud, paint, slurry, and toothpaste. In section 4, the propulsion of a rigid body in a Newtonian fluid is examined. The coupled motion of the solid and the fluid provides useful insight into the flow of pills in the gastronomical duct, kidney stones in the ureter, and blood cells in small blood vessels. In each section, the governing equations are derived and solved using asymptotic theory in the two limits of small and large forcing amplitudes, which give rise to linear and nonlinear responses in tube radius respectively. The theoretical results are complemented by numerical solutions that describe responses to moderate forcing amplitudes.

2 Newtonian fluid

Consider a Newtonian fluid of density ρ and dynamic viscosity μ with pressure p_0, inside a cylindrical tube of constant radius R in its undeformed state. A radial force per unit area of sinusoidal form $F(z-ct) = \eta \sin[(z-ct)2\pi/L]$ is applied on the tube along the axial coordinate z at time t. The imposed force, characterised by its amplitude η, steady speed c, and wavelength L, perturbs the pressure p inside the tube and the tube radius a, as sketch in figure 1. The

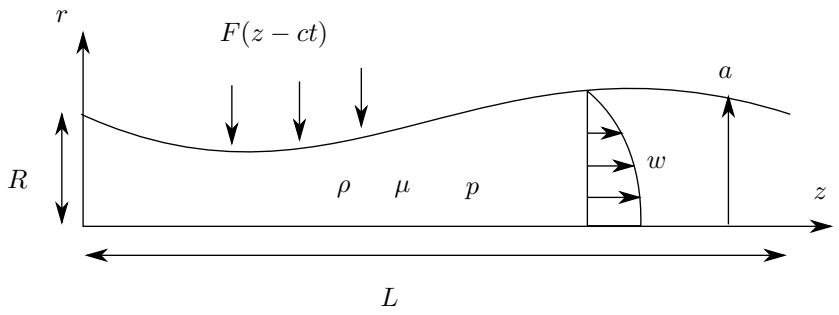


Figure 1: Schematic sketch in cylindrical polar coordinates of a deformed tube filled with Newtonian fluid. A prescribed radial force F perturbs the tube of radius a and induces flow with axial velocity w.

velocity $\mathbf{u}$ of the induced flow of fluid is governed by

$$\rho\frac{D\mathbf{u}}{Dt} = -\nabla p + \mu\nabla^2\mathbf{u}, \tag{1}$$

subject to the condition of incompressibility,

$$\nabla\cdot\mathbf{u} = 0. \tag{2}$$

The elastic properties of the tube are such that the change in pressure across the tube is of the form

$$\Delta p|_{r=a} = D\frac{\partial^n}{\partial z^n}(a-R) + F, \tag{3}$$

where $n = 0, 4$ characterises the type of elastic material constituting the tube and D denotes its stiffness. The case of $n = 0$ corresponds to a linearly elastic tube such that its deformation is proportional to the net radial force. For example, the tube could be attached by springs to a rigid surrounding backing, where D is the spring constant. The case of $n = 4$ corresponds to a thin shell of bending stiffness $D = h^3E/12(1-\nu^2)$, where h is the shell thickness, E the Young modulus and ν the Poisson ratio (Love, 1944). The objective is to determine the shape of the tube and the volumetric flow rate q per wave period, L/c.

It is convenient to introduce dimensionless variables of axial coordinate $z' = z/L$, radial coordinate $r' = r/R$, time $t' = tc/L$, pressure $p' = pR/\mu c$, tube stiffness $D' = DR/\mu c$, amplitude of forcing $\eta' = \eta R/\mu c$, and flow rate $q' = q/\pi R^2 c$. All quantities from here onwards are dimensionless, without the primes for simplicity. In addition, it is convenient to formulate the problem in the reference frame of the wave, moving at speed 1 in the z direction. The advantage of moving with the wave is that the tube radius is steady in time in this reference frame, as justified below. The axial velocity w and radial velocity u are governed by (1)

$$\delta^2 Re\frac{Du}{Dt} = -\frac{\partial p}{\partial r} + \delta^2\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \delta^4\frac{\partial^2 u}{\partial z^2} \tag{4}$$

and

$$Re\frac{Dw}{Dt} = -\frac{\partial p}{\partial z} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial w}{\partial r}\right) + \delta^2\frac{\partial^2 w}{\partial z^2}, \tag{5}$$

where

$$\delta \equiv \frac{R}{L} \tag{6}$$

is the aspect ratio of the region of interest and

$$Re \equiv \frac{\rho c R^2}{\mu L} \tag{7}$$

is the Reynolds number. The deformation of the tube is small provided that $\delta \ll 1$ under the long-wave approximation. Inertia is negligible compared to viscous forces provided that $Re \ll 1$. It is natural to seek steady solutions in the wave frame because the explicit dependence on time is dropped by neglecting effects due to inertia. Under these approximations of lubrication theory, the leading-order equation of (4) indicates that the pressure inside the tube is uniform in the radial direction. Consequently, $\Delta p|_{r=a} = p - p_0$ so

$$\frac{dp}{dz} = D\frac{d^{1+n}a}{dz^{1+n}} + \eta\cos z, \tag{8}$$

by differentiating (3) with respect to z. The leading-order equation of (5) gives rise to a second-order differential equation for the axial velocity w,

$$\frac{dp}{dz} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial w}{\partial r}\right). \tag{9}$$

The associated boundary conditions are $w = -1$ on $r = a$, by the condition of no slip on the tube wall, and $\partial w/\partial r = 0$ on $r = 0$, by either regularity or axisymmetry of the flow. Note that $w = -1$ in the wave frame corresponds to no axial flow in the lab frame. Integrating (9) twice and imposing the two boundary conditions yield the velocity profile in the wave frame,

$$w = \frac{1}{4}\frac{dp}{dz}(r^2 - a^2) - 1. \tag{10}$$

Streamlines are determined by contours of the streamfunction

$$\psi = -\frac{1}{16}\frac{dp}{dz}r^2(2a^2 - r^2) - \frac{1}{2}r^2, \tag{11}$$

which is obtained by integrating $w = r^{-1}\partial\psi/\partial r$.

A measure of the proportion of fluid propagating with a wave is given by the time-averaged flow rate in the lab frame, denoted by q. The flow rate in the lab frame at any axial coordinate is obtained by integrating the axial velocity in the lab frame over the cross-section of the tube, $2\int_0^a (w+1)r\,dr$. Its average in time, which is equivalent to its average in z over 2π, is computed by using (10) and the global conservation of fluid volume

$$\langle a^2 \rangle = 1, \tag{12}$$

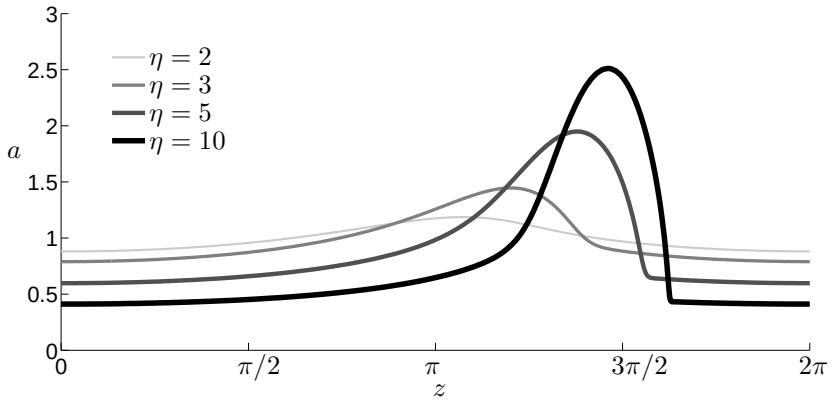


Figure 2: Numerical solutions of the tube radius perturbed by different forcing amplitudes η. The tube is of type $n = 0$ and has stiffness $D = 1$. As η increases, the deformation of the tube increases.

where $\langle\cdot\rangle \equiv (2\pi)^{-1}\int_0^{2\pi}\cdot\, dz$. The time-averaged flow rate in the lab frame is given by

$$q = -\frac{1}{8}\frac{dp}{dz}a^4 - a^2 + 1, \tag{13}$$

where dp/dz is given by (8). When $n = 0$ in (8), the first-order differential equation (13) must be solved subject to the periodic boundary condition $a(0) = a(2\pi)$. When $n = 4$ in (8), the fifth-order differential equation (13) must be solved subject to periodic boundary conditions $d^i a/dz^i(0) = d^i a/dz^i(2\pi)$ for $i = 0, 1, 2, 3, 4$. A boundary-value problem with an eigenvalue to be determined, q, must be solved to determine the radius of the tube.

The system of equations, (12) and (13), can be solved numerically using a built-in function of Matlab called bvp4c. The forcing amplitude η is incremented slowly from 0 with $a = 1$ to obtain a solution during each iteration, which forms an initial guess for solutions with successive increments of η. A representative set of solutions for the tube radius in response to different forcing amplitudes η is plotted in figure 2. The associated solutions of the eigenvalue q are presented later. For small η, the tube deformation is small as expected. As η increases, the tube is occluded except near $z = 3\pi/2$, where its radius peaks due to the imposed force that is maximal and radially outward. Streamlines of the flow induced with forcing amplitudes $\eta = 2$ and $\eta = 3$ are shown in figures 3 and 4, obtained using equation (11). As η increases, a qualitative change in the structure of the streamlines is observed, from axial flow that is everywhere negative in the wave frame to the development of a recirculating zone where the tube peaks in radius. Fluid inside the recirculating zone propagates with the wave.

In the limit as $\eta \to 0$, the tube is almost undeformed and flow inside the tube is expected to be negligible. In the limit as $\eta \to \infty$, the trapped core near

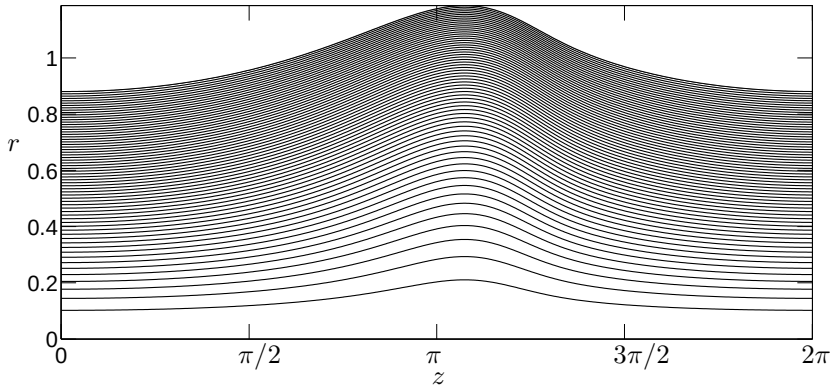


Figure 3: Streamlines of flow driven by a radial force of amplitude $\eta = 2$. In the wave frame, the axial velocity inside the tube is everywhere negative.

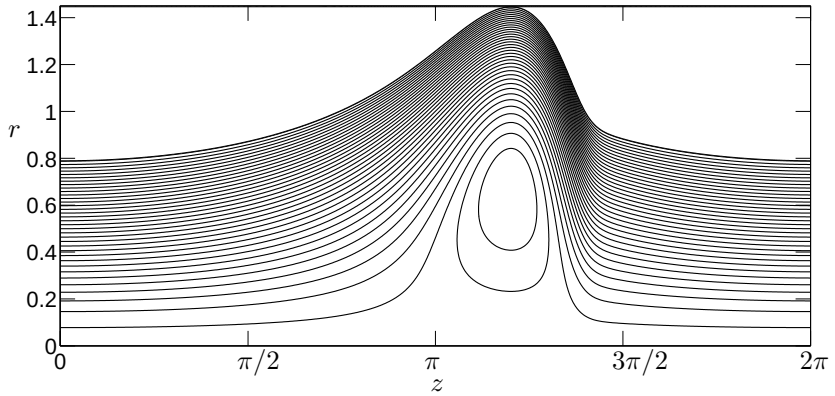


Figure 4: Streamlines of flow driven by a radial force of amplitude $\eta = 3$. In the wave frame, a trapped core, recirculating anti-clockwise, forms in the region where a peak radius of the tube is attained.

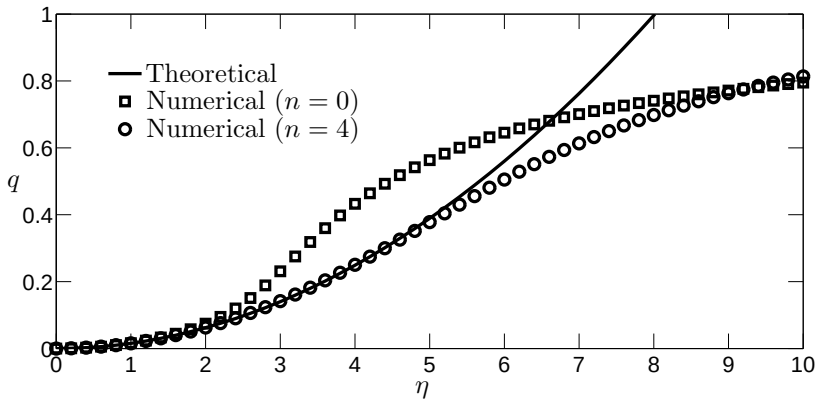


Figure 5: Theoretical prediction and numerical solutions of the time-averaged flow rate in the lab frame for small forcing amplitudes η. The numerical solutions are obtained with tube stiffness $D = 1$ and agree with the theoretical prediction in the asymptotic limit as $\eta \to 0$, independent of the type of tube, $n = 0, 4$.

$z = 3\pi/2$ is expected to increase in size and allow most of the fluid in the tube to propagate with the wave. The two limits of small and large η are investigated separately using asymptotic theory.

In the small-amplitude limit, $\eta \ll 1$, it is fruitful to seek series solutions of a and q about the base state, $a = 1$ and $q = 0$. The subsequent term in a is of order η, which indicates that the leading term in q is of order η^2, by operating $\langle\cdot\rangle$ on (13) and using the integral constraint (12). The radius of the tube responds linearly to the amplitude of the external forcing. Substituting $a = 1 + \eta a_1 + o(\eta^2)$ into (13) gives

$$D\frac{d^{1+n}a_1}{dz^{1+n}} - 16a_1 = \cos z, \tag{14}$$

which is solved subject to periodic boundary conditions. The solution is given by

$$a = 1 - \eta\frac{16\cos z - D\sin z}{16^2 + D^2} + o(\eta^2), \tag{15}$$

independent of the type of tube characterised by $n = 0, 4$ because $\sin z$ and $\cos z$ are invariant under four differentiations. Operating $\langle\cdot\rangle$ on (13), which gives $q = -\langle(dp/dz)a^4/8\rangle$, and imposing periodic boundary conditions at $z = 0, 2\pi$, yield an expression for the time-averaged flow rate,

$$q = \frac{4\eta^2}{16^2 + D^2} + o(\eta^3), \tag{16}$$

which is in agreement with numerical results for small η as shown in figure 5. Note that the theoretical prediction given by (16) in the asymptotic limit as $\eta \to 0$ agrees with numerical results up to $\eta \approx 2$ inside a tube of type $n = 0$

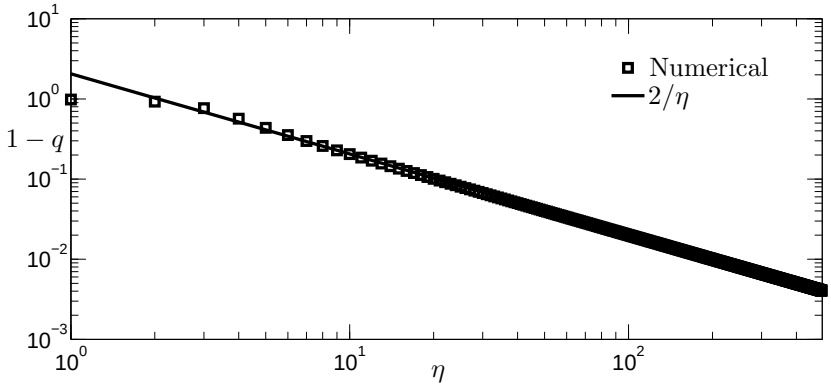


Figure 6: Theoretical prediction and numerical solutions of the characteristic proportion of fluid left behind the wave for large η, plotted on logarithmic scales. The tube is of type $n = 0$ and has stiffness $D = 1$.

and up to $\eta \approx 5$ inside a tube of type $n = 4$. The numerical results in the figure suggest that q approaches 1 as η increases, which is investigated below.

In the large-amplitude limit, $\eta \gg 1$, two qualitatively distinct regions develop, one near $z = 3\pi/2$ where the tube radius peaks, and the other away from $z = 3\pi/2$ where the tube is occluded. Quantitative details of the two regions must be considered separately in the two types of tubes, $n = 0$ and $n = 4$. Of interest is the quantity $1 - q$, a measure of the proportion of fluid left behind the wave.

When $n = 0$, the tube radius is expected to approach 0 in the occluded region so the dominant contribution to dp/dz in (13) arises from the second rather than the first term on the right hand side of (8). The governing differential equation (13) reduces to an algebraic equation

$$-\eta \cos z \, a^4 - 8a^2 + (1 - q) = 0. \tag{17}$$

This quadratic equation in a^2 results in two branches,

$$a^2 = \frac{4}{\eta \cos z} \left(-1 \pm \sqrt{1 + \frac{\eta}{2}(1 - q) \cos z} \right), \tag{18}$$

where $\pm$ is either positive in the lower branch or negative in the upper branch. Only the lower branch corresponds to a real tube radius throughout the domain $0 \le z \le 2\pi$. However, its gradient is unphysically discontinuous at $z = \pi$, which must be resolved by switching smoothly to the upper branch. The condition that the lower and upper branches meet at $z = \pi$ requires

$$q = 1 - 2\eta^{-1} + o(\eta^{-2}). \tag{19}$$

The theoretical prediction for the flow rate given by (19) is in excellent agreement with numerical results, as shown in figure 6. Note that the characteristic

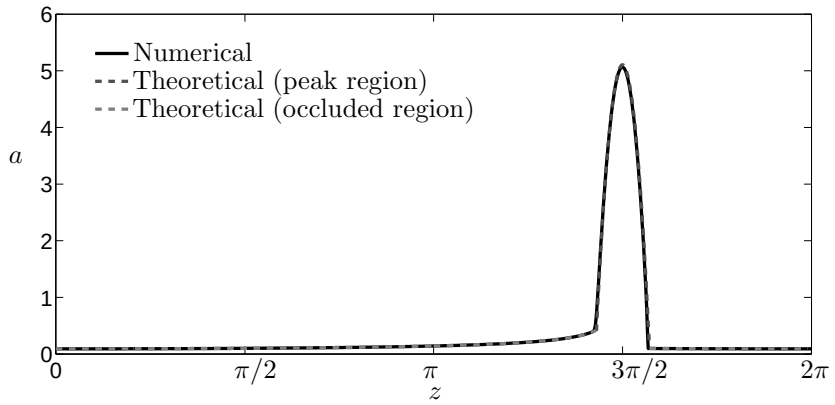


Figure 7: Theoretical and numerical solutions of the tube radius deformed with forcing amplitude $\eta = 200$. The theoretical solution, given by (21) in the peak region and (20) in the occluded region, is in excellent agreement with the numerical solution obtained using input parameters $n = 0$ and $D = 1$.

proportion of fluid left behind the wave is determined solely by the occluded region, independent of the peak region to be examined later. Substituting (19) into (18) gives the solution for the tube radius in the occluded region,

$$a = \begin{cases} 2\left(1 - \sqrt{2}|\cos z/2|\right)^{-1/2}\eta^{-1/2} + o(\eta^{-1}) & \pi \le z < 3\pi/2, \\ 2\left(1 + \sqrt{2}|\cos z/2|\right)^{-1/2}\eta^{-1/2} + o(\eta^{-1}) & 0 \le z < \pi,\ 3\pi/2 < z \le 2\pi, \end{cases} \tag{20}$$

which is continuous at $z = \pi$ but discontinuous at $z = 3\pi/2$. The radius of the tube diverges as $z \to 3\pi/2$ from below and converges to $\sqrt{2/\eta}$ from above. The divergence and discontinuity represent the formation of a shock, which is resolved mathematically by considering a boundary layer near $z = 3\pi/2$, the peak region.

In the peak region, the tube radius is expected to grow arbitrarily with forcing amplitude η, suggesting that only the first term on the right hand side of (13) is dominant. This means that $dp/dz = 0$ so the pressure in the peak region is uniform. It can be shown, by combining equations (8) and (12), where $\cos z$ in (8) is expanded near $z = 3\pi/2$, that the rescaled tube radius is given by $A = \eta^{-1/5}a$ and has a peak A_{max} at $\zeta = 0$, where ζ is the rescaled axial coordinate in the peak region such that $z = 3\pi/2 + \eta^{-2/5}\zeta$. The rescaled equation of (8) reduces to $D\,dA/d\zeta + \zeta = 0$ and is integrated to obtain

$$A = A_{max} - \frac{\zeta^2}{2D}. \tag{21}$$

The tube radius in the peak region has a parabolic profile, provided that the tube has stiffness $D > 0$. The peak radius of the tube is given by $a_{max} = \eta^{1/5}A_{max}$

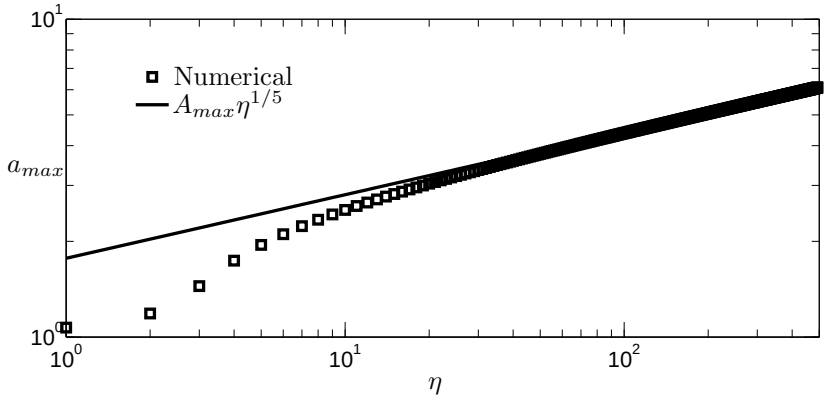


Figure 8: Theoretical prediction and numerical solutions of the peak radius of the tube of type $n = 0$ and stiffness $D = 1$, plotted on logarithmic scales.

with

$$A_{max} = \left(\frac{15^2\pi^2}{2^7 D^5}\right)^{1/5}, \tag{22}$$

which is obtained by solving the rescaled equation of (12), $\int_{-\delta}^{\delta} A^2\, d\zeta = 2\pi$, where $\delta = \sqrt{2DA_{max}}$ corresponds to the half-width of the peak region. The occluded region given by (20) and the peak region given by (21) agree with numerical results, demonstrated in figure 7 for a tube of stiffness $D = 1$, forced with amplitude $\eta = 200$. The scaling $\eta^{1/5}$ and the prefactor (22) for the peak radius of the tube also agree with numerical results as shown in figure 8. The form of a_{max} indicates that the peak radius depends importantly on the tube stiffness but depends weakly on the forcing amplitude for large η.

When the type of tube is characterised by $n = 4$, different scalings of the flow are obtained in both the occluded and peak regions. In the peak region, the pressure is uniform and the rescaled equations are obtained by combining equations (8) and (12), as before. The rescaled tube radius is given by $A = \eta^{-1/13}a$ and satisfies $D\, d^5A/d\zeta^5 + \zeta = 0$, subject to the boundary conditions that A, $dA/d\zeta$, and $d^2A/d\zeta^2$ all vanish near the ends of the peak region, as $\zeta \to \pm\delta$, where $z = 3\pi/2 + \eta^{-2/13}\zeta$. These boundary conditions ensure that the peak region matches smoothly to the occluded region, which is considered later. Integrating the fifth-order differential equation fives times gives the radius of the tube in the peak region,

$$a = \eta^{1/13}\frac{1}{6!D}\left(\delta^2 - \zeta^2\right)^3, \tag{23}$$

where

$$\delta = (\pi D^2 6!^2)^{1/13}\left(\frac{11}{3} - \frac{20}{7} - \frac{6}{11} + \frac{1}{13}\right)^{-1/13} \tag{24}$$

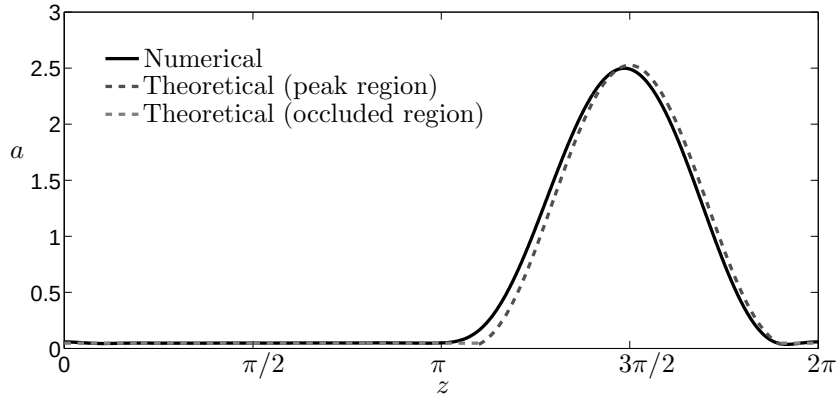


Figure 9: Theoretical and numerical solutions of the tube radius deformed with forcing amplitude $\eta = 200$. The theoretical prediction, given by (23) in the peak region, agrees with the numerical solution obtained using input parameters $n = 4$ and $D = 1$.

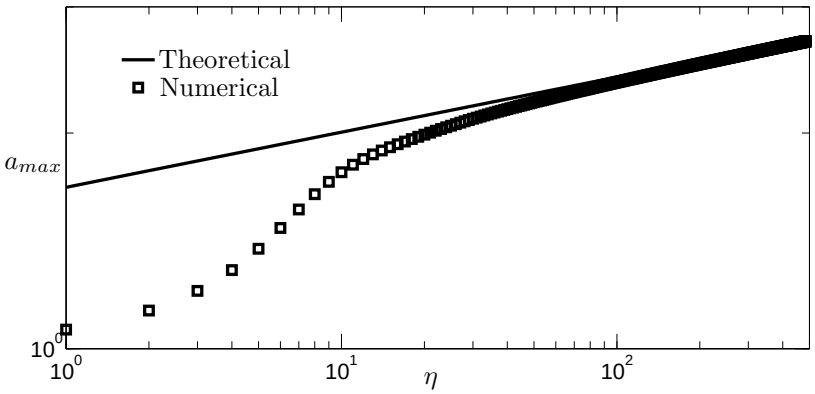


Figure 10: Theoretical prediction and numerical solutions of the peak radius of the tube of type $n = 4$, plotted on logarithmic scales.

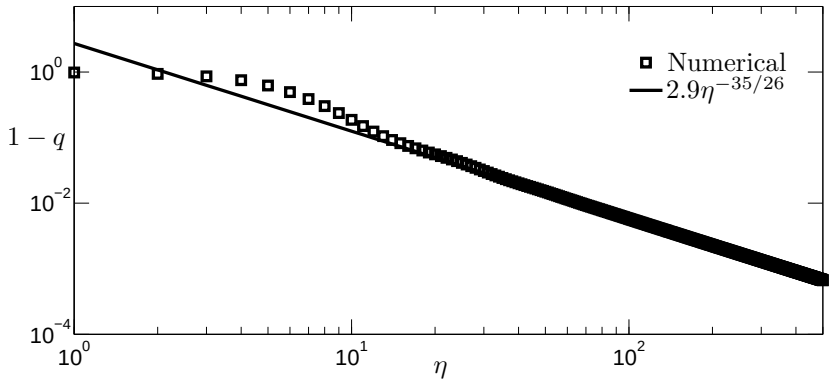


Figure 11: Theoretical prediction and numerical solutions of the characteristic proportion of fluid left behind the wave inside a tube of type $n = 4$, plotted on logarithmic scales.

is obtained using the rescaled integral constraint (12), $\int_{-\delta}^{\delta} A^2\, d\zeta = 2\pi$. The peak region agrees with numerics, demonstrated by figure 9 for $\eta = 200$, $D = 1$. The maximal radius of the tube at $z = 3\pi/2$, $\eta^{1/13}\delta^6/6!D$, agrees with the numerics as shown in figure 10. Note that the forcing amplitude must increase by thirteen orders of magnitude to deform the tube radius by one order of magnitude. A comparison of figures 7 and 9 indicates that for the same value of D and large η, the peak region of the tube of type $n = 4$ is longer and less deformed than that of the tube of type $n = 0$. The extremely weak dependence of the tube deformation on the applied force in the limit of large η is attributed to large amount of energy that is dissipated due to viscous forces in the highly occluded region, instead of doing mechanical work to deflect and expand the peak region of the tube.

In the occluded region away from $z = 3\pi/2$, the numerical solution in figure 9 suggests that the tube radius rapidly approaches zero with increasing η. This implies that keeping the dominant terms of (13) reduces to $a^2 = 1 - q$. The occluded region has a constant radius. In contrast to the previous problem inside a tube of type $n = 0$, in which q was determined by the occluded region, q in a tube of type $n = 4$ must be obtained by matching the occluded and peak regions using matched asymptotic expansions. It can be shown in the matched region near $z = 3\pi/2 \pm \eta^{-2/13}\delta$ that all terms are dominant in (13), except the contribution to dp/dz from $\eta \cos z$, which drops to leading order. A scaling analysis of the reduced equation of (13) and its boundary conditions in the matched region indicates that $1-q = o(\eta^{-35/26})$. The prefactor of this scaling is approximately 2.9 because numerical values of $(1-q)\eta^{35/26}$ quickly approach this value with increasing η. This theoretical prediction agrees well with numerical results as shown in figure 11. Note that the characteristic proportion of fluid left behind the wave approaches 0 more quickly in the limit of $\eta \to \infty$ in a tube of type $n = 4$ than $n = 0$. A tube of type $n = 4$ has a smaller occluded region

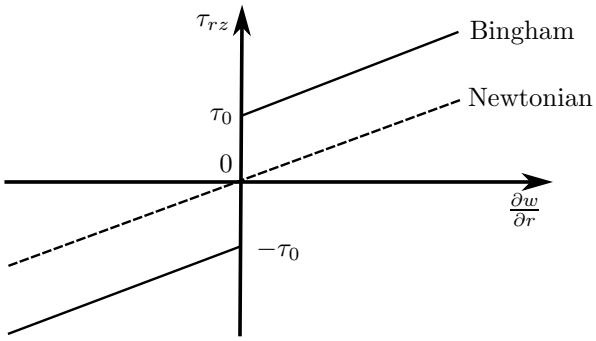


Figure 12: Stress-strain relationships of a Newtonian fluid and a Bingham plastic. The Newtonian curve has a slope corresponding to its viscosity. The Bingham curve has a stress-intercept corresponding to the yield stress and a slope corresponding to the plastic viscosity.

than a tube of type $n = 0$, implying a trapped core of larger size in the peak region that allows more fluid to propagate with the wave.

It has been determined that a non-zero radial force always results in deformation of a tube of Newtonian fluid. This is no longer the case when the tube contains fluid with a yield stress, where sufficient force on the tube must be applied to induce any motion. The problem of pumping a Bingham plastic, which is a type of non-Newtonian fluid with a yield stress, is considered in the following section. For simplicity, the tube is of type $n = 0$ in the remaining sections.

3 Bingham plastic

Consider a Bingham fluid with yield stress τ_0 and plastic viscosity μ. The Bingham fluid behaves either like a Newtonian fluid of viscosity μ in regions where the shear stress exceeds the yield stress, or like a plug without deformation in regions where the shear stress is below the yield stress. The stress-strain relationship of a Bingham fluid is compared with that of a Newtonian fluid in figure 12. The introduction of the yield stress modifies the force balance in the axial direction (9) to the system of equations

$$\frac{dp}{dz} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\tau_{rz}\right) \tag{25}$$

and

$$\frac{\partial w}{\partial r} = \begin{cases} \tau_{rz} - sgn\left(\frac{\partial w}{\partial r}\right) B & |\tau_{rz}| > B, \\ 0 & |\tau_{rz}| < B, \end{cases} \tag{26}$$

where

$$B \equiv \frac{\tau_0 R}{\mu c} \tag{27}$$

is the dimensionless yield stress, also known as the Bingham number. Equation (26) is the one-dimensional approximation of the Bingham constitutive law that applies to the current problem in a slender geometry (Vajravelu et al., 2005). The limit as $B \to 0$ reduces to the Newtonian problem encountered in section 2.

The boundary conditions of the axial velocity w are the same as in the Newtonian problem. The solution, satisfying the no-slip condition on the tube wall and regularity at $r = 0$, is given by

$$w = \begin{cases} -\frac{1}{4}\frac{dp}{dz}[(a^2 - r^2) - 2r_0(a - r)] - 1 & r > r_0 \\ -\frac{1}{4}\frac{dp}{dz}(a - r_0)^2 - 1 & r \leq r_0, \end{cases} \tag{28}$$

where

$$r_0 = \min\left[2B/|dp/dz|, a\right] \tag{29}$$

is the radius of the plug-like region inside the tube. The velocity profile is either parabolic for $r > r_0$ or uniform for $r \leq r_0$. The central region of the tube, bounded by radius r_0, is actually a pseudo-plug because the associated velocity profile is flat in radius and only appears to be below the yield stress to leading order (Balmforth and Craster, 1999). Indeed, this region is not truly rigid because axial velocity variations remain, except if $r_0 = a$, in which case the fluid spanning that section of the tube becomes truly rigid.

The time-averaged flow rate in the lab frame, $q = 2\langle\int_0^a (w + 1)r\,dr\rangle$, is computed from the axial flow velocity w in the wave frame, as in section 2. The expression for q is given by

$$q = -\frac{1}{24}\frac{dp}{dz}(a - r_0)^2\left((a + r_0)^2 + 2a^2\right) - a^2 + 1, \tag{30}$$

where dp/dz is given by (8) with $n = 0$. The expression on the right hand side of (30) without the two final terms represents the steady flow rate in the wave frame and is equivalent to the Buckingham-Reiner equation (Bird et al., 1987). In the limit as $B \to 0$, the flow rate in (30) with $r_0 = 0$ reduces to (13) and recovers the flow rate of Newtonian fluid, as expected.

The tube radius is determined by solving the governing equations as a boundary-value problem in the domain $0 \leq z \leq 2\pi$. Equation (30), subject to the periodic boundary condition $a(0) = a(2\pi)$, contains an eigenvalue q and must be solved subject to the integral constraint (12). A sample solution is shown in figure 13 and features three qualitatively distinct regions. A sheared region, $r_0 < r < a$, forms near the tube wall where the tube is most deformed. A region of pseudo-plug, $r < r_0 < a$, forms inside the sheared region. A solid region, $r < r_0 = a$, occupies the entire cross-section of the tube where its radius is uniform in z. This means that at any extent along the tube in the lab frame, a stationary region develops during an interval of time, which begins after the departure of a sheared region and ends on arrival of another sheared region.

In the limit of small forcing amplitude η, solutions of the form $a = 1 + \eta a_1 + o(\eta^2)$ and $q = \eta^2 q_2 + o(\eta^3)$ are sought, as in section 2. In a tube of type $n = 0$

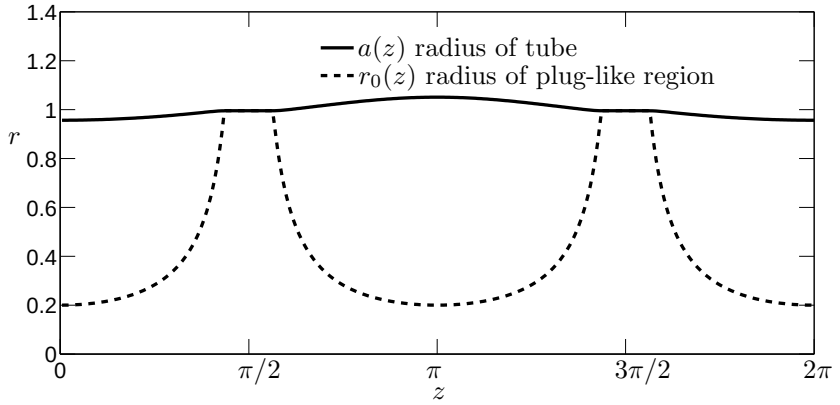


Figure 13: Numerical solution of the tube radius and the plug-like region inside a Bingham fluid. The input parameters are $n = 0$, $D = 0$, $\eta = 1$, $B = 0.1$.

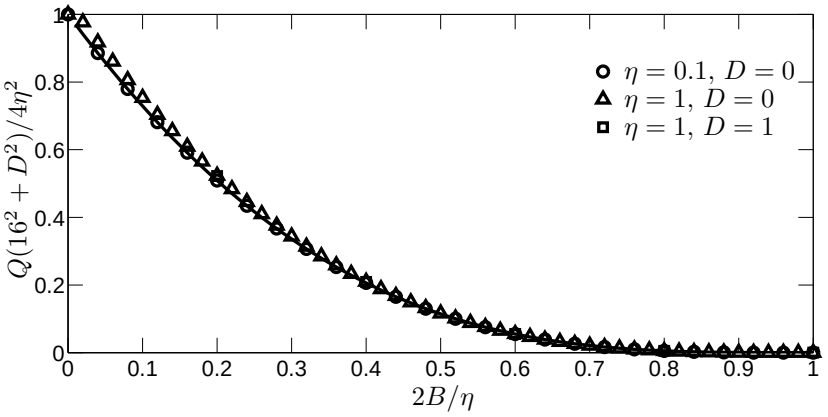


Figure 14: Theoretical prediction for small η and numerical solutions of the time-averaged flow rate of Bingham fluid, scaled by the corresponding flow rate of Newtonian fluid. The theoretical prediction is represented by the solid curve. The numerical solutions, represented by different symbols, are obtained by fixing η and varying the Bingham number given by (27) from 0 to $\eta/2$.

without stiffness, $D = 0$, it follows from (30) that

$$a = 1 - \frac{\eta}{48}\cos z(1 - r_0)^2\left((1 + r_0)^2 + 2a^2\right) + o(\eta^2). \tag{31}$$

The time-averaged flow rate q is computed by operating $\langle\cdot\rangle$ on (30), the method adopted in section 2. It can be shown, after some steps of algebra, that

$$q = \frac{\eta^2}{96}\langle\cos^2 z(1 - r_0)^3\left((1 + r_0)^2 + 2\right)\rangle. \tag{32}$$

There is no contribution to (32) from the solid region, where $r_0 = 1$. The flow rate is determined by the sheared regions, $|z| < \xi$ and $|z - \pi| < \xi$, where $\xi \equiv \cos^{-1}(2B/\eta)$ is the half-width of each region. Substituting $r_0 = 2B/\eta|\cos z|$ into (32) and integrating over the sheared regions gives

$$q = \frac{\eta^2}{48\pi}f(\xi), \tag{33}$$

where

$$\begin{aligned} f(\xi) \quad = \quad & \frac{3\xi}{2} - \frac{11}{4}\sin(2\xi) + 4\xi\cos^2\xi + \tan\xi\cos^4\xi \\ & - \cos^5\xi\left(\tanh^{-1}(\tan\frac{\xi}{2}) + \frac{1}{4(1 - \sin\xi)} - \frac{1}{4(\cos\frac{\xi}{2} + \sin\frac{\xi}{2})^2}\right) \end{aligned} \tag{34}$$

Figure 14 shows the flow rate given by (33), scaled by the corresponding flow rate of Newtonian fluid, (16), as a function of the rescaled Bingham number, $2B/\eta$. The theoretical prediction for $D = 0$ and small η is in excellent agreement with numerical results. The rescaled flow rate is a monotonically decreasing function of the rescaled Bingham number for $2B/\eta \leq 1$, as shown in figure 14, because the plug-like region increases in size and reduces the flow due to a larger yield stress. In the limit as $2B/\eta \to 1$, the flow rate $q \to 0$ because the forcing amplitude η is insufficient to drive much flow. For $2B/\eta \geq 1$, $q = 0$ because the yield stress is not overcome by the imposed force, resulting in no motion. As the Bingham number approaches 0, the flow rate given by (33) approaches the corresponding flow rate of Newtonian fluid, as expected.

In the limit of large forcing amplitude η, solutions are obtained by examining the flow near threshold of no motion, $a = r_0 = 1$, which occurs in the limit as $2B/\eta \to 1$. The small parameter $\epsilon \equiv 1 - 2B/\eta$ is introduced to examine the limit as $\epsilon \to 0$. Equations (29) and (30) respectively suggest solutions of the form $r_0 = 1 - \epsilon r_1 + o(\epsilon^2)$ and $a = 1 + \epsilon^2 a_2 + o(\epsilon^3)$. The $o(\epsilon)$ correction to the radius of the plug-like region

$$r_1 = \begin{cases} 1 - \frac{\zeta^2}{2} & |\zeta| < \sqrt{2},\ |\zeta - \epsilon^{-1/2}\pi| < \sqrt{2}, \\ 0 & \text{otherwise}, \end{cases} \tag{35}$$

where $\zeta = \epsilon^{-1/2}z$ is the rescaled axial coordinate, is obtained by expanding $\cos z$ near $z = 0, \pi$ in (29). Substituting r_1 into (30) gives

$$a_2 = \pm\frac{\eta r_1^2}{8}, \tag{36}$$

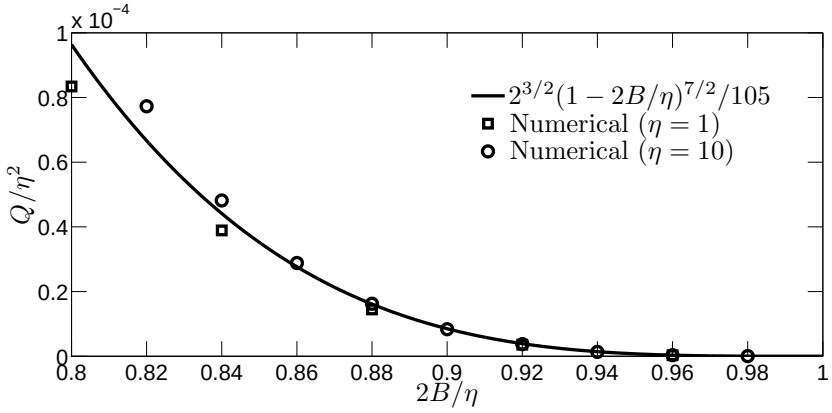


Figure 15: Theoretical prediction in the asymptotic limit as $2B/\eta \to 1$ and numerical solutions of the time-averaged flow rate of Bingham fluid.

where $\pm$ is either $+$ for $|\zeta| < \sqrt{2}$ or $-$ for $|\zeta - \epsilon^{-1/2}\pi| < \sqrt{2}$. Operating $\langle\cdot\rangle$ on (30) gives

$$q = \frac{2^{3/2}\eta^2\epsilon^{7/2}}{105}, \tag{37}$$

which is in agreement with numerical results as shown in figure 15. The flow rate scales like $\epsilon^{7/2}$, indicating that the flow of Bingham plastic increases very weakly with the applied force as it overcomes the yield stress.

It has been determined that the flow rate of both Newtonian fluid and Bingham fluid is a monotonically non-decreasing function of the forcing amplitude. This is no longer the case for the speed of propulsion of a rigid body inside a tube filled with fluid, as investigated in the following section.

4 Rigid body

Consider a rigid body of radius b surrounded by a Newtonian fluid of density ρ and dynamic viscosity μ. For simplicity, the rigid body is considered to be an infinitely-long rod of constant radius b so that steady and periodic solutions can be obtained in the wave frame. The rod has steady axial velocity W in the lab frame, as sketched in figure 16.

The axial velocity profile of the fluid is governed by (9), subject to no-slip conditions on the tube wall and the rigid rod. The boundary conditions that $w = 0$ on $r = a$ and $w = W$ on $r = b$ determine the axial velocity in the lab frame,

$$w = -\frac{1}{4}\frac{dp}{dz}(a^2 - r^2) - \left(\frac{W + \frac{1}{4}\frac{dp}{dz}(a^2 - b^2)}{\log\frac{a}{b}}\right)\log\frac{r}{a}. \tag{38}$$

In a similar manner to before in sections 2 and 3, the flow rate in the lab frame

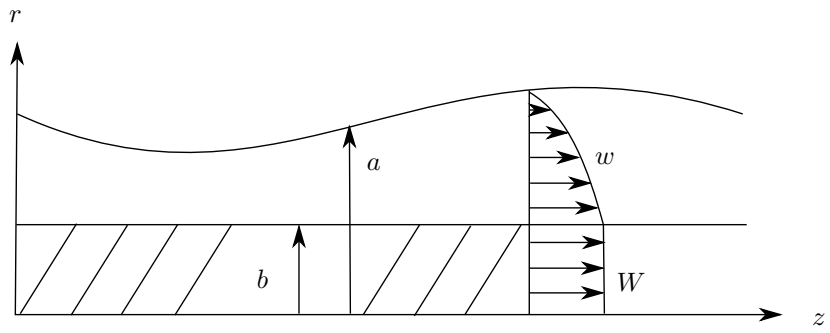


Figure 16: Schematic sketch in cylindrical polar coordinates of an elastic tube filled with Newtonian fluid and a rigid rod. The rod has radius b and propels at a steady axial velocity W in the lab frame.

is determined and given by

$$q = -\frac{1}{8}\frac{dp}{dz}\left(a^4 - b^4 - \frac{(a^2-b^2)^2}{\log(\frac{a}{b})}\right) + \frac{W(a^2-b^2)}{2\log(\frac{a}{b})} - b^2W - a^2 + 1. \qquad (39)$$

In the limit as $b \to 0$, the flow rate given by (39) reduces to the flow rate of Newtonian fluid without any rod given by (13), as expected. Note that the boundary-value problem governed by (39), with the same periodic boundary condition $a(0) = a(2\pi)$ as before, features W as an eigenvalue to be determined in addition to the eigenvalue q.

The two eigenvalues are determined by imposing two integral constraints, one of which is given by the global conservation of fluid, (12). Another integral constraint is obtained from the axial force balance on the rigid rod. Given that the rod is in steady motion, there is no net force exerted on the rod. This is written mathematically as $\langle [r\partial w/\partial r]_{r=b} \rangle = 0$, which reduces to

$$\langle \frac{W + \frac{1}{4}\frac{dp}{dz}(a^2-b^2)}{\log\frac{a}{b}} \rangle = 0. \qquad (40)$$

Figure 17 shows a set of solutions of tubes containing rigid rods of different size, where the forcing amplitude is fixed at $\eta = 10$. The associated eigenvalues, q and W, are presented later. The special case of $b = 0$ corresponds to a tube of Newtonian fluid only, which features an occluded region and a peak region, as investigated in section 2. The peak radius of the tube decreases with increasing b as shown in figure 17, indicating that the presence of the rigid rod reduces the deformation of the tube. The tube is occupied with more solid and less fluid, which partly explains the decreased deformation of the tube with increasing radius of the rigid rod.

The flow rate q increases with η in a qualitatively similar manner, independent of the rod radius b. However, the steady speed of the rod W increases for

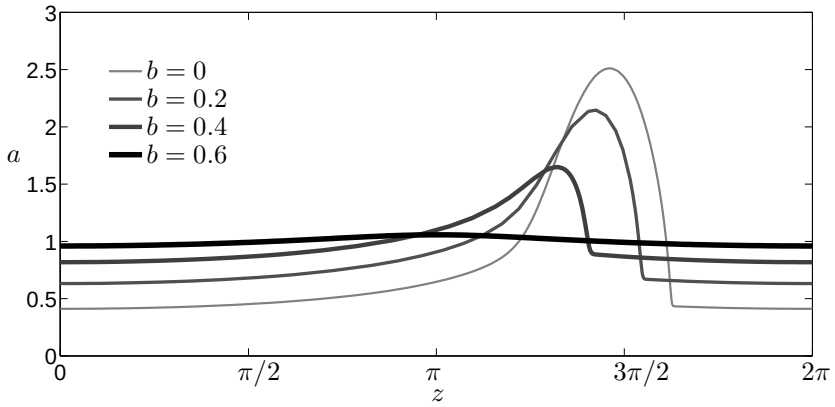


Figure 17: Numerical solutions of the deformed tube containing a rigid rod of different radius b, propelled by a radial force of amplitude $\eta = 10$.

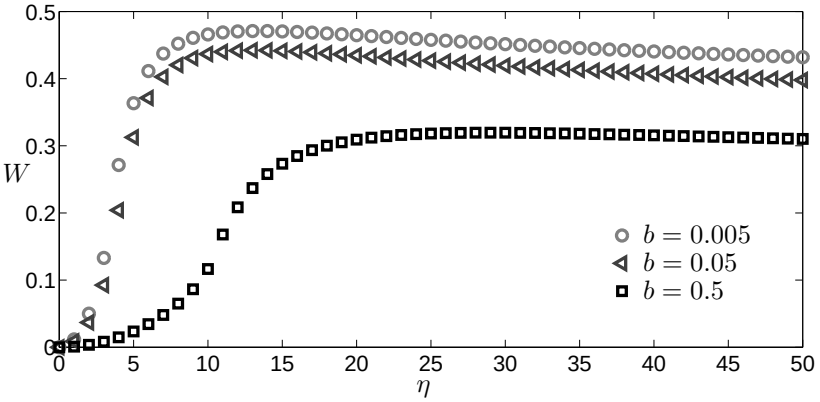


Figure 18: Numerical solutions of the steady velocity of the rigid rod of radius b as a function of the forcing amplitude η.

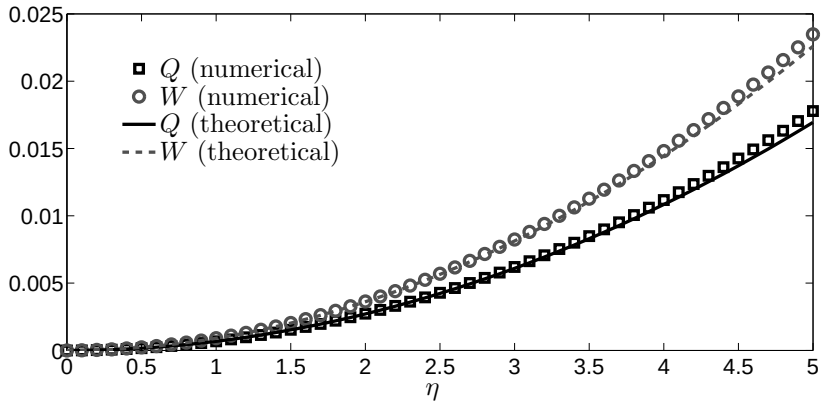


Figure 19: Theoretical predictions in the asymptotic limit of $\eta \to 0$ and numerical solutions of the flow rate q of fluid and the steady speed W of rigid body.

small η and decreases for large η, as shown in figure 18 for three representative values of b. For all values of rod radius, a maximal speed is attained at an intermediate value of η. The maximal attainable speed is less than half the wave speed. The rod speed decreases with η thereafter because the large forcing increases the viscous resistance in the occluded region, making it more difficult for the rod to move relative to the tube. The rod speed generally decreases with increasing b, indicating that it is more difficult to propel a rod of greater size. The quantitative details of the two asymptotic limits of small and large forcing amplitudes are investigated separately below.

In the limit of small η, solutions of the form $a = 1 + \eta a_1 + o(\eta^2)$, $q = \eta^2 q_2 + o(\eta^3)$, and $W = \eta^2 W_2 + o(\eta^3)$ are sought. Equation (39) to order η reduces to $kDda_1/dz + 16a_1 = -k\cos z$, which is integrated subject to the periodic boundary condition $a_1(0) = a_1(2\pi)$ to obtain the solution

$$a = 1 - \eta \frac{16k\cos z + k^2 D \sin z}{(kD)^2 + 16^2} + o(\eta^2), \tag{41}$$

where $k = 1 - b^4 + (1-b^2)^2/\log b$.

The solution given by (41) is substituted into the integral constraint (40) to obtain the leading-order speed of the rigid rod,

$$W = \frac{2k}{(kD)^2 + 16^2}\left(2 + \frac{1-b^2}{\log b}\right)\eta^2 + o(\eta^3). \tag{42}$$

Operating $\langle\cdot\rangle$ on (39) obtains the leading-order flow rate in the lab frame,

$$q = \frac{2k}{(kD)^2 + 16^2}\left(2 + \frac{1-b^2}{\log b}\right)(1-b^2)\eta^2 + o(\eta^3). \tag{43}$$

Note that $q \to W$ in the limit as $b \to 0$, indicating that a rigid wire of negligible radius propagates along the centreline of the tube at a steady speed correspond-

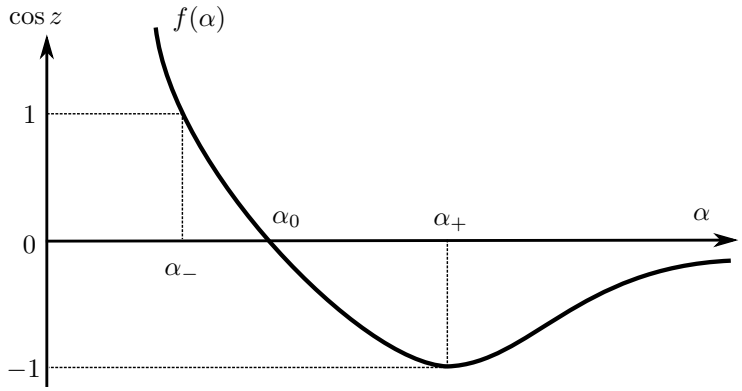


Figure 20: Sketch of $f(\alpha)$ given by (49), which changes sign at α_0 and attains its minimum of -1 at α_+. The value α_- corresponds to the minimal radius of the tube in the occluded region.

ing to the flow rate of the surrounding fluid. The solutions given by (42) and (43) are in agreement with numerical results, as shown in figure 19.

In the limit of large η, the resultant deformation of the tube is qualitatively similar to that of a tube containing Newtonian fluid only, as in section 2. A peak region near $z = 3\pi/2$ and an occluded region away from $z = 3\pi/2$ are expected to develop. The two regions are investigated separately below.

In the peak region near $z = 3\pi/2$, the peak radius is determined in a similar manner to before in section 2. The conditions in the peak region are that the pressure is uniform and that the total volume must be conserved by (12). The peak radius is given by

$$a_{max} = \left(\frac{15^2 \pi^2 (1-b^2)^2 \eta}{2^7 D^5} \right)^{1/5}, \tag{44}$$

which scales like $\eta^{1/5}$ with a prefactor that decreases with increasing b, in agreement with numerical results.

In the occluded region, the radius of the tube is expected to approach the radius of the rigid rod. Solutions of the form

$$a = b + \eta^{-1/2} \alpha + o(\eta^{-1}) \tag{45}$$

are sought, with the two eigenvalues of the form

$$W = W_\infty + o(\eta^{-1/2}) \tag{46}$$

and

$$q = 1 - b^2 - \eta^{-1/2} b \hat{q} + o(\eta^{-1}), \tag{47}$$

where α, W_∞ and $\hat{q}$ are all of order 1 to be determined. Note that $1 - b^2 - q$ is a measure of the proportion of fluid left behind the wave, which is expected

to scale like $(a-b)b$ and diminish in the limit as $\eta \to \infty$. This is because most of the fluid is trapped in a recirculating core of the peak region and propagates with the wave, while the remaining fluid lies in a thin shell of radius b and thickness $a-b$, the gap between the tube and the rod. The choice of the scaling $\eta^{-1/2}$ for corrections to a, q and W is justified below. A cubic equation for α,

$$\cos z\, \alpha^3 + 6(2 - W_\infty)\alpha - 6\hat{q} = 0, \tag{48}$$

is obtained by substituting (45), (46), and (47) into (39), and neglecting terms of order η^{-1}. All three terms in the cubic equation must be of order 1 by demanding that $a \to b + o(\eta^{-1/2})$, $q \to 1 - b^2 + o(\eta^{-1/2})$ and $W \to W_\infty + o(\eta^{-1/2})$, which justifies the earlier choice of the scaling $\eta^{-1/2}$. The solution for α is determined implicitly by a rearrangement of (48),

$$\cos z = f(\alpha) \equiv \frac{6}{\alpha^3}\left(\hat{q} - \alpha\,(2 - W_\infty)\right). \tag{49}$$

The function f has a local minimum at $\alpha_+ = 3\hat{q}/(4-2W_\infty)$ with the asymptotic limits that $f \to \infty$ as $\alpha \to 0$ and $f \to 0$ as $\alpha \to \infty$, as sketched in figure 20. Given that the tube radius must match smoothly with the peak region as $z \to 3\pi/2$, f must attain a local minimum of -1. The equation $f(\alpha_+) = -1$ reduces to

$$\hat{q} = \frac{2^{3/2}}{3}\,(2 - W_\infty)^{3/2}, \tag{50}$$

which provides an equation for the two unknown quantities, W_∞ and $\hat{q}$. A minimal tube radius is attained at $z = 0$, where $\alpha = \alpha_-$ satisfies the cubic equation $f(\alpha_-) = 1$. The function f changes sign at $\alpha_0 = 2\alpha_+/3$.

In addition to equation (50), another equation relating W_∞ and $\hat{q}$ can be obtained by (40). Note that contributions to the integral (40) from the occluded region are dominant because they are negligible from the peak region, where $a \sim \eta^{1/5}$ as $z \to 3\pi/2$. By substituting (45) into (40) and keeping only the leading-order terms of order $\eta^{1/2}$, we obtain

$$\langle \frac{2W_\infty}{\alpha} + \alpha \cos z \rangle = 0. \tag{51}$$

This integral constraint indicates that the tangential-stress balance is between the stress due to the rod speed and the radial forcing in the occluded region. The tangential stress on the rod is independent of the peak region or the stiffness D of the tube. Furthermore, the governing equations for W_∞ and $\hat{q}$, (50) and (51), are independent of b. This indicates that in the limit of large forcing amplitudes, the size of the rod does not influence its speed of propulsion to leading order. The integral constraint (51) reduces to

$$(4W_\infty - 6)\langle \alpha^{-1} \rangle + (4 - 2W_\infty)^{3/2} \langle \alpha^{-2} \rangle = 0, \tag{52}$$

by substituting (49) and (50) into (51). The integral $\langle \cdot \rangle$ in z-space can be evaluated in α-space by using (49), treating the three regions of $0 \le z \le \pi$,

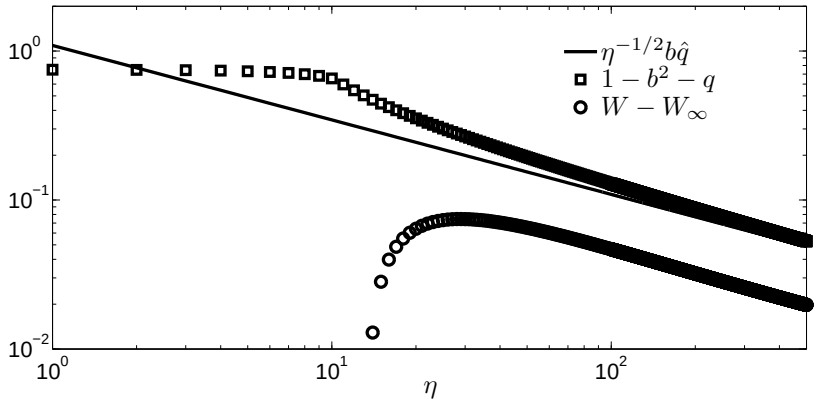


Figure 21: The proportion of fluid left by the wave $1-b^2-q$ and the approach of the rod speed W to $W_\infty \approx 0.247$ for large forcing amplitudes η, plotted on logarithmic scales. The radius of the rod $b = 0.5$ is fixed. Theoretical prediction for the proportion of fluid left by the wave, $\eta^{-1/2}b\hat{q}$, with $\hat{q} \approx 2.19$, agrees with numerical results. The slope of $W - W_\infty$ indicates that W approaches W_∞ like $\eta^{-1/2}$.

$\pi < z < 3\pi/2$ and $3\pi/2 < z < 2\pi$ separately. This gives

$$\langle \cdot \rangle = \int_{\alpha_+}^{\infty} \cdot J \, d\alpha - \int_{\alpha_-}^{\alpha_0} \cdot J \, d\alpha - \int_{\alpha_-}^{\alpha_+} \cdot J \, d\alpha, \tag{53}$$

where

$$J = \frac{df/d\alpha}{2\pi\sqrt{1-f^2}}. \tag{54}$$

Solving equation (52), where $\langle \alpha^{-1} \rangle$ and $\langle \alpha^{-2} \rangle$ are evaluated numerically using (53), gives $W_\infty \approx 0.247$. It follows that $\hat{q} \approx 2.19$ by (50) and completes the theoretical analysis of the problem.

A measure of the proportion of fluid left behind the wave is plotted as a function of the forcing amplitude in figure 21. Numerical values approach slowly to the theoretical prediction with increasing forcing amplitude. The difference between the rod speed from $W_\infty \approx 0.247$, also shown in figure 21, approaches 0 like $\eta^{-1/2}$, which is consistent with (46). In the limit of large forcing amplitude, most of the fluid is carried with the wave while the rod propels at approximately a quarter of the wave speed in the tube.

5 Conclusions

A theoretical analysis of fluid driven along a deformable tube by a prescribed radial force provides important insight into peristaltic motion. For small forcing amplitudes η, the deformation of the tube is of order η and the time-averaged

flow rate is of order η^2, independent of the three different types of fluid considered in the tube. For large forcing amplitudes η, different results are obtained depending on the contents of the tube, as summarised separately below.

A tube of Newtonian fluid features an occluded region and a peak region, where a trapped core of fluid recirculates in the wave frame. The peak region, which is shorter and more deformed in a linearly elastic tube with a spring constant D than a thin shell of constant bending stiffness D, depends importantly on the elastic properties of the tube and weakly on the forcing amplitude. A larger forcing amplitude increases the size of the trapped core in the peak region, allowing more fluid to propagate with the wave.

A tube of Bingham plastic features a sheared region, a pseudo-plug region, and a solid region in the wave frame, provided that the applied force has overcome the yield stress. The existence of the three distinct regions reduces the flow rate of the Bingham plastic considerably compared to that of a Newtonian fluid, which is sheared everywhere. The flow rate of the Bingham plastic increases from 0 extremely slowly with the forcing amplitude as the yield stress is overcome.

The steady propulsion of a rigid rod surrounded by Newtonian fluid in a tube shows that the size of the rod plays an important role. As the radius of the rod increases, the tube deforms less with a smaller speed of propulsion of the rod. The maximal speed of the rod, which is less than half the wave speed, is attained at a moderate forcing amplitude. A larger body, which reduces the maximal attainable speed of propulsion, is a bitter pill to swallow for the hungry python.

The propulsion of a rigid body in a deformable tube could be pursued further by modifying the underlying assumptions of the problem. The theoretical analysis presented here is limited to describing steady and periodic solutions in the wave frame. Time-dependent solutions in a non-periodic domain could be investigated, for example, by prescribing a radial force with a solitary-wave profile to drive a rigid body of finite length. This may provide further insight into the motion of the rigid body, where the regions ahead and behind the body in the tube are respectively relaxed and contracted.

Acknowledgements

The project was supervised by Neil Balmforth and undertaken as part of the Geophysical Fluid Dynamics program at the Woods Hole Oceanographic Institution. I benefited enormously from fruitful discussions with him at Walsh Cottage and on the soccer field. I also benefited from enjoyable discussions with other members of the program, including Joe Keller, John Wettlaufer, Jack Whitehead and Bill Young.

References

M. Argentina, J. Skotheim, and L. Mahadevan. Settling and swimming of flexible fluid-lubricated foils. *Physical Review Letters*, 99(22):224503, 2007.

N. J. Balmforth, D. Coombs, and S. Pachman. A stiff and slimy swimmer. *(submitted)*, 2009.

N. J. Balmforth and R. V. Craster. A consistent thin-layer theory for Bingham plastics. *Journal of Non-Newtonian Fluid Mechanics*, 84(1):65–81, 1999.

A. Bertuzzi, S. Salinari, R. Mancinelli, and M. Pescatori. Peristaltic transport of a solid bolus. *Journal of biomechanics*, 16(7):459, 1983.

R. B. Bird, R. C. Armstrong, and O. Hassager. Dynamics of polymeric liquids, Vol. 1, Fluid mechanics. *John Wiley& Sons, New York*, 1987.

J. G. Brasseur, S. Corrsin, and N. Q. Lu. Influence of a peripheral layer of different viscosity on peristaltic pumping with Newtonian fluids. *Journal of Fluid Mechanics*, 174:495–519, 1987.

E. O. Carew and T. J. Pedley. An active membrane model for peristaltic pumping: part I – Periodic activation waves in an infinite tube. *Journal of biomechanical engineering*, 119:66, 1997.

I. A. Frigaard and D. P. Ryan. Flow of a visco-plastic fluid in a channel of slowly varying width. *Journal of Non-Newtonian Fluid Mechanics*, 123(1): 67–83, 2004.

A. E. H. Love. *A treatise on the mathematical theory of elasticity.* Dover Publications, 1944.

P.S. Lykoudis and R. Roos. The fluid mechanics of the ureter from a lubrication theory point of view. *Journal of Fluid Mechanics*, 43(04):661–674, 1970.

A. H. Shapiro, M. Y. Jaffrin, and S. L. Weinberg. Peristaltic pumping with long wavelengths at low Reynolds number. *Journal of Fluid Mechanics*, 37, 1969.

K. Vajravelu, S. Sreenadh, and V. Ramesh Babu. Peristaltic transport of a Herschel–Bulkley fluid in an inclined tube. *International Journal of Non-Linear Mechanics*, 40(1):83–90, 2005.

J. S. Wettlaufer and M. G. Worster. Dynamics of premelted films: frost heave in a capillary. *Physical Review E*, 51(5):4679–4689, 1995.

J. S. Wettlaufer, M. G. Worster, L. A. Wilen, and J. G. Dash. A theory of premelting dynamics for all power law forces. *Physical review letters*, 76(19): 3602–3605, 1996.

Linear Stability of Su-Gardner Solutions to Small Two-Dimensional Perturbations

Adrienne Traxler

October 1, 2009

1 Introduction

Many wave phenomena of interest in the ocean are strongly affected by mathematically messy factors: variable bottom topography, shocks and wave breaking, and so on. The original motivation for this project begins with the 2004 Indian Ocean tsunami, which among many other difficult to model phenomena included examples of nearby villages (with tiny separations on the scale of the tsunami front) experiencing great differences in the severity of the arriving waves. Purely one-dimensional models of propagation are clearly insufficient for such problems, so from here the focus of the project narrowed to examining a particular model of waves in shallow water and the complications introduced by allowing variation along a second direction of propagation.

A large amount of the material in the principal lectures this summer involved the Korteweg-de Vries equation, a long wavelength and small amplitude treatment of waves in shallow water. Here we advance the treatment a step by using the Su-Gardner equations, which maintain the requirement of long wavelength but allow for larger amplitudes (and thus strongly nonlinear waves). The goal is to understand the one-dimensional system first, then extend it to examine the stability of Su-Gardner solutions to small transverse perturbations via the Green-Naghdi equations, which provide a two-dimensional formulation of Su-Gardner.

2 The 1D Problem

The Su-Gardner equations, which consist of shallow water plus several extra correction terms, are derived here along with their solutions. Before moving on to the two-dimensional case, we show that the Su-Gardner system reduces to the Korteweg-de Vries equation in the limit of small wave amplitude.

2.1 Derivation of Su-Gardner Equations

These equations were first derived by Serre in 1953 [6], more widely known under the later work of Su and Gardner [7]. The derivation below follows the latter (see also the appendix of [2]).

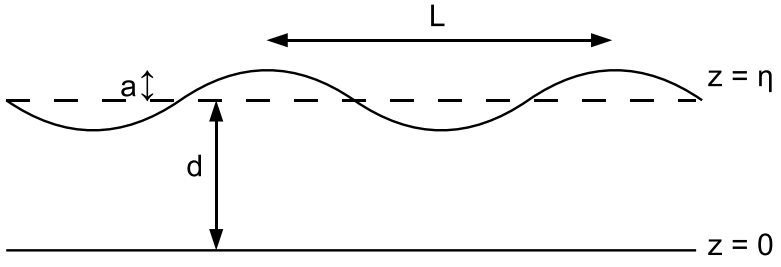


Figure 1: Flat bottom, with $\eta = \eta(x,t)$ or $\eta = \eta(x,y,t)$ for the 1D and 2D problems respectively (in the later case, we also must consider the relative size of the y wavelength, as discussed in 3.2).

We begin with the 2D Euler equations for inviscid and incompressible flow,

$$\begin{aligned} u_t + uu_x + wu_z &= -p_x \\ w_t + uw_x + ww_z &= -p_z - g \\ u_x + w_z &= 0 \end{aligned}$$

where we have assumed a constant fluid density ($\rho = 1$). We consider the domain between a flat bottom at $z = 0$ and a free surface at $z = \eta$ as sketched in Figure 1. The boundary conditions are

$$\begin{aligned} w = 0 \quad &\text{at} \quad z = 0 \\ w = \eta_t + u\eta_x \quad &\text{at} \quad z = \eta \\ p = 0 \quad &\text{at} \quad z = \eta \end{aligned}$$

representing respectively a flat impermeable bottom, and the kinematic and dynamic free surface conditions. As solutions we are seeking the surface height $\eta(x,t)$ and the depth-averaged horizontal velocity $\bar{u}(x,t)$.

From here we begin by integrating the incompressibility condition over the entire depth,

$$\int_0^\eta u_x \, dz + \int_0^\eta w_z \, dz = 0$$

where application of the Liebniz integral rule transforms this to

$$\left(\int_0^\eta u \, dz \right)_x - \eta_x u|_\eta + 0 \cdot u|_0 + w|_0^\eta = 0$$

Use of the kinematic free surface boundary condition gives

$$-\eta_x u|_\eta = \eta_t - w|_\eta$$

and the bottom boundary condition requires

$$w|_0 = 0$$

Applying these reduces the equation to

$$\left(\int_0^\eta u\,dz\right)_x + (\eta_t - w|_\eta) + w|_\eta = 0$$

which rearranges to the final form,

$$\eta_t + (\eta\bar{u})_x = 0$$

where the bar denotes a depth average,

$$\bar{u}(x,t) = \frac{1}{\eta}\int_0^\eta u\,dz$$

Next we integrate the horizontal momentum equation over the water depth. Going term by term, we have

$$\begin{aligned}
\int_0^\eta u_t\,dz &= \left(\int_0^\eta u\,dz\right)_t - \eta_t u|_\eta + 0\cdot u|_0 \\
&= (\eta\bar{u})_t - \eta_t u|_\eta \\
\int_0^\eta uu_x\,dz &= \frac{1}{2}\int_0^\eta (u^2)_x\,dz \\
&= \frac{1}{2}\left[\left(\int_0^\eta u^2\,dz\right)_x - \eta_x u^2|_\eta - 0\cdot u^2|_0\right] \\
\int_0^\eta wu_z\,dz &= \int_0^\eta (uw)_z\,dz - \int_0^\eta uw_z\,dz \\
&= (uw)|_0^\eta + \int_0^\eta uu_x\,dz \\
&= (uw)|_0^\eta + \frac{1}{2}\left[\left(\int_0^\eta u^2\,dz\right)_x - \eta_x u^2|_\eta\right] \\
\int_0^\eta p_x\,dz &= \left(\int_0^\eta p\,dz\right)_x - \eta_x \cancel{p}|_\eta + 0\cdot p|_0
\end{aligned}$$

and putting the pieces together with use of boundary conditions:

$$(\eta\bar{u})_t - (w - u\eta_x)u|_\eta + (uw)|_0^\eta + \left(\int_0^\eta u^2\,dz\right)_x - \eta_x u^2|_\eta = -\left(\int_0^\eta p\,dz\right)_x$$

This reduces to

$$(\eta\bar{u})_t + [\eta(\bar{u^2} + \bar{p})]_x = 0 \tag{1}$$

where the bars over u^2 and p denote depth averages, as above.

To deal with the averaged terms, we need to use the vertical momentum equation. Multiplying by z and then integrating over depth gives

$$\begin{aligned}
\int_0^\eta z\frac{dw}{dt}\,dz &= -\int_0^\eta zp_z\,dz - \int_0^\eta gz\,dz \\
&= -(pz)|_0^\eta + \int_0^\eta p\,dz - \frac{1}{2}g\eta^2
\end{aligned}$$

where the first term on the right hand side is zero when evaluated at the endpoints, and the rest rearranges to

$$\eta\bar{p} = \frac{1}{2}g\eta^2 + \int_0^\eta z\frac{dw}{dt}\,dz$$

After these integrations, we can rewrite equation 1 as

$$(\eta\bar{u})_t + \left(\eta\bar{u}^2 + \frac{1}{2}g\eta^2\right)_x + \left(\int_0^\eta z\frac{dw}{dt}\,dz + \eta\bar{u^2} - \eta\bar{u}^2\right)_x = 0$$

or the shallow water momentum equation plus correction terms.

Instead of neglecting the last set of terms in parentheses, we can continue further to approximate them using the additional assumptions of irrotationality ($u_z = w_x$) and long wavelength ($d \ll L$, see Figure 1). Beginning with the first of the correction terms, expand u and w in a Taylor series at the bottom:

$$\begin{aligned} u(z) &= u(0) + z\frac{\partial u}{\partial z}\bigg|_0 + \frac{1}{2}z^2\frac{\partial^2 u}{\partial z^2}\bigg|_0 + \ldots \\ w(z) &= w(0) + z\frac{\partial w}{\partial z}\bigg|_0 + \ldots \end{aligned}$$

Starting with the second expression, the first term disappears from the bottom boundary condition, while the second can be rewritten using the incompressibility condition,

$$w(z) = -z\tilde{u}_x + \ldots$$

where $\tilde{u} = u(0)$. Looking at the horizontal velocity next, from irrotationality the second term is zw_x, but when evaluated at the bottom $w_x = 0$. This leaves

$$u(z) = \tilde{u} - \frac{1}{2}\tilde{u}_{xx}z^2 + \ldots$$

Now we can approximate the depth averaged velocity:

$$\begin{aligned} \bar{u} &= \frac{1}{\eta}\int_0^\eta \left(\tilde{u} - \frac{1}{2}\tilde{u}_{xx}z^2 + \ldots\right)dz \\ &\approx \tilde{u} - \frac{1}{6}\tilde{u}_{xx}\eta^2 \end{aligned}$$

so $\tilde{u} \approx \bar{u} + \frac{1}{6}\bar{u}_{xx}\eta^2$, and we can write

$$\begin{aligned} u(z) &= \tilde{u} - \frac{1}{2}\tilde{u}_{xx}z^2 + \ldots \\ &= \bar{u} + \frac{1}{6}\bar{u}_{xx}\eta^2 - \frac{1}{2}z^2(\bar{u}_{xx} + \ldots) \\ w(z) &= -z\tilde{u}_x + \ldots \\ &= -z(\bar{u}_x + \frac{1}{6}\bar{u}_{xxx}\eta^2 + \frac{1}{3}\bar{u}_{xx}\eta\eta_x) - \ldots \end{aligned}$$

Finally, these expansions of u and w go into the integral we're working to approximate:

$$\begin{aligned}\int_0^\eta (z d_t w) dz &= \int_0^\eta z(w_t + u w_x + w w_z)\, dz \\ &\approx -\int_0^\eta z^2 \bar{u}_{xt}\, dz - \int_0^\eta z^2 \bar{u}\bar{u}_{xx}\, dz + \int_0^\eta z^2 (\bar{u}_x)^2\, dz \\ &\approx -\frac{1}{3}\eta^3[\bar{u}_{xt} + \bar{u}\bar{u}_{xx} - (\bar{u}_x)^2]\end{aligned}$$

By a similar method, it can be shown that $\bar{u^2} - \bar{u}^2$ is a higher-order correction term. Now the Su-Gardner equations, for mass and momentum conservation, are

$$\eta_t + (\eta\bar{u})_x = 0 \tag{2}$$

$$(\eta\bar{u})_t + \left(\eta\bar{u}^2 + \frac{1}{2}g\eta^2\right)_x = \left\{\frac{1}{3}\eta^3[\bar{u}_{xt} + \bar{u}\bar{u}_{xx} - (\bar{u}_x)^2]\right\}_x \tag{3}$$

2.2 Su-Gardner solutions

To look for solutions, we begin with the momentum equation in conservation form (3). If we look for a traveling wave solution of speed c_0, then the dependence of the velocity and surface height becomes $u(x,t) \to u(x - c_0 t)$, $\eta(x,t) \to \eta(x - c_0 t)$, and the derivatives transform to $\partial_x \to \partial_\zeta$, $\partial_t \to -c_0\partial_\zeta$, where $\zeta = x - c_0 t$ is the moving coordinate. Now the equation can be integrated once:

$$\begin{aligned}-c_0(\eta u) + \eta u^2 + \frac{1}{2}g\eta^2 &= \frac{\eta^3}{3}[-c_0 u_{\zeta\zeta} + u u_{\zeta\zeta} - (u_\zeta)^2] + K \\ (\eta u)(u - c_0) + \frac{1}{2}g\eta^2 &= \frac{\eta^3}{3}[(u - c_0)u_{\zeta\zeta} - (u_\zeta)^2] + K\end{aligned}$$

(From this point forward, we work only with the depth-averaged velocity, and so drop the overbars.)

Next we step back to the continuity equation for a moment, where the same transformation gives

$$-c_0\eta_\zeta + (\eta u)_\zeta = 0$$

which we can integrate over ζ:

$$\begin{aligned}-c_0\eta + \eta u &= K' \\ u - c_0 &= \frac{K'}{\eta}\end{aligned} \tag{4}$$

where K', like K, is an integration constant. To make full use of this relation, first rewrite the momentum equation in terms of $u - c_0$,

$$(\eta u)(u - c_0) + \frac{1}{2}g\eta^2 = \frac{\eta^3}{3}\{(u - c_0)(u - c_0)_{\zeta\zeta} - [(u - c_0)_\zeta]^2\} + K$$

and then simplify the resulting η terms:

$$\begin{aligned}
K'u + \frac{1}{2}g\eta^2 &= \frac{\eta^3}{3}\left\{\frac{K'}{\eta}\left(\frac{K'}{\eta}\right)_{\zeta\zeta} - \left[\left(\frac{K'}{\eta}\right)_{\zeta}\right]^2\right\} + K \\
K'\left(\frac{K'}{\eta} + c_0\right) + \frac{1}{2}g\eta^2 &= \frac{\eta^3}{3}\left[\frac{K'}{\eta}\left(-\frac{K'\eta_\zeta}{\eta^2}\right)_{\zeta} - \left(-\frac{K'\eta_\zeta}{\eta^2}\right)^2\right] + K \\
K'c_0 + \frac{K'^2}{\eta} + \frac{1}{2}g\eta^2 &= \frac{\eta^3}{3}\left[\frac{K'}{\eta}\left(-\frac{K'\eta_{\zeta\zeta}}{\eta^2} + 2\frac{K'\eta_\zeta^2}{\eta^3}\right) - \frac{K'^2\eta_\zeta^2}{\eta^4}\right] + K \\
&= \frac{\eta^3}{3}\left(-\frac{K'^2\eta_{\zeta\zeta}}{\eta^3} + \frac{K'^2\eta_\zeta^2}{\eta^4}\right) + K \\
&= \frac{K'^2}{3}\left(\frac{\eta_\zeta^2}{\eta} - \eta_{\zeta\zeta}\right) + K
\end{aligned}$$

But the first term on the right hand side can be rewritten as another derivative,

$$K'c_0 + \frac{K'^2}{\eta} + \frac{1}{2}g\eta^2 = -\frac{K'^2\eta}{3}\left(\frac{\eta_\zeta}{\eta}\right)_{\zeta} + K$$

Now multiply both sides by η_ζ/η^2, and note that this lets us rewrite both sides as derivatives:

$$\begin{aligned}
-\frac{K'^2}{6}\left[\left(\frac{\eta_\zeta}{\eta}\right)^2\right]_{\zeta} &= \frac{\eta_\zeta}{\eta^2}\left(K'c_0 - K + \frac{K'^2}{\eta} + \frac{1}{2}g\eta^2\right) \\
&= \left(-\frac{K'c_0 - K}{\eta}\right)_{\zeta} + \left(-\frac{K'^2}{2\eta^2}\right)_{\zeta} + \left(\frac{1}{2}g\eta\right)_{\zeta}
\end{aligned}$$

Integrating to remove the ζ derivative, we have

$$\begin{aligned}
\left(\frac{\eta_\zeta}{\eta}\right)^2 &= -\frac{6}{K'^2}\left(-\frac{K'c_0 - K}{\eta} - \frac{K'^2}{2\eta^2} + \frac{1}{2}g\eta + K''\right) \\
(\eta_\zeta)^2 &= -\frac{3g}{K'^2}\eta^3 - \frac{6K''}{K'^2}\eta^2 + \frac{6(K'c_0 - K)}{K'^2}\eta + 3 \\
\eta_\zeta &= \pm\sqrt{\frac{3g}{K'^2}\left(-\eta^3 - \frac{2K''}{g}\eta^2 + \frac{2(K'c_0 - K)}{g}\eta + \frac{K'^2}{g}\right)}
\end{aligned} \tag{5}$$

The solution, available in El et al. [3], is in the form of the Jacobian elliptic function $\mathrm{cn}(\zeta; m)$,

$$\eta(\zeta) = \eta_2 + a\,\mathrm{cn}^2\left(\frac{1}{2}\sqrt{\frac{3g}{K'^2}(\eta_3 - \eta_1)}\,\zeta; m\right) \tag{6}$$

where ζ is the moving coordinate ($\zeta = x - c_0 t$), the η_i are the roots of the cubic in (5), with $\eta_3 \geq \eta_2 \geq \eta_1 > 0$, and the elliptic modulus is $m = \frac{\eta_3 - \eta_2}{\eta_3 - \eta_1}$.

2.3 A limit of the equations

Equations 2 and 3, above, assume long wavelengths but place no restrictions on wave amplitude. In the small amplitude limit—in particular, when $a/d \sim (d/L)^2$—we demonstrate that Su-Gardner reduces to the well-known Korteweg-de Vries equation. Starting from (2) and (3), we follow the method of Su and Gardner [7] and switch to the moving frame, where $\xi = \epsilon^\alpha(x - c_0 t)$, $\tau = \epsilon^{\alpha+1}t$, and thus

$$\begin{aligned} \partial_t &= \epsilon^{\alpha+1}\partial_\tau - \epsilon^\alpha c_0 \partial_\xi \\ \partial_x &= \epsilon^\alpha \partial_\xi \end{aligned}$$

After making the transformation and canceling common powers of epsilon, we have

$$\epsilon\eta_\tau + (u - c_0)\eta_\xi + \eta u_\xi = 0$$

$$\epsilon(\eta u)_\tau - c_0(\eta u)_\xi + +\eta_\xi u^2 + 2\eta u u_\xi + g\eta\eta_\xi - \frac{1}{3}\epsilon^{2\alpha}\left\{\eta^3[\epsilon u_{\xi\tau} + (u - c_0)u_{\xi\xi} - u_\xi^2]\right\}_\xi = 0$$

If we now try expansions of the form $\eta = \eta_0 + \epsilon\eta^{(1)} + \ldots$ and $u = 0 + \epsilon u^{(1)} + \ldots$, where η_0 is the constant equilibrium height ($\eta_0 > 0$ in our coordinate system), then at leading order all terms are zero in both equations. At $O(\epsilon)$,

$$\begin{aligned} -c_0\eta_\xi^{(1)} + \eta_0 u_\xi^{(1)} &= 0 \\ -c_0\eta_0 u_\xi^{(1)} + g\eta_0\eta_\xi^{(1)} &= 0 \end{aligned}$$

These two equations together have a nontrivial solution only if $c_0^2 = g\eta_0$, so by integrating we just have $c_0\eta^{(1)} = \eta_0 u^{(1)}$. Now that we can eliminate $u^{(1)}$, take a look at $O(\epsilon^2)$:

$$\eta_\tau^{(1)} + u^{(1)}\eta_\xi^{(1)} - c_0\eta_\xi^{(2)} + \eta_0 u_\xi^{(2)} + \eta^{(1)}u_\xi^{(1)} = 0$$

$$\eta_0 u_\tau^{(1)} - c_0\eta_0 u_\xi^{(2)} - c_0\eta^{(1)}u_\xi^{(1)} - c_0\eta_\xi^{(1)}u^{(1)} + 2\eta_0 u^{(1)}u_\xi^{(1)} + g\eta_0\eta_\xi^{(2)} + g\eta^{(1)}\eta_\xi^{(1)} - \frac{1}{3}\epsilon^{2\alpha-1}\eta_0^3(-c_0 u_{\xi\xi}^{(1)}) = 0$$

or after eliminating $u^{(1)}$ (and canceling terms),

$$\eta_\tau^{(1)} + 2\frac{c_0}{\eta_0}\eta^{(1)}\eta_\xi^{(1)} - c_0\eta_\xi^{(2)} + \eta_0 u_\xi^{(2)} = 0$$

$$c_0\eta_\tau^{(1)} - c_0\eta_0 u_\xi^{(2)} + g\eta_0\eta_\xi^{(2)} + g\eta^{(1)}\eta_\xi^{(1)} + \frac{1}{3}\epsilon^{2\alpha-1}c_0^2\eta_0^2\eta_{\xi\xi\xi}^{(1)} = 0$$

Finally, multiply the first equation by c_0 and add it to the second to cancel the $\eta^{(2)}$ and $u^{(2)}$ terms, leaving

$$2c_0\eta_\tau^{(1)} + 3g\eta^{(1)}\eta_\xi^{(1)} + \frac{1}{3}\epsilon^{2\alpha-1}c_0^2\eta_0^2\eta_{\xi\xi\xi}^{(1)} = 0$$

Finally, we set $\alpha = 1/2$, so that the dispersion term comes in at the same order as the nonlinear term, which is the balance described by the KdV equation. (See the notes on Lecture 5 in this volume for more about the scaling of the small parameters.) This gives

$$\eta_\tau^{(1)} + \frac{3}{2}\frac{c_0}{\eta_0}\eta^{(1)}\eta_\xi^{(1)} + \frac{1}{6}c_0\eta_0^2\eta_{\xi\xi\xi}^{(1)} = 0 \tag{7}$$

which matches the KdV form of $\eta_\tau + \beta\eta\eta_\xi + \mu\eta_{\xi\xi\xi} = 0$. Note that up to $O(\epsilon^2)$, only one piece of the Su-Gardner correction to shallow water is relevant, specifically the $c_0 u_{\xi\xi\xi}$ term. As the wave amplitude increases, we expect the other parts of the correction term to become non-negligible, but in the KdV range of small amplitudes the two solutions should be indistinguishable.

3 The 2D Problem

With the 1D problem treated analytically, we want to examine the stability of those solutions to small disturbances along the direction perpendicular to propagation. Fortunately, the task of extending the Su-Gardner equations to two dimensions has already been accomplished by Green and Naghdi [4], so we can work from there to linearize around the 1D solutions.

3.1 The Green-Naghdi equations

Originally from Green and Naghdi in 1976 [4], using the updated form of Nadiga et al. [5], we have the Green-Naghdi equations for conservation of mass and momentum:

$$\eta_t + \nabla \cdot (\eta \mathbf{u}) = 0 \tag{8}$$

$$\mathbf{u}_t + \mathbf{u} \cdot \nabla \mathbf{u} + g\nabla\eta = \frac{1}{3\eta}\nabla\left[\eta^2(\partial_t + \mathbf{u}\cdot\nabla)(\eta\nabla\cdot\mathbf{u})\right] \tag{9}$$

Note that the velocity here is the depth-averaged quantity, so $\mathbf{u} = (u, v)$ and $\nabla = \hat{\mathbf{i}}\partial_x + \hat{\mathbf{j}}\partial_y$. It is not completely obvious from inspection that these become the Su-Gardner equations in one dimension,

$$\begin{aligned} \eta_t + (\eta u)_x &= 0 \\ u_t + uu_x + g\eta_x &= \frac{1}{3\eta}\left[\eta^3(u_{xt} + uu_{xx} - u_x^2)\right]_x \end{aligned}$$

but the verification is straightforward. Reducing to one dimension, $\mathbf{u} \to u$ and $\nabla \to \partial_x$. Applied to mass conservation (8), this simply becomes

$$\eta_t + (\eta u)_x = 0$$

while the momentum equation becomes

$$\begin{aligned} u_t + uu_x + g\eta_x &= \frac{1}{3\eta}\left[\eta^2(\partial_t + u\partial_x)(\eta u_x)\right]_x \\ &= \frac{1}{3\eta}\left[\eta^2(\eta_t u_x + \eta u_{xt} + \eta_x u u_x + \eta u u_{xx})\right]_x \\ &= \frac{1}{3\eta}\left[\eta^2(-\eta u_x^2 + \eta u_{xt} + \eta u u_{xx})\right]_x \end{aligned}$$

where the last line follows via substitution from the continuity equation. But this rearranges to

$$u_t + uu_x + g\eta_x = \frac{1}{3\eta}\left[\eta^3(u_{xt} + uu_{xx} - u_x^2)\right]_x$$

which is the second of the Su-Gardner equations, so the match is complete.

3.2 A limit of the equations

In the previous section, we looked at the behavior of the Su-Gardner equations in the small amplitude limit, and showed that they reduce to the KdV equation, providing a useful body of existing solutions to compare with numerical Su-Gardner results. It is reasonable to ask if a similar comparison can be found in two dimensions, and indeed the Kadomtsev-Petviashvili equation provides a weakly two-dimensional extension of KdV. In particular, the KP regime assumes scaling between the x and y derivatives such that the two highest-order spatial terms in the equation, u_{xxxx} and u_{yy}, come in at the same order. For this to happen we need not only the requirements of KdV,

$$\frac{a}{h} = O(\epsilon) \ , \ \left(\frac{h}{L_x}\right)^2 = O(\epsilon)$$

but additionally that the x and y wavelengths compare as

$$\left(\frac{L_x}{L_y}\right)^2 = O(\epsilon)$$

representing a y variation which is even slower than the already slow variation in the x-direction.

If these conditions are met, we conjecture that the Green-Naghdi solutions in this limit should recapture the known stability results for KdV solitons in KP: either linearly stable or unstable to small transverse perturbations, depending on relative signs of two highest derivatives (u_{xxxx} and u_{yy}) in the KP equation.

3.3 Linear stability of 1D solutions

To examine the stability of the known Su-Gardner solutions to transverse disturbances, add small perturbations to the variables of interest:

$$\begin{aligned} \eta &= \eta^{(0)} + \epsilon\eta^{(1)} + \dots \\ u &= u^{(0)} + \epsilon u^{(1)} + \dots \\ v &= 0 + \epsilon v^{(1)} + \dots \end{aligned}$$

where $\eta^{(0)}(x,t)$, $u^{(0)}(x,t)$ represent the one-dimensional Su-Gardner solutions discussed above in (2.2). If we start with the simple case of $\eta^{(0)}$ and $u^{(0)}$ both constant, the equations then substantially reduce in complexity to

$$\begin{aligned} \eta_t^{(1)} + \eta^{(0)}u_x^{(1)} + u^{(0)}\eta_x^{(1)} + \eta^{(0)}v_y^{(1)} &= 0 \\ u_t^{(1)} + u^{(0)}u_x^{(1)} + g\eta_x^{(1)} &= \frac{1}{3\eta^{(0)}}\left[(\eta^{(0)})^3(\partial_t + u^{(0)}\partial_x)(u_x^{(1)} + v_y^{(1)})\right]_x \\ v_t^{(1)} + u^{(0)}v_x^{(1)} + g\eta_y^{(1)} &= \frac{1}{3\eta^{(0)}}\left[(\eta^{(0)})^3(\partial_t + u^{(0)}\partial_x)(u_x^{(1)} + v_y^{(1)})\right]_y \end{aligned}$$

Now look for normal modes, $\left\{\eta^{(1)}, u^{(1)}, v^{(1)}\right\} = \{\hat{\eta}, \hat{u}, \hat{v}\}\exp(\lambda t + ikx + ily)$, and the

equations become

$$\begin{aligned}(\lambda + iku^{(0)})\hat{\eta} + \eta^{(0)}(ik\hat{u} + il\hat{v}) &= 0 \\ (\lambda + iku^{(0)})\hat{u} + ikg\hat{\eta} &= ik\frac{(\eta^{(0)})^2}{3}(\lambda + iku^{(0)})(ik\hat{u} + il\hat{v}) \\ (\lambda + iku^{(0)})\hat{v} + ilg\hat{\eta} &= il\frac{(\eta^{(0)})^2}{3}(\lambda + iku^{(0)})(ik\hat{u} + il\hat{v})\end{aligned}$$

which combines into

$$\left[1 + \frac{1}{3}(\eta^{(0)})^2(k^2 + l^2)\right](\lambda + iku^{(0)})^2 + g\eta^{(0)}(k^2 + l^2) = 0$$

We can rearrange this for the growth rate, λ, as

$$\lambda = -iku^{(0)} \pm i\sqrt{\frac{g\eta^{(0)}(k^2 + l^2)}{1 + \eta^{(0)}(k^2 + l^2)/3}}$$

so the growth rate is entirely imaginary, representing pure oscillation. Additionally, when moving from the system of three equations to the quadratic form above, we factored out an additional root, $\lambda = -iku^{(0)}$, so all roots of the system are purely imaginary. If we rewrite the quadratic root slightly in terms of the angular frequency, we have (with $\lambda = i\omega$ and $K^2 = k^2 + l^2$)

$$\omega = -ku^{(0)} \pm K\sqrt{\frac{g\eta^{(0)}}{1 + \eta^{(0)}K^2/3}} \tag{10}$$

For the special case of constant $\eta^{(0)}$ and $u^{(0)}$, then, the Su-Gardner system plus small transverse perturbations is linearly stable, with weak dispersion for two of the three modes.

Without the assumption of constant leading-order velocity and surface displacement, the $O(\epsilon)$ picture is considerably less pretty:

$$\begin{aligned}\eta_t^{(1)} &+ (u_0\eta^{(1)} + u^{(1)}\eta_0)_x + (\eta_0 v^{(1)})_y = 0 \\ (\eta^{(1)}u_0 + \eta_0 u^{(1)})_t &+ (2u_0u^{(1)}\eta_0 + u_0^2\eta^{(1)} + g\eta_0\eta^{(1)})_x + (u_0v^{(1)}\eta_0)_y \\ &= \frac{1}{3}\Big\{2\eta_0\eta^{(1)}(\partial_t + u_0\partial_x)(\eta_0 u_{0x}) + \eta_0^2(u^{(1)}\partial_x + v^{(1)}\partial_y)(\eta_0 u_{0x}) \\ &\quad + \eta_0^2(\partial_t + u_0\partial_x)(\eta^{(1)}u_{0x}) + \eta_0^2(\partial_t + u_0\partial_x)\left[\eta_0(u_x^{(1)} + v_y^{(1)})\right]\Big\}_x \\ (\eta_0 v^{(1)})_t &+ (u_0\eta_0 v^{(1)})_x + (g\eta_0\eta^{(1)})_y = \frac{1}{3}\{\ldots\}_y\end{aligned}$$

Even assuming normal modes of the form $\{\eta^{(1)}, u^{(1)}, v^{(1)}\} = \{\hat{\eta}(x), \hat{u}(x), \hat{v}(x)\}\exp(\lambda t + ily)$, these equations have not proved tractable to simplify to a dispersion relation comparable to (10) above. The next step is to treat them numerically, discussed below.

4 Numerical solutions

While the one-dimensional Su-Gardner system has been solved, the 2D case is much more difficult to reduce analytically, and to make substantial further progress it seems that numerical solutions will be required. To apply transverse perturbations to the cnoidal solutions of

the Su-Gardner equations, we first need them in numerical form, and so we take a step back to the one-dimensional system to capture that starting point. The first stage of this work seeks solutions of the KdV equation, as a simpler test case which should be comparable with low-amplitude Su-Gardner. The tool used was a Newton-Raphson-Kantorovich solver previously developed by P. Garaud, which treats two-point boundary value problems for systems of ordinary differential equations by the method of relaxing to a solution.

4.1 KdV numerical solution

Starting with the KdV equation, to eliminate the τ derivative and express the problem as a set of ODEs, we shift to a frame that is moving with the speed of the soliton solution, V. (Note that this is the speed of the soliton with respect to the frame that is already moving with the linear wave speed, so that in the notation of the previous sections, the total solution speed in the rest frame is $c = c_0 + \epsilon V$.) Once this is done the problem can be expressed in terms of ordinary differential equations,

$$\begin{aligned} Y_1 &= \eta_1 \\ Y_2 &= \eta_1' \\ Y_3 &= \eta_1'' \\ Y_4 &= V \\ Y_5 &= \int_0^x \eta_1 \, dx \end{aligned}$$

where the primes denote spatial derivatives in the frame moving with speed c, the soliton speed V is solved for as an eigenvalue, and the fifth equation is used to impose a mass conservation boundary condition. The system fed into the solver, then, is

$$\begin{aligned} Y_1' &= Y_2 \\ Y_2' &= Y_3 \\ Y_3' &= Y_2 Y_4 - Y_1 Y_2 \\ Y_4' &= 0 \\ Y_5' &= Y_1 \end{aligned}$$

The boundary conditions supplied are as follows. At the first endpoint, $x_a = 0$, we require that $Y_1 = H$ (the surface height perturbation equals some specified amplitude), that $Y_2 = 0$, and that $Y_5 = 0$. At the second endpoint, $x_b = L$, we again fix the amplitude at $Y_1 = H$ (the solution should be periodic), and also require that $Y_5 = 0$ (the total mass integrated over the domain is zero, which should be the case since we are solving for the perturbation to the equilibrium height η_0).

Finally, we need a set of initial guesses at the solution. For Y_1, Y_2, and Y_3, a cosine and its first two derivatives are used, while for Y_4 and Y_5 nonzero constants are supplied. The system was integrated over a sample domain of horizontal extent $L = 10$ with amplitude $a = 1$ (see Figure 2).

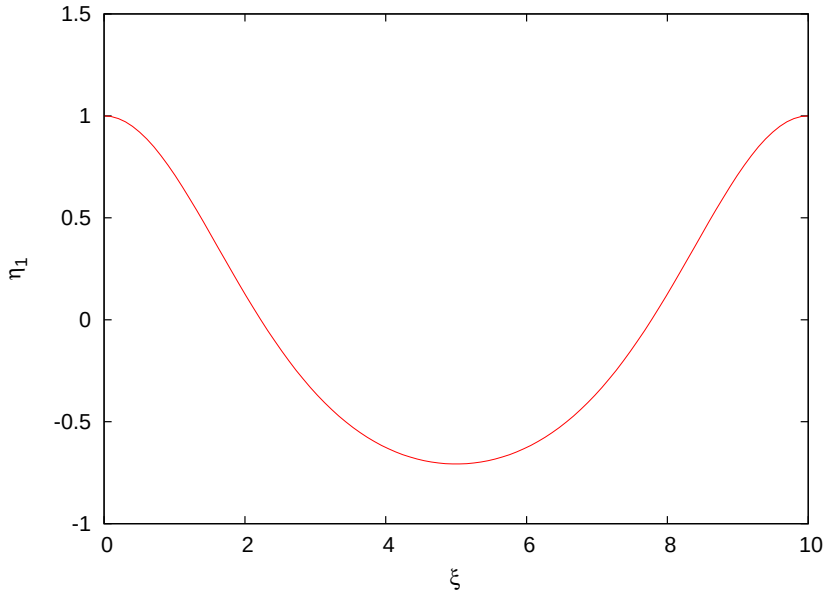


Figure 2: Sample integration of KdV equation for η_1 with the background amplitude η_0 subtracted.

4.2 Su-Gardner numerical solution

With the KdV numerical solution in hand, treatment of the Su-Gardner equations begins with the conservation form of equation 3, shifted to a frame moving with speed c. The continuity equation is integrated to give $u - c = K'/\eta$, but further integrations of the momentum equation in section 2.2 are not carried out. The momentum equation at this point is

$$-c(\eta u)_\zeta + (\eta u^2)_\zeta + \frac{1}{2}(g\eta^2)_\zeta = \left[\frac{1}{3}\eta^3(-cu_{\zeta\zeta} + uu_{\zeta\zeta} - u_\zeta^2)\right]_\zeta$$

where derivatives are in the frame of the moving coordinate, $\zeta = x - ct$. If we rewrite in terms of $(u - c)$,

$$\left[(u-c)\eta u + \frac{1}{2}g\eta^2\right]_\zeta = \frac{1}{3}\left\{\eta^3\left[(u-c)(u-c)_{\zeta\zeta} - (u-c)_\zeta^2\right]\right\}_\zeta$$

then substitute using the integrated continuity equation,

$$\left[\eta\frac{K'}{\eta}\left(\frac{K'}{\eta} + c\right) + \frac{1}{2}g\eta^2\right]_\zeta = \frac{1}{3}\left\{\eta^3\left[\frac{K'}{\eta}\left(\frac{K'}{\eta}\right)_{\zeta\zeta} - \left(\frac{K'}{\eta}\right)_\zeta^2\right]\right\}_\zeta$$

After expanding the derivatives on both sides this leaves

$$-K'^2\frac{\eta_\zeta}{\eta} + g\eta\eta_\zeta = \frac{K'^2}{3}\left(\frac{2\eta_\zeta\eta_{\zeta\zeta}}{\eta} - \frac{\eta_\zeta^3}{\eta^2} - \eta_{\zeta\zeta\zeta}\right)$$

which rearranges to

$$-3\eta_\zeta + \frac{3g}{K'^2}\eta^3\eta_\zeta = 2\eta\eta_\zeta\eta_{\zeta\zeta} - \eta_\zeta^3 - \eta^2\eta_{\zeta\zeta\zeta}$$

Note that η here is the total quantity, so to compare with the KdV solution above, we rewrite in terms of $\eta_1 = \eta - \eta_0 = \eta - 1$ (normalizing the equilibrium level η_0 to 1). The equation is now

$$(\eta_1+1)^2\eta_{1\zeta\zeta\zeta} = 2(\eta_1+1)\eta_{1\zeta}\eta_{1\zeta\zeta} - \eta_{1\zeta}^3 + 3\eta_{1\zeta} - \frac{3}{K'^2}(\eta_1+1)^3\eta_{1\zeta}$$

where we have also set $g = 1$ for convenience. Finally we have the first system of equations tried numerically for Su-Gardner,

$$\begin{aligned} Y_1 &= \eta_1 \\ Y_2 &= \eta_{1\zeta} \\ Y_3 &= \eta_{1\zeta\zeta} \\ Y_4 &= \int_0^x \eta_1\, dx \end{aligned}$$

and its derivatives

$$\begin{aligned} Y_1' &= Y_2 \\ Y_2' &= Y_3 \\ (Y_1+1)^2 Y_3' &= 2(Y_1+1)Y_2Y_3 - Y_2^3 + 3Y_2 - (3/K^2)(Y_1+1)^3Y_2 \\ Y_4' &= Y_1 \end{aligned}$$

With four unknowns we require four boundary conditions, which are to specify the amplitude $\eta_1 = H$ at both endpoints, and to set the first derivative and the integral of η_1 to zero at the first endpoint. The same initial guesses and endpoint amplitude as above were used, and various values of the integration constant K' were tried, but the system failed to converge. A second attempt was made by additionally solving for the eigenvalue $Y_4 = K'$ (making Y_5 the mass integral, as in the KdV case above), but this also did not converge.

This numerical work was begun in the later stage of the summer, and was not successfully completed by the end. To continue in this direction, clues to the lack of convergence might be sought in the recent work of Carter and Cienfuegos [1], who numerically examine the stability of Su-Gardner solutions to small perturbations along the direction of propagation and find that only solutions of sufficiently small amplitude and steepness are stable.

5 Conclusion

By the end of this project, I hoped to have an answer, either analytic or numerical, for the stability of Su-Gardner solutions to small transverse perturbations. The desired end result was a better understanding of large-amplitude waves in shallow water, specifically the stability of known 1D solutions to small transverse perturbations. Toward this goal I made the most progress in the one-dimensional analytics, which are now well understood. Analysis of the two-dimensional case was also begun but not yet complete, with linear

stability results obtained for the simplest case of constant η_0 and u_0, and the more general case most likely needing a numerical treatment. Numerical results in 1D for the purposes of extending to the 2D problem were begun but not completed, with several avenues remaining to explore to track down the non-convergence of the code. To continue from this point will necessitate further attention to the numerical solutions in one dimension, to compare with recent results [1] before extending along the second direction of propagation.

References

[1] J. Carter and R. Cienfuegos, *Periodic solutions of the Serre equations*, (submitted).

[2] G. El, R. Grimshaw, and N. Smyth, *Transcritical shallow-water flow past topography: finite-amplitude theory*, Journal of Fluid Mechanics (submitted).

[3] ——, *Unsteady undular bores in fully nonlinear shallow-water theory*, Physics of Fluids, 18 (2006), p. 027104.

[4] A. Green and P. Naghdi, *A derivation of equations for wave propagation in water of variable depth*, Journal of Fluid Mechanics, 78 (1976), pp. 237–246.

[5] B. Nadiga, L. Margolin, and P. Smolarkiewicz, *Different approximations of shallow fluid flow over an obstacle*, Physics of Fluids, 8 (1996), pp. 2066–2077.

[6] F. Serre, *Contribution à létude des écoulements permanents et variables dans les canaux*, Houille Blanche, 8 (1953), pp. 374–388.

[7] C. Su and C. Gardner, *Korteweg–de Vries equation and generalizations. III. Derivation of the Korteweg–de Vries equation and Burgers equation*, Journal of Mathematical Physics, 10 (1969), pp. 536–539.

Laboratory Experiments on Two Coalescing Axisymmetric Turbulent Plumes in a Rotating Fluid

Hiroki Yamamoto, supervised by Claudia Cenedese

October 20, 2009

1 Introduction

Turbulent plumes from a point source can be observed at various scale. Eruption of a volcano, smoke from a chimney, and seafloor vents are good examples. Such kind of plumes are called axisymmetric turbulent plumes. The presence of ambient rotation and/or stratification affects the behavior of axisymmetric turbulent plumes, and various experiments were carried out (see Table 1 of Bush and Woods (1999) for a summary).

Fernando, Chen, and Ayotte (1998; FCA hereafter) investigated the development of a single axisymmetric turbulent plume in a homogeneous, rotating fluid. They showed experimentally that the rotational effects become important after the plume descends a vertical distance $z_f \approx 3.3(F_0/\Omega^3)^{\frac{1}{4}}$ for a time $0.75T_f$. Here F_0 is the buoyancy flux, Ω the angular velocity of background rotation, and $T_f = \pi/\Omega$ the inertial period. They also showed that the plume's front descends as $z \approx 1.7F_0^{\frac{1}{4}}t^{\frac{3}{4}}$ for $t \leq 0.75T_f$ and $z \approx 1.8F_0^{\frac{1}{4}}t^{\frac{3}{4}}$ for $t > 0.75T_f$.

Meanwhile, the coalescence of axisymmetric turbulent plumes to form a single plume is an interesting problem. Kaye and Linden (2004; KL hereafter) investigated this problem in a non-rotating, homogeneous fluid. They introduced a theoretical model for merging height of two coalescing axisymmetric turbulent plumes in a homogenous fluid. They considered both equal and unequal plumes' cases. Their theory for equal plumes found that the relation between the separation length x_0 and the merging height z_m is $z_m \approx (0.44/\alpha)x_0$, where α is the entrainment constant. They also carried out laboratory experiments and showed that their model is qualitatively correct but over predicts the merging height slightly. However, they mentioned that a quantitative prediction is highly dependent on the value of entrainment constant.

In this study, we investigate the behavior of two coalescing plumes in a homogeneous, rotating fluid. During my project at the GFD Summer School, three kinds of experiments were carried out: a filling tank experiment (Baines and Turner, 1969), two plumes experiments without rotation (KL), and two plumes experiments with rotation. The filling tank experiment was carried out to determine the values of the virtual origin height and the entrainment constant; this is explained in appendix A. The derivation of the theory of KL is explained in appendix B.

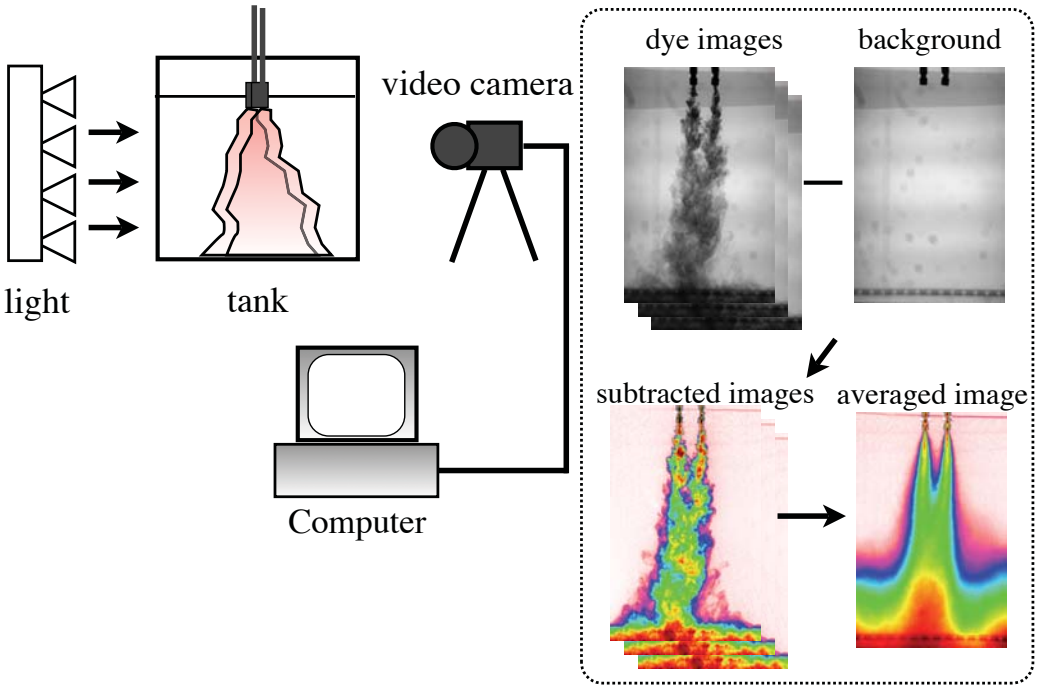


Figure 1: Schematic illustration of the dye attenuation technique.

2 Experimental set up

A transparent tank which has 60 × 60 cm^2 cross section was used. Fresh water (1.0 g/cm^3) was filled up to 45 cm depth for the ambient fluid, and dyed sea water (1.25 g/cm^3) was used for the fluid of the axisymmetric turbulent plume. The plume was generated by 5 mm diameter Cooper nozzle which was positioned just below the surface, and its constant flow rate was 1.7 cm^3/s, namely the buoyancy flux was $F_0 = 41$ cm^4/s^3.

A dye attenuation technique was used to visualize the plumes' behavior. Figure 1 shows a schematic illustration of this technique. A sequence of dye images were captured with side-view video recordings and the background image was subtracted from them using Digiflow. The color tone of the subtracted images shows the snapshots of distribution of dye concentration (i.e. distribution of buoyancy). This is because the intensity of the light recorded is associated with the dye concentration. Then the sequence of subtracted images were averaged to obtain the image of average distribution of buoyancy. The time interval between dye images is shown in table 1. The averaging period was 2 minutes for the non-rotating cases and $0.25T_f$ for the rotating cases. See KL for details of this technique.

For accuracy, we wrote and used an algorithm to determine the merging height. Our algorithm works as follows.

1. Smooths horizontally the time averaged image by box averaging every 7 pixels (if $x_0 > 6$ cm, 13 pixels) to remove noise.

2. Finds a large slope of concentration ($dc/dx > \delta$; c is concentration) starting from the left.

$f = 2\Omega$ (s^{-1})	time interval (s)
0 ($x_0 < 6$ cm)	1
0 ($x_0 > 6$ cm)	0.25
0.05	1
0.1	0.5
0.25	0.25
0.5	0.1
0.75	0.1
1	0.1

Table 1: Time interval between dye images.

3. From this point, finds the location where the slope is negative ($dc/dx < 0$), and defines this point as x_1.

4. Finds a large slope of concentration ($dc/dx < -\delta$) starting from the right.

5. From this point, finds the location where the slope is positive ($dc/dx > 0$), and defines this point as x_2.

6. Comparers x_1 with x_2 at every height, then $x_1 > x_2$ indicates that plumes are merged, $x_1 < x_2$ indicates that plumes are not merged.

In order to visualize the vortices generated by the plumes in the rotating experiments, the free surface was colored with powder dyes. The evolution of the resulting vortices were recorded with top-view video recordings. Note that it is reasonable to consider that the vortices are barotropic, because of the absence of stratification.

For the non-rotating cases, 23 experiments were carried out, and the separation length x_0 varied between 2.4 and 10.3 cm.

For the rotating cases, the Coriolis parameter $f = 2\Omega$, where Ω is the angular velocity of the background rotation, and the separation length x_0 are important parameters. We carried out rotating experiments with combinations of $f = 0.05, 0.1, 0.25, 0.5, 0.75, 1.0$ s^{-1} and $x_0 \approx 3, 5, 8, 10$ cm.

3 Experimental results

3.1 Without rotation

A summary of the two plumes experiments without rotation is shown in figure 2. The solid line and dashed line in figure 2 are the linear fit of our experimental data and the theoretical prediction, $z_m = (0.44/\alpha)x_0$, from the model of KL, respectively. Note that z_m is the merging height measured from the virtual origin. The virtual origin height $z_V = 1.0$ cm and the entrainment constant $\alpha = 0.12$ were determined by the filling tank experiment (Baines and Turner, 1969); see appendix A for details.

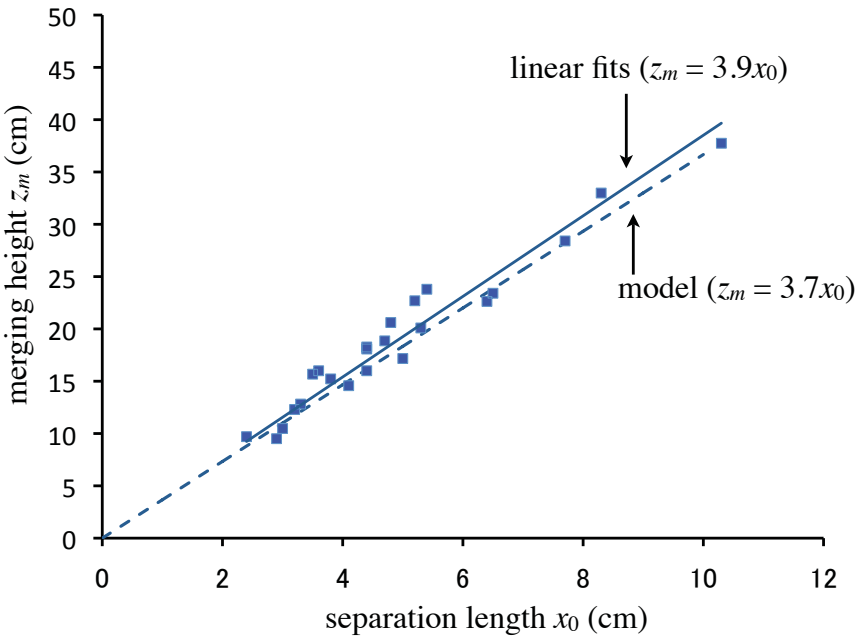


Figure 2: Plume merging height z_m plotted against the initial separation length x_0. The solid line is the linear fit of experimental data and the dashed line is the theoretical prediction, $z_m \approx (0.44/\alpha)x_0$, from the model of Kaye and Linden (2004).

Our experimental results show the linear relationship between x_0 and z_m, which is in agreement with the theoretical prediction. However it should be mentioned that the prediction $z_m = (0.44/\alpha)x_0$ is highly dependent on the value of α.

3.2 With rotation

Figure 3 shows the evolution of two plumes and the single vortex induced by the plumes for $x_0 = 5$ cm, $f = 0.25$ s^{-1}. It was observed that the axes of the plumes tilted off the vertical and a single vortex was generated at the center of two plumes for $t > 0.8T_f$. Figure 4 shows a similar sequence for $x_0 = 5$ cm, $f = 1$ s^{-1}. In this case, two vortices are generated, one from each plume. After $t \sim 3T_f$ the vortices started shedding from the sources (not shown).

In all cases, the plumes did not reach a steady state, because they were advected by the vortex/vortices which was/were generated by the plumes themselves. These processes are so complicated that, in this study, we do not consider the plumes after rotation dominates the dynamics but focus on the vortex/vortices generated by the plumes.

Types of vortex/vortices generated by the plumes for each case are shown in table 2.

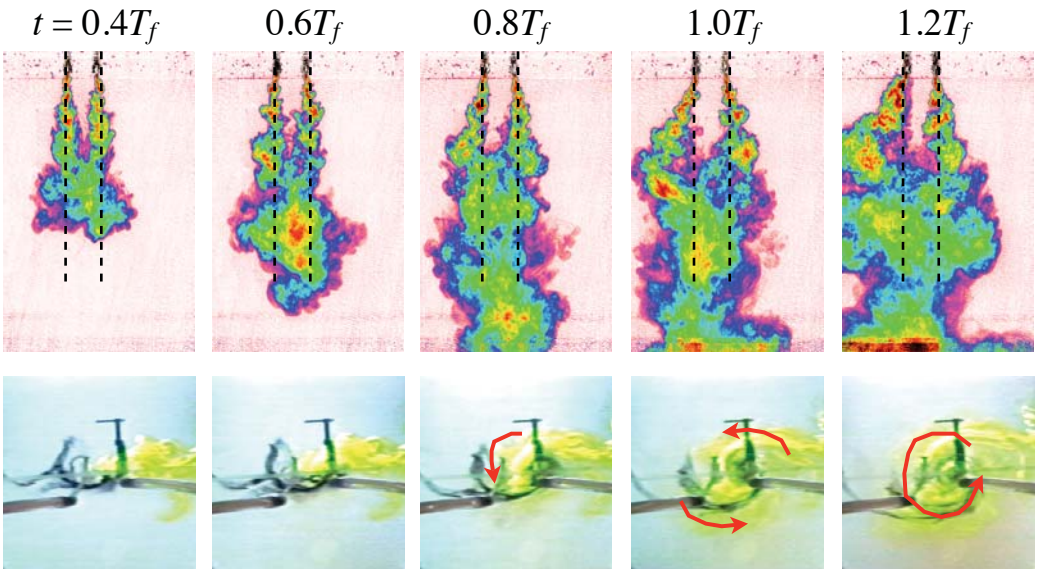


Figure 3: Evolution of two plumes with separation $x_0 = 5$ cm, rotation $f = 0.25$ s^{-1}. Upper pictures are snapshots of side-view dye concentration, and lower pictures are snapshots of top view. $T_f = 2\pi/f$ is the inertial period.

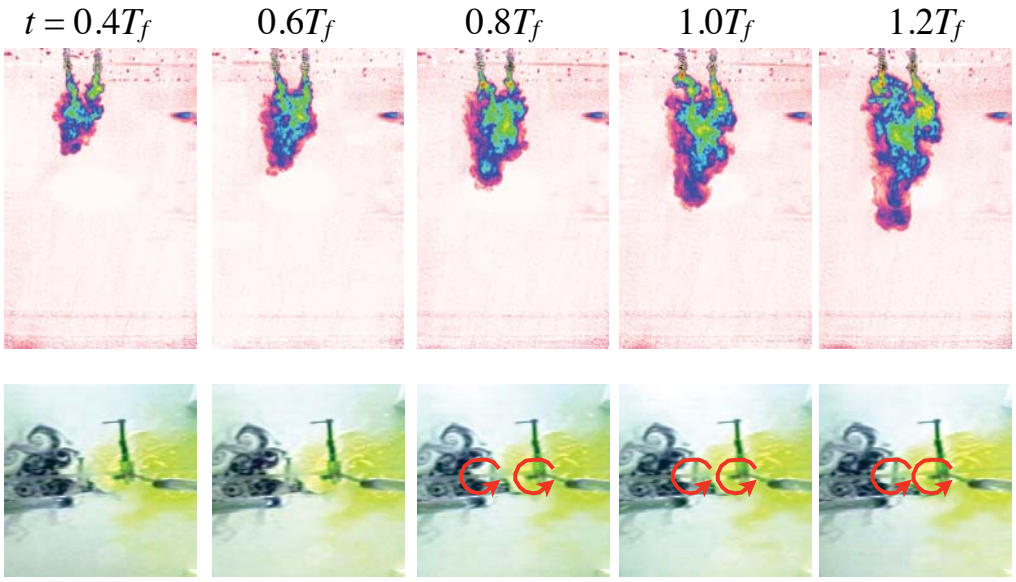


Figure 4: Same as figure 3 but for rotation $f = 1$ s^{-1}.

		x_0 (cm)			
		3	5	8	10
$f(\mathrm{s}^{-1})$	0.05	**1**	**1**	**1**	**1**
	0.1	**1**	**1**	**1**	**1**
	0.25	**1**	**1**	**1**	**2**
	0.5	**1**	**1**	**2**	**2**
	0.75	**1**	**2**	**2**	**2**
	1	**2**	**2**	**2**	**2**

Table 2: Types of vortices generated by the plumes. Number **1** denotes that a single vortex was generated at the center of two plumes, and **2** denotes that two vortices were generated, one from each plume.

4 Quantitative analysis

4.1 Merging height

Figure 5 shows the relation between z_m and x_0 for $0.5T_f < t < 0.75T_f$ for the rotating cases. It is shown that, in this time range, the two plumes do not feel the effect of rotation, so the merging height is same as in the non-rotating cases. In cases of $(f, x_0) = (0.05, 10)$, $(0.25, 3)$, and $(0.5, 3)$, z_m obtained for $0.5T_f < t < 0.75T_f$ are far from the theoretical prediction, but those obtained for $0.25T_f < t < 0.5T_f$ agree well with the prediction. For $(f, x_0) = (0.05, 10)$, this is because the averaging time interval was so long that z_m was affected by dyed water filling up the tank from the bottom. For $(f, x_0) = (0.25, 3)$ and $(0.5, 3)$, the effect of rotation may have appeared earlier.

Figure 6 shows the relation between z_m and x_0 for $0.75T_f < t < T_f$ for the rotating cases. For this averaging time interval, z_m is affected by rotation, and there is not a clear relation between z_m and x_0. The scatter S defined as

$$S = \sqrt{\frac{\sum_i^N (z_{m,i}^{\mathrm{theory}} - z_{m,i}^{\mathrm{experiment}})^2}{N}}, \tag{1}$$

where N is the number of experiments, $z_{m,i}^{\mathrm{theory}}$ the theoretical prediction of the merging height, $z_{m,i}^{\mathrm{experiment}}$ the experimental merging height, for the non-rotation experiments, rotating experiments for $0.5T_f < t < 0.75T_f$, and for $0.75T_f < t < T_f$ is 1.85, 2.89, and 10.98 cm, respectively.

From the above results, we deduce that the effect of rotation becomes important after $t > 0.75T_f$, in agreement with the result of FCA. Our experimental results also show that the merging height for $t < 0.75T_f$ agrees with the theoretical model for no rotation.

4.2 Vortex generated by the plumes

The process of generating a single or two vortices can be described as follows. When the separation length is small and the rotation rate is low, the plumes feel the effects of rotation after they merged and a single vortex will be generated by the resulting single plume. On

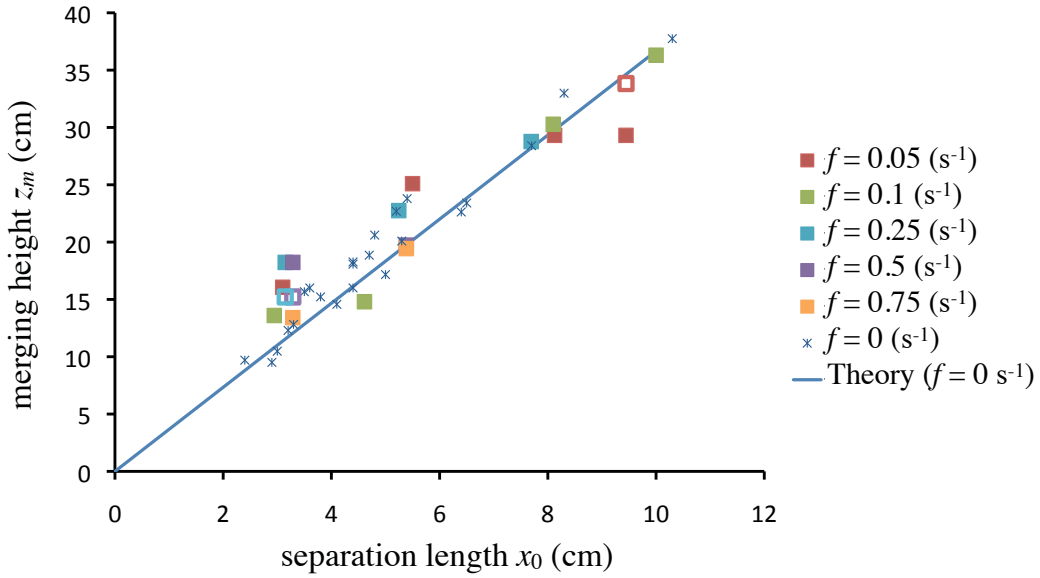


Figure 5: Plume merging height z_m plotted against the separation length x_0 for $0.5T_f < t < 0.75T_f$ (filled square), where T_f is the inertial period. Asterisks denote the results of experiments without rotation, solid line denotes the theoretical prediction for no rotation, $z_m = (0.44/\alpha)x_0$. Open squares are z_m for $0.25T_f < t < 0.5T_f$.

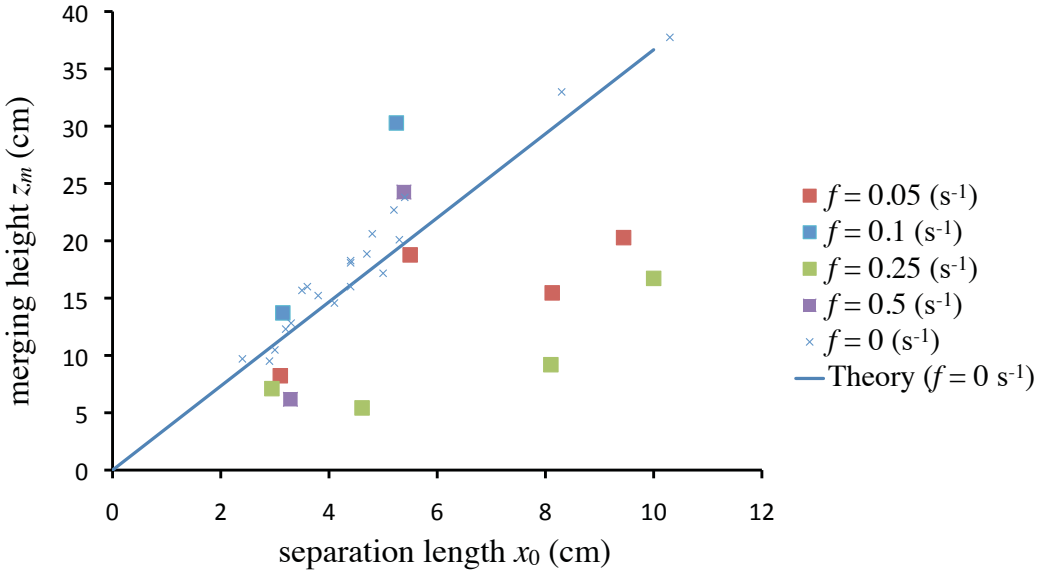


Figure 6: Same as figure 5 but for $0.75T_f < t < T_f$.

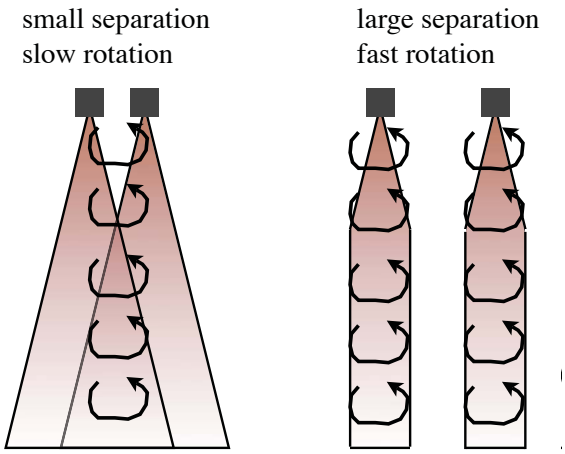


Figure 7: Schematic illustration of the two plumes generating a single vortex after merging (left) and two vortices before merging (right).

the other hand, when the separation length is large and the rotation rate is high, the plumes feel the effects of rotation before merging and two vortices are generated, one by each plume (figure 7).

Considering the process described above, we can predict the number of vortices generated by the two plumes using the results of FCA for the development of a single plume in a homogeneous, rotating fluid, and the theory of KL for the merging height in a non-rotating fluid. FCA showed experimentally that a single axisymmetric turbulent plume's front in a homogeneous, rotating fluid descends as

$$z \approx 1.7 F_0^{\frac{1}{4}} t^{\frac{3}{4}} \quad \text{for} \quad t \leq 0.75 T_f, \tag{2}$$

$$z \approx 1.8 F_0^{\frac{1}{4}} t^{\frac{3}{4}} \quad \text{for} \quad t > 0.75 T_f. \tag{3}$$

We measured the plume's front descent in the rotating and non-rotating experiments (figure 8). Our experimental results agree well with FCA.

Figures 5, 6, and 8 show that the time when rotational effects become important and the descent of the plume's front in our experiments agree with FCA. Therefore, the depth z_f where plumes feel rotation can be written as

$$z_f \approx 1.7 F_0^{\frac{1}{4}} (0.75 T_f)^{\frac{3}{4}} \approx 5.5 F_0^{\frac{1}{4}} f^{-\frac{3}{4}}. \tag{4}$$

Remember that the theoretical prediction of the merging height

$$z_m \approx \frac{0.44}{\alpha} x_0, \tag{5}$$

can be applied to rotating cases for $t < 0.75 T_f$. Therefore, we can predict that a single vortex will be generated if $z_f > z_m$, and two vortices will be generated if $z_f < z_m$. Figure 9 shows the number of generated vortices in the z_m-z_f plane. Our experimental results agree well with the prediction described above. Therefore, we can predict the number of vortices generated by the two plumes from the values of the entrainment constant (α), separation length (x_0), Coriolis parameter (f), and buoyancy flux (F_0).

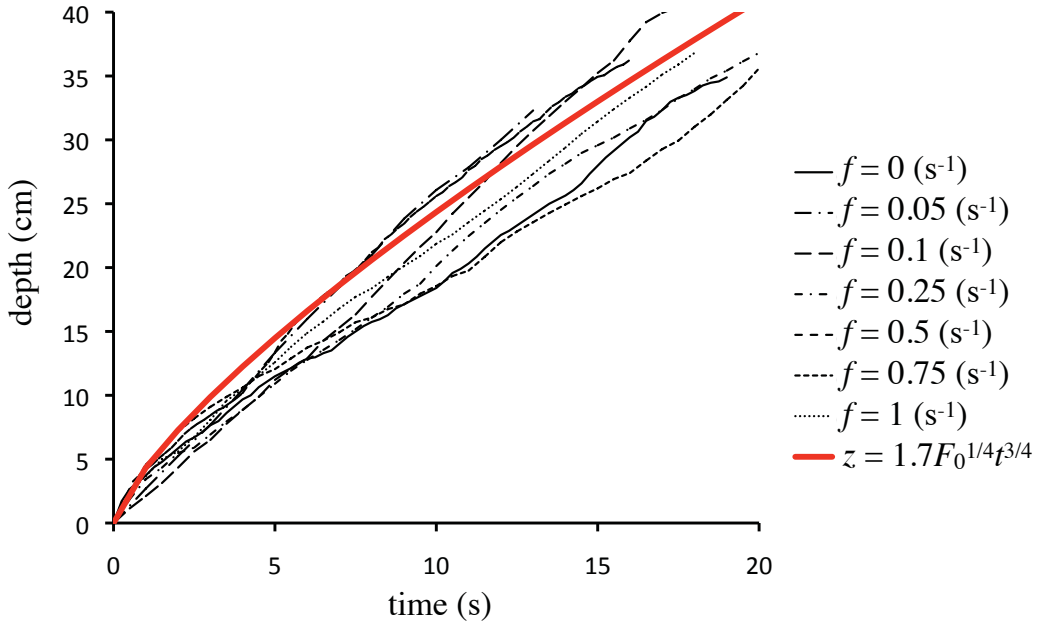


Figure 8: Depth of the plume's front as a function of time. Black lines are experimental results for various rotation rates and the red line is $z \approx 1.7F_0^{\frac{1}{4}}t^{\frac{3}{4}}$(Fernando, Chen, and Ayotte, 1998).

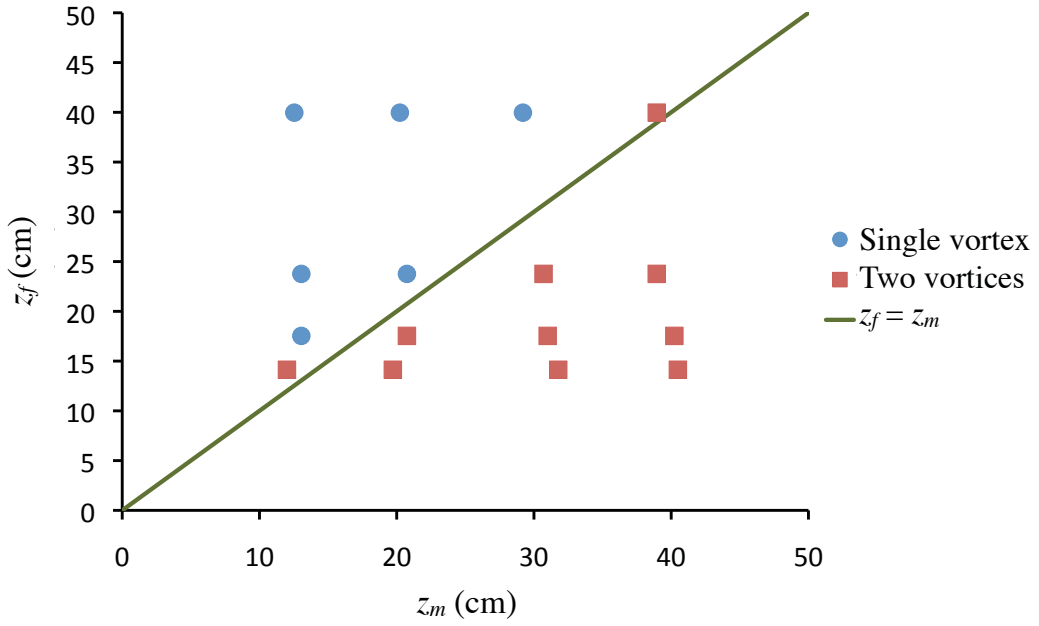


Figure 9: Number of vortices generated by the plumes, in the z_m-z_f plane. Blue circles denote single vortex and red squares denote two vortices. Solid line denotes $z_f = z_m$. The results for $f = 0.05, 0.1$ s^{-1} are not shown.

5 Summary

Laboratory experiments on two coalescing axisymmetric turbulent plumes in a homogeneous, non-rotating and rotating fluids were carried out. The merging height z_m is defined as the height from the virtual origin where the time averaged horizontal buoyancy profile has a single local maximum. The values of the virtual origin height $z_V = 1.0$ cm and entrainment constant $\alpha = 0.12$ were determined by the filling tank experiment (Baines and Turner, 1969).

For the non-rotating experiments, our experimental results (figure 2) agree well with the theory of Kaye and Linden (2004) predicting a merging height $z_m \approx (0.44/\alpha)x_0$, where x_0 is an initial separation length of the two plumes' sources. However, as mentioned in Kaye and Linden (2004), this prediction is highly dependent on the value of α.

For the rotating experiments, our experimental results show that the merging height z_m agrees with the theory of Kaye and Linden (2004) for no rotation, for $t < 0.75T_f$ (figure 5). However, for $t > 0.75T_f$, the plumes were not steady and z_m did not agree with the theory (figure 6). Namely, the effects of rotation became important after $t > 0.75T_f$; this is in agreement with the experimental results of Fernando, Chen, and Ayotte (1998) who studied the development of a single axisymmetric turbulent plume in a homogeneous, rotating fluid. Fernando, Chen, and Ayotte (1998) also obtained experimentally the height of the plume's front as a function of time t as $z \approx 1.7F_0^{\frac{1}{4}}t^{\frac{3}{4}}$, and our results are in agreement with this expression (figure 8).

In the rotating experiments, vortices were generated by the plumes. Two types of vortices were observed: a single vortex generated at the center of the two plumes, and two vortices generated, one from each plume. We showed that the number of vortices can be predicted using the theory for the merging height $z_m \approx (0.44/\alpha)x_0$ and the experimental prediction for the height where the plumes start feeling rotation $z_f \approx 5.5F_0^{\frac{1}{4}}f^{-\frac{3}{4}}$. When $z_f > z_m$, a single vortex is generated, and when $z_f < z_m$, two vortices are generated (figure 9).

Acknowledgments

I would like to thank Claudia Cenedese for her dedicated guidance and advise. I also thank Colm-cille P. Caulfield for his useful comments, and all the other staff and fellows of the Geophysical Fluid Dynamics Program at the Woods Hole Oceanographic Institution in 2009. I participated in this program with a financial support from the Center for Planetary Science.

Appendices

A Filling tank experiment

In order to determine the values of the virtual origin height and the entrainment constant a filling tank experiment was carried out. Theoretically, the origin of an axisymmetric turbulent plume is assumed to be a point. However, in laboratory experiments, the real

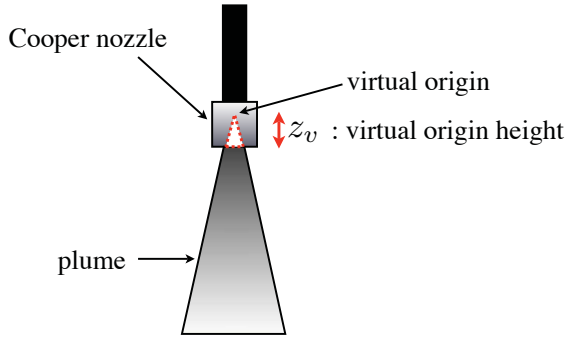


Figure 10: Schematic illustration of virtual origin.

origin of the plume (i.e. Cooper nozzle) has a finite cross section, so it is assumed that there is a virtual origin which is considered to be a point source of the plume (figure 10).

Turbulent plumes entrain the ambient fluid by turbulent mixing, and the horizontal velocity of the entrained flow u_e is assumed to be $u_e = -\alpha w$, where α is the entrainment constant and w is the vertical velocity. This assumption is called the entrainment assumption, and it is a good approximation (Turner, 1973). However the value of α has to be determined experimentally.

The filling tank experiment is a good method to determine both the virtual origin height and the entrainment constant.

A.1 Theory of the filling tank experiment

In this section, the theory of the filling tank experiment (Baines and Turner, 1969) is explained. First, some general properties of an axisymmetric turbulent plume are derived. Then the method for determining both the virtual origin height and the entrainment constant is explained.

A.1.1 General properties of an axisymmetric turbulent plume

It is assumed that the profiles of vertical velocity w and buoyancy g' of the plume are of Gaussian form,

$$w(r,z) = W(z)\exp\left(-\frac{r^2}{b(z)^2}\right), \tag{6}$$

$$g'(r,z) = g\left(\frac{\rho_A - \rho_P(r,z)}{\rho_R}\right) = G'(z)\exp\left(-\frac{r^2}{b(z)^2}\right). \tag{7}$$

Here z is the vertical coordinate, r the radial distance from the axis of the plume, g the gravitational acceleration , $W(z) = w(z,0)$, and $G'(z) = g'(z,0)$. Then ρ_A and ρ_P are the densities of the ambient fluid and the plume, respectively, ρ_R is the reference density for the system. Note that $b(z)$ is the radius where buoyancy and vertical velocity are reduced by a factor $1/e$ from those on the plume's axis.

Volume flux Q of the axisymmetric turbulent Gaussian plume is

$$Q = \int_0^{2\pi} \int_0^{\infty} w r dr d\theta = \pi b^2 W, \tag{8}$$

where θ is the angular coordinate.

The momentum flux M is

$$M = \int_0^{2\pi} \int_0^{\infty} w^2 r dr d\theta = \frac{\pi}{2} b^2 W^2. \tag{9}$$

Then b can be written as

$$b = \frac{1}{\sqrt{2}} \pi^{-\frac{1}{2}} \frac{Q}{M^{\frac{1}{2}}}. \tag{10}$$

The buoyancy flux F is

$$F = \int_0^{2\pi} \int_0^{\infty} g' w r dr d\theta = \frac{G'}{2} Q. \tag{11}$$

The velocity of the entrained flow, u_e, is now assumed to be horizontal,

$$u_e = -\alpha W, \tag{12}$$

where α is the entrainment constant.

The volume flux Q increases with z, because of the entrained flow, so

$$\begin{aligned} \frac{\partial Q}{\partial z} &= -2\pi b u_e, & (13)\\ &= 2\pi \alpha b W, & (14)\\ &= 2\sqrt{2} \pi^{\frac{1}{2}} \alpha M^{\frac{1}{2}}. & (15) \end{aligned}$$

Because the plume is accelerated by buoyancy, the momentum flux changes with z as

$$\begin{aligned} \frac{\partial M}{\partial z} &= \pi b^2 G', & (16)\\ &= \frac{(\pi G' b^2 W)/2 \cdot \pi b^2 W}{(\pi b^2 W^2)/2}, & (17)\\ &= \frac{FQ}{M}. & (18) \end{aligned}$$

We can set $F = F_0$ (constant) because the entrainment not only decrease the buoyancy g' but also increase the flow flux Q. Therefore, equation (18) can be written as

$$\frac{\partial M}{\partial z} = \frac{F_0 Q}{M}. \tag{19}$$

Now, we assume that Q and M can be expressed as

$$\begin{aligned} Q &= Q_0 (z + z_V)^q, & (20)\\ M &= M_0 (z + z_V)^m, & (21) \end{aligned}$$

where z_V is the height of virtual origin, and Q_0, M_0 are constants. This assumption implies that the radius of the plume, b, increases linearly from 0 at its virtual origin $z = -z_V$. Then (15) and (19) become

$$Q_0 q (z + z_V)^{q-1} = 2\sqrt{2}\pi^{\frac{1}{2}} \alpha M_0^{\frac{1}{2}} (z + z_V)^{\frac{m}{2}}, \tag{22}$$

$$M_0 m (z + z_V)^{m-1} = \frac{F_0 Q_0}{M_0} (z + z_V)^{q-m}. \tag{23}$$

From the above equations, we obtain

$$q = \frac{5}{3}, \tag{24}$$

$$m = \frac{4}{3}. \tag{25}$$

Substituting (24) and (25) into (22) and (23), with some algebra, we obtain

$$M_0 = \left(\frac{9\sqrt{2}\pi^{\frac{1}{2}} \alpha F_0}{10} \right)^{\frac{2}{3}}, \tag{26}$$

$$Q_0 = \frac{6\pi^{\frac{1}{2}} \alpha}{5} \left(\frac{18\pi^{\frac{1}{2}} \alpha F_0}{5} \right)^{\frac{1}{3}}. \tag{27}$$

Eventually, the volume flux Q and the momentum flux M are written in the buoyancy flux F_0 and the entrainment constant α.

A.1.2 Method to determine the virtual origin height and the entrainment constant

We consider a tank of cross section A_C filled with ambient fluid, and an axisymmetric turbulent plume which is denser than the ambient fluid and is dyed, descending from the top of the tank ($z = 0$), as in figure 11. After the front of the plume reaches bottom of the tank, a layer of dyed fluid forms at the bottom of the tank and starts filling up the tank. We set the time $t = 0$ when $z_F(t) = z_{F0}$, where z_F is the height of the dyed fluid's front which comes up from the bottom. It is convenient to choose z_{F0} as the height where we can measure z_F accurately enough.

Considering the balance of the volume flux of the dyed fluid at $z = z_F$, the time derivative of $(z_F(t) + z_V)$ can be written as

$$\begin{aligned} \frac{\partial}{\partial t}(z_F(t) + z_V) &\approx -\frac{Q(z_F)}{A_C}, \\ &= -\frac{Q_0}{A_C}(z_F + z_V)^{\frac{5}{3}}. \end{aligned} \tag{28}$$

Here it is assumed that the cross section of the plume is negligible. Integrating (28), we

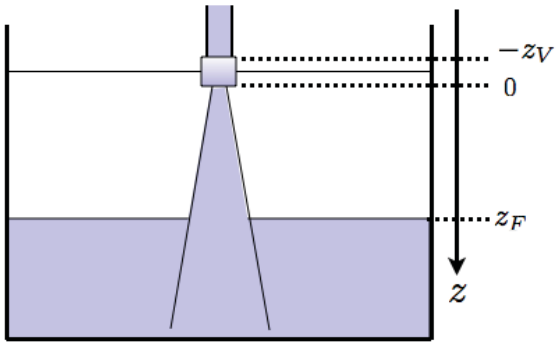


Figure 11: Schematic illustration of the filling tank experiment.

obtain

$$\int_{z_{F0}}^{z_F} (z_F + z_V)^{-\frac{5}{3}} d(z_F + z_V) = -\frac{Q_0}{A_C} \int_0^t dt, \tag{29}$$

$$-\frac{3}{2}\left[(z_F + z_V)^{-\frac{2}{3}} - (z_{F0} + z_V)^{-\frac{2}{3}}\right] = -\frac{Q_0}{A_C} t, \tag{30}$$

$$(z_F + z_V)^{-\frac{2}{3}} = \frac{2}{3}\frac{Q_0}{A_C} t + (z_{F0} + z_V)^{-\frac{2}{3}}. \tag{31}$$

In other words, there is a linear relation between $(z_F + z_V)^{-\frac{2}{3}}$ and t.

We can measure z_F and t from the experiment, so we can determine the value of z_V which produce a straight line in the plot of $(z_F + z_V)^{-\frac{2}{3}}$ against t. After determining z_V, we can estimate the value of α by choosing the value that makes fit equation (31) best to the experimental data with the determined z_V.

A.2 Experimental set up

Experimental equipments and their settings are the same as those explained in section 2, but the depth of the fresh water was 40 cm. We captured the digital images every minute and determined z_F by looking at the the horizontally averaged (excluding the region of the plume) height where the value of concentration of the dye is that of the interface between fresh water and dyed water.

A.3 Experimental results

Figure 12 shows measured values of $(z_F + z_V)^{-\frac{2}{3}}$ against t for various values of z_V. From this figure we can determine $z_V = 1.0$ cm, because this value gives the highest R^2 for the linear fit to the data, where R^2 is the coefficient of determination. Figure 13 shows the measured values of $(z_F + z_V)^{-\frac{2}{3}}$ against t with $z_V = 1.0$ cm, and the theoretical lines calculated from equation (31) with various values of α. From this figure we can estimate that the value $\alpha = 0.12$ fit best the experimental data.

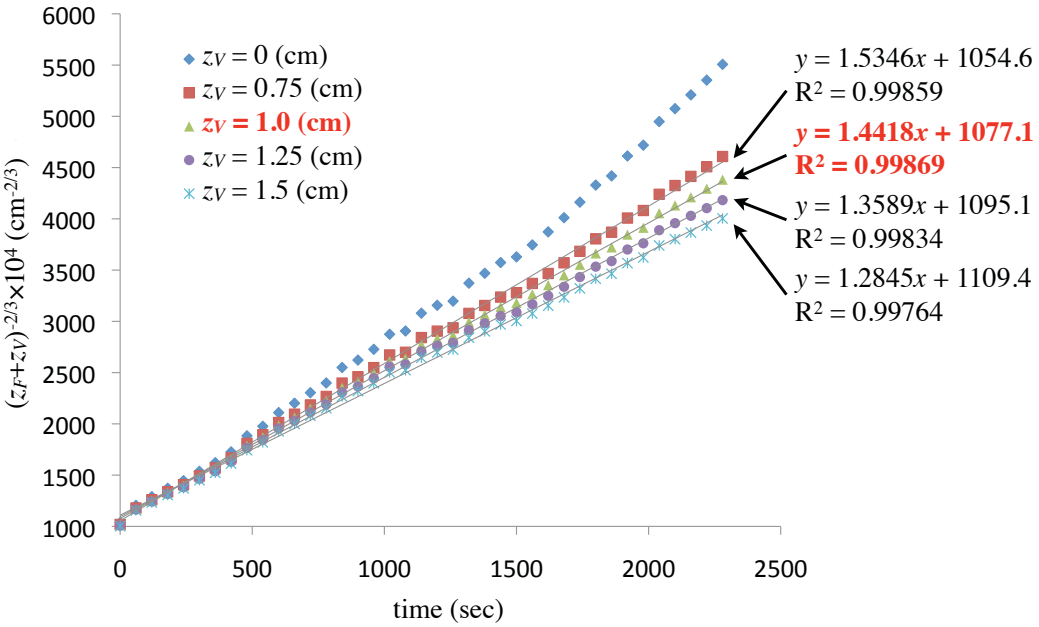


Figure 12: Measured values of $(z_F + z_V)^{-\frac{2}{3}}$ against t for various values of z_V. Solid lines are linear fits of each data set. This figure shows that the best linear fit (highest value of R^2) occurs for $z_V = 1.0$ cm.

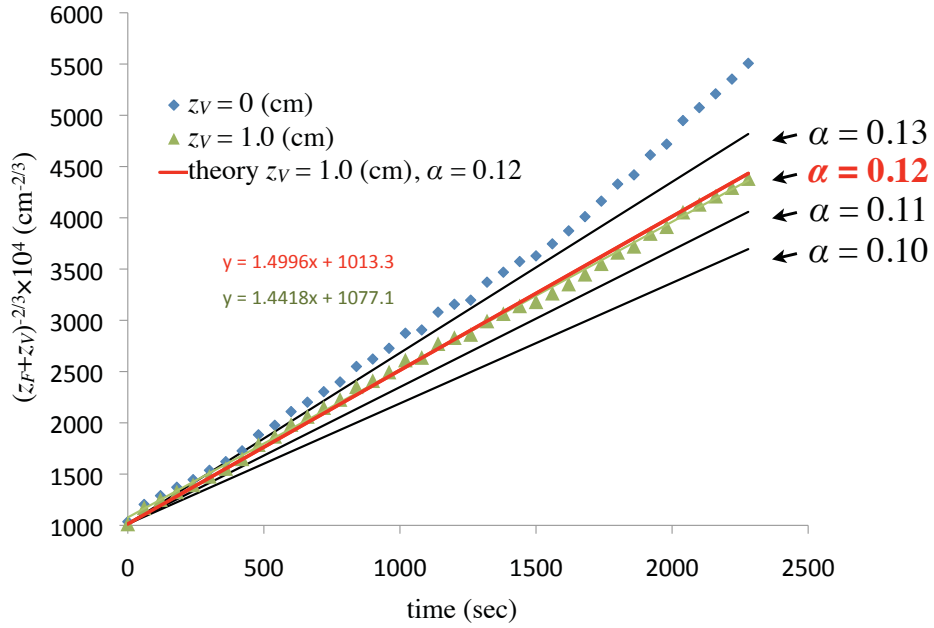


Figure 13: Measured values of $(z_F + z_V)^{-\frac{2}{3}}$ against t for $z_V = 1.0$ cm. Thick solid lines are theoretical lines calculated from equation (31) with various values of α. This figure shows that the value of $\alpha = 0.12$ fit best the experimental data.

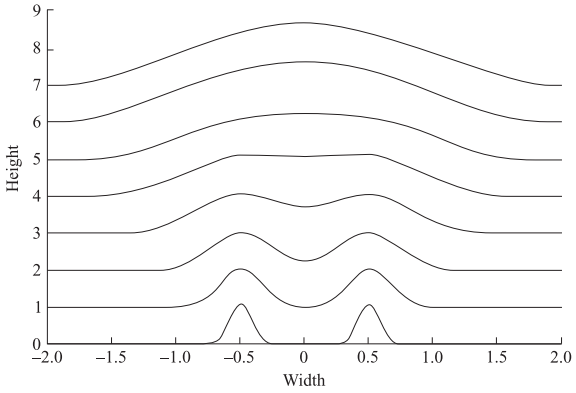


Figure 14: Merging Gaussian functions for a unit sources separation. (Kaye and Linden, 2004)

B Theory of coalescing axisymmetric turbulent equal plumes

In this section the theory of Kaye and Linden (2004) for equal plumes is summarized.

First, we define dimensionless variables as followings,

$$\lambda = \frac{z}{x_0}, \quad \phi = \frac{x}{x_0}, \quad \gamma = \frac{b}{x}, \tag{32}$$

where z is the height above the point plume sources (i.e. including the virtual origin), x_0 the separation of the plume sources, x the separation of the plume axes at any given height, and b the plume radius.

Consider two equal plumes with origins at the same height. The average buoyancy profile of a single turbulent plume can be taken as Gaussian, with a radius given by

$$b = \frac{6}{5}\alpha z, \tag{33}$$

where α is the entrainment constant. Assuming that the plumes do not interact, the buoyancy profile function is given by

$$g'(r,z) \sim G'(z)E(r), \tag{34}$$

$$E(r) = \exp\left[-\frac{(r-\frac{1}{2}x_0)^2}{b^2}\right] + \exp\left[-\frac{(r+\frac{1}{2}x_0)^2}{b^2}\right]. \tag{35}$$

Here r is the distance from the center of the two plumes and g' is the buoyancy.

The function (34) is plotted in figure 14 for $1 < \lambda < 8$ and $x_0 = 1$. Clearly the two Gaussians coalesce when λ is large enough. We define the merging height to be the height where the trough in the buoyancy profile disappears, that is,

$$\frac{d^2E}{dr^2} = 0 \quad \text{at} \quad r = 0. \tag{36}$$

For non-interacting plumes (i.e. $x(z) = x_0$), (36) is easily solved to give

$$x_0 = \sqrt{2} b_m. \tag{37}$$

Here the subscript m denotes the value at the point where plumes merge. Dividing (37) by x, we obtain

$$\gamma_m = \frac{1}{\sqrt{2}}, \tag{38}$$

so, in terms of the non-dimensional height we obtain

$$\hat{\lambda}_m = \frac{1}{\sqrt{2}} \frac{5}{6\alpha}. \tag{39}$$

Here $\hat{(\cdot)}$ denotes the upper bound on the value. As shown by Bjorn and Neilsen (1995) this estimate of λ_m is poor. In order to model the drawing together of two equal plumes we need to consider the entrainment of one plume by another.

Based on experimental results (e.g. Rouse, Yih, and Humpherys, 1952) it is reasonable to take the velocity field outside the plumes, created by entrainment, as horizontal. The mean entrainment velocity field, over a horizontal plane across the two plumes, may be approximated by two sinks of strength $-m(z)$ placed at $(-\frac{1}{2}x, z)$ and $(\frac{1}{2}x, z)$. The complex velocity potential in this horizontal plane is given by

$$\Psi = -\frac{m}{2\pi} \left[\ln\left(Z + \frac{1}{2}x\right) + \ln\left(Z - \frac{1}{2}x\right) \right], \tag{40}$$

where Z is the complex variable $re^{i\theta}$. The velocity field is given by

$$U = \frac{\partial \Psi}{\partial Z} = -\frac{m}{2\pi} \left(\frac{1}{Z + \frac{1}{2}x} + \frac{1}{Z - \frac{1}{2}x} \right). \tag{41}$$

The sink strength of the plume is

$$m = \int_0^{2\pi} b\alpha w d\theta. \tag{42}$$

Along the line joining the sources of the plume ($\theta = 0$) the velocity is given by

$$U_{\theta=0} = -b\alpha w \left(\frac{1}{r + \frac{1}{2}x} + \frac{1}{r - \frac{1}{2}x} \right). \tag{43}$$

On the plume axis ($r = -\frac{1}{2}x, \frac{1}{2}x$), the value of horizontal entrained velocity due to that plume is zero, and only the term resulting from the other plume needs to be considered. Therefore, (43) gives an expression for the mean horizontal velocity u on the plume axis

$$u = -\frac{b\alpha w}{x}. \tag{44}$$

Assuming that each plume is passively advected by the entrainment field of the other, the rate of change of the plumes' separation with height is given by the ratio of the vertical and horizontal velocities at the plume axis. As both plumes are deflected equally, the rate is

$$\frac{dx}{dz} = -2\frac{1}{w}\frac{\alpha b w}{x}. \tag{45}$$

Substituting $b = 6\alpha z/5$, $\lambda = z/x_0$, and $\phi = x/x_0$ gives

$$\frac{d\phi}{d\lambda} = -\frac{12\alpha^2}{5}\frac{\lambda}{\phi}, \tag{46}$$

which can be integrated to give

$$\phi^2 - \phi_0^2 = \frac{12\alpha^2}{5}(\lambda_0^2 - \lambda^2), \tag{47}$$

where $\phi_0 = 1$ and $\lambda_0 = 0$. Using (38) to (47) we obtain

$$\lambda_m = \frac{1}{\alpha}\sqrt{\frac{25}{132}} \approx \frac{0.44}{\alpha}. \tag{48}$$

In other word, the merging height is predicted as

$$z_m \approx \frac{0.44}{\alpha}x_0. \tag{49}$$

Note that the use of (38) derived for the non-interacting model is appropriate here as the correction for drawing together of the plumes assumes that they are passively advected only. The radial growth rate of each plume is assumed to be unaffected by this process.

References

[1] W. D. Baines and J. S. Turner, *Turbulent buoyant convection from a source in a confined region*, Journal of Fluid Mechanics, 37 (1969), pp. 51–80.

[2] E. Bjorn and P. V. Nielsen, *Merging thermal plumes in the internal environment*, in Proc. Healthy Buildings 95, M. Maroni, ed., 1995.

[3] J. W. M. Bush and A. W. Woods, *Vortex generation by line plumes in a rotating stratified fluid*, Journal of Fluid Mechanics, 388 (1999), pp. 289–313.

[4] H. J. S. Fernando, R.-r. Chen, and B. A. Ayotte, *Development of a point plume in the presence of background rotation*, Physics of Fluids, 10 (1998), pp. 2369–2383.

[5] N. B. Kaye and P. F. Linden, *Coalescing axisymmetric turbulent plumes*, Journal of Fluid Mechanics, 502 (2004), pp. 41–63.

[6] H. Rouse, C. S. Yih, and H. W. Humphreys, *Gravitational convection from a boundary source*, Tellus, 4 (1952), pp. 201–210.

[7] J. S. Turner, *Buoyancy Effects in Fluids*, Cambridge University Press, Cambridge, U.K., 1973.