Stochastic modelling of atmospheric gravity waves

Jonathan Powell

Doctor of Philosophy
University of Edinburgh
2004
Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.
Abstract

Internal gravity waves have an important effect on the large-scale circulation of the middle atmosphere, which is conditioned by the deposition of momentum due to their breaking. The propagation of gravity waves is influenced by the properties of the background wind. This thesis examines this influence: it uses stochastic techniques to study gravity wave propagation through a randomly fluctuating background wind.

It begins by describing general features of the atmosphere and gravity wave propagation. The basic equations of fluid flow within the atmosphere are derived. These lead via the WKB approximation to a dispersion relation and to ray equations for gravity wave propagation.

Propagation equations, such as the ray equations and the dispersion relation, are derived in a general context. The notion of a Wigner matrix is introduced, and this is used to derive transport equations for a general Hamiltonian system that may contain random components. These results generalise earlier works by Ryzhik et al. and Guo and Wang. Atmospheric gravity waves are described as an application and the equations derived via the WKB approximation are recovered.

The major factor influencing the distribution of gravity waves is the spread of their wavenumber as they propagate through a wind. This is described by the Doppler spreading model. A one-dimensional system with a randomly fluctuating background wind, dependent on altitude only, is considered. The model revisits that of Souprayen et al. by using an Ornstein-Uhlenbeck process to describe the wind. Simple equations for the energy spectrum induced by gravity waves are derived. Analytic forms of the energy spectrum are given and features of the spectrum such as the $m^{-3}$ (where $m$ is the vertical wavenumber) spectral tail, central wavenumber and scaling with the Brunt-Väisälä frequency are found to be consistent with observations. An equation for the force on the background, induced by gravity wave breaking, is also derived. The analytical results are backed up by numerical simulations.

Caustics, that is, singularities caused by the intersection of neighbouring rays, are often overlooked as a possible mechanism for wave breaking. A two-dimensional model is considered in which the background wind is constructed from several Fourier modes, with random phase and amplitude to represent the complexities of a real atmospheric wind. A single parameter controls the amplitude of the random fluctuations of the wind. Numerical techniques are used to detect caustics by computing the ray-tube area. Probability distribution functions for the altitude of caustic events are then obtained. A scaling of the altitude of caustic formation with the amplitude parameter of the wind is proposed.
Acknowledgements

I would like to thank my supervisor Jacques Vanneste for the large amounts of time and effort he has spent discussing topics and reading work. His enthusiasm and helpful suggestions provided me with motivation.

I would like to thank the University of Edinburgh, the Mathematics department and all its staff for providing support, assistance and a productive working environment.

I would also like to thank all the office mates I have had throughout my time as a research student. Their help and stimulation has been invaluable. In particular, thanks go to Saul Nsumbuga, Nikolas Geogakarakos and Thomas Lagland.

Thanks also to Stuart for giving up valuable time to print and bind this thesis as well insisting I take time off to go walking and climbing.

Last but certainly not least I would like to thank Catherine for her encouragement and moral support, not to mention many hours spent proof reading drafts of this thesis.
# Table of Contents

## Chapter 1  General introduction

1.1 A brief description of the atmosphere .................... 6

1.2 Gravity waves in the atmosphere ........................ 7

1.3 Observations of gravity waves .......................... 11
   1.3.1 Measurement methods ............................ 11
   1.3.2 Method of data processing ...................... 12
   1.3.3 Physical values of parameters for the atmosphere .... 13

1.4 Global circulation of the atmosphere .................... 13

1.5 Basic equations .................................. 17
   1.5.1 WKB approximation and ray equations ............. 19

1.6 Wave-breaking .................................. 22
   1.6.1 Wave-steepening mechanisms ..................... 22
   1.6.2 Wave-breaking mechanisms ....................... 24
   1.6.3 Wave breaking models .......................... 26

1.7 Description of Chapters ........................... 27

## Chapter 2  Transport equations for waves in randomly perturbed Hamiltonian systems

2.1 Introduction .................................. 30

2.2 Formulation .................................. 34

2.3 Wigner matrix .................................. 36
   2.3.1 Reduction to operators used in Ryzhik et al. (1996) .... 40

2.4 Derivation of the transport equations .................... 40
2.4.1 $O(1)$ equation ........................................ 41
2.4.2 $O(\varepsilon^{1/2})$ equation .......................... 42
2.4.3 $O(\varepsilon)$ equation ............................ 43
2.5 Application to gravity waves ............................ 46
2.6 Discussion and summary .................................. 50

Chapter 3 Doppler spreading of gravity waves  .... 53
3.1 Introduction ........................................... 53
  3.1.1 Doppler Spreading .................................. 54
  3.1.2 Chapter description .................................. 57
3.2 Stochastic model ....................................... 58
3.3 Basic equations ........................................ 61
  3.3.1 Boundaries in wavenumber space ...................... 64
3.4 Analytical solutions of (3.8) .......................... 65
  3.4.1 Solution in infinite domain .......................... 65
  3.4.2 Solution in finite domain ............................ 66
  3.4.3 Approximation for $\lambda$ ............................ 69
3.5 Numerical solution of (3.8) ........................... 71
3.6 Energy PSD curves ..................................... 73
  3.6.1 Infinite domain ..................................... 73
  3.6.2 Finite domain ....................................... 78
3.7 Wave-induced force ..................................... 79
3.8 Density variations ...................................... 84
3.9 Summary and discussion ................................ 87

Chapter 4 Development of caustics ................. 91
4.1 Introduction .......................................... 91
  4.1.1 Gravity wave caustics ............................. 92
  4.1.2 Chapter description ................................ 96
4.2 Stochastic formulation of White & Fornberg (1998) .... 96
Chapter 5 Conclusion

Appendix A Calculations for Chapter 2

A.1 Details concerning the functions $H$ and $J$ .......................... 123
A.2 Correlation functions ......................................... 124
A.3 Energy and Wigner matrix .................................... 125
A.4 Dirac delta function ........................................... 125
A.5 Calculation of equation for $W_e$ ............................... 126
A.6 Span of null space of $Q_{00}$ .................................. 127
A.7 Calculation of $P(s,t)$ .......................................... 127
A.8 Deterministic contributions to the transport equations ............ 127
A.9 Derivation of $I_J$, $I_H$ and $I_K$ .............................. 129

Appendix B Critique of work by C.O. Hines

Appendix C Basic stochastics

C.1 Stochastic processes .......................................... 134
C.2 The correlation formula ....................................... 135
C.3 Fokker–Planck equations ..................................... 136
C.4 The normal distribution ..................................... 136
Chapter 1

General introduction

Go on till you come to the end; then stop.
(Lewis Carroll)

This section contains a basic introduction to the Earth’s atmosphere and some important processes within it. It includes a description of gravity waves, their generation and their effects on the atmosphere. The basic equations for gravity wave propagation in the atmosphere are presented. Some mathematical and physical ideas that may not be familiar to the reader and will be needed in further chapters are also given. We start with a brief description of the atmosphere’s vertical structure.

1.1 A brief description of the atmosphere

The Earth’s atmosphere is about 100 km thick and is separated into four distinct sections, (see for example Gill 1982 or any other standard text on the atmosphere for a description of this). The lowest 10 km contains 80 % of the total mass of the atmosphere and is called the troposphere (the upper boundary of which is the tropopause). Temperature decreases with altitude here at about 7 K km$^{-1}$. The troposphere is characterised by strong vertical mixing and it is here that most of the Earth’s weather features occur.

The region between 10 km and 50 km is called the stratosphere (the upper boundary of which is called the stratopause) and contains about 19.9 % of the atmosphere’s mass. This is a region of pronounced stability, where temperatures increase with altitude at about 2 K km$^{-1}$. Vertical motions are much weaker in the stratosphere compared to the troposphere due to this stability. The transition between the troposphere and stratosphere is marked by an abrupt change in concentration of trace constituents, such as water vapour and ozone. Strong concentration gradients indicate very little mixing between the two layers and even the most vigorous thunderstorm updraft will only marginally penetrate the stratosphere.

The remaining 0.1 % of the atmosphere’s mass is contained in the upper atmosphere. This is split into the mesosphere, between 50 km and 100 km, where temper-
Figure 1.1: Schematic of the temperature and pressure structure of the atmosphere as a function of altitude, from ground level to over 100 km. The different layers of the atmosphere are separated into the troposphere, stratosphere, mesosphere and thermosphere. Temperature steadily decreases with altitude, and the thermosphere, 100 km and upwards, where temperature steadily increases with altitude. In the thermosphere molecules become sufficiently separated so that the modelling assumptions of ideal gases or simple fluids can no longer be applied. The region of the atmosphere from the tropopause up to about 100 km, containing the stratosphere and mesosphere, is also called the middle atmosphere. Figure 1.1 shows a schematic of the overall structure of the atmosphere.

1.2 Gravity waves in the atmosphere

More than 50 years ago, irregular winds in the upper atmosphere known as travelling ionospheric disturbances (TIDs) were observed through radio observations (e.g beynon 1948, Munro 1948) and it was suggested that they might be due to propagating waves.
Indeed, the atmosphere is capable of sustaining many wave phenomena at various scales. More than twenty different types are documented in Beer (1974). Further observations of TIDs were made by Munro (1953) and it was shown that they were characterised by phase velocities in the range 50 ms⁻¹ to 200 ms⁻¹, horizontal wavelengths of several hundred kilometres and periods of 10 minutes to several hours. Explanations for the irregular winds were given in terms of vertically propagating internal waves known as gravity waves², by, for example, Martyn (1950) and Hines (1955).

Gravity waves propagate in any fluid whose density is stably stratified, such as the Earth's oceans or atmosphere. The system in which light fluid overlays heavy fluid is stable because the strength of the gravitational force is balanced by the restoring buoyancy force. The mechanism that generates gravity waves is as follows: consider a parcel of air in a stably stratified atmosphere. If the parcel is forced to rise, it will move to a region of lower density. As it is then heavier than its new surroundings, the parcel will sink. The momentum that the parcel possesses will now cause it to overshoot its original position, so it will move into a lower region that has higher density. The parcel will then be lighter than its surroundings and it will rise and start the whole process again. Hence a wave is formed that propagates through the stratosphere. Figure 1.2 shows a schematic of the mechanism producing gravity waves.

If the wave fields have a sinusoidal dependence of the form \( \exp(i(kx + mz - \omega t)) \) where \( i = \sqrt{-1} \), \( k \) is the horizontal wavenumber³, \( m \) is the vertical wavenumber and \( \omega \) is the frequency, then for two-dimensional gravity waves in a non rotating fluid, the frequency and wavenumber are related by a dispersion relation (e.g. Gill 1982):

\[
\omega^2 = \frac{N^2 k^2}{k^2 + m^2},
\]

where the restoring buoyancy force acting on a vertically displaced fluid particle is characterised by the Brunt-Väisälä frequency \( N \), defined by

\[
N^2(z) = -\frac{g}{\rho_0(z)} \frac{\partial \rho_0(z)}{\partial z},
\]

in which \( \rho_0 \) is the background density of the atmosphere and \( g \) is the gravitational acceleration. In a stable medium \( N \) is real and has the dimensions of frequency (s⁻¹). It is the frequency of small oscillations of a particle in purely vertical motion. It gives a measure of the degree of stability of a medium: the larger \( N \) is, the more stable the medium.

Gravity waves are produced in the lower atmosphere (troposphere and lower stratosphere). For gravity waves to form, a trigger mechanism is required: a force that

---

¹The phase velocity is the velocity of the wave-crests and is described mathematically in Section 1.3.
³The wavenumber is defined as 2π/wavelength.
displaces parcels of air. This can be produced, for example, by a wind flowing over topography (see Figure 1.2) or by currents in convection clouds. Waves generated by topography are characterised by a single prominent phase speed and frequency, whereas convection clouds generate waves with a full spread of wavenumbers and frequencies (e.g. Fritts & Alexander 2003).

One effect of gravity waves can be seen by looking at cloud formations in the sky. The upward moving region of a gravity wave is the most favourable region for cloud development, whereas the sinking region is more favourable for clear skies. That is why rows of clouds with intervening clear areas can frequently be seen.

Why are we interested in gravity waves? They are not generally of importance in synoptic-scale (short time-scale) weather forecasting and are in fact omitted from some models. They are, however, important in long time-scale global circulation models (GCMs): models of the circulation of air masses within the atmosphere over periods of months or years. In fact, the middle atmosphere circulation is thought to be largely controlled by the forcing effects caused by gravity waves, (this is described in more detail in Section 1.4). Gravity waves are short time and short space scale fluctuations and are far more complicated than can be resolved in GCMs. For this reason it is important to gain an understanding of the propagation of gravity waves, so that the effect they have on the atmosphere can be accounted for within the GCMs.
Wavefronts (surfaces of constant phase)

Motion of fluid particles

Parcels motions are represented by the black dot with attached arrows. The direction of group-velocity (velocity of an isolated wave packet) is given by $c_g$ and phase velocity (velocity of wave crests) by $c_p$. These concepts are more fully explained in Section 1.5.1.

Figure 1.2: Schematic showing the mechanism that causes gravity waves to form when air is forced up over a topography by an incident horizontal wind $U(z)$. Lines of constant phase (for example all the parts with minimum speed have the same phase) are represented by the dashed lines.
1.3 Observations of gravity waves

The purpose behind the theory of gravity waves is to try to explain the mechanism that causes them and so gain a better understanding of their propagation. This can be done by comparing theories to observations of gravity waves. One method of making this comparison is to look at the energy spectral density at a particular altitude and time. This is the distribution of energy between the various wavenumbers in the disturbance.

Observations of gravity waves have shown several distinct features of the energy spectrum. At large vertical wavenumber $m$, the slope of the energy spectrum is often independent of time, place and altitude and is proportional to $m^{-3}$. Observations also display a pronounced peak of the energy spectrum at smaller wavenumber. This is known as the central wavenumber and is usually denoted by $m_*$, (see below for references concerning observations).

Before observations can be compared to theory it is necessary to answer some questions about the methods used to observe gravity waves. Is it the case that a wavenumber distribution is obtained by measuring at a place for a long period of time and observing many single gravity waves and then Fourier transforming to obtain the distribution in wavenumber? If so, how is a single gravity wave isolated? Or is it the case that a broad spectrum of waves is present at any one time?

1.3.1 Measurement methods

There are several different ways in which characteristics of the atmosphere such as temperature, density or wind speed can be measured. This information is used to infer the existence of gravity waves and the different measurement methods will now be described.

Note that gravity waves can be distinguished from Rossby waves by measuring the vertical and horizontal components of velocity and checking they satisfy the polarization relations.

Balloon measurements: rising balloons are launched into the atmosphere and vertical velocity profiles of the horizontal wind relative to the balloon are collected with an anemometer. Typically the balloon has a vertical velocity of a few metres per second and several measurements are taken per second. See for example, Barat (1982) and Sidi (1988).

Remote sensing using radar: radar sensing measures the radial wind velocity. Altitudes visible to radar are the lower stratosphere, mesosphere and lower thermosphere. A typical example of data processing is described by Tsuda (1989): wind velocity was measured every 60 seconds and averaged over each hour. A background profile was
obtained by averaging 24 hourly profiles. The background wind was then subtracted from each hour profile to obtain the perturbation velocities. Endlich et al. (1969) also used this technique. Observations are generally restricted to a limited number of sites over a short period of time.

**Remote sensing using lidar:** a lidar operates on similar principles to a radar but emits radiation in the visible spectrum. Relative perturbations of density and temperature in the middle atmosphere can be measured directly using lidar. Lidar measurements typically have a vertical resolution of tens of metres and simultaneously measure parameters at all altitudes. See for example, Wilson et al. (1991) and Beatty et al. (1992).

**Airglow observations:** these use cameras to observe atmospheric density variations over a time interval. These observations are of an “all sky” nature: the cameras have a field of view of hundreds of kilometres. Wave patterns appear as several bright and dark bands. These observations often exhibit single gravity wave structures consisting of either a single wavelength component or a few different wavelength components. Observed gravity waves generally have horizontal wavelengths in the range 10 km to 70 km and phase velocity around 30 m s\(^{-1}\). Examples of these studies can be found in Swenson (1994), Taylor & García (1995) and Taylor (1998).

### 1.3.2 Method of data processing

The method used to analyse this data and obtain energy spectral densities is important in determining the question posed above. In general, radar and lidar methods provide a more global coverage but involve indirect data processing, whereas balloons only take very localised measurements but provide direct velocity profiles. A good description of the most common method of data analysis can be found in Endlich et al. (1969) and a short description is given here.

Whatever method of data collection is used the result is a wind speed profile over altitude \(U(z_i)\). This is measured at discrete, equally spaced height intervals \(z_i\). A transformation is then made from the altitude domain to the frequency domain to obtain the following sine and cosine coefficients:

\[
a_k = \frac{2}{n} \sum_{j=1}^{n-1} U(z_j) \cos\left(\frac{2\pi j k}{n}\right) \quad \text{and} \quad b_k = \frac{2}{n} \sum_{j=1}^{n-1} U(z_j) \sin\left(\frac{2\pi j k}{n}\right) \quad \text{for} \quad k = 0 \ldots n/2,
\]

where \(n\) is the number of altitude steps. The raw energy spectrum is then given by

\[
E_k = \frac{1}{2}(a_k^2 + b_k^2).
\]
The method of data processing given here relies on there being a spectrum of gravity waves present at any one time and most of the observation methods also reflect this. So, it seems a broad spectrum of gravity waves is present at any one time. This is the same conclusion as that drawn by Staquet & Sommeria (2002). However several other studies suggest that the spectra are often composed of a single (or a few) isolated gravity waves. These isolated gravity waves are most apparent in the airglow observations where regular cloud patterns are observed. Thus, the answer to the question "broad spectrum or isolated gravity wave?" posed above, is really that attempts to answer it are ongoing and a general consensus as to the answer has yet to be reached.

1.3.3 Physical values of parameters for the atmosphere

It is important to know the value of the physical parameters characterising the atmosphere and gravity waves so that the correct modelling assumptions and approximations can be made. There are several sources of real data for the atmosphere, for example, Sidi (1988) and other references described in the previous section. Typically, horizontal wavenumbers \( k \) have been observed in the range \( 1 \times 10^{-7} \text{ m}^{-1} \) to \( 1 \times 10^{-4} \text{ m}^{-1} \) and vertical wavenumbers \( m \) have been observed in the range \( 5 \times 10^{-5} \text{ m}^{-1} \) to \( 1.6 \times 10^{-3} \text{ m}^{-1} \). Density in the atmosphere varies exponentially with altitude from \( 1 \text{ kg m}^{-3} \) at ground level to about \( 10^{-6} \text{ kg m}^{-3} \) at 100 km. Wind speeds in the lower stratosphere are generally less than about 50 m s\(^{-1}\). The scale height \( H \) of the atmosphere represents the characteristic vertical dimension of mass distribution in the atmosphere. It is defined as the increase in altitude necessary to reduce the pressure by a factor \( e \). It varies from about 8 km near the surface to about 6 km in very cold regions of the atmosphere, such as the summer mesopause. The Brunt-Väisälä frequency \( N(z) \) takes the approximate values 0.01 s\(^{-1}\) in the troposphere and 0.017 s\(^{-1}\) in the stratosphere (e.g. Gill 1982). To compare with the ocean, \( N \) is approximately 0.01 s\(^{-1}\) in the upper ocean and 0.001 s\(^{-1}\) or smaller in the deep ocean.

1.4 Global circulation of the atmosphere

We have so far described what gravity waves are and the setting in which they propagate. We have also detailed methods of observing gravity waves and the characteristics of the energy spectrum that these observations display (i.e., \( m^{-3} \) tail and central wavenumber \( m_* \)). The questions we must now ask are: (i). Why are gravity waves important? (ii). Do theories exist to explain the characteristics of the energy spectrum? We answer the first of these questions by describing the global circulation of the atmosphere (e.g. Gill 1982). The second of these questions will be answered in the following section.

Without the presence of an atmosphere the surface temperature of the Earth would
be due to solar radiation only. The solar radiation absorbed would cause the Earth to warm up to a critical temperature where it radiates as much energy as it absorbs. This state is called radiative equilibrium and it is stable with the poles being cold (150 K for South Pole, 170 K for North Pole) and the equator hot (270 K) (e.g. Gill 1982). The difference between the Pole and Equator temperatures in the radiative equilibrium model is much greater than that for the actual Earth. This difference must be due to the presence of an atmosphere.

With the addition of an atmosphere that is heated from the ground, a vertical temperature gradient is introduced. This temperature gradient is unstable and so convection currents are produced which try to reduce the gradient. In addition there is a horizontal, latitudinal temperature gradient that is due to the fact that the sun’s rays are more intense at the equator than at the poles. These temperature gradients might be expected to produce a large circulation cell with rising air in the tropics and falling air at the poles. This is not the case however and the actual circulation can be seen as a schematic in Figure 1.3.

So what causes this difference? In the troposphere, at mid-latitudes, where the Coriolis force (force produced by the rotation of the Earth) is strongest, the motion produced by the horizontal temperature gradient is forced into a zonal (east–west) motion and so there is little or no meridional (north–south) component of motion. The accumulation of air at mid-latitudes is compensated for by air moving downwards producing a Hadley cell. A similar mechanism produces the Ferrel Cell at mid-latitudes and the Polar Cell at high-latitudes.

In the stratosphere and mesosphere (the middle atmosphere), the picture is much simpler and it is here that gravity waves take on their importance. Temperature gradients in the middle atmosphere are much weaker than in the troposphere and are overpowered by the effect of wave-forcing on the atmosphere. Wave-forcing is produced by "breaking" waves of various different scales, among them, gravity waves. Like ocean waves, gravity waves break when their amplitude gets large and turbulence follows. When this happens they deposit their momentum, typically in regions far removed from their source. For example, ocean waves generated by a storm may lose their momentum on the opposite side of the world when they break on a beach. Breaking provides a force on the atmosphere known as wave-drag. This force is the main factor that drives circulation within the middle atmosphere by a mechanism known as "gyroscopic pumping": when the force acts on a parcel of air it drives it either westwards or eastwards (depending on the type of wave). In the westwards case, the parcel of air is

---

4First proposed by Halley (1686). Due to his incorrect explanation of the easterly trade winds it is now named after Hadley (1735) who appealed to the conservation of angular momentum for an explanation.

5It is actually the wave pseudomomentum that is the quantity that defines the wave-drag and not momentum. A description of pseudomomentum can be found in Section 1.5.
turned polewards by the Coriolis force; in the eastwards case, it is turned equatorwards.

In the stratosphere, the Brewer-Dobson\(^6\) (B–D) circulation (e.g. Shepherd 2002) is a direct manifestation of wave-drag. The B–D circulation is a global-scale cell in the stratosphere in which air rises in the tropics and then moves polewards in both hemispheres (force is westwards in both hemispheres, see Figure 1.3). Since air is being continually pushed polewards it must eventually go somewhere and so it moves downwards in the extra-tropics. The B–D circulation is responsible for the exchange of air between the troposphere and stratosphere. It is characteristic of the winter season as it requires that waves propagate deep into the stratosphere: a feature permitted by a westerly zonal flow (e.g. Fritts & Alexander 2003). One of the most important effects of the B–D circulation is that it controls the rate at which man-made CFCs and other chemical species are transported from the troposphere to the stratosphere.

The mesosphere is characterised by a single pole-to-pole transport, in which air rises at the summer pole and falls at the winter pole. This is called the Murgatroyd–Singleton circulation. This is due to the fact that the force is eastward in the summer hemisphere, driving equator-ward motion and westward in the winter hemisphere, driving pole-ward motion.

We have suggested that wave-breaking is an important feature that drives the circulation in the middle atmosphere. So what are the mechanisms that cause wave-breaking and how are they modelled? We first describe the basic equations governing gravity waves in the atmosphere.

\(^6\)Named after Alan Brewer and Gordon Dobson who suggested it to explain humidity and helium tracer observations over England in 1949.
Figure 1.3: Schematic of the circulation of the Earth's atmosphere. The horizontal coordinate is latitude (i.e., poles on left and right, equator in the centre) and the vertical coordinate is altitude. In the troposphere there are six circulation cells. Two Hadley cells in the tropics, two Ferrel cells in the mid-latitudes and a polar cell at each pole. Gravity waves have their biggest effect in the middle atmosphere: the forcing effect of the waves on the background flow is responsible for the global circulation cells. The stratosphere is characterised by the Brewer-Dobson circulation in which air rises in the tropics and falls at the poles. The winter equator-to-pole transport is stronger than the summer transport. The mesosphere is characterised by a single pole-to-pole transport where air rises at the summer pole and falls at the winter pole. Gravity wave forcing is denoted by the large plus and minus signs. A plus sign denotes eastward forcing and a minus sign denotes westward forcing.
1.5 Basic equations

In this section, we derive the basic equations describing motion in an atmosphere for the three different regimes which will be used in Chapters 3 and 4. The equations for mass conservation and momentum conservation for an atmosphere with background wind $U(x)$ are given. Here, $U(x) = (U(x,z),0,W(x,z))$, where $U$ is the horizontal wind speed in the $x$ direction and $W$ is the vertical wind speed. Assuming that $U$ and $W$ vary slowly in space, we use these conservation equations to derive equations for wave propagation in the atmosphere: the ray equations. We detail the approximations that are made.

Let us denote the density of the atmosphere by $\rho = \rho(x,y,z,t)$, pressure by $p = p(x,y,z,t)$, velocity by $u = u(x,z,t)$ and the rotation rate of the Earth about the vertical axis by $\Omega$. The equations for conservation of mass and conservation of momentum in Cartesian coordinates are given by

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0$$

and

$$\frac{Du}{Dt} + 2\Omega \hat{k} \times u = \frac{1}{\rho} \nabla p - \nabla \Phi$$

respectively, where $\frac{D}{Dt} = \frac{\partial}{\partial t} + u \cdot \nabla$

is the derivative following the motion, $\hat{k}$ is a unit vector in the vertical direction and $\Phi$ is the gravitational potential. Since gravity waves are much smaller scale than the background wind, we can assume that the background is incompressible and stationary. The first assumption leads to

$$\nabla \cdot u = 0,$$

(derivations for (1.2)—(1.4) can be found in most basic texts on fluid dynamics, e.g. Acheson 1990). The second assumption leads to ignoring wave motions due to potential vorticity. Using (1.4) to simplify (1.2) we obtain

$$\frac{\partial \rho}{\partial t} + u \cdot \nabla \rho = 0.$$  

For small-scale high-frequency gravity waves, we can neglect the effect of rotation as it is insignificant compared to the gravitational force. We also assume that the gravitational force acts in the vertical coordinate $z$ only, so (1.3) becomes

$$\rho \frac{Du}{Dt} = -\nabla p - \rho g \hat{k},$$

where $g$ denotes the vertical gravitational force. When we expand this into component form with $u = (u,v,w)$, we obtain

$$\rho \frac{\partial u}{\partial t} + \rho u \cdot \nabla u = -\frac{\partial p}{\partial x},$$
\[ \frac{\partial v}{\partial t} + \rho u \cdot \nabla v = -\frac{\partial p}{\partial y} \quad (1.7) \]

and

\[ \frac{\partial w}{\partial t} + \rho u \cdot \nabla w = -\frac{\partial p}{\partial z} - \rho g \quad (1.8) \]

Equations (1.5)-(1.8) are now linearized for small-amplitude wave motions, by making the substitutions

\[ u = U + \hat{u}, \quad \rho = \rho_0(z) + \hat{\rho}(x, y, z, t), \quad p = p_0(z) + \hat{p}, \quad v = \hat{v} \quad \text{and} \quad w = W + \hat{w} \]

and neglecting all terms quadratic in hatted quantities (e.g. Acheson 1990). Dropping hats, we obtain

\[ u_t + U u_x + U_x u + U_z w + W w_x = -\frac{p_x}{\rho_0}, \quad (1.9) \]

\[ v_t + U v_x + W v_z = -\frac{p_y}{\rho_0}, \quad (1.10) \]

\[ w_t + U w_x + W_x u + W_z w + W w_x = -\frac{1}{\rho_0}(p_z + \rho g) \quad (1.11) \]

and

\[ \rho_t + U \rho_x + w \rho_0 + W \rho_z = 0. \quad (1.12) \]

Equations (1.4) and (1.9)-(1.12) are the basic equations of small vertical scale gravity waves in the atmosphere. There are several further approximations that can be used to simplify these equations. The use of these depends on the particular situation which we are describing. We first list typical approximations and then describe three separate regimes in which the approximations may be used. The approximations are as follows:

1. The hydrostatic approximation states that, since pressure decreases with height, an upward force able to accelerate a particle of air is generated. This force must be balanced by gravity. Ideally this requires no motion, but unless we are in a region of large vertical acceleration such as a thunderstorm then the amount of motion is negligible anyway, so the approximation is a good one. The equation of hydrostatic balance replaces (1.8) and is given by

\[ \frac{\partial p}{\partial z} = -\rho g. \]

This is equivalent to the mid-frequency approximation, i.e., making the assumption that \( \omega \ll N \) or alternatively, \( m \gg k \) (e.g. Fritts & Rastogi 1985). Thus, the hydrostatic approximation is limited in that it assumes vertical wave motions have much shorter scale than horizontal wave motions.

2. For short vertical distances where the density may be assumed to vary very little, we can use the Boussinesq approximation. This consists of assuming that \( \rho_0(z) \) is constant in the momentum equations, apart from where density variations give rise to buoyancy forces (which occurs if there is a multiplying factor of \( g \) in the vertical component). This approximation does have its limitations: it is invalid
over large vertical distances because density decreases significantly and it cannot be used to study breaking gravity waves due to density decreases.

3. This thesis will only be concerned with two-dimensional motion, i.e., we repress one horizontal dimension so in all equations, terms involving $v$ or $y$ can be neglected.

In this thesis there will be several different regimes studied:

i. The main bulk of Chapter 3 models the atmosphere in two dimensions, with a horizontal wind dependent on the vertical coordinate only and varying slowly in that coordinate, i.e., $U = U(z)$ and $W = 0$. Short vertical distances are considered, so approximations (2) and (3) are applied.

ii. In Section 3.8 of Chapter 3, we model the atmosphere with a slowly varying wind as in regime (i), but larger vertical distances are considered and so only approximation (3) is applied. We let the density be a slowly varying function in altitude given by $\rho_0(z) = e^{-z/H}$, where $H$ is the scale height of the atmosphere, (see Section 1.3.3).

iii. Chapter 4 models the atmosphere with horizontal and vertical winds dependent on both horizontal and vertical coordinates, i.e., $U = U(x, z)$ and $W = W(x, z)$. We make use of approximations (1) – the use of which is verified in Chapter 4 – and also (2) and (3).

In the following section we apply to (1.9)–(1.12) an approximation called the WKB approximation, to obtain the ray tracing equations and dispersion relation for the propagation of waves in the atmosphere.

### 1.5.1 WKB approximation and ray equations

Consider (1.9)–(1.12) in two-dimensions so that all dependence on $y$ is dropped. If (1.9)–(1.12) have constant coefficients, i.e., constant $N$, $U$ and $W$, then they admit wave-like solutions in the form $(u, v, w, \rho, p) = (u', v', w', \rho', p')e^{i\phi}$, where $u'$, $v'$, $w'$, $\rho'$ and $p'$ are constants and $\phi = kx + mz - \omega t$ is the phase that represents the position within a wave cycle. Here, $\omega$ is the constant frequency and $k = (k, m)$ is the constant wavevector, whose direction is normal to planes of constant phase and whose components give the average number of crests per $2\pi$ units of distance in each direction.

When $N$, $U$ and $W$ are not constant, solutions to (1.9)–(1.12) can, in general, only be solved numerically or by using an approximation.

---

\footnote{Named after G. Wentzel H.A. Kramers and L. Brillouin after papers published in 1926 it is however thought that the original inventors were J. Liouville and G. Green in 1837.}
**WKB approximation**: If $N$, $U$ and $W$ vary slowly with $x$ and $z$ through the background medium then it is expected that locally the solution will be similar to that for constant $m$, $k$ and $\omega$, but that $m$, $k$ and $\omega$ will depend only on $z$ and $x$. We develop now the corresponding approximation in the case that the slow dependence is upon $z$ only.\(^8\) We proceed as follows (details can be found in, for example Hinch (1991)): we let the scale separation between the waves and the background media be characterised by a small parameter $\varepsilon$; we let $N = N(\varepsilon z)$, $U = U(\varepsilon x, \varepsilon z)$, $W = W(\varepsilon x, \varepsilon z)$, $t \rightarrow t/\varepsilon$, $z \rightarrow \varepsilon$ and $x \rightarrow x/\varepsilon$. Then we have

$$\omega = -\varepsilon^{-1} \frac{\partial \phi}{\partial t}, \quad k = \varepsilon^{-1} \frac{\partial \phi}{\partial x} \quad \text{and} \quad m = \varepsilon^{-1} \frac{\partial \phi}{\partial z}.$$  

So when $\varepsilon$ is small, we are in the large wavenumber, high frequency regime. We now look for solutions in the form\(^9\) $(u, v, w, \rho, p) = A(t, x)e^{i\phi - z/H}$ and substitute into (1.9)–(1.12). At leading order, we obtain a relationship between frequency and wavenumber, i.e., the dispersion relation (e.g. Gill 1982, Fritts & Alexander 2003).

$$\omega = \hat{\omega} + k \cdot U,$$

where $U = (U, W)$ and

$$\hat{\omega} = \pm \frac{kN}{\sqrt{k^2 + m^2 + 1/4H^2}}$$

is the relative frequency, i.e., the frequency relative to the background wind and $H$ is the scale height of the atmosphere (see Section 1.3.3). With the addition of approximation (1), this reduces to regime (iii)

$$\omega = \pm \frac{kN}{m} + k \cdot U.$$

Taking $U = U(z)$ and $W = 0$ we obtain for regime (ii)

$$\omega = \pm \frac{kN}{\sqrt{k^2 + m^2 + 1/4H^2}} + kU.$$  

In addition, applying approximation (2), we obtain for regime (i)

$$\omega = \pm \frac{kN}{\sqrt{k^2 + m^2}} + kU.$$

**Ray equations**: The slow change in $k$, $m$ and $\omega$ with $z$ and $x$, is governed by the ray equations. We define a ray as the trajectory of a point moving with the local group-velocity $c_g = \partial \omega/\partial k$. The group-velocity is the speed at which an isolated wave packet (and energy) travels as a whole (e.g. Acheson 1990). For example, if a stone is dropped\(^8\) An elegant derivation of the WKB equations for general Hamiltonian systems can be found in Chapter 2.

\(^9\)This is the geometrical approximation to the WKB approximation and strictly speaking is not an asymptotic expansion. However we are only concerned with the first order approximation of the solution and so this will suffice.
into a pond then the group-velocity is the velocity at which an observer must travel to continually see waves of the same wavelength $\lambda = 2\pi/k$ (Acheson 1990). The equation of the ray is given by

$$\frac{dx}{dt} = \frac{\partial \omega}{\partial k},$$

(1.13)

where

$$\frac{d}{dt} = \frac{\partial}{\partial t} + c_g \cdot \nabla.$$

Applying the identity $\phi_{xt} = \phi_{ix}$ to the phase, we obtain

$$\frac{\partial k}{\partial t} = -\frac{\partial \omega}{\partial x} = -k \cdot U_x \pm \omega_x.$$

(1.14)

This implies that the rate of change (following the group-velocity) of the local wavevector depends on the spatial variation of $\omega$, itself due to variations in the background wind and changing fluid properties (i.e., $N$). Equation (1.14) determines the refraction of the ray through the background medium.

For given initial conditions, the ray-tracing equations (1.13) and (1.14) can be solved for a wave whose wavenumber and frequency vary according to the dispersion relation along the rays. A recent review of ray-tracing methods can be found in Broutman et al. (2004).

Gravity waves have the remarkable property that the group-velocity vector is perpendicular to the wavevector in the absence of a background wind,\(^{10}\) (See Figure 1.2.) Since energy propagates at the group-velocity then energy must propagate parallel to wave crests and troughs. This is in stark contrast to acoustic waves or shallow water waves in which energy propagation is perpendicular to wave troughs and crests.

Values of the group-velocity for gravity waves, using the typical values for $k$ and $m$ given in Section 1.3, are between 0.1 m s\(^{-1}\) and 200 m s\(^{-1}\). These are calculated in Section 4.4.

**Pseudoenergy and pseudomomentum:** The equations derived in this section provide no information about the amplitude of the fields that characterise a gravity wave. To reconstruct these fields we require some kind of evolution equation, (see Chapter 2). For time-independent systems, this takes the form of energy conservation for the linear perturbation. It is not always the case however that the usual energy for the perturbation is conserved. For example, energy is not conserved for gravity waves in a shear flow. There does exist however, a conserved quadratic quantity known as the pseudoenergy (or wave energy). Similarly, the analogue of momentum in spatially invariant flows is called pseudomomentum. The conservation equations for pseudomomentum and pseudoenergy are discussed in more detail in Andrews & McIntyre (1978).

\(^{10}\)In fact this is the case for all conical waves, i.e., with dispersion relation of the form $\tilde{\omega} = f\left(\frac{k}{|k|}\right)$, which is proved in Section 4.2.
and McIntyre (1981). Vanneste & Vial (1994) provide expressions for the pseudoenergy and pseudomomentum for gravity waves in a background shear.

1.6 Wave-breaking

Just as an ocean wave can break if it becomes too steep, so can a gravity wave. The mechanisms behind the "steepening" and "breaking" will now be discussed. We also detail some of the theories that have been used to model these mechanisms. Chapters 3 and 4 will also be concerned with this.

1.6.1 Wave-steepening mechanisms

Stratification: in a stratified atmosphere, density decreases with altitude, and so the amplitude of a vertically propagating gravity wave increases with altitude. This is because there will be less resistance to particle motions in a low density atmosphere than in a high density one. Correspondingly the amplitude of vibrations will be larger. As amplitude increases then the wave must correspondingly become steeper.

Resonant interaction: at small wave amplitudes wave steepening may occur via resonant wave interactions where energy is exchanged between waves having different wavevectors. Resonant interactions require a match of wavenumbers and frequencies of the three waves taking part: \( k_1 = k_2 + k_3 \) and \( \omega_1 = \omega_2 + \omega_3 \). For example, an upward propagating wave with wavenumber \( k_2 \) may be backscattered into a wave of comparable wavenumber \( k_3 \) by a mean flow of twice the wavenumber \( k_1 \), (e.g. McComas & Bretherton 1977). The significance and importance of resonant wave interactions in gravity wave breaking is not fully understood and much work remains to be done in this field (e.g. Staquet & Sommeria 2002).

Critical levels: if we consider gravity wave propagation in an incompressible atmosphere with a vertically stratified wind, an equation for the amplitude of a wave called the Taylor-Goldstein equation can be derived. For example, for a Boussinesq system with a horizontal background wind dependent on altitude only (regime (i), Section 1.5), some manipulation of (1.9)–(1.12) leads to an equation for \( w \)

\[
(\partial_t + U \partial_x)^2(w_{xx} + w_{zz}) + N^2w_{zz} - U_{zz}(\partial_t + U \partial_x)w_x = 0. \tag{1.15}
\]

Looking for solutions of the form \( w = A(z,t)e^{it\phi} \) we obtain

\[
A'' + \left( \frac{N^2(z)}{(U(z) - c_pz)^2} - \frac{U''(z)}{U(z) - c_pz} - (k^2 + m^2) \right) A = 0,
\]

where the phase velocity vector of a phase surface is defined by

\[
c_p = (c_pz, c_{pz}) = (\omega/k, \omega/m).
\]
Figure 1.4: Schematic of a ray approaching a critical level in a horizontal background wind $U(z)$. The dashed line represents the critical level $U(z) = c_{pz}$. $c_p$ and $c_g$ are the directions of the phase velocity and group-velocity respectively. The velocity profile of the wind is shown on the right.

When the horizontal phase velocity $c_{pz}$ matches the horizontal velocity of the background wind, ($c_{pz} = U(z)$) the Taylor-Goldstein equation becomes singular. Provided the Richardson number $\text{Ri} = \frac{N(z)^2}{U'(z)^2}$ exceeds 1/4 (this is the condition for shear stability, see Section 1.6.2), the function $A(z)$ oscillates and the wavelength reduces more and more rapidly as the critical level is approached. Hence the wave becomes steeper.

Critical levels can arise in any problem involving wave propagation through a fluid. In the ocean, for example, small scale waves encounter a critical level within a few wave periods, (e.g. Henyey 1986). 12

We can interpret the singularity at the critical level as being a wave absorber: as the vertical wavenumber increases towards the critical level, the vertical group-velocity diminishes and there is infinite time for the wave (for a steady flow) to dissipate before the critical level is reached (see Figure 1.4). This theoretical limit however, is never reached and in reality other mechanisms become more significant.

To rectify the singularity in the theory, previously neglected processes must be re-introduced in the vicinity of the critical level. This region, often small, is called a critical layer. Viscous effects were considered by Hazel (1967). Non-linear effects were considered by Brown & Stewartson (1982), Maslowe (1986) and Winters & D’Asaro (1994) and many others. A time-dependent background wind also dramatically changes the effects seen at a critical level (e.g Broutman et al. 1997).

---

11 Named after Lewis Fry Richardson (1881–1953), a Newcastle-upon-Tyne born meteorologist and the founder of numerical weather prediction. He published a paper on the topic in 1922. Due to lack of computing power he was forced to propose the notion of bands of messengers on motorcycles cruising around the Royal Albert Hall to communicate arithmetic results between banks of clerks in order to obtain the necessary numerical solutions!

12 These critical levels are however, time dependent and so do not have a very large impact (e.g. Broutman et al. 1997).
Figure 1.5: Rays traced through a random wind in the atmosphere. All the rays have the same initial altitude but different horizontal positions. Caustic formation can quite clearly be seen. The method used to obtain this plot is described in Chapter 4.

**Caustics:** a caustic is the locus of a singularity that develops when two infinitesimally close rays converge, (see Section 4.1 for a fuller explanation of caustics). The superposition of linear waves through constructive interference causes exceptionally large (but finite) wave amplitudes and hence wave steepening. Caustics in the atmosphere are sometimes overlooked, but they could make a significant contribution towards the overall dissipation of gravity waves. Caustics will be fully discussed in Chapter 4. Figure 1.5 shows rays being traced through a background wind with a random component. The formation of caustics can quite clearly be seen by the dark lines denoting convergence of rays.

### 1.6.2 Wave-breaking mechanisms

The stability of the atmosphere is quantified by the Richardson number $R_i$ given by

$$
R_i = \frac{N^2}{U^2}.
$$

$R_i$ is the ratio between the stabilizing mechanism of the stratification and the destabilising mechanism of the shear (a shear flow is one in which the velocity varies in a direction at right angles to the flow direction). If the Richardson number is small enough, in fact if $R_i < 1/4$, the destabilising shear may overcome the stabilizing stratification and shear instabilities may be produced and cause wave breaking (Dewan & Good 1986). If $R_i > 1/4$ then the atmosphere is too stable for shear instabilities. Convective instability occurs when $R_i < 0$ (e.g. Whitham 1974, Gill 1982).

---

13 This was first proposed by Taylor (1998). Further explanation can be found in Turner (1973).
Shear instability: consider a shear flow with opposite velocities as given in Figure 1.6(a). Suppose that a disturbance causes the boundary between the opposing velocities to become slightly wavy as in Figure 1.6(b). Fluid on the convex side of the boundary ('-' signs in Figure 1.6(b)) moves slightly faster than fluid on the concave side of the boundary ('+' signs in Figure 1.6(b))\(^{14}\). By Bernoulli's theorem (e.g. Acheson 1990)) this results in increased pressure at the '+' signs and decreased pressure at the '-' signs. This pressure difference acts to increase the disturbance. If this destabilising mechanism is stronger than the stabilising mechanism of the stratification \((Ri < 1/4, \text{ see Acheson (1990))}\) then the instabilities will grow and cause wave breaking (e.g. Fritts & Rastogi 1985, Dunkerton 1984, 1997). This discussion only really applies to steady situations. However we can extrapolate to non-steady situations by considering the background flow with a superimposed wave to be the basic flow. We redefine \(N\) to be

\[ N^2 = -\frac{d\rho_0}{dz} - \frac{d\rho'}{dz}, \]

where \(\rho_0\) is the background density and \(\rho'\) is the perturbation density of the wave.

Convective instability: if the amplitude of a gravity wave increases, (by one of the above wave steepening mechanisms for example) until a critical amplitude where non-linear effects become important, then the wave may become convectively unstable: the wave becomes so steep that heavy fluid over-lies light fluid and the wave collapses, see Figure 1.7. We can derive a condition for convective instability by referring to Figure 1.7. We consider layers of fluid with densities \(\rho_1\) and \(\rho_2\), where \(\rho = \rho_0(z) + \rho'\), (the background density \(\rho_0\) plus the perturbation density \(\rho'\) due to the wave). When the wave is not steep, so that \(\rho_1 > \rho_2\) and \(z_1 < z_2\), \(d\rho/dz\) is less than zero. However when the wave becomes too steep, as in the top curve of Figure 1.7, then \(z_1 > z_2\) and \(d\rho/dz\)

\(^{14}\)An argument for faster flow on the convex side of the displacement than on the concave side can be made in terms of vorticity. We refer the reader to Batchelor (1967) or Drazin (2002) for a detailed explanation
is greater than zero. So, we have the following condition for convective instability:

$$\frac{d\rho}{dz} > 0.$$  

1.6.3 Wave breaking models

Nonlinear waves of any amplitude are in fact unstable, (e.g. Klostermeyer 1991, Vanneste 1995) although the growth rates for small amplitude waves may not be very fast since the non-linear effects are insignificant. Shear and convective instabilities are much faster instabilities and hence dominate wave breaking. If we imagine a wave starting from a stable region ($R_i > 1/4$) and propagating such that the Richardson number is decreasing then the criteria for shear instability ($R_i < 1/4$) is met first. Shear instabilities however, take a long time to form and so it is quite likely that the criteria for convective instability ($R_i < 0$) will be met. For example, (e.g. Dewan & Good 1986) claim that the amplitude required for convective instability is four times smaller than that required for shear instability. Since convective instability is a much stronger event than shear instability it is natural to only consider convective instability as being significant in causing wave breaking.

The consequence of wave instability is that the growth in amplitude of the wave is terminated. As was described in Section 1.3, the energy spectra at large wavenumber display a quasi-universal form: the spectral tail is proportional to $m^{-3}$. There have been various theories and models that try to explain the existence of this spectral tail.
using wave instabilities. Saturation models and Doppler spreading models will be described here.

**Saturation models:** when instability occurs, no further amplitude growth takes place as the wave propagates. The wave amplitude is then assumed to remain at this maximum amplitude and any excess wave energy is presumed to degrade into turbulence. The waves are said to be saturated.

The earliest theory describing saturation was proposed by Hodges (1967), who noted that the composition of the upper atmosphere cannot be explained if only molecular diffusion is present. He pointed out the possibility that gravity waves were capable of producing localised instabilities which may be a dominant source of turbulence. He devised a convective instability model to limit the propagation of a gravity wave with height. It was first suggested by VanZandt (1982) that a universal spectrum might exist. Dewan et al. (1984) and Dewan & Good (1986) backed the suggestion of VanZandt and proposed a theory consisting of a succession of wave groups with a range of initial values for $m$. They used dimensional analysis to show that (via convective or shear instabilities) the vertical wavenumber spectra should scale as $m^{-3}$. (We refer the reader to Chapter 3 or Dewan & Good (1986) for a more precise explanation of the argument involved).

**Doppler spreading models:** Doppler spreading models first proposed by Hines (1955) describe the change in the energy spectrum. As these waves propagate vertically through a near uniform background wind, they are advected and refracted by it. As a result their wavenumbers change. When an ensemble of waves, encountering different background conditions, is considered, there is a spread of the energy spectrum with wavenumber as altitude increases. The $m^{-3}$ tail is in general not necessarily produced by Doppler models, (see Appendix B and Section 3.9). The smallest scale gravity waves (those with large wavenumber) can break and deposit momentum to the background flow. This results in the transport of energy from larger to smaller scales. Doppler spreading models will be fully discussed in Chapter 3.

### 1.7 Description of Chapters

A description of the remaining chapters is now given.

**Chapter 2: Transport equations in generalised random media.** The dispersion relation, ray equations and wave transport equation (governing the evolution of the wave amplitude) of a system are traditionally derived by applying a WKB expansion to the basic equations of motion, (See Section 1.5.1). However, there is a different
Chapter 3: Doppler spreading models of gravity waves. The Doppler spreading of gravity waves will be examined by developing and analysing an extremely simple model of wave propagation in a random flow. The chapter begins with a description of Doppler spreading and previous theories (including that by Hines 1991b, of which a critique is given in Appendix B) that predict the $m^{-3}$ tail of the energy spectrum. We revisit the paper by Souprayen et al. (2001) and consider a one-dimensional model. Simple models are used in which the effect of fluctuations in the background velocity field encountered by the waves is modelled by a simple random process: an Ornstein-Uhlenbeck process, (see Appendix C). The simplicity of these models allow many results to be derived analytically, but key features of gravity wave propagation are still reproduced. We derive simple closed form expressions in terms of altitude for the energy spectrum induced by gravity waves and the wave-induced force of gravity waves on the background wind. Many of the results are backed up by numerical simulations. The aim is to improve the existing models of the atmosphere.

Chapter 4: Model of gravity wave caustics. The formation of caustics, due to the interaction of atmospheric gravity waves propagating through a random background wind, is another possible mechanism for gravity wave steepening and hence breaking, that is often overlooked. Caustic surfaces cause an increase in wave intensity and hence breaking may follow. The chapter begins with a detailed explanation of caustics and a literature review of work done on atmospheric gravity wave caustics. Most of this has been carried out using one-dimensional time dependent models. Two-dimensional models in random background media have been considered by B. White and his coworkers. In particular White & Fornberg (1998) (which is mainly concerned with oceanic surface waves but is set in a general context) use an analytic approach to the problem of caustic formation on rays. They state that after propagating long distances, rays develop caustics and the probability distribution of the distance along a ray to first caustic
formation is given by a universal curve: the statistics of the random medium do not influence the shape of the caustic probability curve and only contribute a scale factor. Atmospheric gravity waves however, do not fit into the general context of this paper. The reasons for this and the difficulties involved in a similar analytic approach are discussed in detail. A mainly numerical approach is taken to study atmospheric gravity waves in a two-dimensional time-independent random background wind. Probability density functions are provided for the altitude at which a caustic forms as a function of the strength of the random part of the background wind. We obtain relations for the height of caustic formation as a function of parameters controlling the background wind.
Chapter 2

Transport equations for waves in randomly perturbed Hamiltonian systems

A mathematician is a device for turning coffee into theorems!
(P. Erdos)

2.1 Introduction

In Chapter 1 we derived the ray equations and dispersion relation for gravity waves in the atmosphere. However, these equations tell us nothing about the amplitude of the various wave fields. To determine this amplitude, we need to carry out the WKB calculation to the next order in the small parameter $\epsilon$ and derive transport equations governing the change in wave amplitude along the rays.

In conservative (or more precisely Hamiltonian) systems, such as that governing gravity wave propagation, the transport equations have a special structure. For time-independent systems, wave energy (or pseudoenergy, see Section 1.5 or Shepherd 1990) conservation provides the required transport equations. For time-dependent systems, wave energy is not conserved. Instead, it is the adiabatic invariance (i.e., the approximate conservation) of the wave action, given by the ratio energy/frequency, that leads to transport equations (Vanneste & Shepherd 1999). It should be noted, however, that neither wave-energy conservation nor wave-action conservation are sufficient to specify the WKB solution fully. This is because they determine only the modulus of the wave amplitude, not its phase. When knowledge of the phase is needed, an additional evolution equation needs to be obtained (e.g. Bretherton 1971, Vanneste & Shepherd 1999).

Transport equations are generally obtained in the WKB setting of waves propagating in a slowly varying medium. They can be generalized to include the scattering effect of random perturbations of the medium, at least for a certain scaling of the perturbation scale and amplitude. Specifically, if $\epsilon \ll 1$ characterizes the scale separation between
the medium and the waves, then one needs to assume that the random perturbations have an \(O(\epsilon^{1/2})\) amplitude and spatial scales comparable to the wavelengths. (This is important as it allows strong interaction between the waves and the media, see Section 2.6 for more details.) Figure 2.1 shows a comparison between the different scalings of the background medium and the wave. Transport equations of the form

\[
\frac{\partial a(x, k, t)}{\partial t} + \nabla_k \omega(x, k) \cdot \nabla_x a(x, k, t) - \nabla_x \omega(x, k) \cdot \nabla_k a(x, k, t) = \int \sigma(x, k, k') a(x, k', t) \, dk' - \Sigma(x, k) a(x, k, t)
\]

(2.1)

are obtained at leading order in \(\epsilon\), where \(a(x, k, t)\) is a scalar energy density, \(\omega\) is the frequency, \(\sigma\) is the differential scattering cross-section, i.e., the rate at which wave energy with wavevector \(k'\) is converted to wavevector \(k\) and

\[
\Sigma(x, k) = \int \sigma(x, k, k') \, dk'
\]

is the total scattering cross-section. This equation describes the transport of energy in \((x, k)\) space. The second term on the left is the spatial transport of energy at the group velocity \(\nabla_k \omega\); the third term on the left is the analogous transport term in wavenumber space; the first term on the right is an energy source – the total amount of energy being converted from all wavenumbers to energy with wavenumber \(k\). The second term on the right is an energy sink – the total amount of energy with wavenumber \(k\) converted to all other wavenumbers.

Equation (2.1) is valid when the dispersion relation relating frequency \(\omega\) to wavenumber \(k\) is single valued\(^1\). This is equivalent to excluding polarised waves. Thus, the equation is not valid for polarised waves. When the dispersion relation has several branches, i.e., several wave modes coexist, the wave propagation is governed by a set of transport equations similar to (2.1) but with additional terms describing the evolution of the polarisation state.

Our aim in this chapter is to derive transport equations of the form (2.1) for gravity wave propagation in a randomly perturbed atmosphere. Such equations have been derived for a variety of systems, for example for seismic waves (Wu 1985) and for acoustic waves in oceanic sediments (Besieris et al. 1982), but, to our knowledge, not for gravity waves. Ryzhik et al. (1996) developed a general theory for a class of symmetric hyperbolic systems, of the form

\[
A(x) \partial_t u + \sum_i D^i \partial_{x_i} u = 0,
\]

(2.2)

\(^1\)Here, single valued refers to there being distinct eigenvectors, which will be defined later. Also note that the wave mode due to potential vorticity is neglected. However, given that gravity waves are much smaller scale than the background, we can consider the background to be stationary, hence the potential vorticity mode is negligible.
where $A$ and $D^i$ are symmetric matrices and the $D^i$ are space independent. They presented applications to several systems: acoustic waves, elastic waves and electromagnetic waves. Ryzhik et al. used a powerful formalism, based on the Wigner function\(^2\), which lends itself naturally to further generalizations. Guo & Wang (1999) extended their results by considering vector Schrödinger equations of the form

\[ i\partial_t u = H(x, \partial_x) u, \tag{2.3} \]

where $H(x, \partial_x)$ is a pseudodifferential operator.

However, neither of these general theories apply directly to the models of atmospheric gravity wave propagation. Since a number of different models can be used to that effect, we will not limit our derivation of transport equations to a specific system. Rather, we will further generalize the work of Ryzhik et al. and Guo & Wang and derive transport equations for a general class of Hamiltonian systems, which will include most models describing atmospheric gravity waves. Specifically, in this chapter we derive transport equations for waves in systems governed by pseudodifferential equations of the form

\[ \partial_t u = J(x, \partial_x)H(x, \partial_x) u, \tag{2.4} \]

\(^2\)First introduced by Wigner (1932) for use in semi-classical quantum mechanics.
where $J$ and $H$ are, respectively, skew-adjoint and self-adjoint pseudodifferential (matrix) operators, i.e.,

$$
\int g^*(x)H(x, \partial_x) f(x) \, dx = \int f^*(x)H(x, \partial_x)g(x) \, dx
$$

and

$$
\int g^*(x)J(x, \partial_x) f(x) \, dx = -\int f^*(x)J(x, \partial_x)g(x) \, dx,
$$

for arbitrary functions $f$ and $g$. Equation (2.4) is the most general form for linear systems conserving an energy-like quadratic quantity $\mathcal{H}$ given by

$$
\mathcal{H} = \frac{1}{2} \int u^* H(x, \partial_x) u \, dx,
$$

with a non-degenerate $H(x, \partial_x)$. $\mathcal{H}$ will generally be pseudoenergy which we term here wave energy. Written as

$$
\partial_t u = J(x, \partial_x) \frac{\delta \mathcal{H}}{\delta u^*},
$$

equation (2.4) is recognized as a general linear non-canonical Hamiltonian (or Poisson) system. In particular, for canonical Hamiltonian systems, $u$ is real and $J$ is given by the canonical $2n \times 2n$ matrix

$$
J = \begin{pmatrix}
0 & -I \\
I & 0
\end{pmatrix},
$$

where $I$ is the $n$-dimensional identity matrix, while for (conservative) Schrödinger equations $J = -iI$.

The plan of this chapter is as follows: in Section 2.2 we formulate the general problem of wave propagation in a Hamiltonian system. We describe the scale separation between waves and the deterministic properties of the media in terms of a parameter $\epsilon$. We also discuss the properties (scaling in particular) of the random perturbation. We detail properties of Fourier transforms, pseudodifferential operators and correlation tensors that will be used in deriving the transport equations. Section 2.3 defines the Wigner matrix and describes how the conserved energy can be written in terms of this matrix. We derive an evolution equation for the Wigner matrix which can be solved perturbatively using a multiple-scale method. This method is described in Section 2.4, where we finally derive transport equations for the energy density. The principle of energy conservation is exploited in deriving the deterministic part of the transport equations. Section 2.5 applies the general results to atmospheric gravity waves; transport equations are obtained and the ray equations derived in Section 1.5.1 are recovered. Finally, Section 2.6 contains a discussion of further developments and extensions to the theory that could be carried out. The contents of this chapter have been submitted as a paper to the Journal of Wave Motion.
2.2 Formulation

We consider evolution equations of the form (2.4) for a \( n \)-dimensional complex variable \( u(x,t) \), with \( x \in \mathbb{R}^d \). In this expression, the operators \( J(x, \partial_x) \) and \( H(x, \partial_x) \) are the pseudodifferential operators associated with the corresponding \( n \times n \) matrices \( J(x, ik) \) and \( H(x, ik) \). We use the standard Kohn–Nirenberg correspondence (e.g. Folland 1989) in which the differentiations are on the right of the \( x \)-dependence. Thus, for instance,

\[
H(x, \partial_x)u(x,t) = \int H(x, ik)\hat{u}(k,t)e^{ik \cdot x} dk = \int \int \hat{H}(l,k)\hat{u}(k,t)e^{i(k+l) \cdot x} dk dl,
\]

where \( \hat{u}(k,t) \) denotes the Fourier transform of \( u(x,t) \), with

\[
\hat{u}(k,t) = \frac{1}{(2\pi)^d} \int u(x,t)e^{-ik \cdot x} dx.
\]

Notice that the \( t \) has been removed from the definition of \( \hat{H} \) to make later calculations more compact. We have chosen this interpretation of pseudodifferential operators (rather than the Weyl correspondence chosen by Guo & Wang 1999) for its simplicity, even though it makes our derivation of the transport equation somewhat less elegant. It is straightforward, if sometimes cumbersome, to translate between the two interpretations (e.g. Folland 1989).

With the assumption that \( J(x, \partial_x) \) and \( H(x, \partial_x) \) are, respectively, skew-adjoint and self-adjoint, the conservation of the energy (2.5) is readily established:

\[
\frac{\partial H}{\partial t} = \frac{1}{2} \int [u^*Hu + u^*(Huu + Htu)] dx = \frac{1}{2} \int [u^*H^*Ju + u^*HJHu + u^*HJu] dx = 0,
\]

since \( H = H^* \), \( J = -J^* \) and \( H_t = 0 \).

Our interest is in wave-like solutions to (2.4) with wavelengths much shorter than the typical scale of deterministic variations of \( J \) and \( H \) (see Figure 2.1). To make this explicit, we introduce the scale-separation parameter \( \epsilon \ll 1 \) and redefining time and space variables according to \( t \mapsto t/\epsilon \) and \( x \mapsto x/\epsilon \), we rewrite (2.4) in terms of slow (wave) variables as

\[
\epsilon \partial_t u = J(x, \epsilon \partial_x)H(x, \epsilon \partial_x)u. \tag{2.7}
\]

In addition to the \( O(1) \) slowly-varying contribution to \( J \) and \( H \) indicated in (2.7), we consider a second contribution that is random with zero average and varies over the spatial scale of the waves \( x/\epsilon \). If the amplitude of the random fluctuations is strong, then scattering will dominate and waves will be localised. (See Section 2.6 for further discussion on localization.) If the amplitude is weak, the random fluctuations will not have a big effect on the energy transport. The distinguished limit in which this random contribution has an effect comparable to that of the slow variations of \( J \) and \( H \) is
achieved when the amplitude of the random perturbation scales like $\epsilon^{1/2}$ (cf. Ryzhik et al. 1996). We adopt this scaling here and, correspondingly, expand $J$ and $H$ according to

$$J(x, ik) = J_0(x, ik) + \epsilon^{1/2}J_{1/2}(x/\epsilon, ik) + \epsilon J_1(x, ik) + \ldots$$

and

$$H(x, ik) = H_0(x, ik) + \epsilon^{1/2}H_{1/2}(x/\epsilon, ik) + \epsilon H_1(x, ik) + \ldots,$$  

where $J_{1/2}$ and $H_{1/2}$ are zero-average random matrices. For simplicity we have assumed here that $J_{1/2}$ and $H_{1/2}$ depend on space only through $x/\epsilon$; an additional, slow dependence on $x$ could in fact be included without significant changes. Note that the (deterministic) $O(\epsilon)$ terms $J_1$ and $H_1$ appear automatically when $J$ and $H$ are expanded. These terms are crucial, in particular to ensure energy conservation, and they are related to $J_0$ and $H_0$. Indeed, the self-adjointness of $H$ and skew-adjointness of $J$ imply at leading order that

$$H_0(x, ik) = H_0^*(x, ik) \quad \text{and} \quad J_0(x, ik) = -J_0^*(x, ik),$$

where $*$ denotes the Hermitian adjoint and at $O(\epsilon)$ that

$$H_1(x, ik) - H_1^*(x, ik) = -i\nabla_x \cdot \nabla_k H_0(x, ik)$$  

and

$$J_1(x, ik) + J_1^*(x, ik) = -i\nabla_x \cdot \nabla_k J_0(x, ik).$$

The last two equalities determine the skew-adjoint and self-adjoint parts of $H_1$ and $J_1$ respectively (see Appendix A.1, cf. Vanneste & Shepherd (1999)).

The transport equations to be derived depend on the random operators $J_{1/2}$ and $H_{1/2}$ only through their correlation tensors, which we now define. Denoting by

$$\hat{J}_{1/2}(l, k) = \frac{1}{(2\pi)^d} \int J_{1/2}(\xi, ik)e^{-i\xi \cdot \xi} d\xi$$

the Fourier transform of $J_{1/2}(x/\epsilon, ik)$ with respect to its first argument, (again notice that the $i$ has been dropped from the Fourier transformed matrix as a convention) and similarly for $H_{1/2}$, we define the correlation 4-tensors $J$, $H$ and $K$ by

$$\langle J^{\alpha\beta}(l, k + m)J^{\gamma\delta}_{1/2}(m, n) \rangle = -J^{\alpha\beta\gamma\delta}(l, k, n)\delta(l + m),$$

$$\langle H^{\alpha\beta}(l, k + m)H^{\gamma\delta}_{1/2}(m, n) \rangle = H^{\alpha\beta\gamma\delta}(l, k, n)\delta(l + m)$$  

and

$$\langle J^{\alpha\beta}(l, k + m)H^{\gamma\delta}_{1/2}(m, n) \rangle = K^{\alpha\beta\gamma\delta}(l, k, n)\delta(l + m),$$

where $\langle \rangle$ denotes ensemble average and $\alpha, \beta, \gamma, \delta = 1, 2, \ldots, n$ denote the components. We have here assumed that $J_{1/2}$ and $H_{1/2}$ are statistically homogeneous, hence the presence of the Dirac distributions on the right-hand sides. For a physical interpretation of these relations we note that for example (Ryzhik et al. (c.f. 1996))

$$\langle J^{\alpha\beta}_{1/2}(y, k)J^{\gamma\delta}_{1/2}(x + y, n) \rangle = R^{\alpha\beta\gamma\delta}(x, k, n) = \int R^{\alpha\beta\gamma\delta}(l, k, n)e^{i\xi \cdot \xi} d\xi,$$
where $\hat{R}$ is relabeled as $J$.

For future reference, we note that the skew-adjointness and self-adjointness of $J_{1/2}$ and $H_{1/2}$ imply that (see Appendix A.2)

$$
\hat{J}_{1/2}(l,k) = -\hat{J}_{1/2}(-l,k + l) \quad \text{and} \quad \hat{H}_{1/2}^*(l,k) = \hat{H}_{1/2}(-l,k + l).
$$

(2.14)

Simple manipulations then show that (see Appendix A.2)

$$
\begin{align*}
J^{\alpha\beta\gamma\delta*}(l,k,n) &= J^{\delta\gamma\beta\alpha}(l,n,k), \\
H^{\alpha\beta\gamma\delta*}(l,k,n) &= H^{\delta\gamma\beta\alpha}(l,n,k), \\
\text{and} \quad K^{\alpha\beta\gamma\delta*}(l,k,n) &= -K^{\beta\alpha\delta\gamma}(-l,k - l,n - l).
\end{align*}
$$

(2.15)

### 2.3 Wigner matrix

If we have a function $u$ that depends on time and one spatial dimension, then we can define an energy spectral density as the Fourier transform of a correlation function of $u$, i.e.,

$$
E(k) = \int u(x + y)u(x)e^{iky}dy.
$$

However, this expression does not include any information about the fast spatial variation. Generalizing to the multidimensional case and following Ryzhik et al. (1996), we define the $n \times n$ (Hermitian) Wigner matrix associated with $u(x,t)$ by

$$
W(x,k,t) = \frac{1}{(2\pi)^d} \int u(x - \epsilon y/2,t)u^*(x + \epsilon y/2,t)e^{iky}dy,
$$

(2.16)

and we note the dual expression

$$
W(e(x,k,t)) = -\int \hat{u}(k/\epsilon + l/2,t)\hat{u}^*(k/\epsilon - l/2,t)e^{il\cdot y}dl
$$

(2.17)

in terms of the Fourier transform $\hat{u}$ of $u$. The Wigner matrix is a localised power spectrum, i.e., the Fourier transform of a correlation between components of $u$. The scale separation allows us to keep the spatial dependence. It is useful to relate the conserved energy $\mathcal{H}$ given by (2.5) to the Wigner matrix: a short calculation detailed in Appendix A.3 gives

$$
\mathcal{H} = \frac{1}{2} \text{Tr} \int H(x,ik + \epsilon \partial_x/2)W_e(x,k,t)dxdk,
$$

(2.18)

where $\text{Tr}$ denotes the trace of a matrix.

Let us now derive an evolution equation for the Wigner matrix. Taking the time derivative of (2.16) we obtain

$$
\partial_t W_e(x,k,t) = \frac{1}{(2\pi)^d} \int \left[ u_t(x - \epsilon y/2,t)u^*(x + \epsilon y/2,t) \\
+ u(x - \epsilon y/2,t)u^*_t(x + \epsilon y/2,t) \right]e^{iky}dy.
$$
By using (2.7) this becomes
\[
\epsilon \partial_t W_\epsilon(x, k, t) = \frac{1}{(2\pi)^d} \int \left[ J(x - \epsilon y/2, t) H(x - \epsilon y/2, t) u(x - \epsilon y/2, t) u^*(x + \epsilon y/2, t) 
+ u(x - \epsilon y/2, t) u^*(x + \epsilon y/2, t) H^*(x + \epsilon y/2, t) J^*(x + \epsilon y/2, t) \right] e^{ik \cdot y} dy,
\]
which when expanded using (2.8), gives
\[
\epsilon \partial_t W_\epsilon = S_0 + \epsilon^{1/2} S_{1/2} + \epsilon S_1 + \ldots,
\]
where the terms on the right-hand side can be written as
\[
S_p = \frac{1}{(2\pi)^d} \int \left[ u(x - \epsilon y/2, t) (L_p u)^*(x + \epsilon y/2, t) 
+ (L_p u) (x - \epsilon y/2, t) u^*(x + \epsilon y/2, t) \right] e^{ik \cdot y} dy,
\]
for \( p = 0, 1/2 \) and 1. Here, the pseudodifferential operators \( L_p = L_p(x \pm \epsilon y/2, \epsilon \partial_x) \) are defined by the corresponding matrices
\[
L_0(x, \epsilon k) = J_0(x, \epsilon k) H_0(x, \epsilon k),
\]
\[
L_{1/2}(x, \epsilon k) = J_0(x, \epsilon k) H_{1/2}(x/\epsilon, \epsilon k) + J_{1/2}(x/\epsilon, \epsilon k) H_0(x, \epsilon k)
\]
and \( L_1(x, \epsilon k) = J_0(x, \epsilon k) H_1(x, \epsilon k) + J_1(x, \epsilon k) H_0(x, \epsilon k) + J_{1/2}(x/\epsilon, \epsilon k) H_{1/2}(x/\epsilon, \epsilon k) \).

Note that whereas the matrix \( L_0 \) is independent of \( \epsilon \), the term \( S_0 \) is not, due to the \( \epsilon \) dependence of the arguments \( x \pm \epsilon y/2 \). Furthermore, \( L_{1/2} \) and \( L_1 \) depend on \( \epsilon \) through the argument \( x/\epsilon \) of \( J_{1/2} \) and \( H_{1/2} \).

To reduce (2.19) to a closed equation for the Wigner matrix, we introduce the Fourier transforms of \( u \) and of the matrices \( L_p \) and express products of \( \hat{u} \) using (2.17). We now detail this calculation for \( S_0 \) (the argument \( t \) we leave out to make the calculation more compact). Starting from (2.20) we obtain
\[
S_0 = \frac{1}{(2\pi)^d} \iiint \left\{ \hat{u}(p) e^{ip \cdot (x - \epsilon y/2)} \left[ \hat{J}_0(m, \epsilon(l + q)) \hat{H}_0(l, \epsilon q) \hat{u}(q) e^{i(m+l+q) \cdot (x+\epsilon y/2)} \right]^* 
+ \hat{J}_0(m, \epsilon(l + p)) \hat{H}_0(l, \epsilon q) \hat{u}(q) e^{i(m+l+p) \cdot (x-\epsilon y/2)} \hat{u}^*(q) e^{-q \cdot (x+\epsilon y/2)} \right\} \times 
\times e^{ik \cdot y} dy dp dq dl dm
\]
by integrating over \( y \). We now integrate over \( p \) to obtain
\[
S_0 = \left( \frac{2}{\epsilon} \right)^d \iiint \left\{ \hat{u}(2k/\epsilon - m - l - q) \hat{u}^*(q) \left[ \hat{J}_0(m, \epsilon(l + q)) \hat{H}_0(l, \epsilon q) \right]^* \times 
\times e^{2i(k/\epsilon - m - l - q) \cdot x} 
+ \hat{J}_0(m, 2k - \epsilon(m + q)) \hat{H}_0(l, 2k - \epsilon(l + m + q)) \hat{u}(2k/\epsilon - m - l - q) \times 
\times \hat{u}^*(q) e^{2i(k/\epsilon - q) \cdot x} \right\} dq dl dm.
\]
Substituting \( q = \frac{k}{\epsilon} - \frac{l}{2} - \frac{m}{2} - n \) into the above equation we obtain

\[
S_0 = \left( \frac{2}{\epsilon} \right)^d \iint \left\{ \hat{u}(k/\epsilon - (l + m)/2 + n)\hat{u}^*(k/\epsilon - (l + m)/2 - n) \times 
\right. \\
\times \left[ \tilde{J}_0(m, k + \epsilon(l/2 - m/2 - n))\tilde{H}_0(l, k - \epsilon(l/2 + m/2 + n)) \right]^* e^{2(n-l/2-m/2)x} \\
+ \tilde{J}_0(m, k + \epsilon(l/2 - m/2 + n))\tilde{H}_0(l, k - \epsilon(l/2 + m/2 - n)) \times \\
\times \hat{u}(k/\epsilon - l/2 - m/2 + n)\hat{u}^*(k/\epsilon - l/2 - m/2 - n)e^{2(n+l/2+m/2)x} \right\} dn \, dl \, dm.
\]

Making the transformations \( l \to -l \) and \( m \to -m \) in the second term of the integral and \( n \to n/2 \) in both terms and using the following expression

\[
\hat{u}(k/\epsilon + n)\hat{u}^*(k/\epsilon - n) = \left( \frac{\epsilon}{2\pi} \right)^d \int W_\epsilon(y, k)e^{-2y^\epsilon n} \, dy,
\]

(which is derived in Appendix A.5), we find

\[
S_0 = \frac{1}{(2\pi)^d} \iint \left[ W_\epsilon(y, k - \epsilon(l + m)/2)[\tilde{J}_0(m, k + \epsilon(l - m - n)/2) \times \\
\times \tilde{H}_0(l, k - \epsilon(l + m + n)/2)]^* \\
+ \tilde{J}_1/2(m, k + (l - \epsilon m - \epsilon n)/2)\tilde{H}_1/2(l, k - (l + \epsilon m + \epsilon n)/2) \times \\
\times W_\epsilon(y, k + (l + \epsilon m)/2)]e^{i[n(x-y)-(l+\epsilon m)x]} \, dy \, dl \, dn.
\]

Similar calculations provide \( S_{1/2} \) and \( S_1 \) in the forms

\[
S_{1/2} = \frac{1}{(2\pi)^d} \iint \left\{ W_\epsilon(y, k - \epsilon(l + m)/2)[\tilde{J}_0(m, k + (l - \epsilon m - \epsilon n)/2) \times \\
\times \tilde{H}_1/2(l, k - (l + \epsilon m + \epsilon n)/2) \\
+ \tilde{J}_1/2(m, k + (l - \epsilon m - \epsilon n)/2)\tilde{H}_1/2(-l, k + (l + \epsilon m + \epsilon n)/2) \times \\
\times W_\epsilon(y, k + (l + \epsilon m)/2)]e^{i[n(x-y)-(l+\epsilon m)x]} \, dy \, dl \, dn
\]

and

\[
S_1 = \frac{1}{(2\pi)^d} \iint \left\{ W_\epsilon(y, k - \epsilon(l + m)/2)[\tilde{J}_0(m, k + (l - m - n)/2) \times \\
\times \tilde{H}_1/2(l, k - (l + m + n)/2) \\
+ \tilde{J}_1(m, k + (l - m - n)/2)\tilde{H}_1(-l, k - \epsilon(l + m + n)/2) \times \\
\times W_\epsilon(y, k + (l + m)/2)]e^{i[n(x-y)-(l+\epsilon m)x]} \\
+ \left\{ W_\epsilon(y, k - (l + m)/2)[\tilde{J}_1/2(m, k + (l - m - \epsilon n)/2) \times \\
\times \tilde{H}_1/2(l, k - (l + m + \epsilon n)/2)]^* \\
+ \tilde{J}_1/2(-m, k - (l - m - \epsilon n)/2)\tilde{H}_1/2(-l, k + (l + m + \epsilon n)/2) \times \\
\times W_\epsilon(y, k + (l + m)/2)]e^{i[n(x-y)-(l+\epsilon m)x/\epsilon]} \, dy \, dl \, dn.
\]

38
It is now straightforward to expand $S_p$ in powers of $\epsilon$. Expanding $S_0$, using (2.22) yields

$$S_0 = Q_{00}W_\epsilon(x,k,t) + \epsilon Q_{01}W_\epsilon(x,k,t) + O(\epsilon^2),$$

(2.25)

where

$$Q_{00}W = WL_0^* + L_0W$$

(2.26)

and

$$Q_{01}W = \frac{i}{2} [\nabla_x W \cdot \nabla_k L_0^* - \nabla_k W \cdot \nabla_x L_0^* - W \nabla_k \cdot \nabla_x L_0^*]$$

$$- \frac{i}{2} [\nabla_k L_0 \cdot \nabla_x W - \nabla_x L_0 \cdot \nabla_k W - \nabla_k \cdot \nabla_x L_0 W],$$

(2.27)

for any $W = W(x,k,t)$. Using (2.23) we obtain an expansion for $S_{1/2}$

$$S_{1/2} = Q_{1/2,0}W_\epsilon(x,k,t) + O(\epsilon),$$

(2.28)

with $Q_{1/2,0}$ given by

$$Q_{1/2,0}W = 2^d \int \left[ W(x,k+l,t)V^*(x,-2l,k+l) + V(x,2l,k-l)W(x,k-l,t) \right] e^{2i\frac{x}{\epsilon}} \, dl,$$

(2.29)

where we have introduced the matrix

$$V(x,l,k) = \hat{J}_{1/2}(l,k)H_0(x,ik) + J_0(x,ik + ul)\tilde{H}_{1/2}(l,k).$$

(2.30)

Finally, using (2.24), we have

$$S_1 = Q_{10}W_\epsilon(x,k,t) + O(\epsilon),$$

where

$$Q_{10}W = W [J_1(x,ik)H_0(x,ik) + J_0(x,ik)H_1(x,ik)]^*$$

$$+ \left[ J_1(x,ik)H_0(x,ik) + J_0(x,ik)H_1(x,ik) \right] W$$

$$+ 4^d \int \left\{ W(x,k+l+m,t) \left[ \hat{J}_{1/2}(-2m,k-l+m)\tilde{H}_{1/2}(-2l,k+l+m) \right]^*$$

$$+ \left[ \hat{J}_{1/2}(2m,k+l-m)\tilde{H}_{1/2}(2l,k-l-m) \right] W(x,k-l-m,t) \right\} \times$$

$$\times e^{2i(l+m)\frac{x}{\epsilon}} \, dl \, dm.$$
2.3.1 Reduction to operators used in Ryzhik et al. (1996)

The operators $Q$ above reduce to those in Ryzhik et al. (1996). In particular if $H_0 \equiv k_3 \partial_1$ and $J_0 \equiv -iA^{-1}$, where $A(x)$ is a symmetric matrix and we denote the corresponding operators in Ryzhik et al. – given by (3.16)–(3.18) in their paper – by $Q^R_2$, $Q^R_2$ and $Q^R_1$, then our operators $Q_00$ and $Q_{01}$ reduce as follows:

$$Q_{00}W = W[JH]^* + JHW = W[-iA^{-1}k_1D^i] - iA^{-1}k_1D^iW$$

$$= - (iA^{-1}k_1D^iW - iWk_1D^iA^{-1}) = -Q^R_2$$

and

$$Q_{01}W = \frac{i}{2}(W_{x_1}[JH]^*_{x_1} - W_{x_1}[JH]_{x_1}^* - W(JH)_{x_1}^*_{x_1}$$

$$= \frac{i}{2}(-W_{x_1}[-iA_{x_1}^{-1}k_1D^i]^* - iA_{x_1}^{-1}k_1D^iW_{x_1})$$

$$= \frac{i}{2}(W_{x_1}[-iA^{-1}D]_x^* + iA^{-1}D_jW_{x_1} - W[-iA_{x_1}^{-1}D]_x^* + [-iA_{x_1}^{-1}D]_xW)$$

$$= -\frac{1}{2}(-A_{x_1}^{-1}k_1D^iW_{x_1} - W_{x_1}k_1D^iA_{x_1}^{-1})$$

$$-\frac{1}{2}(A^{-1}D_jW_{x_1} + W_{x_1}D_jA^{-1} - A_{x_1}^{-1}D_jW - WD_jA_{x_1}^{-1}) = -Q^R_{21} - Q^R_1.$$

2.4 Derivation of the transport equations

We expand the Wigner matrix according to

$$W_\epsilon(x, k, t) = W_0(x, k, t) + \epsilon^{1/2}W_1(x, \xi, k, t) + \epsilon W_2(x, \xi, k, t) + O(\epsilon^{3/2}). \quad (2.31)$$

Here, we have introduced the fast spatial variable $\xi = x/\epsilon$ and we have anticipated the fact that the (deterministic) leading order approximation $W_0$ to $W_\epsilon$ is independent of $\xi$. When applied to differential operators, the multiple-scale method is efficiently implemented using the substitution $\partial_x \mapsto \partial_x + \epsilon^{-1}\partial_\xi$ which follows from the chain rule (e.g. Hinch 1991). The analogous substitution for pseudodifferential operators is conveniently carried out in the Fourier representation. If an operator $Q(\partial_x)$ is given by

$$Q(\partial_x)u(x) = \frac{1}{(2\pi)^d} \int \int Q(\Im n)u(y)e^{im(x-y)} dy dm$$

when acting on $u(x)$, its multiple-scale version, which we will denote by a tilde for clarity, acts on $u(x, \xi)$ and is given by

$$\tilde{Q}(\partial_x + \epsilon^{-1}\partial_\xi)u(x, \xi)$$

$$= \frac{1}{(2\pi)^{2d}} \int \int \int \int Q(\Im m + \Im n/\epsilon)u(y, \eta)e^{im(x-y) + n(\xi-\eta)} dy dm \eta dn. \quad (2.32)$$
This provides a simple rule to compute the multiple-scale versions of $S_p$ and, by expansion in powers of $\epsilon$, of the operators $Q_{pq}$.

Introducing (2.31) into (2.19) leads to a sequence of equations for the Wigner matrices $W_p$ for $p = 0, 1/2, 1, \ldots$ We now detail each of these equations, up to that for $W_1$ whose solvability conditions yield transport equations of the form (2.1).

### 2.4.1 $O(1)$ equation

At leading order, we find that

$$Q_{00}W_0(x, k, t) = 0. \tag{2.33}$$

To solve this equation, we consider the eigenvalues $-\omega(s)(x, k)$ and (right) eigenvectors $e_{(s)}(x, k)$ of $L_0(x, k)$. (The brackets in the subscript here are just used to distinguish the eigenvalue index from the matrix indices.) These satisfy

$$L_0(x, k)e_{(s)}(x, k) = -\omega_{(s)}(x, k)e_{(s)}(x, k), \quad \text{for } s = 1, 2, \ldots, n \tag{2.34}$$

and respectively define the dispersion relation and polarisation relations for waves propagating in the system (2.7). The sign convention for the frequencies is the usual one.

Seeking approximate solutions to the leading order approximation

$$\epsilon \partial_t u = J_0(x, \partial_x)H_0(x, \partial_x)u$$

of (2.7) in the WKB form (see Section 1.5.1)

$$u(x, t) \propto e^{i\phi(x, t)/\epsilon} e$$

leads to the eigenvalue problem (2.34), with the usual relations

$$k = \nabla_x \phi \quad \text{and} \quad \omega = -\partial_t \phi.$$ 

We assume (i) that the frequencies $\omega_{(s)}(x, k)$ are all real and (ii) that they all have multiplicity one (eigenvectors are distinct). The first assumption, which amounts to the linear stability of the system, is satisfied in particular if $H_0(x, \iota k)$ is sign definite; the second assumption excludes polarised waves (they are not important for atmospheric gravity waves). We now review some properties of the eigenvalue problem (2.34) which will be needed in what follows. The left eigenvectors $\hat{e}_{(s)}(x, k)$ of $L_0(x, k)$, which satisfy

$$L_0^*(x, k)\hat{e}_{(s)}(x, k) = i\omega_{(s)}(x, k)\hat{e}_{(s)}(x, k), \tag{2.35}$$

can be related to the right eigenvectors according to

$$\hat{e}_{(s)}(x, k) = H_0(x, \iota k)e_{(s)}(x, k). \tag{2.36}$$

41
The relationship
\[ \hat{e}^*_s(x, k)J_0(x, ik) = -\omega_s(x, k)e^*_s(x, k) \] (2.37)
follows readily. We choose to normalize the eigenvectors so that there is orthogonality between left and right eigenvectors:
\[ \hat{e}^*_s(x, k)e(t)(x, k) = \delta_{ts}, \] (2.38)
where \( \delta_{ts} = 1 \) for \( t = s \) and 0 otherwise. With this normalization, the \( n \)-dimensional identity matrix has the expansion
\[ I = \sum_s e_s(x, k)\hat{e}^*_s(x, k). \] (2.39)

Returning to (2.33), we note that the null space of \( Q_{00} \) is spanned by the matrices \( e_s(x, k)e^*_s(x, k) \), (see Appendix A.6). Thus, the solution of (2.33) takes the general form
\[ W_0(x, k, t) = \sum_s a_s(x, k, t)e_s(x, k)e^*_s(x, k), \] (2.40)
for some amplitudes \( a_s(x, k, t) \). These amplitudes, whose transport equations we seek to derive, can be interpreted as the energy density of mode \( s \) in the \((x, k)\) phase space. Indeed, introducing (2.31) into (2.18) and taking (2.38) and (2.40) into account gives the approximation for the energy
\[ \mathcal{H} = \frac{1}{2} \sum_s \iint a_s(x, k, t) \, dx \, dk + O(\epsilon^{1/2}). \] (2.41)

### 2.4.2 \( O(\epsilon^{1/2}) \) equation

At \( O(\epsilon^{1/2}) \), the evolution equation (2.19) for the Wigner matrix gives
\[ \tilde{Q}_{00}W_{1/2}(x, \xi, k, t) + Q_{1/2, 0}W_0(x, k, t) = 0, \] (2.42)
where, by applying (2.32) to (2.26) and then expanding in powers of \( \epsilon \) we obtain
\[ \tilde{Q}_{00}W_{1/2}(x, \xi, k, t) = \frac{1}{(2\pi)^d} \int \left[ W_{1/2}(x, \eta, k, t)L_0(x, k - n/2) \right. \\
+ L_0(x, k + n/2)W_{1/2}(x, \eta, k, t)] e^{i(m(\xi - \eta))} \, d\eta \, dn. \]

To solve this equation, we define the \( n \times n \) matrix \( F(x, m, k, t) \) by
\[ W_{1/2}(x, \xi, k, t) = \int F(x, m/2, k, t)e^{im\cdot\xi} \, dm \]
(the argument \( m/2 \) has been used to simplify future calculations) and introduce this expression into (2.42) to find that
\[ F(x, l, k, t)L_0(x, k - l) + L_0(x, k + l)F(x, l, k, t) \]
\[ = -[W_0(x, k + l, t)V^*(x, -2l, k + l) + V(x, 2l, k - l)W_0(x, k - l, t)]. \] (2.43)
We then expand $F$ according to

$$ F(x, l, k, t) = \sum_{s,l} P_{s,t}(x, l, k, t) e_{s}(x, k + l) e_{s}(x, k - l), \quad (2.44) $$

where the scalar coefficients $P_{s,t}$ remain to be determined. This is achieved by introducing the expansion (2.44) into (2.43), left-multiplying by $e_{s}(x, k + l)$ and right-multiplying by $e_{s}(x, k - l)$. Calculations detailed in Appendix A.7 yield the result

$$ 1 \left[ \omega_{s}(x, k + l) - \omega_{s}(x, k - l) \right] \mathcal{P}_{s,t}(x, l, k, t) \\
= - \left[ a_{s}(x, k + l, t) - a_{s}(x, k - l, t) \right] \hat{e}_{s}(x, k + l) \hat{H}_{1/2}(2l, k - l) \hat{e}_{s}(x, k - l) \\
+ i \left[ \omega_{s}(x, k - l) a_{s}(x, k + l, t) - \omega_{s}(x, k + l) a_{s}(x, k - l, t) \right] \times \\
e_{s}(x, k + l) \hat{H}_{1/2}(2l, k - l) \hat{e}_{s}(x, k - l), \quad (2.45) $$

which completes the determination of $W_{1/2}$.

### 2.4.3 $O(\epsilon)$ Equation

The $O(\epsilon)$ equation for the Wigner matrix reads

$$ \partial_{t} W_{0}(x, k, t) = \hat{Q}_{00} W_{1}(x, \xi, k, t) + (Q_{01} + Q_{10}) W_{0}(x, k, t) + \hat{Q}_{1/2,0} W_{1/2}(x, \xi, k, t), \quad (2.46) $$

where using (2.28) and (2.32) and then expanding in powers of $\epsilon$ we obtain

$$ \hat{Q}_{1/2,0} W_{1/2}(x, \xi, k, t) \\
= \left( \frac{2}{\pi} \right)^{d} \int_{\mathbb{R}^{d}} \int_{\mathbb{R}^{d}} [W_{1/2}(x, \eta, k + l, t) V^{*}(x, -2l, k + l - m) \\
+ V(x, 2l, k - l + m) W_{1/2}(x, \eta, k - l, t)] e^{2i(m(\xi - \eta) + t \xi)} d\eta dl dm \\
= 2^{d} \int_{\mathbb{R}^{d}} \int_{\mathbb{R}^{d}} [F(x, m, k + l, t) V^{*}(x, -2l, k + l - m) \\
+ V(x, 2l, k - l + m) F(x, m, k - l, t)] e^{2i(m+l)\xi} dl dm. $$

The transport equations are deduced as solvability conditions for (2.46). We first take the ensemble average of this equation and note that in a homogeneous media an ensemble average is equivalent to a spatial average. The term $L_{0}(x, k - n/2)$ contains a fast wave variation $n$ superimposed on a slow wave variation $k$. When we take the spatial average of $L_{0}$, the fast variation is smoothed out. This implies that

$$ \langle \hat{Q}_{00} W_{1}(x, \xi, k, t) \rangle = Q_{00}(W_{1}(x, \xi, k, t)). $$

The solvability conditions follow by left- and right-multiplying by $\hat{e}_{s}(x, k)$ and $\hat{e}_{s}(x, k)$ respectively for $s = 1, 2, \ldots, n$. This cancels the first term on the right-hand side of (2.46), leading to

$$ \partial_{t} a_{s}(x, k, t) = \hat{e}_{s}(x, k) \left[ (Q_{01} + (Q_{10})) W_{0}(x, k, t) + (\hat{Q}_{1/2,0} W_{1/2}(x, \xi, k, t)) \right] \hat{e}_{s}(x, k), \quad (2.47) $$
after using (2.38) and (2.40). $W_{1/2}$ can be expressed in terms of $W_0$ and hence of $a_{(s)}(x,k,t)$, so these $n$ equations are closed and, when simplified, provide the sought transport equations. The simplification starts by considering the deterministic terms, which are obtained by setting $J_{1/2} = H_{1/2} = 0$; in this case, $Q_{1/2,0}$ vanishes and $Q_{10}$ simplifies. We show in Appendix A.8 that (2.47) then reduces to transport equations of the Liouville form

$$\partial_t a_{(s)}(x,k,t) + \nabla_k \omega(x,k) \cdot \nabla_x a_{(s)}(x,k,t) - \nabla_x \omega(x,k) \cdot \nabla_k a_{(s)}(x,k,t) = 0. \quad (2.48)$$

We now turn to the random contributions to (2.47). We write the last term as

$$\tilde{e}_{(s)}^*(x,k)(\tilde{Q}_{1/2,0}^+ W_{1/2}(x,\xi,k,t)) \tilde{e}_{(s)}(x,k) =$$

$$= 2d \tilde{e}_{(s)}^*(x,k) \left( \int \int V(2l,k-l+m) F(x,m,k-l,t) e^{2\pi i (m+l) \xi} dl dm \right) \tilde{e}_{(s)}(x,k) + \text{c.c.}$$

$$= I_J + I_H + I_K + \text{c.c.}, \quad (2.49)$$

where c.c. denotes the complex conjugate of the previous terms in the expression. In this expression we have separated the three terms involving, respectively, products of the form $\tilde{J}_{1/2} \otimes \tilde{J}_{1/2}$, $\tilde{H}_{1/2} \otimes \tilde{H}_{1/2}$, and $\tilde{J}_{1/2} \otimes \tilde{H}_{1/2}$; thus, in terms of the correlation tensors defined by (2.11)–(2.13), $I_J$ involves $J$, $I_H$ involves $H$ and $I_K$ involves $K$. We detail in Appendix A.9 the derivation of each of these terms. The first is found to be

$$I_J = -\pi \sum_t \int \Phi^J_{(s,t)}(x,l,k) \delta [\omega(t)(x,k-l) - \omega_{(s)}(x,k)] \times$$

$$\times [a(t)(x,k-l,t) - a_{(s)}(x,k,t)] \, dl, \quad (2.50)$$

where

$$\Phi^J_{(s,t)}(x,l,k) = J^{\alpha \beta \gamma \delta}(l,k,k) \tilde{e}_{(s)}^*(x,k) \tilde{e}_{(t)}^\gamma(x,k-l) \tilde{e}_{(t)}^\delta(x,k-l) \tilde{e}_{(s)}^\delta(x,k) \quad (2.51)$$

and a summation over the repeated superscripts is used. The second term is given by

$$I_H = \pi \omega_{(s)}^2(x,k) \sum_t \int \Phi^H_{(s,t)}(x,l,k) \delta [\omega(t)(x,k-l) - \omega_{(s)}(x,k)] \times$$

$$\times [a(t)(x,k-l,t) - a_{(s)}(x,k,t)] \, dl, \quad (2.52)$$

where

$$\Phi^H_{(s,t)}(x,l,k) = H^{\alpha \beta \gamma \delta}(l,k,k) \tilde{e}_{(s)}^\alpha(x,k) \tilde{e}_{(t)}^\beta(x,k-l) \tilde{e}_{(t)}^\gamma(x,k-l) \tilde{e}_{(s)}^\delta(x,k). \quad (2.53)$$

Finally, the third term is given by

$$I_K = \pi \omega_{(s)}(x,k) \sum_t \int \Phi^K_{(s,t)}(x,l,k) \delta [\omega(t)(x,k-l) - \omega_{(s)}(x,k)] \times$$

$$\times [a(t)(x,k-l,t) - a_{(s)}(x,k,t)] \, dl$$

$$- \left[ a_{(s)}(x,k,t) \tilde{e}_{(s)}^\alpha(x,k) \tilde{e}_{(s)}^\delta(x,k) \int K^{\alpha \beta \gamma \delta}(l,k,k) \, dl + \text{c.c.} \right], \quad (2.54)$$

44
where
\[
\Phi^K_{(s,t)}(x,l,k) = \int \left[ K^{\alpha\beta\gamma\delta}(x,l,k,k)\tilde{e}^\alpha_{(s)}(x,k)e^\beta_{(t)}(x,k-l)e^{\gamma*}_{(t)}(x,k-l)e^{\delta*}_{(s)}(x,k) \\
+ K^{\gamma\delta\alpha\beta}(-l,k-l,k-l)e^{\alpha}_{(s)}(x,k)e^\beta_{(t)}(x,k-l)e^{\gamma*}_{(t)}(x,k-l)e^{\delta*}_{(s)}(x,k) \right].
\]

(2.55)

There is a further random contribution in (2.46), stemming from the random term in $Q_{10}W_0$. This contribution is given by
\[
4^d\tilde{e}^*_{(s)}(x,k)\int \left\{ W_0(x,k+l+m,t)\langle \hat{J}_1/2(-2m, k-l+m)\hat{H}_1/2(-2l, k+l+m) \rangle^* \\
+ \langle \hat{J}_1/2(2m, k-l-m)\hat{H}_1/2(2l, k-l-m) \rangle W_0(x,k-l-m,t) \right\} \times e^{2i(l+m)\xi} \, dl \, dm \, \tilde{e}(x,k).
\]

(2.56)

Taking (2.13) into account, this term can be shown to be identical, up to the sign, to the second line in the expression (2.54) for $I_K$. Therefore, combining (2.50), (2.52) and (2.54) leads to the transport equations of the form
\[
\partial_t a_{(s)}(x,k,t) + \nabla_x \omega(x,k) \cdot \nabla_x a_{(s)}(x,k,t) - \nabla_x \omega(x,k) \cdot \nabla_k a_{(s)}(x,k,t) = \sum_t \int a_{(s,t)}(x,k,k')a_{(t)}(x,k',t) \, dk' - \Sigma_{(s)}a_{(s)}(x,k,t).
\]

(2.57)

These equations represent the transport of a scalar energy density in $(x,k)$ space, described fully in Section 2.1. The differential scattering cross-section is given by
\[
\sigma_{(s,t)}(x,k,k') = 2\pi \delta \left[ \omega(t)(x,k') - \omega(s)(x,k) \right] \phi^H_{(s,t)}(x,k-k',k) + \omega(s)(x,k)\Phi^K_{(s,t)}(x,k-k',k),
\]

where the functions $\Phi^H_{(s,t)}$, $\Phi^I_{(s,t)}$ and $\Phi^K_{(s,t)}$ are defined by (2.53), (2.51) and (2.55). A factor of 2 appears here because of the complex conjugate terms in J, H and K. Substituting these into (2.57) we arrive at
\[
\sigma_{(s,t)}(x,k,k') = 2\pi \delta \left[ \omega(t)(x,k') - \omega(s)(x,k) \right] \phi^H_{(s,t)}(x,k-k',k) + \omega(s)(x,k)\Phi^K_{(s,t)}(x,k-k',k).
\]

(2.58)

The total scattering cross section is
\[
\Sigma_{(s)} = \sum_t \int \sigma_{(s,t)}(x,k,k') \, dk',
\]

(2.59)

ensuring energy conservation. Although we have not proved this, it should be possible to show that $\sigma > 0$ \footnote{Ryzhik et al. (1996) makes a statement to the effect that $\sigma > 0$ can be proved by using Bochner's theorem (e.g. Gihman & Skorohod 1962).} and indeed this will be shown to be the case for the gravity waves example given in Section 2.5.

45
The transport equations (2.56), together with the explicit expressions (2.58) and (2.59) for the differential and total scattering cross-sections, are the main results of this chapter. They generalize the results of Ryzhik et al. (1996) and of Guo & Wang (1999) (in the conservative case) to the large class of randomly perturbed linear Hamiltonian systems of the form (2.4). With these results, the derivation of transport equations for particular systems is reduced to the straightforward, algorithmic computation of the left and right eigenvectors \( \hat{e}_{(s)}(x,t) \) and \( e_{(s)}(x,t) \) and of their products with the correlation 4-tensors defined in (2.11)–(2.13).

In the absence of random perturbations, i.e., in the absence of the scattering terms on the right-hand side, (2.56) can be used to recover all the equations traditionally derived using a WKB expansion: the ray equations (cf. Section 1.5.1)

\[
\frac{dx}{dt} = \nabla_k \omega \quad \text{and} \quad \frac{dk}{dt} = -\nabla_x \omega.
\]

are nothing other than the characteristics of (2.56), and the transport equation for the wave-packet energy (or pseudoenergy)

\[
\partial_t A(x,t) + \nabla_x \cdot (c_g A(x,t)) = 0
\] (2.60)

where \( c_g \) is the group velocity vector, is obtained by noting that

\[
a(x,k,t) = A(x,t) \delta(k - k(t))
\]

and integrating (2.56) with respect to \( k \).

### 2.5 Application to gravity waves

We now use the above theory to derive a transport equation for gravity waves in the atmosphere. For two dimensional flow with no rotation and a one-dimensional background wind \( U(z) = U_0(\epsilon z) + \epsilon^{1/2}U_{1/2}(z) \) where, for \( \epsilon \ll 1 \), \( U_0(\epsilon z) \) is the deterministic slowly varying part of the wind and \( U_{1/2}(z) \) is the rapidly varying random perturbation. The linearised Boussinesq equations of motion reduce to (see Section 1.5 or Gill 1982)

\[
\begin{align*}
 u_x + w_z &= 0, \\
 u_t + (U_0 + \epsilon^{1/2}U_{1/2})u_x + (\epsilon U_0' + \epsilon^{1/2}U_{1/2}')w &= -\frac{p_z}{\rho_0}, \\
 w_t + (U_0 + \epsilon^{1/2}U_{1/2})w_x &= -\frac{p_z + \rho g}{\rho_0}
\end{align*}
\]

and

\[
\rho_t + (U_0 + \epsilon^{1/2}U_{1/2})\rho_z + \rho_0 w = 0,
\]

where the prime denotes differentiation with respect to \( z \). Since we have a two dimensional system, the incompressibility condition can be eliminated by expressing the velocity components \( u \) and \( w \) in terms of a stream-function \( \psi \), with

\[
u = -\psi_z \quad \text{and} \quad w = \psi_x.
\]
The equations of motion now become
\[
(\nabla^2 \psi_t + (U_0 + \epsilon^{1/2}U_{1/2}) \nabla^2 \psi_x - (\epsilon^2 U_0'' + \epsilon^{1/2}U_{1/2}'') \psi_x + \frac{g \rho_x}{\rho_0} = 0
\]
and \( \rho_t + (U_0 + \epsilon^{1/2}U_{1/2}) \rho_x + \rho_{0x} \psi_x = 0. \)

If we let \( \rho = \dot{\rho}/g \) and \( \xi = \nabla^2 \psi \) and drop hats we obtain
\[
\xi_t + (U_0 + \epsilon^{1/2}U_{1/2}) \xi_x - (\epsilon^2 U_0'' + \epsilon^{1/2}U_{1/2}'') \psi_x + \rho_x = 0
\]
and \( \rho_t + (U_0 + \epsilon^{1/2}U_{1/2}) \rho_x - N^2 \psi_x = 0. \) (2.61)

These can be written in matrix form as
\[
\mathbf{u}_t = JH \mathbf{u},
\]
where \( \mathbf{u} = (\rho, \xi). \) We expand
\[
J = J_0 + \epsilon^{1/2}J_{1/2} + O(\epsilon^2) \quad \text{and} \quad H = H_0 + \epsilon^{1/2}H_{1/2} + \epsilon H_1 + O(\epsilon^2).
\]

However, it is not immediately clear what the matrices \( J \) and \( H \) should be. We can calculate them by first writing down an expression for the wave energy (or pseudoenergy) \( \mathcal{H} \) (see Section 1.5). For a Hamiltonian system (2.62), the wave energy \( \mathcal{H} \) is given by (2.5):
\[
\mathcal{H} = \frac{1}{2} \int \mathbf{u}^* H \mathbf{u} \, dx.
\] (2.63)

For the system (2.61), we have (e.g. Vanneste & Vial 1994)
\[
\mathcal{H} = \frac{1}{2} \int \left\{ -\psi \xi + \frac{\rho^2}{N^2} + \frac{2}{N^2}(U_0 + \epsilon^{1/2}U_{1/2})\xi \rho - \frac{1}{N^4}(U_0 + \epsilon^{1/2}U_{1/2})(\epsilon^{1/2}U_{1/2}')\rho^2 \right\} dx + O(\epsilon^2). \tag{2.64}
\]

Using (2.63) and (2.64) we can now calculate \( \mathcal{H} \), which is given by
\[
\mathcal{H}(x, \partial_x) = \left( N^{-2} - (U_0 + \epsilon^{1/2}U_{1/2})\epsilon^{1/2}U_{1/2}'N^{-4} \begin{pmatrix} (U_0 + \epsilon^{1/2}U_{1/2})N^{-2} & -\nabla^2 \end{pmatrix} \right) + O(\epsilon^2), \tag{2.65}
\]
where \( \mathcal{H} \) is symmetric and \( \nabla^2 \) is the inverse Laplacian operator such that \( \nabla^2 (\nabla^2 \phi) = \phi. \) Therefore using (2.61), (2.62) and (2.65) we can calculate \( J \), which is given by
\[
J(x, \partial_x) = \begin{pmatrix} 0 & -N^2 \partial_x \\ -N^2 \partial_x & -\epsilon^{1/2}U_{1/2}' \partial_x \end{pmatrix} + O(\epsilon^2),
\]
where \( J \) is skew symmetric\(^4\). Expanding \( J \) and \( H \) in powers of \( \epsilon \) we find
\[
H_0(x, \partial_x) = \begin{pmatrix} N^{-2} & U_0N^{-2} \\ U_0N^{-2} & -\nabla^2 \end{pmatrix}, \quad H_1(x, \partial_x) = -\frac{1}{N^4} \begin{pmatrix} U_{1/2}U_{1/2}' & 0 \\ 0 & 0 \end{pmatrix},
\]
\(^4\)Note that this may not appear to be so, but is true due to the fact that the operator \( \partial_x \) is skew.
\[ H_{1/2}(x/\epsilon, \partial_x) = \frac{1}{N^2} \begin{pmatrix} -U_0 U''_{1/2} N^{-2} & U_{1/2} \\ U_{1/2} & 0 \end{pmatrix}, \]

\[ J_0(x, \partial_x) = \begin{pmatrix} 0 & -N^2 \partial_x \\ -N^2 \partial_x & 0 \end{pmatrix} \quad \text{and} \quad J_{1/2}(x/\epsilon, \partial_x) = \begin{pmatrix} 0 & 0 \\ 0 & -U''_{1/2} \partial_x \end{pmatrix}. \]

Note that \( H_{1/2} \) has a slow spatial dependence on \( z \) through the deterministic \( U_0(z) \), in addition to the fast dependence on \( z/\epsilon \) through the random \( U_{1/2}(\hat{z}) \). As mentioned earlier, the transport equation (2.56) remains valid in the presence of such a dependence. Replacing the operators \( H_0 \) and \( J_0 \) by their matrix form and taking the Fourier transform of \( H_{1/2} \) and \( J_{1/2} \) with respect to \( x/\epsilon \) gives

\[ H_0(x, ik) = \begin{pmatrix} N^{-2} & 0 \\ U_0(z) N^{-2} (k_1^2 + k_2^2)^{-1} \end{pmatrix}, \]

\[ \hat{H}_{1/2}(l, k) = \frac{2\pi \hat{U}_{1/2}(l_2) \delta(l_1)}{N^2} \begin{pmatrix} -l_2^2 N^{-2} U_0(z) & 1 \\ 1 & 0 \end{pmatrix}, \]

\[ J_0(x, ik) = -i N^2 k_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \hat{J}_{1/2}(l, k) = -2\pi i k_1 l_2^2 \delta(l_1) \hat{U}_{1/2}(l_2) \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \]

Now \( L_0 = J_0 H_0 \) and the eigenvalues \( \omega \) (of \( L_0 \)) are found by calculating the characteristic equation as follows:

\[ \det(L_0 + \omega I) = 0 \implies \det \begin{pmatrix} -i k_1 U_0(z) + i \omega & -i N^2 k_1 (k_1^2 + k_2^2)^{-1} \\ -i k_1 & -i k_1 U_0(z) + i \omega \end{pmatrix} = 0 \]

\[ \implies \omega = \pm N k_1 \left( \frac{1}{(k_1^2 + k_2^2)^{1/2}} + k_1 U_0 \right), \]

which is the well-known dispersion relation for two-dimensional gravity waves in a slowly varying background wind (see Section 1.5.1 or Gill 1982). Two wave modes are present here, with the negative sign representing upward propagating waves and the positive sign representing downward propagating waves. If we now calculate the derivatives of \( \omega \) and substitute into the deterministic part of the transport equation (2.56), then we obtain two equations for the two wave modes

\[ \partial_t a(s)(x, k, t) \pm |k|^{-3} N k_2 (k_2 \partial_{x_1} - k_1 \partial_{x_2}) a(s)(x, k, t) + U_0(z) \partial_{x_1} a(s)(x, k, t) - U_0'(z) \partial_{k_2} a(s)(x, k, t) = \sum \int a(s,t)(x, k', t) a(t)(x, k', t) dk' - \Sigma(s) a(s)(x, k, t), \]

where \( k = (k_1, k_2) \). The two wave modes are represented by \( s = + \), corresponding to the positive sign in the dispersion relation and \( s = - \), corresponding to the negative sign. We now need to calculate the scattering part of the transport equation, i.e., \( \sigma(s,t) \) which is given by (2.58).

The eigenvectors are determined from the equations \((L_0 + \omega I)e(s) = 0\) and \(\hat{\epsilon}(s) = H_0 e(s)\) and the normalisation condition \(\hat{\epsilon}^*(s) e(t) = \delta_{st}\). A simple calculation leads
\[ e(\pm) = \frac{1}{2^{1/2} (1 + U_0 |k| N^{-1})^{1/2}} \left( \frac{N}{\pm |k|} \right) \quad \text{and} \quad \tilde{e}(\pm) = \frac{(1 + U_0 |k| N^{-1})^{1/2}}{2^{1/2}} \left( \frac{N^{-1}}{\pm |k|^{-1}} \right). \]

Note that we will assume that \(|U_0 |k| N^{-1}| < 1\), which ensures that there are no critical levels. Also needed for the calculation of the scattering cross-sections are the correlation 4-tensors defined in (2.11)–(2.13). With the notation

\[ \delta_{\alpha_1 \alpha_2 \ldots \alpha_n} = \begin{cases} 1 & \alpha_1 = \alpha_2 = \ldots = \alpha_n \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \nu_{\alpha_1 \alpha_2} = \begin{cases} 1 & \alpha_1 \neq \alpha_2 \\ 0 & \text{otherwise} \end{cases} \]

and denoting by \( R(l_2) \) the power spectrum of the \( U_{1/2} \), with

\[ R(l_2) = \langle \hat{U}_{1/2}(l_2) \hat{U}_{1/2}(-l_2) \rangle, \]

these tensors can be written

\[
\begin{align*}
J^{\alpha \beta \gamma \delta}(l, k, n) &= 4\pi^2 \delta(l_1) R(l_2) k^2 \delta_{\alpha \beta} \delta_{\gamma \delta}, \\
H^{\alpha \beta \gamma \delta}(l, k, n) &= 4\pi^2 N^{-4} \delta(l_1) R(l_2) [H^2 N^{-4} U_0^2(\tau) \delta_{\alpha \beta} \delta_{\gamma \delta} + \nu_{\alpha \beta} \nu_{\gamma \delta} \\
&\quad - l_2 N^{-2} U_0(\tau) (\delta_{\alpha \gamma} \nu_{\delta \tau} + \nu_{\alpha \beta} \delta_{\gamma \delta}) ] \quad \text{and} \\
K^{\alpha \beta \gamma \delta}(l, k, n) &= 4\pi^2 N^{-2} \delta(l_1) R(l_2) k^2 \delta_{\alpha \beta} \delta_{\gamma \delta} \left[ l_2 N^{-2} U_0(\tau) \delta_{\gamma \delta} - \nu_{\gamma \delta} \right].
\end{align*}
\]

It is now a straightforward, if tedious, matter to compute the scattering cross-sections. We first note that there is no scattering between the + and – modes because their frequencies never coincide. Thus the only non-vanishing scattering coefficients are \( \sigma_{(+,+)} \) and \( \sigma_{(-,-)} \). From (2.58), these are found in the form

\[
\sigma_{(+,\pm)}(x, k, k') = 2\pi^2 \delta(k_1 - k'_1) R(k_2 - k'_2) \delta[\omega_\pm(x, k) - \omega_\pm(x, k')] \times
\left\{ \begin{align*}
\omega_\pm^2(x, k) N^{-4} U_0^2(k_2 - k'_2)^2 + N^{-2} (|k| + |k'|)^2 &\pm 2 N^{-3} U_0(|k| + |k'|)(k_2 - k'_2)^2 \\
(1 \pm U_0 |k| N^{-1})(1 \pm U_0 |k'| N^{-1}) &|k|^2 |k'|^2 \\
- \omega_\pm(x, k) &\left( (k_1 + k'_1)(k_2 - k'_2)^2 [N^{-2} U_0(k_2 - k'_2)^2 + N^{-1} (|k| + |k'|)] \right) \bigg| |k||k'|
\end{align*} \right\}.
\]

Note that \( \sigma(x, k, k') = \sigma(x, k', k) \). Using the definition of the frequency and the equality \( \omega_\pm(x, k) = \omega_\pm(x, k') \), this simplifies

\[
\sigma_{(+,\pm)}(x, k, k') = \frac{2\pi^3 k_1^2}{|k|^2 |k'|^2} \delta(k_1 - k'_1) R(k_2 - k'_2) \delta[\omega_\pm(x, k) - \omega_\pm(x, k')] \times
\left[ (k_2 - k'_2)^2 - 2|k'|(|k| + |k'|)(k_2 - k'_2)^2 + |k||k'|(|k| + |k'|)^2 \right].
\]

This can be simplified considerably by noticing that (using properties in Appendix A.4)

\[
\delta(k_1 - k'_1) \delta[\omega_\pm(x, k) - \omega_\pm(x, k')] = \frac{|k|^3}{k_1 k_2 N} \delta(k_1 - k'_1) [\delta(k_2 - k'_2) + \delta(k_2 + k'_2)], \quad (2.70)
\]
where we have made a change of variable in the $\delta$ function by using

$$
\delta[\omega(x, k) - \omega(x, k')] = \delta(k_2 + k_2') \left| \frac{d\omega(x, k)}{dk_2} \right|^{-1}.
$$

The corresponding factor $\delta(k_1 - k_1')\delta(k_2 - k_2')$ in (2.70) implies that $R(k_2 - k_2') = 0$ and so contributes nothing to the scattering cross section. In which case we obtain

$$
\sigma_{(\pm, \pm)}(x, k, k') = \frac{8\pi^3 k_1(k_2^2 - k_1^2)^2}{k_2|k|N}\delta(k_1 - k_1')\delta(k_2 + k_2')R(2k_2).
$$

This indicates that the only possible scattering is between modes whose wavevectors are symmetric with respect to the horizontal and, correspondingly, whose group velocities are symmetric with respect to the vertical. And so we arrive at the final transport equation for gravity waves propagating through a random background wind

$$
\partial_t a(s)(x, k, t) + ik^3 k_2(k_2 \partial_{k_1} - k_1 \partial_{k_2}) a(s)(x, k, t) + U_0(z) \partial_{k} a(s)(x, k, t) -
\int \frac{k_1 U_0' (z) \partial_{k} a(s)(x, k, t) = \frac{8\pi^3 k_1(k_2^2 - k_1^2)^2}{k_2|k|N}R(2k_2) \sum s (a(s)(x, k, t) - a(s)(x, k, t)),
$$

where $k^p = (k_1, -k_2)$ and the scattering terms have been simplified by using (A.9).

### 2.6 Discussion and summary

There are several developments and extensions to the theory presented in this chapter that it may be interesting to consider, in particular localization, the diffusion approximation and time-dependent media. These are now discussed, followed by a summary of the chapter.

**Localization:** in random media, waves may become exponentially trapped in the direction of propagation, i.e., after an infinitely long time the waves can be found in a finite region, within a finite distance of the starting point, with non-zero probability. This is known as Anderson localization, (see Molchanov (1991) for a detailed review of Anderson localization and its applications). Whether localization takes place or not is dependent on the isotropy of the inhomogeneities of the medium and on their magnitude (Ryzhik et al. 1996). If the medium is isotropic (or weakly anisotropic) and the fluctuations of inhomogeneities are weak then both wave propagation and localization are possible. If the fluctuations are strong then even if the medium is completely isotropic, all wave motion is localised and there is no wave propagation. In strongly anisotropic media such as the atmosphere, with a one-dimensional background shear (as described in Section 2.5), localization can take place even for weak fluctuations.

With the assumptions made in this chapter, localization does not appear explicitly. It

---

5First proposed by P.W. Anderson (1958), to show that quantum mechanical motion of a particle in a random potential can be localised in space, turning a conductor into an insulator.
might be that the localization length is asymptotically large in the limit $\epsilon \to 0$, so that the effect is not captured by the multiple-scale approach employed. This point may be worth investigating in the context of a Hamiltonian system by further asymptotic analysis.

**Diffusion approximation:** at propagation distances that are long compared to a typical transport mean free path $|\nabla_k \omega|/\Sigma$ and propagation times that are long compared to a typical transport mean free time $1/\Sigma$, (2.1) can be reduced to a diffusion equation in space, with wavevector-dependent diffusivity (Ryzhik et al. 1996). The derivation of this approximation for several systems of the form (2.2) is given in Ryzhik et al. (1996). A similar derivation could probably be carried out for the class of Hamiltonian systems given in this chapter.

**Time-dependent media:** in this chapter and most previous works on the topic, only time-independent media have been considered, in which case the wave-energy density $a(x, k, t)$ satisfies the conservation equation (2.1). However, it may be of interest to consider background media whose properties vary slowly in time (Vanneste & Shepherd 1999). If the time-scale separation between medium and waves is characterised by the small parameter $\epsilon$, then it is expected that a conservation law of the form (2.1) holds, but only approximately, with an $O(\epsilon)$ error. In this case, $a(x, k, t)$ would be the density of wave action rather than of wave energy. A proof of this could be given for the Hamiltonian systems of this chapter.

**Polarisation:** In the presence of polarised waves, i.e., waves for which the eigenvalues satisfying (2.34) have multiplicity greater than one, the transport equation (2.56) is not complete. When polarisation exists additional terms must be added to the transport equation to capture the cross polarisation effects. This should be fairly easy to do within the context of this chapter (cf. Ryzhik et al. 1996).

**Numerical simulations:** the transport equation derived for gravity waves (2.68) (or indeed any other system) could actually be used in numerical simulations of gravity waves. The transport equation is entirely deterministic, which is a great advantage as random effects can be modelled without having to undertake random simulations. A numerical simulation could also be used to verify the results of the previous section, however, it is not clear what the practical cost of doing this might be.

**Rotation:** the effect of rotation on gravity waves could be included by the addition of another term in the basic equations. This will complicate the analysis slightly, but not
impossibly so. It might be interesting to see what effect rotation has on gravity waves.

Rossby waves: numerous works have been undertaken to model the interaction of oceanic Rossby waves with a random two-dimensional bottom topography (e.g. Klyatskin et al. 1998, Vanneste 2003). Scaling methods have been used to examine this interaction asymptotically. However, so far none have considered the scattering scaling considered in this chapter. Powell & Vanneste (2004) obtain a transport equation of the type (2.56) for this problem.

Summary: we have studied waves propagating through a random inhomogeneous media whose deterministic properties vary slowly compared with typical wavelengths. The most general form of linear system conserving an energy like quantity is considered: a general linear non-canonical Hamiltonian system. We consider the intermediate regime where the random perturbations of the media have spatial scale comparable to the wavelengths. Based on the work by Ryzhik et al. (1996) and Guo & Wang (1999), we use multiple scale expansions for the Wigner distribution to formulate transport equations that describe the evolution of a scalar wave energy density function for the waves. The effects of scattering of the wavenumber by the random inhomogeneities are included in this equation. As an example of the application of this theory, we consider the system of atmospheric gravity waves propagating through a random background wind in two dimensions. We have also discussed some further developments that could be made to the existing theory presented in this chapter: in particular, Anderson localization, the diffusion approximation and time-dependent media.
Chapter 3

Doppler spreading of gravity waves

Big waves have little waves that feed on deformation, and little waves have lesser waves, to turbulent dissipation.
(E. Dewan 1979)

3.1 Introduction

The large-scale circulation of the middle atmosphere is partly controlled by the driving effect that breaking gravity waves have on the vertical transfer of energy and momentum (or pseudomomentum, see Section 1.5), from low levels in the troposphere to higher levels in the stratosphere. An example where gravity waves have a significant effect is the Brewer–Dobson circulation (see Section 1.4). Middle atmosphere computer models to predict the circulation generally have too large a grid scale to incorporate the effect of gravity waves. It is therefore important to be able to parameterize the effect of gravity waves within the models. For this, it is essential to understand the mechanisms behind gravity wave propagation, breaking, and the effect of the breaking on the large-scale flow. There are two different ways in which gravity waves are modeled:

Empirical models: observations of temperature and velocity profiles in the atmosphere have been obtained by using conceptually different experiments, for example, balloons, radar and lidar (see Section 1.3 for a description of these different techniques). These observations are used to build empirical models of gravity wave activity. They use parameters such as: energy level (the scale of the energy spectra); central wavenumbers (also known as cut-off wavenumbers – the value of the vertical wavenumber that marks a transition between two areas of the spectra, displaying different characteristics) and spectral indices (power-law exponents for physical parameters that might be assumed to obey a power law.) The power calculated by fitting the model to real data). See VanZandt (1982) and Sidi (1988) for examples of empirical models. These models aim
to synthesize observations by using these parameters. They are useful in parameteri-
zations of the long term effects of wave–mean flow interactions, but they provide little
information on the mechanisms involved in energy transfer.

A feature of empirical models well supported by observations, is the $m^{-3}$ spectral
tail of the energy spectrum; that is, for large values of vertical wavenumber $m$, the en-
ergy spectrum scales like the power law $m^{-3}$. It has been observed that this tail varies
little with altitude, season and geographical location, i.e., it is universal (see Section
1.3). Examples of some empirical spectra that demonstrate the universality of the tail
are shown in Figure 3.1.

Theoretical models: several conceptually different models have been developed to
help understand the form of the observed spectrum and in particular, the $m^{-3}$ spectral
tail. Dewan & Good (1986) model saturation (see Section 1.6.3) by means of shear
instabilities. For example, if the energy spectrum $E$ depends on saturation due to shear
instabilities then $E$ must depend on the critical $R_i$ number condition $U_b^2 = 4N^2$ (see
Section 1.6.2). In the case of a background wind composed of superimposed sine waves,
the only other variable $E$ will depend on is the vertical wavenumber $m$. If we require
that the dimensions of $E$ $(L^3 T^{-2})$ and $m^3 U_b^2$ are equivalent then we obtain the scaling
$E \sim m^{-3} N^2$.

Weinstock (1990) and Gardner (1994) assume that the non-linear damping effects
that molecular viscosity, turbulence and off resonant wave–wave interactions (effects
that are neglected in linear theory) have on gravity waves can be modeled by a “dif-
fusivity” parameter which increases with altitude. Gravity waves are assumed to grow
exponentially with increasing altitude due to density decrease until they are removed
by diffusive damping. Gardner obtains an energy spectrum, scaling at large $m$ like,
$4\pi N^2/\text{Ri} \ln (N/\Omega)m^3$ where quantities are defined in Chapter 1.

Another mechanism that can be invoked to explain the form of the energy spectrum
is Doppler spreading which will be the focus of this chapter and will now be discussed.

3.1.1 Doppler Spreading

In a time and/or space varying background wind (e.g. produced by long waves), the
frequency and wavenumber of short waves are modified by Doppler shifting. The back-
ground wind is assumed to be weakly inhomogeneous and weakly unsteady, i.e., nearly
uniform. When an ensemble of waves, encountering different background conditions,
is considered, there is therefore a spread of the energy spectrum with wavenumber, as
altitude increases.

The short–long wave interaction implies a scale separation and hence the WKB
approximation (see Section 1.5.1, also known as eikonal theory) can be made. In this
Figure 3.1: Example of an empirical spectra. The figure reproduced from Smith et al. (1987) shows plots of data collected at various levels of the atmosphere (e.g. Smith et al. 1983, Dewan et al. 1984, Endlich et al. 1969). The theoretical limit $N^2/2m^3$ predicted using a saturation theory is shown by a broken line. This figure has been used in many research articles to show the universality of the spectral shape between different regions of the atmosphere. However, it can be argued that this figure is not very convincing. For example, the troposphere and thermosphere spectrums are not very straight. Further comments are made concerning the universality of the spectral tail in 3.9.
model, action and energy are advected along rays (wave paths in \((x, k)\) space, see Section 1.5.1) in the time and space varying background produced by the long waves.

Eikonal theory was first proposed for oceanic gravity waves by Henyey & Pomphrey (1983). Eikonal theory has been used in the atmosphere by, for example, Dunkerton (1984) and Dunkerton & Butchart (1984). An atmosphere constructed from actual data measurements was used to describe the effects of planetary waves\(^1\) and sudden warmings\(^2\) on the propagation and dissipation of gravity waves in the stratosphere.

**Stochastic models:** Doppler spreading through a random wind was first tackled by Hines (1991\(a\)) who considered individual waves propagating upward through a wind produced statistically by all the waves. The model uses a Gaussian distribution of wave-induced winds. It treats the Doppler shift imposed on each wave as if it were propagating through a stationary wind with the same statistics. For a given arbitrary input spectra, Hines derives energy spectra at a higher altitude and claims to produce the \(m^{-3}\) spectral tail. This work has been well quoted by many papers, (e.g. Staquet & Sommeria 2002). The result however is not as compelling as it might first appear. He predicts an \(m^{-3}\) spectra for intermediate values of \(m\) but a \(m^{-1}\) spectra for large \(m\). In fact there is a mistake in his paper. A slight correction to this theory will produce the \(m^{-3}\) tail for large \(m\), and a critique of this work together with a correction is presented in Appendix B; see also Broutman & Grimshaw (2003).

Hines (1992) derived an analytical form for the cutoff wavenumber (see Section on empirical models). These cutoff numbers were used in Hines (1996\(a,b\)) to derive parameterizations of gravity wave momentum deposition in the atmosphere, which have performed well (e.g. Mengel 1995).

In Souprayen et al. (2001) an approach is taken in which only some aspects of the wave–wave and wave–mean flow interaction, which are simple enough for an analytic formulation of the spectrum to be derived, are considered. A one-dimensional system is considered with a horizontal background wind that is a function of altitude only. The shear of the background wind is modeled by a random Ornstein–Uhlenbeck (O-U) process (e.g. Gardiner (1985) and Appendix C). A system of coupled stochastic differential equations (SDEs) is obtained. Firstly a non-rotating hydrostatic approximation is made and the SDEs are solved analytically and numerically for the energy spectra in

---

\(^1\)Also known as Rossby waves, after Carl-Gustav Rossby (1898–1957) an influential Swedish meteorologist. These are waves with frequencies considerably lower than gravity waves and driven by the fact that the potential vorticity on the curved surface of the earth is not a constant, but varies with latitude (e.g. Gill (1982)).

\(^2\)In a major event of this type, the temperature at about the 30 Km altitude level over the North Pole may increase by about 40–60 K in less than a week. This is caused by a negative feedback mechanism where the polar vortex breaks down due to negative planetary wave drag which in turn makes the pole more susceptible to planetary waves. In other words the vortex becomes unstable in an adiabatically changing environment. The increased downwelling causes a sudden warming.
the infinite $m$ domain (i.e., no upper boundary on $m$). The energy spectra display an $m^{-3}$ tail and only weak dependence on altitude at the upper levels.

Secondly an atmosphere with a decreasing density is considered to model the atmosphere more precisely. An exponential decrease in density is included in the O-U process. Again this system is solved analytically and numerically in the infinite $m$ domain. This time the energy spectra show a strong dependence on altitude, particularly in the low wavenumber part of the spectra. The central wavenumber decreases exponentially with altitude, while the $m^{-3}$ tail remains fixed. This is consistent with observations (e.g. Smith et al. 1987). An extension is made to the Boussinesq approximation and the results obtained are similar.

The breaking of atmospheric gravity waves is modeled by prescribing a cutoff limit for the vertical wavenumber at which the wave is assumed to transfer its momentum to the background flow. The wavenumber $m$ is considered a random process in a finite domain and so the cutoff limit is modeled as an absorbing boundary in $m$. A simple analytical expression for the solution in closed form is not possible. A complicated series expansion is presented in Marshall & Watson (1987) and Selinger & Titulaer (1984) and a truncated version of these expansions is used in Souprayen et al. (2001) to obtain an asymptotic solution to the bounded problem.

The results of Souprayen et al. (2001) are complemented and confirmed in Herzog et al. (2002) by using numerical simulations in real three-dimensional time-varying background winds. The background wind and temperature field for the troposphere and lower stratosphere were provided by the European Center for Medium-Range Weather Forecasts (ECMWF). It was shown that for various source spectra, (made up from randomly launched wave packets) at lower tropospheric altitudes, the energy spectra predicted at higher altitudes are similar to those inferred from observations. The asymptotic $m^{-3}$ tail and the variation with altitude of the central wavenumber were recovered. The model identified the vertical shear of the horizontal wind to be an important parameter. However any estimate of energy transfer cannot be regarded as reliable due to the poorly resolved nature of the ECMWF data. The results in Herzog et al. (2002) have been obtained without including any wave breaking effects due to large vertical wavenumber which is an obvious lack of realism.

3.1.2 Chapter description

Several processes (e.g. wave breaking, resonant triad interaction and Doppler spreading) have been proposed to explain the $m^{-3}$ spectra tail. All of these processes have been studied separately and in detail at different spectral and altitude ranges (e.g. Fritts & Alexander 2003). However a convincing theory for a generic shape of the spectrum is still missing.
The objectives of this chapter are to derive simple closed form expressions for the energy spectra and for the wave-induced force of gravity waves propagating in a random background wind. To this end, we consider a modification to the model in Souprayen et al. (2001) and represent the velocity of the wind (rather than its shear) by an O-U process. This enables us to obtain simple closed form solutions both in a bounded and unbounded wavenumber domain.

The plan of this chapter is as follows. Section 3.2 discusses the application of a one-dimensional time-independent stochastic model for an atmosphere that is realistically three-dimensional and time-dependent. Section 3.3 describes the basic non-dimensional equations and boundary conditions that will be used to model atmospheric gravity wave propagation. Section 3.4 describes analytic solutions for probability density functions for finding a wave-packet at a particular altitude and a particular wavenumber. We treat both the cases of infinite-domain (no upper boundary for \( m \) in spectral space) and finite-domain (with an upper boundary for \( m \) in spectral space). The infinite domain can be integrated directly to find a PDF. The finite-domain problem requires the solution of a transcendental eigenvalue equation and it is shown that to good approximation only the first eigenfunction needs to be used. Asymptotics are given for an approximation to the value of the first eigenvalue in section 3.4.3. Section 3.5 describes a numerical solution for the PDF in the infinite and finite domains. Benchmark tests are made between numerical and analytic solutions and are shown in the following section. Section 3.6 obtains expressions for the energy power spectral density curves. Physical characteristics that are found from observed energy spectra are present in the theoretical energy spectra derived here. Among these characteristics is the existence of the \( m^{-3} \) spectral tail. Comparisons between numerical and analytic solutions are made. Analytical expressions for the central wavenumber of the energy spectra are also obtained. Asymptotics are used to derive simplified analytic expressions. Section 3.7 derives an analytic expression for the wave-induced force (the force that the waves induce on the mean flow, see section 1.4) as a function of altitude for various different initial conditions. Section 3.8 investigates density variations in the atmosphere and achieves similar results to previous sections using a slowly varying density with altitude. Section 3.9 contains a discussion and summary of the results and also suggests further work that could be undertaken.

3.2 Stochastic model

What we wish to model is the propagation of rays through a three-dimensional, time-dependent background wind. However, if analytical solutions are to be derived in closed form, then this is sadly not possible without additional hypotheses. Figure 3.2 shows an example of ray propagation through a time dependent wind in one spatial dimension.
Figure 3.2: Schematic showing rays launched at equal time intervals propagating through a background wind that is time-dependent. This is equivalent to rays propagating through a time-independent random wind where each ray corresponds to a different realization of the wind.

where rays are launched at equal time intervals. If the propagation times are short compared to the typical time scale of the wind, this can be modeled by a time-independent random wind, where each ray propagates through a different realization of the wind. The same argument can be applied for using a horizontally-independent random wind to model a horizontally-dependent wind. In this scenario, rays arriving at particular altitude have initiated from many different places and times. Figure 3.3 (reproduced from Hertzog et al. (2002)) illustrates this point well. Thus, we model propagation through a time-dependent, horizontally varying wind by a time-independent, columnar stochastic model. Each ray samples a different realization of the wind.

What we are ultimately interested in is the amount of action (defined in Section 2.1) per unit altitude at a particular time with a particular wavenumber $A(m; z)$. Let us suppose that a number of wave-packets are launched. For each of them, the conservation of action $A_0(z, t)$ states that

$$\frac{\partial A_0}{\partial t} + \frac{\partial}{\partial z}(c_g A_0) = 0,$$

where $c_g$ is the vertical group velocity. This is a consequence of the transport equation without scattering derived in Chapter 2, with $a(z, m, t) = A_0(z, t)\delta(m - m(t))$ (equation (2.60)). Here, $c_g$ does not depend on $z$ and so

$$\frac{\partial A_0}{\partial t} + c_g \frac{\partial A_0}{\partial z} = 0,$$

implying that a ray conserves its action along the ray path. We define the density of rays $\rho_r$, to be the number of rays at a particular time over a unit altitude $\Delta z$. So

$$A(m; z) = A_0\rho_r p(m; z),$$

where $p(m; z)$ is the probability of finding a ray at a particular altitude with a particular wavenumber. The density of rays is inversely proportional to the initial separation of the rays $\Delta t$ and the average group velocity of rays. This is because the magnitude of
the group velocity controls the steepness of the rays and hence the number of rays found in a particular altitude interval (see Figure 3.4). Hence, $\rho_r \sim 1/\Delta t(c_g)$. Provided the initial ray separation is constant and the rays are densely packed then $\rho_r$ is a constant.

If rays are launched with an initial wavenumber $m_0$ then the total action may contain contributions from rays with many different initial wavenumbers and so

$$ A(m; z) = \rho_r \Delta z \int A_0 p(m; z|m_0, z_0) \, dm_0. $$

Contrast this scenario (in which each wave-packet experiences a different background wind), to the scenario where rays propagate through a time-independent wind, but where each wave-packet experiences the same realization of the wind, (see Figure 3.4). In this scenario the density of rays is not constant, but inversely proportional to $c_g$, (see Figure 3.4. In this case the action is given by

$$ A(m; z) = \frac{\rho_r \Delta z A_0}{c_g(m)} p(m; z). $$

The difference is crucial when it comes to determining the $m$-dependence of the action and energy spectra.

Figure 3.3: Rays traced in a three-dimensional model using an observed velocity field. At a particular altitude, it is possible to find rays that have originated from many different places. Reproduced from Hertzog et al. (2002).
Figure 3.4: Schematic showing rays launched at equal time intervals propagating through a background wind that is time-independent and each ray experiences the same realization of the wind.

### 3.3 Basic equations

We now derive a single stochastic differential equation that describes gravity waves propagating through a random background wind. Consider gravity waves in one dimension (the vertical) with a purely horizontal wind $U(z)$ that is a random function of altitude only. We assume that the wind is slowly varying in altitude compared to the vertical wavelength. Thus, there is a scale separation between the the background wind and the propagating waves.\(^3\) This allows the WKB approximation to be used (see Section 1.5.1). The dispersion relation is given by

\[
\omega = \dot{\omega} + kU, \\
\text{where } \dot{\omega} = -\frac{kN}{\sqrt{k^2 + m^2}}.
\]

A minus sign, which ensures that the vertical group velocity is positive for $k > 0$, has been chosen here as we are considering upward propagating waves.

As in Souprayen et al. (2001) $N$ is taken to be a constant. The energy spectra obtained in Souprayen et al. (2001) by using a constant $N$ did not differ significantly from the energy spectra obtained by Hertzog et al. (2002) where real three dimensional winds with variations of $N$ were used. This justifies the assumption made here. We consider a linear model in which the non-linear feedback of the waves on the background flow and the wave–wave interactions are neglected. This is a greatly simplified model of the real atmosphere, but it will be seen that it provides good results to first order.

The ray-tracing equations describing the evolution over time of wave packets in $(x,k)$ space are (see Section 1.5.1 or Lighthill (1978) for more details).

\[
\frac{dx}{dt} = \frac{\partial \omega}{\partial k} = -m^2 N (k^2 + m^2)^{-3/2} + U, \quad (3.1) \\
\frac{dz}{dt} = \frac{\partial \omega}{\partial m} = k m N (k^2 + m^2)^{-3/2}, \quad (3.2)
\]

\(^3\)This step may perhaps be viewed as slightly heuristic since the wind is, in part, determined by the gravity waves themselves. However experience tells us that this approximation yields good results and hence must be taken on trust.
\[
\begin{align*}
\frac{dm}{dt} &= -\frac{\partial \omega}{\partial z} = -kU'(z) \quad (3.3) \\
\text{and} \quad \frac{dk}{dt} &= -\frac{\partial \omega}{\partial x} = 0, \quad (3.4)
\end{align*}
\]

where the prime denotes a derivative with respect to \( z \). Equation (3.4) implies that the horizontal wavenumber \( k \) is constant, \( k = k_0 \). For a steady background flow the frequency \( \omega \) is constant along rays since by using (3.1)-(3.4)

\[
\frac{d\omega}{dt} = \frac{\partial \omega}{\partial k} \frac{dk}{dt} + \frac{\partial \omega}{\partial m} \frac{dm}{dt} + \frac{\partial \omega}{\partial z} \frac{dz}{dt} = 0.
\]

That is, following a ray, frequency observed will be constant.

An Ornstein–Uhlenbeck process (e.g. Gardiner 1985 and Appendix C) is chosen to model the background wind. This is the simplest Gaussian process with a finite correlation length; it is obtained by adding a linear drift term to a Wiener process. The stochastic differential equation

\[
dU(z) = -\gamma U(z)dz + \sqrt{\beta}dW(z) \quad (3.5)
\]

is satisfied by the wind velocity, where \( \gamma \) is the inverse correlation length, \( W \) is a Wiener process, and \( \sigma^2 = \beta/2\gamma \) is the variance of the wind (see Appendix C or Gardiner 1985). The inverse correlation length \( \gamma \) of an Ornstein–Uhlenbeck process is defined by (see Section C or Gardiner 1985)

\[
E[U(z)U(z + \delta)] = \sigma^2 e^{-\gamma|\delta|},
\]

where \( E \) denotes the expected value. The physical significance of the correlation length is illustrated in Figure 3.5.

Values of these parameters that will be used in later calculations will be close to those used by Souprayen et al. (2001). Specifically, we choose, \( N = 0.02 \text{ s}^{-1} \) (see Section 1.3); \( \gamma = 1/400 \text{ m}^{-1} \) which is typical of large scale atmospheric flows and \( \sigma = 4 \text{ m s}^{-1} \). Both \( \gamma \) and \( \sigma \) are consistent with a dynamically stable atmosphere with Richardson number \( \text{Ri} \geq O(1) \) (see Section 1.6.2). Combining (3.2) and (3.3), the time dependence can be eliminated to obtain

\[
\frac{dm}{mN} = -\frac{U'(k_0^2 + m^2)^{3/2}}{dz}.
\]

Equations (3.5) and (3.6) form a closed system describing gravity wave propagation in a random background wind. These are now non-dimensionalised using the following scaling arguments.

We wish \( U \) to have the scale of the typical variations in the wind, so we choose the standard deviation of the O-U process as a non-dimensional parameter for \( U \). We will
use the inverse correlation length of the wind as a non-dimensional parameter for $z$.

Scale separation between $U$ and the waves is ensured if

$$\frac{1}{m} \frac{dm}{dz} \ll m.$$  

Using (3.6) and the values of $N$, $\gamma$ and $\sigma$ given above we have

$$\frac{\sigma \gamma}{N} = \frac{1}{2}.$$  

We deem this sufficiently less than 1 for there to be scale separation between the wind and the waves.

For a Wiener process we have the following relation (see Appendix C or Gardiner 1985)

$$E|W(z) - W(z')|^2 = |z - z'|.$$  

Since $z$ is scaled like $\gamma^{-1}$ we must scale $W$ like $\gamma^{-1/2}$ for this relation to hold. We also use an inverse non-dimensional dependent variable $\eta$ given by

$$\eta = \frac{N}{\sqrt{2\sigma}} (k_0^2 + m^2)^{-1/2}. \hspace{1cm} (3.7)$$

Using this and the transformations

$$z = \frac{z}{\gamma}, \hspace{0.5cm} U = \sqrt{2\sigma} \hat{U}, \hspace{0.5cm} W = \frac{1}{\gamma^{1/2}} \hat{W} \hspace{0.5cm} \text{and} \hspace{0.5cm} \beta = 2\sigma^2 \gamma,$$

(3.5) and (3.6) become

$$d\eta = dU \hspace{0.5cm} \text{and} \hspace{0.5cm} dU = -Udz + dW,$$

where the hats have been dropped for convenience. Combining these two equations, we obtain

$$d\eta = -(\eta - \eta_0)dz + dW, \hspace{1cm} (3.8)$$

where $\eta_0$ is an integration constant which is the value $\eta$ takes when $z = 0$ and is given in terms of $m_0$ (the value the vertical wavenumber $m$ takes when $z = 0$, see
Section 1.3), i.e.,
\[ \eta_0 = N(k_0^2 + m_0^2)^{-1/2}/\sqrt{2\sigma}. \]
So we have a single non-dimensional stochastic differential equation to describe gravity waves in a random background wind. Equation (3.8) will now be solved for \( m \) in the infinite and the finite domains for various different boundaries (explained below). This is done both analytically in Section 3.4 and numerically in Section 3.5. The results are compared to those of Souprayen et al. (2001) in Section 3.6.

As explained above, our stochastic model describes the following scenario of wave propagation. An ensemble of rays are launched with the same initial wave numbers \( k_0 \) and \( m_0 \) at the same initial altitude \( z_0 \). Waves are launched continuously at equal time intervals, with the same distribution of action and hence we have a stationary system. At any altitude that is greater than the initial altitude we can find the probability of a ray being located at this altitude with a specified vertical wavenumber, i.e., \( p(m; z) \).

### 3.3.1 Boundaries in wavenumber space

In linear instability models, the gravity wave breaking mechanism is due to convective instabilities. These occur when the growth in wave amplitude causes heavy fluid to overlay light fluid. For convective instabilities we have the critical condition (see Section 1.6.2)
\[ \frac{\partial \rho_0}{\partial z} + \frac{\partial \rho}{\partial z} = 0. \]  
(3.9)

Now since \( A = \mathcal{E}/\dot{\omega} \), where \( A \) is wave action and is constant then \( \rho \sim (\dot{\omega} A)^{1/2} \). Using (3.9) we can derive a condition for convective instability in terms of the vertical wavenumber, i.e.,
\[ m_c \sim \frac{d\rho_0}{dz} \frac{1}{(\dot{\omega} A)^{1/2}}. \]

Wave breaking is assumed to occur to waves with the shortest vertical scales only and to result in total wave dissipation. This can be modeled by the addition of an absorbing boundary on the vertical wavenumber \( m \) that prevents the propagation of a wave if it reaches a certain value, \( m_c \), say. A schematic of this is shown in Figure 3.6.

So, the picture we have in mind is that of waves being launched at a low level source, propagating upwards through the atmosphere and breaking at some higher level. Wave action is dissipated at this point. In addition, an absorbing boundary at zero vertical wavenumber, \( m_l = 0 \), is added to prevent back reflection of waves. This ensures a one-one correspondence between altitude and time. The case with upper and lower boundaries will be called the finite domain; the case with no upper or lower boundaries will be called the infinite domain. It is worth pointing out at this stage that the solution in the infinite domain will not be correct in a physical sense, since it allows for solutions with \( \eta < 0 \), which when transforming back to \( m \) space would imply an imaginary \( m \). However, it is valuable as it will match other methods of solution for large enough \( \eta \).
Figure 3.6: Schematic showing the mechanism controlling gravity wave breaking in the Doppler spreading model. Waves propagate vertically with their vertical wavenumber $m$ varying with altitude. If $m$ exceeds some critical wavenumber $m_c$ then breaking is assumed to occur and the wave deposits its momentum. $m_l$ is a lower boundary (usually set to be zero) that prevents back reflection of waves. The equivalent scenario in the $(\eta, z)$ plane is also shown.

Large values of the critical boundary $m_c$ in the finite-domain model will not produce significantly different results to those of the infinite domain model. See Section 3.6.2 for a quantification of this large wavenumber limit.

3.4 Analytical solutions of (3.8)

We now calculate PDFs, $p(m; z)$ and $p(\eta; z)$ for (3.8) in both the infinite (Section 3.4.1) and finite domains (Section 3.4.2).

3.4.1 Solution in infinite domain

Substituting $x = \eta - \eta_0$ in (3.8) and integrating gives

$$x(z) = x_0 e^{-z} + e^{-z} \int_0^z \exp(v) dW(v) = e^{-z} \int_0^z \exp(v) dW(v),$$

where $x_0$ is a constant of integration which by definition vanishes. Thus, $x$ is a Gaussian process since it is the integral of $U$, which is Gaussian. Therefore, all that is needed to calculate the probability density function $p(x; z)$ of $x(z)$ are its mean and variance. Now, the mean $\bar{x}$ of $x(z)$ is given by

$$\bar{x} = x_0 e^{-z} = 0,$$

and so the variance takes the form

$$(x - \bar{x})^2 = e^{-2z} \left( \int_0^z e^v dW(v) \right)^2.$$
The variance can be evaluated using the correlation formula for Wiener processes (see Appendix C), and it is found to be

$$\frac{(x - \bar{x})^2}{2} = 1 - e^{-2z}.$$

With the mean and variance, the Gaussian PDF for \( \eta \) is given by (see Appendix C)

$$p(\eta; z) = p(x; z) \left| \frac{dx}{d\eta} \right| = [\pi(1 - e^{-2z})]^{-1/2} \exp \left\{ \frac{-(\eta - \eta_0)^2}{1 - e^{-2z}} \right\}.$$  (3.10)

Hence the probability in \( m \) space is given by

$$p(m; z) = p(\eta; z) \left| \frac{d\eta}{dm} \right| = \frac{\omega m (k_0^2 + m^2)^{-3/2}}{\sqrt{\pi}(1 - e^{-2z})^{1/2}} \exp \left\{ \frac{-[a(k_0^2 + m^2)^{-1/2} - \eta_0]^2}{1 - e^{-2z}} \right\},$$  (3.11)

where \( \omega = N/\sqrt{2\sigma} \)  (3.12)

and we have maintained the use of \( \eta_0 \) for simplicity. The PDF in the large \( z \) limit is given by

$$p(m) \approx \frac{\omega m (k_0^2 + m^2)^{-3/2}}{\sqrt{\pi}} \exp \left\{ -[a(m^{-1} - \eta_0)^2] \right\}. $$

In the large \( z \) and large \( m \) limit we obtain

$$p(m) \approx \frac{\omega}{\sqrt{\pi m^2}} \exp \left\{ -[am^{-1} - \eta_0]^2 \right\}. $$

### 3.4.2 Solution in finite domain

Let us now consider absorbing boundaries for small and large \( m \). In this case (3.8) will no longer be directly integrable as in the infinite domain case, and so a different approach is needed to determine \( p(m; z) \). This can be done by solving the Fokker-Plank equation corresponding to (3.8). See Appendix C for the relationship between Fokker-Plank equations and stochastic differential equations. The Fokker-Plank equation reads

$$p_\varepsilon(\eta; z) = \left[ (\eta - \eta_0)p(\eta; z) \right]_\eta + \frac{1}{2} p_{\eta\eta}(\eta; z).$$  (3.13)

We now place absorbing boundaries (e.g. Gardiner 1985) on \( m \), at \( m = m_l \) and \( m = m_c \), some lower and critical higher boundaries respectively. In terms of \( \eta \), these give the boundary conditions

$$p(\eta; z) = 0 \quad \text{and} \quad p(\eta_c; z) = 0.$$  (3.14)

We also have as initial condition at the altitude \( z = z_0 \), that all the waves have the same initial wavenumber \( m_0 \). The equivalent condition in \( \eta \) space is given by

$$p(\eta; z_0) = \delta(\eta - \eta_0).$$  (3.15)
where $\delta$ is the Dirac delta function (see Appendix A.4). To solve (3.13), we separate variables by letting $p(\eta; z) = T(z)\phi(x)$, with $x = \eta - \eta_0$, and obtain the following two equations and boundary conditions

\begin{align*}
T_z &= -\lambda T, \tag{3.16} \\
x\phi_x + \phi + \frac{1}{2}\phi_{xx} &= -\lambda \phi, \quad \phi(x_l) = 0 \quad \text{and} \quad \phi(x_c) = 0, \tag{3.17}
\end{align*}

where $\lambda$ is a positive constant. Equation (3.16) has solution $T(z) = e^{-\lambda z}$. Now, to solve (3.17) we make the change of variables $\phi(x) = e^{-u^2/4}q(u)$, where $u = \sqrt{2}x$ and obtain

\[ q_{uu} - \left( \frac{1}{4}u^2 - \lambda - \frac{1}{2} \right) q = 0, \tag{3.18} \]

which has the standard solution (e.g. Abramowitz & Stegun 1965)

\[ q(u) = C_1 U(-\lambda - 1/2, u) + C_2 V(-\lambda - 1/2, u), \tag{3.19} \]

where $C_1$ and $C_2$ are arbitrary constants, and $U$ and $V$ are the parabolic cylinder functions. Then,

\[ \phi(x) = e^{-x^2/2}[C_1 U(-\lambda - 1/2, \sqrt{2}x) + C_2 V(-\lambda - 1/2, \sqrt{2}x)]. \tag{3.20} \]

Applying the lower boundary condition gives

\[ C_2 = -C_1 \frac{U(-\lambda - 1/2, \sqrt{2}x_l)}{V(-\lambda - 1/2, \sqrt{2}x_l)}. \]

Substituting this back into (3.20) for $\phi$ and applying the upper boundary condition leads to the solvability condition

\[ U(-\lambda - 1/2, \sqrt{2}x_c)V(-\lambda - 1/2, \sqrt{2}x_l) - V(-\lambda - 1/2, \sqrt{2}x_c)U(-\lambda - 1/2, \sqrt{2}x_l) = 0. \tag{3.21} \]

To obtain the eigenvalues $\lambda_i$, we solve this equation using Maple (see Appendix D). The full solution of (3.13) is then given by

\[ p(x; z) = \sum_i K_i \phi_i(x) e^{-\lambda_i z}, \tag{3.22} \]

where $K_i$ are constants to be determined and $\phi_i$ are the eigenfunctions corresponding to the eigenvalues $\lambda_i$, given by (3.20). In order to calculate the constants $K_i$, we consider the solutions $\phi^\dagger$ to the adjoint of (3.17), namely,

\[ -x\phi^\dagger_x + \frac{1}{2}\phi^\dagger_{xx} = -\lambda \phi^\dagger. \]

The solutions to this equation are given by

\[ \phi^\dagger_i(x) = e^{x^2}\phi_i(x) \]
and are orthogonal to the $\phi_i(x)$. With a suitable normalization, we have
\[
\int \phi_i(x)\phi_j^*(x) \, dx = \delta_{ij}. \tag{3.23}
\]
By using the initial condition, (3.15), we find
\[
p(x; z_0) = \sum_i K_i \phi_i(x)e^{-\lambda_i z_0} = \delta(x - x_0).
\]
Multiplying by $\phi^*_j$ and integrating, using (3.23) we find
\[
K_i = e^{\lambda_i z_0} \int \phi^*_i(x)\delta(x - x_0) \, dx = \phi^*_i(x_0)e^{\lambda_i z_0}. \tag{3.24}
\]
For large $z$, only the first eigenvalue makes a significant contribution to the solution provided $\lambda_2 \gg \lambda_1$. The tables in Appendix D show the eigenvalues for various different critical boundaries $m_c$ and initial conditions $m_0$ and $k_0$; it is clear that in general $\lambda_2 \gg \lambda_1$. So the probability density function using only the first eigenvalue value is
\[
p(\eta; z) = K e^{-\left(\eta - \eta_0\right)^2/2} \left[ U(-\lambda - 1/2, \sqrt{2}(\eta - \eta_0)) - A(\eta) V(-\lambda - 1/2, \sqrt{2}(\eta - \eta_0)) \right] e^{-\lambda z}, \tag{3.25}
\]
where $\lambda$ represents the first eigenvalue with the subscript dropped, similarly $K$ represents $K_1$ and $A$ is given by
\[
A(\eta) = \frac{U(-\lambda - 1/2, \sqrt{2}(\eta - \eta_0))}{V(-\lambda - 1/2, \sqrt{2}(\eta - \eta_0))},
\]
a constant defined by the lower boundary. Converting back to $(m, z)$ space, we get
\[
p(m; z) \equiv p(\eta; z) \left| \frac{d\eta}{dm} \right| = \omega K m^2 (m^2 + k_0^2)^{-3/2} \exp \left[ -\frac{1}{2} \left( \frac{\omega}{\sqrt{m^2 + k_0^2}} - \eta_0 \right)^2 \right] \times \tag{3.26}
\]
\[
\times \left[ U\left(-\lambda - 1/2, \frac{\sqrt{2}\omega}{\sqrt{m^2 + k_0^2}} - \sqrt{2}\eta_0 \right) - A\left(\frac{\omega}{\sqrt{m^2 + k_0^2}} \right) V\left(-\lambda - 1/2, \frac{\sqrt{2}\omega}{\sqrt{m^2 + k_0^2}} - \sqrt{2}\eta_0 \right) \right] e^{-\lambda z},
\]
Note that if we want $p(m; z)$ for a small value of $z$ then the full expansion (3.22) is required with the necessary number of eigenfunctions. The smaller the value of $z$ the more eigenfunctions that are necessary.

This solution implies that the probability of finding a ray at a particular altitude is roughly proportional to $e^{-\lambda z}$: the larger the eigenvalue the less likely we are to find a ray at a particular altitude since more rays will have been absorbed by the boundary.

So, we have obtained a closed form solution for the probability of finding a ray at a particular altitude and wavenumber. This is an improvement on the work by Hertzog et al. (2002) where the shear of the wind was modeled by an O-U process and closed form solutions were not obtained.
3.4.3 Approximation for $\lambda$

The first eigenvalue governs the decay rate of the PDF (3.26) for large $z$. It is therefore useful to have an analytic approximation for $\lambda$ that does not rely on solving a transcendental equation involving parabolic cylinder functions.

We will proceed here to define asymptotic expansions for the parabolic cylinder functions $U$ and $V$ which in combination with (3.21) will lead to an expression for $\lambda$.

For large $u$, i.e., $u \gg |\lambda + 1/2|$ and $\lambda = O(1)$ the following asymptotic expansions hold for the parabolic cylinder functions (e.g. Abramowitz & Stegun 1965).

$$U(-\lambda - 1/2, u) = e^{-u^2/4} u^{\lambda} \left[ 1 + \frac{\lambda(1 - \lambda)}{2u^2} + O(1/u^4) \right]$$

(3.27)

and

$$V(-\lambda - 1/2, u) = \sqrt{\frac{2}{\pi}} e^{u^2/4} u^{-(1+\lambda)} \left[ 1 + \frac{(1 + \lambda)(2 + \lambda)}{2u^2} + O(1/u^4) \right].$$

(3.28)

Now since $\lambda$ is defined by the solvability condition (3.21) which contains the functions $U$ and $V$ evaluated at the boundaries $u_l$ and $u_c$, we require, for validity of the expansions, that $|u_l| \gg 1$ and $|u_c| \gg 1$. The range of validity of these assumptions for different initial values $m_0$ and $k_0$ are given in Table 3.1. Whether or not the approximation holds depends on the values of $k_0$, $m_0$ and $m_c$ and the following points can be noted: as the upper boundary $m_c$ gets further away from the initial value $m_0$ the approximation improves; the approximation is not valid for large values of $k_0$, i.e., $k_0 \geq 7 \times 10^{-4}$ m$^{-1}$, irrespective of the values of $m_0$ and $m_c$; the approximation is better for smaller values of $m_0$.

<table>
<thead>
<tr>
<th>$k_0$ (m$^{-1}$)</th>
<th>$m_0$ (m$^{-1}$)</th>
<th>$m_c/m_0 = 2$</th>
<th>$m_c/m_0 = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$u_l$</td>
<td>$u_c$</td>
<td>$u_l$</td>
</tr>
<tr>
<td>$7 \times 10^{-4}$</td>
<td>7 \times 10^{-4}</td>
<td>2.09</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>7 \times 10^{-3}</td>
<td>6.43</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>7 \times 10^{-2}</td>
<td>7.07</td>
<td>0.0036</td>
</tr>
<tr>
<td>$7 \times 10^{-5}$</td>
<td>7 \times 10^{-5}</td>
<td>20.9</td>
<td>18.6</td>
</tr>
<tr>
<td></td>
<td>7 \times 10^{-4}</td>
<td>64.3</td>
<td>3.54</td>
</tr>
<tr>
<td></td>
<td>7 \times 10^{-3}</td>
<td>70.7</td>
<td>0.36</td>
</tr>
<tr>
<td>$7 \times 10^{-6}$</td>
<td>7 \times 10^{-6}</td>
<td>209</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>7 \times 10^{-5}</td>
<td>643</td>
<td>35.4</td>
</tr>
<tr>
<td></td>
<td>7 \times 10^{-4}</td>
<td>707</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Table 3.1: The range of validity of the eigenvalue approximation for different initial values of $m_0$ and $k_0$ (dimensions shown in brackets). The table shows values for upper and lower value of $u$, i.e., $u_c$ and $u_l$ for two different ratios of $m_c/m_0$.

Now since the expansions (3.27) and (3.28) are only valid for $u > 0$ we need the following connection relations to apply the expansions to $u < 0$. These can be found in for example Abramowitz & Stegun (1965).

$$\pi V(-\lambda - 1/2, u) = \Gamma(-\lambda) \left\{ \sin[\pi(-\lambda - /2)] U(-\lambda - 1/2, u) + U(-\lambda - 1/2, -u) \right\}$$

(3.29)
and \( \Gamma(-\lambda)U(-\lambda - 1/2, u) = \pi \sec^2[\pi(-\lambda - 1/2)] \times \{V(-\lambda - 1/2, -u) - \sin[\pi(-\lambda - 1/2)]V(-\lambda - 1/2, u)\}. \) (3.30)

The eigenvalue \( \lambda \) is the smallest root of

\[
U(-\lambda - 1/2, u_c)V(-\lambda - 1/2, u_l) - V(-\lambda - 1/2, u_c)U(-\lambda - 1/2, u_l) = 0.
\]

Substituting (3.27-3.30) into this, taking the series expansion in \( \lambda \) to \( O(\lambda) \), neglecting terms proportional to \( e^{-u^2} \) in the denominator and rearranging in terms of \( \lambda \), we obtain

\[
\lambda \approx \frac{1}{\sqrt{2\pi}} \frac{e^{-u_c^2/4}e^{u_l^2/4}u_c^3(u_l^2 + 1) - e^{u_l^2/4}e^{-u_c^2/4}u_l^3(u_c^2 + 1)}{e^{u_l^2/4}e^{u_c^2/4}(u_c^2 + 1)(u_l^2 + 1)}.
\]

Now since \( |u_l| \gg 1 \) and \( |u_c| \gg 1 \) then this approximation can be simplified to

\[
\lambda \approx \lambda_{ap} = \frac{u_l e^{-u_l^2/2} - u_c e^{-u_c^2/2}}{\sqrt{2\pi}}.
\] (3.31)

Re-expressing \( \lambda \) in terms of \( m_c, m_l \) and \( m_0 \) we get

\[
\lambda_{ap} = \frac{1}{\sqrt{\pi}} \left[ \left( \frac{\omega}{(m_l^2 + k_0^2)^{1/2}} - \eta_0 \right) e^{-\left( \frac{\omega(m_l^2 + k_0^2)^{-1/2} - \eta_0}{\sqrt{2\pi}} \right)^2} - \left( \frac{\omega}{(m_c^2 + k_0^2)^{1/2}} - \eta_0 \right) e^{-\left( \frac{\omega(m_c^2 + k_0^2)^{-1/2} - \eta_0}{\sqrt{2\pi}} \right)^2} \right].
\] (3.32)

where \( \omega \) is given by (3.12). The accuracy of this approximation is illustrated in Figure 3.7. The log of the approximated first eigenvalue \( \lambda_{ap} \), compared to the exact eigenvalue (obtained from by solving (3.4.3) numerically) is plotted against \( m_c/m_0 \), the ratio of the upper boundary to the initial value of vertical wavenumber. It is clear that the approximation gets better as the upper boundary increases in distance from the initial condition.

This approximation for \( \lambda \) can be simplified further for the regimes where the upper boundary \( m_c \) is close to the initial value \( m_0 \), and when \( m_c \) is much greater than \( m_0 \). If we assume that \( m_0 \gg k_0 \), then we obtain the following limits

\[
\lambda \sim \frac{\omega}{\sqrt{\pi}k_0} e^{-\omega^2/k_0^2} \quad \text{for} \quad m_c \downarrow m_0, \quad m_l = 0 \quad \text{and} \quad m_0 \gg k_0
\]

and \( \lambda \sim \frac{\omega}{\sqrt{\pi}m_0} e^{-\omega^2/m_0^2} \quad \text{for} \quad m_c \to \infty, \quad m_l = 0 \quad \text{and} \quad m_0 \gg k_0, \) (3.33)

In Figure 3.8, we have plotted the first eigenvalue \( \lambda \) against the ratio \( m_c/m_0 \) for various different initial conditions. Several things are observed.

• \( \lambda \) only approaches zero in case (1). This is because in cases (2)–(4) \( u_l \) and \( u_c \) are small.

70
Figure 3.7: Comparison of value of first eigenvalue $\lambda$ calculated using (3.4.3) shown by solid line and asymptotic value of first eigenvalue $\lambda_{\text{ap}}$ shown by dashed line. Initial values are: (1) $k_0 = 0.0007$ m$^{-1}$ and $m_0 = 0.0007$ m$^{-1}$; (2) $k_0 = 0.0002$ m$^{-1}$ and $m_0 = 0.0002$ m$^{-1}$.

Figure 3.8: The first eigenvalue against the ratio $m_c/m_0$: the upper boundary wavenumber to the initial wavenumber. Initial values are $k_0 = 0.0007$ m$^{-1}$ and: (1) $m_0 = 0.001$ m$^{-1}$; (2) $m_0 = 0.005$ m$^{-1}$; (3) $m_0 = 0.007$ m$^{-1}$ and (4) $m_0 = 0.05$ m$^{-1}$ the other parameters are $\gamma = 1/400$ m$^{-1}$, $N = 0.02$ s$^{-1}$ and $\sigma = 4$ ms$^{-1}$.

- The limits for $\lambda$ given by (3.33) do not give the correct limits for $m_c \to \infty$ in Figure 3.8. This is because for (3.33) to apply we require that $u_t \gg 1$ and $u_c \gg 1$ as well as $m_0 \gg k_0$, which is not the case in Figure 3.8. In fact when both these conditions hold it becomes very difficult to calculate the exact eigenvalue using (3.21) (see Appendix D) and we must rely on the asymptotic expansion (3.32) or (3.33).

3.5 Numerical solution of (3.8)

Equation (3.8) is also solved numerically. This solution will be used as a check on the analytical results of the previous sections. We use the first order finite-difference
operator (e.g. Conte & de Boor 1983, Burden & Faires 1993)

\[ \frac{(\eta_{i+1} - \eta_i)}{\Delta}, \]

where \( \Delta \) is a small step in \( z \), to approximate the derivative \( \eta_z \). We also need to discretise the derivative of the Wiener process.

Now \( W(s) - W(t) \) is normally distributed with mean zero and variance \( t - s \) (e.g. Breźniak & Zastawniak 2000). So

\[ \text{Var} \left[ \frac{1}{\sqrt{\Delta}} [W(z_{i+1}) - W(z_i)] \right] = \frac{1}{\Delta} (z_{i+1} - z_i) = 1, \]

since the difference in two adjacent \( z \) steps is \( \Delta \). Therefore,

\[ \frac{W(z_{i+1}) - W(z_i)}{\sqrt{\Delta}} \]

is normally distributed with mean 0 and variance 1. We can therefore discretise the derivative of the Wiener process in (3.8) as

\[ \frac{dW}{dz} \approx \frac{1}{\Delta} [W(z_{i+1}) - W(z_i)] = \frac{\sqrt{\Delta}}{\Delta} \left[ \frac{W(z_{i+1}) - W(z_i)}{\sqrt{\Delta}} \right] = \Delta^{-1/2} N[0, 1], \]

where \( N[0, 1] \) is a random number distributed normally with mean 0 and variance 1, (see Appendix C). Hence, we obtain the following finite-difference scheme.

\[ \eta_{i+1} = \eta_i (1 - \Delta) + N[0, 1] \sqrt{\Delta} + \eta_0 \Delta. \quad (3.34) \]

This equation is solved in practice by using a Fortran code. The code is run for a large number of realizations, each starting with initial value \( \eta_0 \). At each step in \( \eta_i \) we test to see whether the upper or lower boundaries specified have been crossed. If a boundary is crossed then the particular realization is stopped. If in any realization the value of the altitude at which we wish to calculate the PDF is reached then the value of \( \eta \) is recorded there.

The PDF for \( p(m; z) \) can be approximated by a histogram with a discrete number of bars, where, as the number of bars increases the approximation approaches the exact PDF. Statistics for the histogram are obtained by a binning procedure where \( N \) bins are spread over the range of wavenumbers, i.e., \((\delta i, \delta (i + 1)) \) for \( i = 0 \ldots N \), where \( m_c < \delta(N + 1) \). If a ray reaches the required altitude then the number of realizations falling in the appropriate bin is incremented by one. The probability of finding a ray at altitude \( z \) with wavenumber \( m \) is approximated by

\[ p(m; z) = \frac{\text{number of realizations with wavenumber in } (\delta i, \delta (i + 1))}{\delta \times \text{total number of realizations}} \]

We wish to be able to identify features of the PDF such as the maximum value, with a reasonable degree of accuracy. To do this we require the number of realisations is much larger than the number of bins.
3.6 Energy PSD curves

The energy density $\mathcal{E}$ possessed by the waves is given by $\mathcal{E} = A \hat{\omega}$, where $A$ is the wave action, (see Section 3.2). The action at a given time over a unit altitude in $(\eta, z)$ space is given by

$$A(\eta; z) = C \int A_0(\eta_0)p(\eta; z) \, d\eta_0,$$

where $C$ is a constant that depends on the density of wavepackets (see Section 3.2); so the energy in $(\eta, z)$ space is given by

$$\mathcal{E}(\eta; z) = C\omega(\eta) \int \frac{\mathcal{E}_0(\eta_0)}{\omega(\eta_0)} p(\eta; z) \, d\eta_0$$  (3.35)

If we let the initial energy be given by $\mathcal{E}_0(\eta_0) = \hat{\omega}_0 \delta(\eta_0 - \hat{\eta}_0)$ so that all rays start with the same initial inverse wavenumber $\hat{\eta}_0$ then we obtain

$$\mathcal{E}(\eta; z) = C\hat{\omega}_0 \frac{\omega(\eta)}{\omega(\eta_0)} p(\eta; z).$$

To obtain the energy in $(m, z)$ space we multiply (3.35) by the Jacobian $d\eta/dm$ and define $\mathcal{E}_0$ by $\mathcal{E}_0(m_0) = \hat{\omega}_0 \delta(m_0 - \hat{m}_0)$, and so we find

$$\mathcal{E}(m; z) = C\hat{\omega}_0 \frac{\omega(m)}{\omega(m_0)} p(m; z).$$  (3.36)

We can now calculate the energy spectrum for the waves in the infinite and finite domains in turn, using both analytical and numerical techniques.

3.6.1 Infinite domain

Using the above relations for the energy, (3.35) and (3.36), and the PDFs (3.10) and (3.11), we find that

$$\mathcal{E}(\eta; z) = \frac{C\mathcal{E}_0\eta}{\eta_0[\pi(1 - e^{-2z})]^{1/2}} \exp \left\{ \frac{-(\eta - \eta_0)^2}{1 - e^{-2z}} \right\}$$  (3.37)

and in terms of dimensional variables,

$$\mathcal{E}(m; z) = \frac{C\mathcal{E}_0m(k_0^2 + m_0^2)^{1/2}}{\sqrt{2\sigma(k_0^2 + m_0^2)^2[\pi(1 - e^{-2z})]^{1/2}}} \exp \left\{ \frac{-[\omega(k_0^2 + m^2)^{-1/2} - \eta_0]^2}{1 - e^{-2z}} \right\}.$$  (3.38)

For large values of $z$ we get

$$\mathcal{E}(m) \simeq \frac{C\mathcal{E}_0m(k_0^2 + m_0^2)^{1/2}}{\sqrt{2\sigma(k_0^2 + m_0^2)^2[\pi(1 - e^{-2z})]^{1/2}}} \exp \left\{ -[\omega(k_0^2 + m^2)^{-1/2} - \eta_0]^2 \right\}.$$

For large $m$, i.e., $m \gg m_0 \gg k_0$ we obtain

$$\mathcal{E}(m) \simeq \frac{C\mathcal{E}_0 m_0 e^{-N^2/(2\sigma^2m_0^2)}}{\sqrt{2\pi \sigma m_0^3 e^{-N^2/(2\sigma^2m_0^2)}}}. $$  (3.39)
We note that the dimensions of $\mathcal{E}$ are

$$\mathcal{E}(m) \sim \frac{\mathcal{E}_0 s^{-1} L^{-1}}{L s^{-1} L^{-3}} \sim \mathcal{E}_0 L.$$  

This is correct since $\mathcal{E}_0$ is an energy, while $\mathcal{E}$ is an energy per unit wavenumber.

Figures 3.9 and 3.10 show analytical curves for the evolution of the energy PSD with altitude for different initial conditions in both linear coordinates and log coordinates. The value of $\mathcal{E}_0$ was chosen the same as in Souprayen et al. (2001) to be $\mathcal{E}_0 = 0.012/m_0^2 \text{m}^2 \text{s}^{-2}$. For the numerical code a large but finite value for the boundary ($m_c = 1000 m_0$) was chosen to simulate the infinite domain. Of course in the finite-domain all rays will eventually hit the boundary but this may only occur at a very large altitude. This point was also observed in Souprayen et al. (2001). So when $m_c$ is sufficiently large compared to the initial value $m_0$, the finite domain is essentially equivalent to the infinite domain except for $m$ small enough for the presence of the lower boundary to cause differences. Using (3.37) we can quantify the value of the central wavenumber $m_*$ (or equivalently $\eta_*$), i.e., the wavenumber for which the energy takes its maximum value as a function of altitude. (We use the variable $\eta$ and not $m$ as the expressions are much simpler to work with.) This is obtained by solving $\mathcal{E}_\eta = 0$ and is given by

$$\eta_* = \frac{1}{2} \left[ \eta_0 + \sqrt{\eta_0^2 + 2 - 2e^{-2z}} \right].$$  

Thus, $m_*$ is given by

$$m_* = \sqrt{\frac{2\omega^2}{\eta_0^2 + 1 - e^{-2z} + \eta_0(\eta_0^2 + 2 - 2e^{-2z})^{1/2}} - k_0^2}.$$  

For large values of altitude, specifically for $e^{-2z} \ll 1$, this tends to the limit

$$m_* \approx \sqrt{\frac{2\omega^2}{\eta_0^2 + 1 + \eta_0(\eta_0^2 + 2)^{1/2}} - k_0^2}. \quad (3.40)$$

**Physical conclusions:** the energy spectrum for the infinite domain solution of the previous section exhibits several features that are consistent with atmospheric observations.

i. Equation (3.38) gives an expression for the energy spectral density and it is clear that for realistic parameter values, the $m^{-3}$ spectral tail for large $m$, inferred by many observations (see Figure 3.1 and Section 1.3), is a feature. This result has been obtained from a very simple model and suggests that Doppler spreading is a good candidate for explaining the $m^{-3}$ spectrum. In fact this model is possibly the most simple model to date that predicts an $m^{-3}$ spectral tail.

ii. The energy level scales like $N$. It could be argued that this is a lack of realism, since many observations predict an $N^2$ scaling, (e.g. Sidi 1988, Smith et al. 1987).

74
Figure 3.9: Energy PSD spectrum as a function of the vertical wavenumber $m$ at several altitudes in the infinite domain, using both analytical and numerical solutions. The upper graph shows a linear coordinate plot and the lower graph shows a logarithmic coordinate plot. Three different altitudes are displayed: (1) $z = 0.1$ (40 m); (2) $z = 0.5$ (200 m) and (3) $z = 10$ (4 km). (4) shows a $m^{-3}$ power law. The following initial values were used: $m_0 = 0.0007$ m$^{-1}$ and $k_0 = 0.0007$ m$^{-1}$ with parameters $\gamma = 1/400$ m$^{-1}$, $N = 0.02$ s$^{-1}$ and $\sigma = 4$ ms$^{-1}$. 

75
Figure 3.10: Same as Figure 3.9 but with $k_0 = 0.00007$ m$^{-1}$
However, there is not much significance in discussing the $N$ dependence in a model with constant $N$. Extra factors involving $N$ can come from $\hat{E}_0$, from the density of wave-packets $C$, etc. At first sight this $N$ scaling also seems to contradict the work of Souprayen et al. (2001) who obtained the following energy spectrum when the velocity shear is an O-U process

$$E(m;z) = \frac{2C\gamma^2 N^2 \hat{E}_0 m(k_0^2 + m^2)^{1/2}}{\sigma^2 \sqrt{\pi f(\gamma z)(k_0^2 + m^2)^2}} \exp \left\{ \frac{-[2\gamma^2 N^2((k_0^2 + m^2)^{-1/2} - \eta_0)^2]}{\sigma^2 f(\gamma z)} \right\},$$

where $f(\gamma z) = 2\gamma z - 3 + 4e^{-\gamma z} - e^{-2\gamma z}$.

However, an error has been made in this calculation. This is evident from the fact that the dimension of $E(m;z)$ has the same dimension as $\hat{E}_0$. If we adapt the argument that gives (3.36) to the model of Souprayen et al. (2001), this should read

$$E(m;z) = \frac{\sqrt{2C\gamma N \hat{E}_0 m(k_0^2 + m^2)^{1/2}}}{\sigma \sqrt{\pi f(\gamma z)(k_0^2 + m^2)^2}} \exp \left\{ \frac{-[2\gamma^2 N^2((k_0^2 + m^2)^{-1/2} - \eta_0)^2]}{\sigma^2 f(\gamma z)} \right\}.$$ 

This now scales correctly, i.e., $E \sim \hat{E}_0 l$.

iii. Equation (3.40) gives an expression for the central wavenumber $m_*$ in the large $z$ limit. This limit is true for values of $z$ greater than about 2, which corresponds to a dimensional altitude of 800m. So there is not really any significant variation of $m_*$ with altitude. This is in contrast to the work by Souprayen et al. (2001) where significant variations in $m_*$ are obtained to higher altitudes. However, the result of Souprayen et al. (2001) has little physical significance: it results from the (rather arbitrary) use of an O-U process for the shear $U'$ with as a consequence, an increase of $U$ with altitude.

Table 3.2 gives values for the large $z$ limit of $m_*$ for various initial values of $k$ and $m$. These values are within the range of those that have been found in observations. See for example: Tsuda (1991), who recorded an average value $m_* = 5 \times 10^{-4} \text{ m}^{-1}$; Allen & Vincent (1995), who recorded values in the range $m_* = 3.5 - 4.8 \times 10^{-4} \text{ m}^{-1}$ and Fritts & Chou (1987) who found $m_* = 5.48 \times 10^{-4} \text{ m}^{-1}$. The values of $m_*$ also agree well with those predicted by the model of Souprayen et al. (2001). They derive a formula for $m_*$ provided $m_0$ is large enough. In that paper, in the large $z$ limit, $m_*$ is written as

$$m_* \approx \frac{2N\gamma}{\sigma} (6\gamma z - 9)^{-1/2}.$$

Here, $\sigma$ and $\gamma$ in this formula are defined for the O-U process modeling the shear of the wind not the velocity of the wind as is the case in this chapter. Using the values given by Souprayen et al., $\gamma = 1/400 \text{ m}$ and $\sigma = 0.01 \text{ s}^{-1}$, we find that $m_* = 4.08 \times 10^{-3} \text{ m}^1$ at 1 km and $m_* = 8.42 \times 10^{-4} \text{ m}^1$ at 10 km.
Table 3.2: Table showing the values for the large z limit of $m_*$ for various initial values of $k$ and $m$. Dimensions are shown in brackets.

### 3.6.2 Finite domain

Using the relations for the energy, (3.37) and (3.36) and the PDFs (3.25) and (3.26), we find

$$E(\eta; z) = \frac{C\varepsilon_0 K \eta}{\eta_0} e^{-(\eta-\eta_0)^2/2} \left[ U(-\lambda-1/2, \sqrt{2}(\eta-\eta_0)) - A(\eta) V(-\lambda-1/2, \sqrt{2}(\eta-\eta_0)) \right] e^{-\lambda z},$$

(3.41)

where this expression contains a contribution from a single eigenfunction and so is only valid for suitably large values of $z$. For small values of $z$ a simple generalization to include more eigenvalues can be made. $A$, here, is a constant defined by the lower boundary and is given by

$$A(\eta_\ell) = \frac{U(-\lambda-1/2, \sqrt{2}(\eta_\ell-\eta_0))}{V(-\lambda-1/2, \sqrt{2}(\eta_\ell-\eta_0))}.$$ 

Rewriting in terms of $m$ and $m_0$ we find

$$E(m; z) = \frac{CN K \varepsilon_0 m (k_0^2 + m_0^2)^{1/2}}{\sqrt{2\sigma(k_0^2 + m^2)^2}} \exp \left[ -\frac{1}{2} \left( \frac{\omega}{\sqrt{(m^2 + k_0^2)}} - \eta_0 \right)^2 \right] \times \left[ U(-\lambda-1/2, \frac{\sqrt{2}\omega}{\sqrt{m^2 + k_0^2}} - \sqrt{2}\eta_0) - A\left(\frac{\omega}{\sqrt{m^2 + k_0^2}}\right) V(-\lambda-1/2, \frac{\sqrt{2}\omega}{\sqrt{m^2 + k_0^2}} - \sqrt{2}\eta_0) \right] e^{-\lambda z}.$$ 

This can be simplified when $m_0 \gg k_0$ (which implies that $A \ll 1$) to

$$E(m; z) \approx \frac{CN K \varepsilon_0 m_0}{\sqrt{2\sigma} m^3} \exp \left[ -\frac{\omega^2}{2} \left( \frac{1}{m} - \frac{1}{m_0} \right)^2 \right] U(-\lambda-1/2, \sqrt{2}\omega \left( \frac{1}{m} - \frac{1}{m_0} \right)) e^{-\lambda z}.$$ 

(3.42)

In the limit $m_c > m_0$ so that $\lambda \to 0$ and by using (3.27), (3.28), (3.29) and (3.30) we can recover the infinite domain limit (3.39). We demonstrate this graphically by comparing the finite-domain solution to the infinite-domain solution in the limit of an infinite boundary, i.e., $m_c \gg m_0$. Plots for the energy power spectral density (PSD) for the two solutions are shown for two different sets of values of $m_0$ and $k_0$ in Figure 3.11. Each graph shows four different plots of the finite domain solution with increasing value of the upper boundary $m_c$ from (1) to (4). (5) shows the Gaussian solution. We can see that as $m_c$ increases the finite domain solution becomes a better approximation.
to the infinite domain solution. For very small values of $E$ the two solutions do not fit very well. This is due to inaccuracies in Maple’s calculations. The poor match of the solutions at small values of $m_c/m_0$ is due to the finite domain having a boundary at zero whereas the infinite domain does not. We can also compare the finite-domain solution with the numerical solution obtained by solving the stochastic differential equation (3.8). Figure 3.11 also shows a comparison between the numerical data and analytic solution in the finite domain (parabolic cylinder solution) for the boundary $m_c = 2m_0$ and again a good match is obtained.

**Physical conclusions:** several physical conclusions can be drawn about the energy spectrum in the finite domain by using (5.1) and Figure 3.11.

i. In the large $z$ and $m$ limit when the boundaries on $m$ are much larger than the initial value ($m_c \gg m_0$), the energy spectrum takes an $m^{-3}$ form. This is not the case for boundaries that are close to the initial value, which is clearly demonstrated by Figure 3.11. As can be seen, the finite domain energy spectrum is equivalent to that for the infinite domain apart from close to a boundary where its behavior is influenced. This Figure is slightly misleading as it appears that an $m^{-3}$ tail does not exist for the finite domain. This discrepancy is put down to the inaccuracies of calculation in Maple at very small values. Indeed, in the large $m$ limit, (3.41) can be shown to have an $m^{-3}$ form.

ii. As in the infinite domain case, the energy spectrum does not scale like $N^2$ but rather with $N$.

iii. A central wavenumber $m_*$ exists. We do not calculate this value for the finite domain because the good match between the finite and infinite domains around the value of $m_*$ implies that it will be equivalent in both finite and infinite domains.

### 3.7 Wave-induced force

Gravity waves, as they propagate vertically, transport momentum from lower levels in the atmosphere to upper levels. The process of breaking causes these waves to deposit their momentum. This results in a force being exerted by the gravity waves on the background wind (known as the mean flow). This force does not just perturb the system but causes features that would otherwise not be present, (see Section 1.4). There are two problems concerning wave-mean flow interaction: the change in the mean flow due to the waves, and the reaction of the mean flow back on the waves. We will be mainly concerned with the first point here.

In this section, we first derive an equation for the wave-induced force acting on the background wind in terms of the probability functions given in Sections 3.4.1 and 3.4.2.
Figure 3.11: Energy PSD curves comparing finite domain, parabolic cylinder solution with infinite domain, Gaussian solution in log–log coordinates. Boundaries for finite domain solutions are set at $m_l = 0$ and: (1) $m_c = 2m_0$; (2) $m_c = 8m_0$; (3) $m_c = 64m_0$ and (4) $m_c = 2048m_0$. (5) is a Gaussian solution with no boundaries. A numerical simulation is also shown for $m_c = 2m_0$ in (i) but this is not very easy to see. Initial values for $m$ and $k$ we take as follows: (i) $m_0 = 0.0007 \text{ m}^{-1}$, $k_0 = 0.0007 \text{ m}^{-1}$ and (ii) $m_0 = 0.0003 \text{ m}^{-1}$, $k_0 = 0.0003 \text{ m}^{-1}$. The altitude is $z = 10$ (4 km), and the parameters are $\gamma = 1/400 \text{ m}^{-1}$, $N = 0.02 \text{ s}^{-1}$ and $\sigma = 4 \text{ m s}^{-1}$. 
This is then used to derive analytic expressions for the wave induced force in terms of the initial wavenumbers and the value of the boundary wavenumber.

The wave induced force is the divergence of total pseudomomentum flux (e.g. Warner & McIntyre 1996, and references therein). The pseudomomentum (see Section 1.5.1) and pseudomomentum flux of a wave traveling along a ray path in \((m, z)\) space are given by

\[
P(m, z) = kA(m, z) \quad \text{and} \quad P_{\text{flux}}(m, z) = c_g(m)P(m, z).
\]

For an ensemble of waves the wave induced force is given by

\[
\mathcal{F}(z) = \text{pseudomomentum flux of dissipated wave packets} \times \text{number of wave packets dissipated per unit height.} \tag{3.43}
\]

The rate of change of dissipated wave-packets (or wave-packets dissipated per unit height) is given by

\[
\frac{d}{dz} \int_{m_t}^{m_e} p(m; z) dm. \tag{3.44}
\]

Now, by integrating (3.13) with respect to \(\eta\) we get

\[
\frac{d}{dz} \int_{\eta_t}^{\eta_e} p(\eta; z) d\eta = (\eta - \eta_0)p(\eta; z) \bigg|_{\eta_t}^{\eta_e} + \frac{1}{2} \frac{\partial}{\partial \eta} p(\eta; z) \bigg|_{\eta_t}^{\eta_e}, \tag{3.45}
\]

but the first term on the right hand side is zero since \(p(\eta; z) = 0\) at \(\eta = \eta_t\) and \(\eta = \eta_e\). We also have

\[
\int p(\eta; z) d\eta = \int p(m; z) \frac{dm}{d\eta} d\eta = \int p(m; z) dm. \tag{3.46}
\]

Using (3.44), (3.45) and (3.46) we obtain

\[
\frac{d}{dz} \int_{m_t}^{m_e} p(m; z) dm = \frac{\partial p(\eta; z)}{\partial \eta} \bigg|_{\eta_e}. \tag{3.47}
\]

Using this together with the definition of the wave induced force (3.43) we obtain

\[
\mathcal{F}(z) = \frac{\sqrt{2}}{2} A_0 \left( c_g(\eta_e) \frac{\partial p(\eta; z)}{\partial \eta} \bigg|_{\eta_e} - c_g(\eta_t) \frac{\partial p(\eta; z)}{\partial \eta} \bigg|_{\eta_t} \right). \tag{3.48}
\]

Notice that the group velocity \(c_g(\eta_t) = 0\) since \(m_t = 0\) and so

\[
\mathcal{F}(z) = \frac{1}{2} A_0 c_g(\eta_e) \frac{\partial p(\eta; z)}{\partial \eta} \bigg|_{\eta_e}. \tag{3.49}
\]

For simplicity we transform this into \((u, z)\) space and obtain

\[
\mathcal{F}(z) = \frac{\sqrt{2}}{2} A_0 c_g(u_e) \frac{\partial p(u; z)}{\partial u} \bigg|_{u_e}. \tag{3.49}
\]

Now, from (3.25) we have

\[
p(u; z) = K_1 2^{-1/2} e^{-u^2/4} \left[ U( -\lambda - 1/2, u) - A(u_t) V( -\lambda - 1/2, u) \right] e^{-\lambda z},
\]

81
where \( A(u) = \frac{U(-\lambda - 1/2, u)}{V(-\lambda - 1/2, u)} \),

and hence

\[
\frac{\partial p}{\partial u} = K 2^{-1/2} e^{-u^2/4} \left\{ U'(-\lambda - 1/2, u) - A(u)V'(-\lambda - 1/2, u) \right. \\
\left. - \frac{u}{2} \left[ U(-\lambda - 1/2, u) - A(u)V(-\lambda - 1/2, u) \right] \right\} e^{-\lambda z},
\]

(3.50)

where the primes denote differentiation with respect to the second argument (in this case \( u \)). The derivatives of \( U \) and \( V \) can be written in terms of \( U \) and \( V \) by using the following recurrence relations (e.g. Abramowitz & Stegun 1965).

\[
U'(a, x) + xU(a, x)/2 + (a + 1/2)U(a + 1, x) = 0
\]

and

\[
V'(a, x) + xV(a, x)/2 - V(a + 1, x) = 0.
\]

Substituting these into (3.50) we obtain

\[
\frac{\partial p}{\partial u} = K 2^{-1/2} e^{-u^2/4} \left[ - uU(-\lambda - 1/2, u) + A(u)uV(-\lambda - 1/2, u) \right. \\
\left. + \lambda U(-\lambda + 1/2, u) - A(u)V(-\lambda + 1/2, u) \right] e^{-\lambda z}
\]

(3.51)

and so the wave induced force is given by (3.49) and (3.51) as

\[
F(z) = \frac{A_0 K}{2} c_g(u_c)e^{-u_c^2/4}\left[ - u_cU(-\lambda - 1/2, u_c) + A(u)uV(-\lambda - 1/2, u_c) \right. \\
\left. + \lambda U(-\lambda + 1/2, u_c) - A(u)V(-\lambda + 1/2, u_c) \right] e^{-\lambda z}.
\]

(3.52)

This can be simplified when \( m_0 \gg k_0 \), in which case \( A(\eta) \ll 1 \). However, we do not assume that \( u_c \gg 1 \) because we require the boundary to be close enough to the initial value that some dissipation takes place: without dissipation there is no force. Hence,

\[
F(z) = \frac{A_0 K}{2} c_g(u_c)e^{-u_c^2/4}\left[ - u_cU(-\lambda - 1/2, u_c) + A(u)uV(-\lambda - 1/2, u_c) \right. \\
\left. + \lambda U(-\lambda + 1/2, u_c) - A(u)V(-\lambda + 1/2, u_c) \right] e^{-\lambda z}.
\]

(3.53)

Equation (3.52) or (3.53) gives the first order decay rate of the force with altitude, (see Figure 3.12(1)). If we add consecutive eigenfunctions to this expression, we should obtain a profile such as that shown in Figure 3.12(2), i.e., the wave induced force is concentrated at an altitude \( z_c \). The full eigenfunction expansion is given by

\[
F(z) = \frac{A_0 K}{2} c_g(u_c)e^{-u_c^2/4}\sum_{i=1}^{p} \left\{ K_i G(\lambda_i, u_c)e^{-\lambda_i z} \right\},
\]

(3.54)

where \( G \) is given by the square braces of (3.52) or (3.53), \( p \) controls the desired level of accuracy, \( K_i \) is given by (3.24) and \( \lambda_i \) is given by solutions to (3.21). Figure 3.13 shows some typical force profiles for a boundary value \( m_c = 3m_0 \) using 1, 2 and 3 eigenfunctions. As we can see, in this case, the maximum \( z_c \) is resolved when using two eigenfunctions. For values of the boundary closer to the initial value \( m_0 \), the number of eigenfunctions required to resolve this maximum dramatically increases.

82
We expect that as the boundary is placed further away the force occurs at higher altitudes. This is demonstrated by Figure 3.14. Here, the wave induced force as a function of altitude is plotted for three different values of the ratio $m_c/m_0$. In each case the critical altitude $z_c$ is marked by a spot on the curve.

We can parameterize the force by considering how the critical altitude $z_c$ changes as the the boundary $m_c$ is placed further away from the initial value $m_0$. Figure 3.15 shows $z_c$ for 10 different values of the ratio $m_c/m_0$. for $m_c/m_0 < 11$ this curve can be fitted by the quartic polynomial

$$z_c = ax^4 + bx^3 + cx^2 + dx + e,$$

where $a = -0.001658354414$, $b = 0.05262349852$, $c = -0.6171933519$, $d = 3.223581537$ and $e = -3.675496584$.

It must be stressed that this approximation is only valid for $m_c/m_0$ smaller than 11. For values of $m_c/m_0$ larger than 11, higher order polynomials are likely to be needed to fit to the data.
Figure 3.14: The wave induced force (using three eigenfunctions) as a function of altitude for \( m_0 = 0.0007 \text{ m}^{-1} \), \( k_0 = 0.0007 \text{ m}^{-1} \) and (1) \( m_c = 3m_0 \); (2) \( m_c = 4m_0 \) and (3) \( m_c = 5m_0 \).

Figure 3.15: The critical altitude \( z_c \) for ten values of the ratio \( m_c/m_0 \) in the range (1,11).

### 3.8 Density variations

The Boussinesq approximation (taking the density \( \rho \) as constant) is clearly not realistic over large altitude scales, (see Section 1.3.3). In this section, we aim to quantify the impact that using the Boussinesq approximation has on the characteristics of the energy profile.

Consider a background density profile that is decreasing exponentially with altitude \( (\rho_0 = e^{-z/H}) \), where \( H \) is the scale height of the atmosphere (defined in Section 1.3), we obtain the dispersion relation (see Section 1.5.1)

\[
\omega = \pm \frac{kN}{\sqrt{k^2 + m^2 + 1/4H^2}} + kU. \tag{3.55}
\]

Compared to the dispersion relation obtained by using the Boussinesq approximation in Section 3.3, the only difference is the \( 1/4H^2 \) term. This change to the dispersion relation on its own is not very interesting as the energy equations are only modified very slightly. If we assume that the background is composed of a superposition of
waves, the exponential decrease means that with wave energy conserved, the amplitude (of \( u', v', w', \rho' \) and \( p' \)) increases exponentially like \( e^{z/2H} \), so breaking occurs faster as \( z \to \infty \).

In Souprayen et al. (2001) this change in the background wind was modeled by letting the amplitude (i.e., the variance) of the background shear increase slowly with altitude. This was introduced into the hydrostatic model by adding a non-dimensional factor \( e^Rz \) to the Wiener process amplitude, where \( R \) is a constant that controls the rate of increase of the shear and is given as \( R = (2\gamma H)^{-1} \), where \( H = 8 \) km and \( \gamma \) is the inverse correlation length of the background wind (defined in Section 3.3). This led to a strong dependence of the energy spectra on altitude. The central wavenumber \( m_0 \) was also seen to decrease with altitude. These variations however are only seen in the small wavenumber part of the spectra. At large wavenumbers, the \( m^{-3} \) tail does not change significantly with altitude.

The exponentially increasing wind can also be introduced in our model, where the velocity rather than the its shear is described by an O-U process. For the infinite domain described in Section 3.4.1, this is done as follows. We first add a factor \( e^Rz \) to the amplitude of the Wiener process in (3.5) and proceed as in Section 3.4.1 with the non-dimensional equations,

\[
d\eta = dU \quad \text{and} \quad dU = -Udz + e^{Rz}dW,
\]

where to take into account the extra term in the denominator of (3.55) we take (cf. 3.7)

\[
\eta = \frac{N}{\sqrt{2\sigma}} (k^2 + m^2 + 1/4H^2)^{-1/2}.
\]

Integrating this system and calculating the mean and variance of \( \eta \) gives the following probability distribution

\[
p(\eta; z) = \frac{\sqrt{R+1}}{[\pi(e^{2Rz} - e^{-2z})]^{1/2}} \exp \left[ \frac{-(R+1)(\eta - \eta_0)^2}{e^{2Rz} - e^{-2z}} \right].
\]

In the infinite domain, the energy spectra is then given by

\[
\mathcal{E}(\eta; z) = \frac{\mathcal{E}_0 \eta_0 \sqrt{R+1}}{\eta_0 [\pi(e^{2Rz} - e^{-2z})]^{1/2}} \exp \left[ \frac{-(R+1)(\eta - \eta_0)^2}{e^{2Rz} - e^{-2z}} \right].
\]

In \((m, z)\) space the energy can be written as

\[
\mathcal{E}(m; z) = \frac{\mathcal{E}_0 Nm \sqrt{R+1}(k_0^2 + m_0^2 + 1/4H^2)^{1/2}}{\sqrt{2\sigma} (k_0^2 + m^2 + 1/4H^2)^{1/2} [\pi(e^{2Rz} - e^{-2z})]^{1/2}} \times \exp \left[ \frac{-(R+1)(\pi(k_0^2 + m^2 + 1/4H^2)^{-1/2} - \eta_0)^2}{e^{2Rz} - e^{-2z}} \right].
\]

Figure 3.16 shows a comparison between the energy density for the constant background density and the exponentially decreasing background density for altitudes \( z = 400 \) m.
Figure 3.16: Energy PSD curves comparing for (i) constant background density; (ii) exponentially decreasing background density and (iii) the spectral limit $m^3$, in the infinite domain at altitudes: (1) $z=400$ m and (2) $z=4$ km. The initial wavenumbers are $m_0 = 7 \times 10^{-4}$ m$^{-1}$ and $k_0 = 7 \times 10^{-5}$ m$^{-1}$ and parameters $\gamma = 1/400$ m$^{-1}$, $N = 0.02$ s$^{-1}$ and $\sigma = 4$ m s$^{-1}$. 
and $z = 4\text{ km}$. The central wavenumber can also be calculated by solving $E_\eta = 0$ and we find

$$\eta_* = \frac{1}{2} \left[ \eta_0 + \sqrt{\eta_0^2 + 2 \left( \frac{e^{2Rz} - e^{-2z}}{R + 1} \right)} \right].$$

So $m_*$ is given by

$$m_* = \sqrt{\frac{2\sigma^2}{\eta_0^2 + (e^{2Rz} - e^{-2z})(R + 1)^{-1} + \eta_0(\eta_0^2 + 2(e^{2Rz} - e^{-2z})(R + 1)^{-1})^{1/2} - k_0^2}}.$$

For large values of $z$ this tends to the limit

$$m_* = \sqrt{\frac{2\sigma^2}{\eta_0^2 + e^{2Rz}(R + 1)^{-1} + \eta_0(\eta_0^2 + 2e^{2Rz}(R + 1)^{-1})^{1/2} - k_0^2}}.$$

A demonstration of this result can be seen in figure 3.17. This shows a plot of the central wavenumber $m_*$ against altitude for a background wind without density variations and one with density variations. As we can see, for a background wind without density variations $m_*$ is constant for large $z$, whereas for a background wind with density variations $m_*$ decreases with $z$ at large $z$. This result was also found by Souprayen et al. (2001).

### 3.9 Summary and discussion

In this chapter we model the atmosphere in one dimension (the vertical). First an atmosphere with constant background density field is considered. The background wind is modelled by an Ornstein–Uhlenbeck process: a simple Gaussian process with
a finite correlation length. Wave evolution in position–wavenumber space is modeled using the ray-tracing equations. The derivation of these ray-equations relies on the WKB approximation which has some limitations:

1. Near caustics, where constructive wave interference causes large wave amplitudes, the assumption that wave amplitude and wavenumber are slowly varying with altitude breaks down, (e.g. Broutman & Young 1986). The problem of caustics will be addressed in Chapter 4.

2. WKB theory makes the assumption that the background wind is due to the waves. WKB also assumes that there is a scale separation between the waves and the background wind. Thus, WKB theory is limited in that winds generated by waves with the same scale cannot be considered. In this chapter we use the WKB approximation to model winds that, in part, are determined by the gravity waves themselves. This step may perhaps be viewed as slightly heuristic (see Souprayen et al. 2001). However experience tells us that this approximation yields good results and hence must be taken on trust.

The ray-equations and O-U process are combined to obtain a single stochastic ordinary differential equation to describe gravity wave propagation through a random background wind. This equation is solved by direct integration, in the infinite (no-boundaries) domain, to obtain a PDF for the the probability of finding a ray at a given position with a given wavenumber.

Gravity wave breaking is included in the model by the addition of an upper vertical wavenumber boundary. In the finite domain (with a wavenumber boundary), a Fokker-Plank equation is obtained and solved by an eigenfunction expansion. Again we obtain a PDF, which is expressed in terms of parabolic cylinder functions. For large altitude it is shown that only the first eigenfunction makes a significant contribution to the PDF. An asymptotic expansion for the first eigenvalue is obtained.

It is possible that modeling the dissipation of wave-momentum at a single wavenumber is too crude an approximation. A more plausible model would ensure that dissipation occurs over a range of wavenumbers centered about a critical wavenumber. In this case it will no longer be possible to use an absorbing boundary on the stochastic model. Instead, a convective instability condition could be included and evaluated along the ray.

Expressions for the energy contained in the waves are obtained for both the infinite and finite domains. Both analytical and numerical solutions are considered and good matches between them obtained. The energy spectrum derived here exhibit features that are characteristic of observed spectra: the existence of a prominent central wavenumber and the large altitude limit $m^{-3}$ spectral tail.
In the final section, we drop the Boussinesq approximation and model gravity waves in an exponentially decreasing density profile. This preserves the constant buoyancy frequency. We note that the $m^{-3}$ tail is still observed although the energy $E$ in the non-Boussinesq model is slightly higher than that in the Boussinesq model. The value of the central wavenumber $m_*$ is also noted to decrease exponentially at large altitude in the non-Boussinesq model whereas it constant in the Boussinesq model.

In most theoretical works the $m^{-3}$ tail obtained from observations is generally assumed to be universal. That is, it does not vary with altitude, season or location. However, there have been measurements at different locations and altitudes that have revealed a large $m$ spectral tail different from the $m^{-3}$ spectral tail predicted by most models. Tsuda (1991) showed that in the summer stratosphere, spectrums of $m^{-2.2}$ and $m^{-2.4}$ can be obtained. Some theoretical attempts have been made to explain the variability in the spectral tail. In particular, Hines (1993) showed that the spectral variations may occur in the Doppler spreading model by considering input source spectra of different forms.

With this in mind, it is important to view Doppler theory as one of a host of particular mechanisms that can be used to predict the form of atmospheric gravity wave spectra. It is by no means a definitive approach, however, it does have its place and the stochastic model of this chapter can be viewed as one of the simplest models to date that predicts an $m^{-3}$ spectral tail.

In Chapter 3.7 we calculated an expression for the wave induced force induced at a critical wavenumber $m_c$. Some examples of force profiles in altitude were given. If a sufficient number of eigenfunctions are used, then the force is concentrated around a critical altitude. The number of eigenvalues needed increases as the boundary $m_c$ approaches the initial value $m_0$. It was found that the further $m_c$ is away from $m_0$ the greater the value of $z_c$. It can also be shown that for values of $m_c/m_0$ less than 11, $z_c(m_c/m_0)$ can be approximated by a quartic polynomial.

The model described in this chapter is a severe idealisation of the real atmosphere. Several extensions and generalizations may be made:

1. Further realism could be added to the model by the inclusion of a randomly fluctuating density field. This adds significantly more complications to the analysis. In the infinite domain it will no longer be possible to model the system with a single integrable stochastic differential equation. In the finite domain, a separable Fokker-Planck equation will not be obtained. It may therefore not be possible to obtain analytical closed form solutions; approximations or numerics may have to be relied on.

2. The WKB approximation is limited in that there must be a scale separation
between the wave and the random background fluctuations. Non-WKB effects could be examined where the fluctuations and the waves are of a similar scale. This would require solving the full set of linear stochastic partial differential equations. This would obviously not be possible analytically and would require the development of a new specialized code.
Chapter 4

Development of caustics

It is nice to know that the computer understands the problem.
But I would like to understand it too.
(Eugene P. Wigner 1993)

4.1 Introduction

The idea of caustics is probably best explained using an example: imagine a plane parabolic wavefront propagating in two dimensions; assume that the wavefront is propagating at a constant velocity in a direction normal to the wavefront. Figure 4.1 shows two successive parabolic wavefronts, $C(0)$ and $C(t)$, separated at each point by a distance $t$. As this wavefront evolves further, the lowest part of the curve becomes more and more bunched up. Eventually points on the curve start to overlap and produce singularities. This is demonstrated in Figure 4.2. This form of overlap is known as a

![Figure 4.1: Two successive parabolic wavefronts, $C(0)$ and $C(t)$, separated at each point by a distance $t$. Further details on the calculation of these curves can be found in Appendix F.](image)

---

1 This is just Huygen's principle for non-dispersive waves: each point on a wave front may be regarded as a new source of waves (e.g. Richards & Williams 1972).
Figure 4.2: Successive parabolic wavefronts demonstrating the existence of swallow tails. The caustic curve joining the singularities is shown by the broken line. Further details on the calculation of these curves can be found in Appendix F.

swallow tail due to its similar shape. The envelope of the singularities formed as the wavefront progresses is called a caustic curve and is shown by the broken line in Figure 4.2. Further details on the calculation of these curves can be found in Appendix F. Due to the fact that rays overlap on caustic curves (or surfaces in three dimensions), constructive interference causes wave amplitudes to become large. This can be seen for example when sunlight shines into a coffee cup: a cusp shaped bright area can often be observed as the light reflects off the inside of the cup onto the surface of the coffee. This bright area is due to large wave amplitudes on a caustic curve. A good demonstration of this can be found at Williams (2004)

Caustics appear in many branches of physics and over a large range of physical scales. Some examples are: stress waves in the earth's crust; interaction of radio waves with the earth's atmosphere and freak waves in the earth's oceans or atmosphere. The study of caustics in the atmosphere will be the focus of this chapter.

4.1.1 Gravity wave caustics

Caustics can play an important part in the breaking of stratospheric gravity waves. They cause an increase in wave amplitudes – which in turn may lead to instabilities – and gravity wave breaking, (see Section 1.6.2). Potentially, caustics have a significant effect on the flux of action to the mean flow, affecting the global circulation of the atmosphere, (see Section 1.4). However, they are often overlooked: the effect of caustics
are not generally included in gravity wave parameterisations in GCMs.\(^2\)

In the setting of ray theory, we can define caustics as the envelope of singularities in a ray field, appearing when two infinitesimally close rays converge, (see Broutman et al. 2004 for a recent review on ray methods for caustic modelling). At these caustic curves, ray theory predicts infinite wave amplitudes. (A calculation of this in the setting of ray theory can be found in Appendix E.) There is, in fact, a breakdown in the modelling assumptions (because the amplitude and wavenumber are not slowly varying near caustics) and although the amplitudes are large, in reality they are not infinite and can be calculated, (see (ii) below).

The nature of variations in the background wind (through which gravity waves propagate) is crucial in determining the form of caustics. In the atmosphere caustics can appear because of spatial and temporal variations in the wind. Three regimes have been considered by the current literature:

1. One-dimensional time independent background wind – caustics are not possible.
   All rays are congruent and so no ray convergence can take place and no caustics form.

2. One-dimensional time dependent background wind – caustics are possible. Consider, for example, gravity waves governed by the dispersion relation \(\omega = \omega' + k \cdot U\) in a wind given by \(U = (U(z - c_{pz}t), 0, 0)\), in a fixed reference frame, where \(c_{pz}\) is the vertical phase velocity of the wind. Transforming to a vertical coordinate moving with the vertical phase velocity, i.e., by letting \(z = z' - c_{pz}t\), we find \(U = (U(z), 0, -c_{pz})\). Rays change their direction of propagation when \(\omega_m = 0\), i.e., when \(\omega_m = c_{pz}\), the vertical group velocity of the short waves matches the vertical phase velocity of the wind. Figure 4.3 demonstrates that when this condition is met we get an intersection of rays, i.e., caustics. This is exactly the model considered by Broutman (1986) to model gravity waves. He chose the background wind to be represented by a purely sinusoidal infinite wave and constructed a solution for the wave amplitude that is valid close to a caustic. Based on the work by Peregrine & Smith (1978), this is achieved by defining the function \(G(\omega, k, z) = \omega' - \omega\) where \(\omega'\) is defined such that \(G = 0\) gives the dispersion relation. When substituted into the wave equation (the Taylor-Goldstein equation) and expanded about the caustic point, this yields an Airy equation. Matching this to the ray solution yields a function that is valid both away from the caustic and close to the caustic.

\(^2\)In the Lindzen (1981) parameterisation scheme, for example, the altitude at which gravity waves break in the model is higher than that inferred from observations. This results in a much stronger gravity wave drag than might be realistic. The inclusion of caustics, causing breaking at lower levels, is a possible remedy to this weakness of the scheme.
Broutman et al. (1997) also consider ray propagation in a time dependent background similar to that in Broutman (1986). They conclude that time dependence in the background wind can make a significant difference to gravity wave behaviour.

Sonmor & Klaassen (1999) extended the theory of Broutman. They studied caustic focusing in time dependent background winds, consisting of more than a single phase speed, using numerical techniques. They illustrated a few points. In particular that non-critical caustic events, i.e., with \( c_{gz} \neq c_{pz} \), are possible. They also showed that the presence of a mean flow causes caustics to form at lower altitudes than would occur without a mean flow.

iii. Two-dimensional time independent wind – caustics are possible. Hydrostatic inertia-gravity waves (i.e., with rotation effects included) have been studied by Dunkerton (1984). Rays were traced numerically through an analytic background wind designed to match observed wind velocity data: the wind was dependent on altitude and latitude and purely horizontal. The formation of caustic envelopes was detected visually by plotting multiple rays. However, no verification was made on the altitude of caustic formation.

In this chapter, we consider gravity wave caustics appearing in a two-dimensional time-independent background flow. The inhomogeneities in the background are modelled by a random process. The waves refract and scatter from the inhomogeneities which act as lenses with random properties (see Figure 1.5 in Section 1.6.1). Consequently, a random distribution of caustics results.\(^3\) This model is

\[^3\]It is also possible that the waves may interact strongly with the medium and produce more inho-
highly appropriate as in many natural media, such as the atmosphere, there exist spatial fluctuations that can be modelled by random processes.

Caustic formation in random backgrounds have been considered by B. White and his coworkers for several systems. We now give a brief review of some of their results. Further details of the stochastic formulation they use can be found in Section 4.2.

Kulkarny & White (1982) investigated the propagation of high-frequency non-dispersive waves in a two-dimensional medium with small, isotropic random inhomogeneities. After propagating long distances, the rays are shown to develop caustics and the probability distribution of the distance along a ray to first caustic formation is found to be given by a universal curve: the statistics of the random medium do not influence the shape of the probability distribution and only contribute a scale factor. This theory has been backed up by the numerical simulations of Hesslink & Stutevant (1988). White (1984) considered a similar theory in three dimensions and obtained analogous results to the two-dimensional work of Kulkarny & White (1982).

The paper most relevant for our purpose is that by White & Fornberg (1998), which describes the propagation of deep water surface gravity waves in the ocean. The purpose of this work is to seek to explain the existence of freak waves (waves of extreme amplitude), which could be caused by caustics. Although only ocean waves are considered, this analysis is quite general in that it makes the calculation for an (almost – see Section 4.2) arbitrary dispersion relation. Results similar to those of Kulkarny & White (1982) for the distance to a first caustic along a ray, are derived using analytical methods and backed up with numerical results. We had hoped to use this result and apply it directly to atmospheric gravity waves, however, it turns out the results are not applicable. This is further discussed in Section 4.2.

A brief mention should be made of the work by Hertzog et al. (2002), part of which concerns atmospheric caustics. As part of their Doppler spreading calculations through a wind obtained from ECMWF analyses, they made a calculation of caustic events. However, they make no quantification of caustic events on wave dissipation, arguing that the increase in amplitude of a gravity wave due to caustic focusing is weak enough that dissipation does not occur (this argument is based on the work by Broutman (1986)).
4.1.2 Chapter description

In this chapter, we will model the propagation of gravity waves in a two-dimensional time independent background wind. We use numerical techniques that solve the ray equations and their derivatives to predict the occurrence of caustics. Using a stochastic description of the wind, we will find statistics for the altitude of caustic development. The main aim is to quantify the dependence that the altitude at which caustics occur has on the strength of random inhomogeneities in the background wind. We expect that the average height of caustics decreases as the strength of the inhomogeneities increases and aim to find the scaling of this correlation.

Section 4.2 describes in some detail the theory of White & Fornberg (1998) for caustic formation in a two-dimensional random media. We explain why this theory does not apply to atmospheric gravity waves and hence we adopt an entirely numerical approach. Section 4.3 describes the basic two-dimensional equations that govern atmospheric gravity wave propagation in both dimensional and non-dimensional forms. Approximations to the hydrostatic regime are made and some numerical benchmarks run to check the accuracy of this approximation. The notion of “derived” ray equations (e.g. Hayes (1970)) is introduced: these allow us to monitor the distance between rays and hence detect the formation of caustics. Section 4.4 describes the numerical implementation of the ray and derived ray equations. “Many ray” codes are developed in which rays from multiple different initial conditions are to be traced to determine the location of a caustic; “single ray” codes are developed, using the derived ray equations, where only one ray need be traced to determine the location of a caustic. Benchmarks between the two codes are conducted. Details of the random background wind are also given and the methods used to control the error in the code (quite a delicate matter) are described. Section 4.5 describes the results obtained from running codes for various different initial conditions and background winds. We provide probability density functions (PDFs) for the altitude at which a caustic forms as a function of the strength of the random part of the background wind. A summary and discussion is made in Section 4.6.


Consider the following scenario: a wave enters a random medium. The medium is assumed to have a constant (possibly zero) mean. The relative amplitudes of the random variations of the medium are assumed to be uniform and isotropic and are characterised by their standard deviation $\sigma$. The random fluctuations have an intrinsic length scale $l$, which can be thought of as a correlation length. To obtain a PDF for the distance along a ray until a first caustic is observed, the following steps are taken:
1. If we assume that the characteristic length scale of the waves $\lambda$ is much greater than the correlation length of the inhomogeneities, i.e., $\lambda \gg l$, then we can apply the WKB approximation and hence ray theory (see Section 1.5.1) everywhere except very close to a caustic. A dispersion relation is obtained in a general context in the form

$$\omega(x,k) = F(k) + \sigma G(x,k).$$

For example, for wave motion through a medium with a random velocity field, we have: $F = \omega + k \cdot U_0$ and $G = k \cdot \hat{U}$, where $U_0$ is a constant, $\hat{U}$ is the random part of the medium and $\omega$ is the frequency relative to the background medium.

2. To locate a caustic some information about the amplitude of a wave is required since WKB predicts infinite amplitudes at caustics. The amplitude is governed by a transport equation as was discussed in Chapter 2. For time-independent systems this equation takes on the form of energy conservation; for time-dependent systems, action conservation. These transport equations are only valid assuming that the background flow is an exact solution of the fluid equations. This will not be the case for the random velocity used in this chapter, (see Section 4.4.4). Thus, these equations are not applicable here. However, an alternative method of locating a caustic (without relying on a transport equation) is available, using a geometric formulation and the notion of ray-tube area, which is now defined.

Consider a single ray and a small collection of other rays in its near vicinity forming a tube (known as a ray-tube, e.g. Lighthill (1978)). The point at which the area of this tube tends to zero (where the rays intersect) will give the location of a caustic. An evolution equation can be derived for the ray-tube area (without using a transport equation, demonstrated in Appendix E) which can be used to locate the position of a caustic.

3. The equations are scaled to the distance in which caustic formation is expected to be observed. White assumes that this is $O(\sigma^{-2/3}l)$ and so rescales time: $\tau = \sigma^{2/3}t$. Hence the following scaled ray equations are obtained:

$$\frac{dx}{d\tau} = \frac{1}{\epsilon^2} F_k(k) + \epsilon G_k(x,k) \quad \text{and} \quad \frac{dk}{d\tau} = -\epsilon G_k(x,k),$$

where $\epsilon = \sigma^{1/3}$.

4. If the standard deviation of the random variations is assumed small, $\sigma \ll 1$, so that a typical inhomogeneity produces only a small wave deflection, then an asymptotic theory can be applied, in which rays are considered to be small random perturbations from their deterministic trajectories. The position vector $x$ and wavevector $k$ are decomposed into a deterministic part and a smaller random
part. The random part is further decomposed into perturbations in the direction of the group velocity along a unit vector \( t \) and perpendicular to it along a unit vector \( n \). See Figure 4.4.

5. With some manipulations the equations can be put into the form

\[
\frac{dW}{dt} = \frac{1}{\epsilon} H_1 \left( \frac{\tau}{\epsilon^2}, W \right) + H_2 \left( \frac{\tau}{\epsilon^2}, W \right), \quad W(0) = W_0, \quad W \in \mathbb{R}^d. (4.1)
\]

Here, \( \epsilon \) is a small parameter and for each fixed non-random value of \( W \), \( H_1 \) and \( H_2 \) are random functions with \( \langle H_1 \rangle = 0 \) satisfying a mixing condition. The probabilistic limit theorem of Papanicolaou & Kohler (1974) states that in the limit \( \epsilon \to 0 \), \( W \) is well approximated by the solutions of the stochastic differential equation

\[
dW = b(W)dt + \sqrt{2}\Sigma(W)d\beta,
\]

where \( \beta \) is a vector of independent Brownian motions. The vectors \( b \) and \( \Sigma \) have complicated dependencies on \( H_1 \) and \( H_2 \); exact definitions can be found in White & Fornberg (1998). \( \Sigma \) only depends on the time integrated correlation functions of \( H_1 \) and \( H_2 \) and so it is apparent that the probability distribution of time until the formation of a caustic, depends only on a distance scale parameter and not on the statistics of the background medium.

6. A single partial differential equation for the probability density of the time at which the ray-tube area is zero can then be derived from (4.2). This is solved to give a functional form for the probability density.
The "general" nature of the work by White & Fornberg (1998) does not apply to that of atmospheric gravity waves. This can be explained as follows: with no background wind, gravity waves belong to the class of conical waves, i.e., those of the form

\[ F(k) = f \left( \frac{k}{|k|} \right), \]

where \( F \) is defined as the deterministic part of the dispersion relation. For conical waves, the group velocity vector is perpendicular to the wave vector, i.e., \( c_g \cdot k = 0 \). This is easily proved if we let \( n = k/|k| \):

\[ c^i k_i = \frac{\partial f}{\partial k_i} k_i = \frac{\partial f}{\partial n_j} \frac{\partial n_j}{\partial k_i} k_i = \frac{\partial f}{\partial n_j} \left( \delta_{ij} - \frac{k_i k_j}{|k|^2} \right) k_i = 0, \quad (4.3) \]

since \( \frac{\partial}{\partial k_i} |k|^{-1} = -|k|^{-2} \frac{\partial |k|}{\partial k_i} = -\frac{k_i}{|k|^3} \).

In White's theory, one component of (4.1) is an evolution equation for the ray-tube area \( A \), (see Appendix E) and is given by

\[ \frac{dA}{d\tau} = \Delta B + O(\epsilon), \quad (4.4) \]

where \( B \) is a quantity given by White & Fornberg (1998), equation (2.25) (but is not important for the explanation here) and

\[ \Delta = n^T \cdot F_{kk}(k) \cdot n, \]

where \( T \) denotes the transpose.

For non-conical waves (those considered in White & Fornberg 1998) \( \Delta \) is non-zero. However, for gravity waves an easy, but tedious, calculation (using (4.3)) gives \( \Delta = 0.4 \).

Equation (4.4) then implies that at first order, the ray-tube area is constant, hence no caustics can form for gravity waves. In practice what this means is that there are no caustics over a \( \sigma^{-2/3} \) time (and distance). We would nevertheless expect to see caustics over longer distances. The asymptotic theory, although possible in principle, would be very complicated and so we will investigate the occurrence of caustics numerically. We first describe the basic equations for gravity wave propagation in a two-dimensional background wind.

### 4.3 Governing equations

We derive the basic equations for gravity wave propagation in two dimensions with a background wind given by \( U = (U(x, z), W(x, z)) \). We start by applying the Boussinesq approximation (there is no decrease in density with height in this model, but we are not interested in breaking due to density decreases so is not a major simplification)\(^4\). This is probably also true for conical waves in general but the calculation has not been verified.
and then show that, by an appropriate choice of non-dimensionalisation, the equations reduce naturally to the hydrostatic regime (see Section 1.5.1).

Using the Boussinesq approximation, the dispersion relation for internal gravity waves in a two-dimensional atmosphere with a background wind is given by

$$\omega = \frac{-kN}{(k^2 + m^2)^{1/2}} + k \cdot U,$$

where \( k = (k, m) \) is the two-dimensional wavenumber in \((x, z)\) space and \( N \) is the buoyancy (Brunt-Väisälä) frequency. The ray equations (see Section 1.5.1) are given by

$$\frac{dx}{dt} = \frac{\partial \omega}{\partial k} \quad \text{and} \quad \frac{dk}{dt} = -\frac{\partial \omega}{\partial x},$$

which, when expanded, give:

$$\frac{dx}{dt} = \omega_k = -m^2N(k^2 + m^2)^{-3/2} + U,$$

$$\frac{dz}{dt} = \omega_m = kmN(k^2 + m^2)^{-3/2} + W,$$

$$\frac{dk}{dt} = -\omega_x = -kU_x - mW_x$$

and

$$\frac{dm}{dt} = -\omega_z = -kU_z - mW_z.$$

Let us non-dimensionalise these equations. We define the dimensionless Froude number (e.g. Gill 1982) as the ratio between the speed of the background wind and a typical internal wave speed (the horizontal phase velocity), i.e., \( F = U_0m/N \), where \( U_0 \) is a typical wind speed. The vertical wavenumber is then scaled so that the Froude number is order 1, i.e., we use a reference wavenumber \( m_0 = N/U_0 \). We now define the aspect ratio: \( \delta = k_0/m_0 \), where \( k_0 \) is a typical horizontal wavenumber, and take \( l \) to be a typical length scale for the basic flow. Scaling the ratio \( W/U \) like \( \delta \), we introduce the following transformations:

$$x = l\hat{x}, \quad z = l\delta\hat{z}, \quad k = k_0\hat{k}, \quad m = \frac{N}{U_0}\hat{m}, \quad t = \frac{l}{U_0}\hat{t}, \quad U = U_0\hat{U},$$

$$\omega = U_0k_0\hat{\omega} \quad \text{and} \quad W = \delta U_0\hat{W}.$$ 

Dropping hats, we obtain the system of non-dimensional equations

$$\omega = \frac{-k}{(\delta^2k^2 + m^2)^{1/2}} + k \cdot U,$$

$$\frac{dx}{dt} = -m^2(\delta^2k^2 + m^2)^{-3/2} + U,$$

$$\frac{dz}{dt} = km(\delta^2k^2 + m^2)^{-3/2} + W,$$

$$\frac{dk}{dt} = -kU_x - mW_x$$

and

$$\frac{dm}{dt} = -kU_z - mW_z.$$

100
For the atmosphere, the physical parameters obtained using the values given in Section 1.3.3

\[ 10^{-7} \text{ m}^{-1} < k_0 < 10^{-4} \text{ m}^{-1} \quad \text{and} \quad 5 \times 10^{-5} \text{ m}^{-1} < m_0 < 1.6 \times 10^{-3} \text{ m}^{-1}, \]

are realistic. These give an approximate value for the aspect ratio

\[ 10^{-4} < \delta < 10^{-1}. \]

A reasonable approximation is therefore to take \( \delta = 0 \) – equivalent to making the hydrostatic approximation, (see Section 1.5(1)). In this case, we obtain the ray equations

\[
\begin{align*}
\omega &= \frac{-k}{m} + kU, \\
\frac{dx}{dt} &= -m^{-1} + U, \\
\frac{dz}{dt} &= km^{-2} + W, \\
\frac{dk}{dt} &= -kU_x - mW_x, \\
\text{and} \quad \frac{dm}{dt} &= -kU_z - mW_z. \tag{4.6}
\end{align*}
\]

A numerical check was made on the affect of the hydrostatic approximation on ray trajectories and the comparison between hydrostatic and non-hydrostatic rays can be seen in Figure 4.5. We observe that for \( \delta < 0.05 \), there is a good match between the hydrostatic and the non-hydrostatic regime, while for \( \delta = 0.5 \) the two regimes give significantly different results. Equations (4.5)–(4.6) can easily be solved analytically when \( U \) has a simple form. These cases are useful as benchmark tests for the numerical codes that will be developed in Section 4.4. Solutions for three separate cases of the background wind where initial conditions are taken to be \( \{x_0, z_0, k_0, m_0\} = \{1, 1, 1, 1\} \) are listed below. (The choice of \( k > 0 \) means the waves will be leftwards propagating. This is of no significance).

1. \( U = 0, \quad W = 0, \quad \{x, z, k, m\} = \{1 - t, 1 + t, 1, 1\} \)
2. \( U = -z, \quad W = 0, \quad \{x, z, k, m\} = \{1 - 2t, 2 - (1 + t)^{-1}, 1, 1 + t\} \)
3. \( U = z, \quad W = x, \quad \{x, z, k, m\} = \{\cosh t, 3e^t/2 + e^{-t}/2, e^{-t}, e^t\} \)

A feature displayed by case 2 is a critical level. As discussed in Section 1.6.1, critical levels are found when the phase velocity of a wave equals the horizontal wind velocity and are important as mechanisms for wave steepening and hence breaking. At critical levels the equation \( k \cdot U = \omega \) holds. This gives

\[ \frac{k}{m} = 0 \quad \text{hence} \quad m \to \infty \quad \text{and} \quad c_{gz} = \frac{1}{m^2} \to 0. \]
Figure 4.5: Comparison between hydrostatic and non-hydrostatic regimes for different values of the aspect ratio $\delta$: top left, $\delta = 0$; top right, $\delta = 0.05$ and bottom, $\delta = 0.5$. The background wind used was $U(x, z) = -z + \sin x/5$. 100 rays were traced with initial values of $x$ in the range $(0,50)$. It can clearly be seen that the graphs for $\delta = 0$ and $\delta = 0.05$ are almost identical. The graph for $\delta = 0.5$ is obviously sufficiently different to the other two that the hydrostatic approximation would not hold.
Thus, as a ray approaches a critical level, the vertical group velocity approaches 0 and vertical propagation ceases. The analytic solution for \( U = -z \) and \( W = 0 \) demonstrates the existence of a critical level at \( z = 2 \) as \( t \to \infty \). This can clearly be seen in the ray trajectory plotted in Figure 4.6.

### 4.4 The numerical model

Our objective is to determine the distribution of caustics for rays governed by (4.5)-(4.6) with random \( U \) and \( W \). The question that we still need to answer is: how can we locate the altitude position of caustics in a given background wind? This question will now be addressed using numerical methods.

Equations (4.6) form a system of four ordinary differential equations with independent variable \( t \). For an arbitrary background wind these can be solved numerically using a Runge-Kutta scheme, (see appendix G). In practice we have written a Fortran code for the solution of differential equations, relying on the nag_ivp_ode_rk module from the Numerical Algorithms Group (Nag 1995). This code - from now on referred to as the “multi-ray code” - fixes initial values for \( k, m \) and \( z_5 \) and traces a number of rays for various initial values of \( x \). Caustics are then identified as the lines of intersection of neighbouring rays.

A check of the multi-ray code was made by comparing numerical results with the three analytical cases in Section 4.3; satisfactory matches were obtained. For example, Figure 4.7 shows analytical plots for the parameters \( x, z, k \) and \( m \) as functions of \( t \), for the background wind \( U = -z \) and \( W = 0 \). The numerical solution is indistinguishable from these. Hence we assume reliability of the multi-ray code and use it to explore more interesting situations.

For a wind that is a function of altitude \( z \) only, caustics do not appear since all the

---

\(^4\)It is possible that rays could originate from different altitudes and still cross to form caustics. However, the restriction to a single initial value of \( z \) will be maintained here for simplicity.
Figure 4.7: Parameters $x$, $z$, $k$ and $m$ as functions of $t$ given by the lines: (1) $m = 1 + t$; (2) $z = 2 - (1 + t)^{-1}$; (3) $k = 1$ and (4) $x = 1 - 2t$ for the background wind $U = -z$ and $W = 0$. Analytical and numerical solutions are shown and are indistinguishable from each other.

To show an example of caustic detection using the multi-ray code, we consider a simple wind that exhibits caustic behaviour. We take

$$U(x, z) = -\Delta z + af(x, z) \quad \text{and} \quad W(x, z) = -bf(x, z),$$

where $f(x, z) = \sin(x/a + z/b)$ is such that the velocity field is divergence free. We take $a \ll \Delta$ and $b \ll \Delta$ so that the wind consists of a shear flow (dependent on $z$) plus a small perturbation (dependent on $x$ and $z$). With $\Delta = -1$, $a = 0.1$ and $b = 0.01$ a numerical multi-ray code is run. Figure 4.8 shows a plot of 70 rays traced numerically by the multi-ray code. An initial separation distance of 0.02 in the $x$ direction and initial conditions $k = 1$, $m = 1$ and $z = 1$ are used. The formation of caustics can be clearly seen by the darkened areas where the rays cross. This method of locating a caustic is not very efficient since many rays have to be displayed to gain any information about the position of the caustics. It is also very difficult to obtain statistics for the position of a caustic. There is an alternative however, which is to locate caustics by calculating the ray-tube area as described in Section 4.2.

The evolution of the ray-tube area can be computed by solving a system of nine ordinary differential equations (ODEs), called the derived-ray equations, (e.g. Hayes 1970). The first four of these equations are the ray equations given in Section 4.3. Another four consist of the evolution equations for $k_x$, $k_z$, $m_x$ and $m_z$, which define the evolution of the variation of the wavenumbers between neighbouring rays. The ninth equation is that for the evolution of the ray-tube area. The derivation of these equations is given in Appendix E. The nine ODEs obtained here are equivalent to (18) and (19) of Hayes (1970), or to (2.26), (2.29) and (2.30) of White & Fornberg (1998). If we solve these ODEs numerically, for a particular value of $x$, we only need to trace a single ray to determine the location of a caustic, since every ray traced also has a value for the ray-tube area attached to it. This code – from now on referred to as the
"single-ray code" – requires initial values for the variables $k_x, k_z, m_x$ and $m_z$ in terms of the initial values for $k, m, x, z, U$ and $W$. These are calculated using the fact that

$$\frac{d k}{d t} = \frac{\partial k}{\partial x} \frac{d x}{d t} + \frac{\partial k}{\partial z} \frac{d z}{d t}. \tag{4.7}$$

If all rays start with same initial values for $k$ then partial derivatives of $k$ with respect to $x$ are zero. Using the ray equations (4.6) and (4.7) we obtain:

$$k_x = 0, \quad m_x = 0, \tag{4.8}$$

$$\frac{\partial k}{\partial z} = \frac{d k}{d t} \frac{d t}{d z} = \frac{-kU_x - mW_x}{km^{-2} + W},$$

and

$$\frac{\partial m}{\partial z} = \frac{d m}{d t} \frac{d t}{d z} = \frac{-kU_z - mW_z}{km^{-2} + W}.$$

To determine the location of a caustic, we monitor the value of the ray-tube area $J$. This is determined by the evolution equation given in Appendix E. The initial value we use is $J_0 = 1$ and a caustic event will be recorded if $J(t)$ is much smaller than $J_0$, i.e., $J < 0.01$. Technically a caustic event only occurs when $J = 0$, but at $J = 0$ there are singularities in the equations so numerical operations have a greater time cost as $J \to 0$.

It is possible that a ray might hit a critical level (see Section 4.3) before it reaches a caustic; these events must be accounted for as our objective is to quantify the affect.
that caustics have on gravity wave breaking. The code contains a criteria that if for any ray \( k/m < 0.01 \), we determine a critical level to have been reached, the realisation is stopped and the altitude recorded as a critical level event, not a caustic event.

The single-ray code requires an upper limit for \( t \), \( t_{\text{end}} \) say (note that this is a non-dimensional parameter). We need to make sure that this value is large enough so that rays have a chance to propagate through the stratosphere, but not so large that we make unnecessary computations. We find an upper bound for \( t_{\text{end}} \) as follows: the typical time taken for a gravity wave to propagate through the stratosphere is given by \( t_s = H_s/c_{gz} \), where \( H_s \) is the height of the stratosphere and \( c_{gz} \) is a typical vertical group velocity.

Now \( t_{\text{end}} \) is the non-dimensional equivalent of \( t_s \). By using the scaling for \( t \) given in Section 4.3 we have \( t_{\text{end}} \sim t_s U/l \), where \( l \) is a typical horizontal length scale of the basic flow – which can be thought of as the horizontal distance a wave will travel while it travels through the stratosphere, i.e., \( l = c_{gz} t_s \), where \( c_{gz} \) is a typical horizontal group velocity and \( U \) is a typical horizontal wind velocity. And so, \( t_{\text{end}} \) is given by

\[
t_{\text{end}} = \frac{U t_s}{l} = \frac{U H_s}{c_{gz} c_{gz} t_s} = \frac{U c_{gz}}{c_{gz} c_{gz}}.
\]

If we take large estimates for values on the numerator and small estimates for values on the denominator then we will obtain a maximum estimate for \( t_{\text{end}} \). We let \( U = 50 \text{ m s}^{-1} \) and referring to Table 4.1, we find

\[
t_{\text{end}} = \frac{50 \times 198}{0.1 \times 30} = 3300.
\]

Just to be on the conservative side, we take \( t_{\text{end}} = 5000 \).

<table>
<thead>
<tr>
<th>( k ) (m(^{-1}))</th>
<th>( m ) (m(^{-1}))</th>
<th>( c_{gz} ) (m s(^{-1}))</th>
<th>( c_{gz} ) (m s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^{-7} )</td>
<td>( 10^{-5} )</td>
<td>20.1</td>
<td>1950</td>
</tr>
<tr>
<td>( 10^{-4} )</td>
<td>( 0.3 )</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>( 10^{-5} )</td>
<td>( 0.1 )</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>( 10^{-6} )</td>
<td>( 10^{-5} )</td>
<td>198</td>
<td>1930</td>
</tr>
<tr>
<td>( 10^{-4} )</td>
<td>( 2.1 )</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>( 10^{-3} )</td>
<td>( 0.12 )</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>( 10^{-5} )</td>
<td>( 10^{-4} )</td>
<td>19.9</td>
<td>148</td>
</tr>
<tr>
<td>( 10^{-3} )</td>
<td>( 0.3 )</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>( 10^{-4} )</td>
<td>( 10^{-3} )</td>
<td>2.1</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 4.1: Values of the vertical group velocity \( c_{gz} \) and the horizontal group velocity \( c_{gz} \), calculated using realistic values of \( k_0 \) and \( m_0 \) namely, \( 10^{-7} \text{ m}^{-1} < k_0 < 10^{-4} \text{ m}^{-1} \) and \( 5 \times 10^{-5} \text{ m}^{-1} < m_0 < 1.6 \times 10^{-3} \text{ m}^{-1} \) given in Section 1.3.3 and wind velocities \( U = 10 \text{ m s}^{-1} \) and \( W = 0.1 \text{ m s}^{-1} \) (in keeping with the hydrostatic approximation) are used. Units are given in brackets.

### 4.4.1 Error control

In simulations of random processes, numerical errors arise for particular realisations of the background wind for example, for extreme values of the random function. If such
an error is detected for a particular realisation then further calculation for that ray is stopped, no results are recorded and the code moves on to the next realisation. Errors for which this will occur are as follows:

1. The back reflection condition: the change in sign of $m$ is preceded by a breakdown of the WKB approximation, since $m \to 0$. To prevent downward propagation of gravity waves we restrict $m > 0$.

2. The Nag routine Nag (1995) employed requires the specification of constants for relative local error control. These are given as a tolerance (tol) and a threshold (thresh) and are explained as follows. The magnitude of the local error on any step will not be greater than $\text{tol} \times \max(\eta, \text{thresh})$, where $\eta$ is the average magnitude of the solution over the step. The choice of the threshold depends on the general size of the solution in the course of the integration, i.e., the solution is of no interest if it is smaller than the threshold. Given that we are trying to obtain trends in the value of altitude at a caustic depending on certain parameters, we do not require solutions to a very high degree of accuracy. For this reason, we take quite a large value for the threshold, i.e., large relative to the range of values for the threshold allowed by the Nag routines.

The tolerance defines the local error allowed during the integration. Since we are taking statistics over a large number of realisations, it is not important to achieve a low level tolerance; the biggest value allowed by the Nag routines is therefore chosen – a larger tolerance enables a faster code. The values used for tol and thresh are

$$\text{thresh} = 0.01 \text{ or } 0.1 \quad \text{and} \quad \text{tol} = 0.01,$$

where the smaller value for thresh is used for smaller values of $\epsilon$.

### 4.4.2 Stiffness

A system of differential equations is called stiff if there are processes in the system that behave with significantly different time scales. For example, if the solution of a differential equation is given by

$$y(t) = C_1 e^{-t} + C_2 e^{-100t},$$

where $C_1$ and $C_2$ are constants, then the second component of the solution will decay much more rapidly than the first component as $t$ increases. Stiff systems create difficulties in numerical routines in terms of stability and time costs: the time step must be small enough to resolve the system for the fast time scale over the whole problem even when only a small part of the problem requires fine resolution.
Although the Nag codes are very proficient, for some realisations of the velocity field the problem can become stiff. A Runge-Kutta order 4 method was chosen as the best method to deal with stiffness, (see Appendix G). This is at an increased time cost compared to an order 2 method, which would be sufficient for our accuracy requirements.

The adaptive time step methods built into the Nag routines were still unable to cope with the stiffness observed for some realisations as they only reduce the time step by small amounts. The first option to solve this problem may have been to use routines specifically designed for solving stiff problems (i.e., implicit schemes). However, given that stiffness is only a problem in a few realisations, another method was applied. This involved attempting to predict when the system would become stiff and compensating by reducing the time step accordingly (the time step is reduced to a tenth of its original value). This was done by making trial runs of the code for a small number of time-steps and realisations and noting the value of the ray-tube area when numerical instabilities occurred. It was then possible to reduce the time step before this event in further simulations. Hence it was possible to maximise the efficiency of the code and this method was able to reduce the amount of errors to around 1% of the total realisations.

4.4.3 Computing times

The computer being used for the numerics was a Sun Enterprise 450 server running Solaris 7 with a 400 MHz Ultra Sparc II CPU with 4 Mbs of built-in cache memory and 2.25 Gb of RAM. The specs are: 16.1 specint95 and 39.2 specfp95. Due to the complex nature of the derived ray equations, each of which requires several calculations the background wind (which is a sum of five Fourier modes), a standard run for a single value of $\alpha$ and a single value of $\epsilon$, for 100 realisations takes in the order of about 1 day to complete. On the given machine this is a very time consuming code.

4.4.4 The random wind

Much experimental work has been done on the measurement of the horizontal wind shear over a large range of altitudes in the atmosphere using various methods (e.g. Endlich et al. 1969, Barat 1982). In this caustic model, we require a value for the wind at each time step. It would be possible to take actual data for a wind profile and implement this in simulations (e.g. Hertzog et al. 2002). However, this would require interpolation between data points and would not result in a very elegant or efficient code. The method adopted here is to represent the wind by an analytical function that is consistent with real data.

\footnote{To put this into perspective for the reader who may not be familiar with these specifications, this machine is far inferior to the average home desktop computer.}
Actual wind data can be represented by the superposition of a small number of sine waves as a function of altitude and horizontal position. See Dunkerton & Butchart (1984), Dunkerton (1984) and Sonmor & Klaassen (1999) for examples of real winds modelled by analytical functions.

We introduce a general expression for the background wind (which has a random phase) given by the following stream function

$$\Phi = \sigma \sum_{r,n}^{R,N} \frac{1}{(r^2 + n^2)^\alpha} \cos (p n x + q r z + \phi_{r,n}),$$

where the $\phi_{r,n}$ are random variables, distributed uniformly between 0 and $2\pi$ (e.g. Grimmett & Welsh 1991); $p$ and $q$ are fixed parameters that control the scale of the wind in the horizontal and vertical respectively; $\alpha$ controls the rate of decay of consecutive cosine waves; $R$ and $N$ determine the number of cosine waves in the directions $z$ and $x$ respectively and $\sigma$ governs the amplitude of the fluctuations. These parameters can be varied to obtain as close a match to realistic winds as possible. We let $U$ and $W$ be given by a superposition of a vertical shear and the random velocity field obtained with $\Phi$ as a stream function, i.e.,

$$U = \Delta z + \Phi_z = \Delta z - \sigma \sum_{r,n}^{R,N} \frac{r q}{(r^2 + n^2)^\alpha} \sin (p n x + q r z + \phi_{r,n})$$

and

$$W = -\Phi_x = \sigma \sum_{r,n}^{R,N} \frac{n p}{(r^2 + n^2)^\alpha} \sin (p n x + q r z + \phi_{r,n}),$$

where $\Delta$ is the amplitude of the shear. The velocity field satisfies the incompressibility condition $\nabla \cdot U = 0$.

Plots of the wind (for $\Delta = 0$) showing the effect of the number of Fourier modes and that of the parameters $\alpha$ and $q$ can be found in Figure 4.9. The table gives the values of the parameters chosen for each plot. These parameters are not necessarily those that would be chosen to model a wind, but are used here for illustrative purposes. (1) is a reference graph with all parameters equal to 1. (2) and (3) demonstrate that as $q$ increases then so does the number of wavelengths of the wind in a given altitude range. (4) and (5) demonstrate the changing nature of the wind as we add more Fourier modes. (6) and (7) show that as $\alpha$ increases then the first Fourier mode becomes more dominant and adding subsequent modes has less effect. The correlation length of the background winds shown in Figure 4.9 is between about 1 m and 10 m. This correlation length is much smaller than the scale of the waves which have wavelengths of the order of a few kilometres.

We can make the wind more realistic by adding an amplitude $\psi$ for each $r$ and $n$, which is a random variable normally distributed with mean 0 and variance 1. With
Figure 4.9: Graphs showing random part of wind without shear ($\Delta = 0$) for $\sigma = 1$ and various different values of $\alpha$, $q$, $\mathcal{R}$ and $\mathcal{N}$ given in the table. (1) is a reference graph with all parameters equal to 1. (1), (2) and (3) demonstrate that as $q$ increases then so does the number of wavelengths of the wind in a given altitude. (1), (4) and (5) demonstrate the changing nature of the wind as we add more Fourier modes. (4), (6) and (7) show that as $\alpha$ increases then the first Fourier mode becomes more dominant and adding subsequent modes has less effect.
this amplitude, $U$ and $W$ are given by

$$U = \Delta z - \sigma \sum_{r,n}^{\mathcal{N}} \frac{\psi_{r,n} nq}{(r^2 + n^2)^2} \sin (pnx + qrz + \phi_{r,n})$$

and

$$W = \sigma \sum_{r,n}^{\mathcal{N}} \frac{\psi_{r,n} np}{(r^2 + n^2)^2} \sin (pnx + qrz + \phi_{r,n}).$$

### 4.4.5 Benchmark test

The single-ray code is tested against the multi-ray code as a benchmark. The benchmark is constructed as follows. The multi-ray code is used to generate two rays with a small separation distance. From these two rays the ray-tube area is estimated by calculating the area of a small quadrilateral contained between two close rays at two adjacent time steps. Referring to Figure 4.10, $p$ and $q$ are points (in $x-z$ space) along the left hand ray for the first two time steps; $r$ and $s$ are the same for the right hand ray. Therefore, the area $A_1$ is given by:

$$A_1 = \frac{1}{2} (p \times q + q \times r + r \times s + s \times p),$$

where for $p = (x_1, z_1)$ and $q = (x_2, z_2)$, the cross product $\times$ is given by

$$p \times q = \det \begin{pmatrix} x_1 & z_1 \\ x_2 & z_2 \end{pmatrix}.$$

At every time step a similar area is calculated. The ratio of the ray-area at a particular time step to the initial ray-area (for example $A_1/A_0$ in Figure 4.10) is then an equivalent parameter to the ray-tube area in the multi-ray code. The two codes are run for 100 different values of $\sigma$ in the range $(0, 1)$ using common initial values. The background wind is non-random: $\phi_{r,n} = 0.5$ and $\psi_{r,n} = 1$ and the parameters $\alpha = 1$, $\mathcal{R} = \mathcal{N} = 5$, $\mathcal{V} = 5$. 

Figure 4.10: Diagram showing calculation of ray-tube area for two close rays. $A_0$ is the initial ray-tube area. $A_1$ is the ray-tube area at a later time step.
Figure 4.11: Schematic showing the propagation of three rays.

Figure 4.12: Bench mark test for the single ray code and multi-ray code. Altitude level of caustic $z_c$ is plotted for 100 different values of $\sigma$ in the range $(0, 1)$ for both single-ray and multi-ray codes.

$q = 10$ and $p = 1$ are fixed. Both codes are run using a Runge-Kutta order 4 method, (see Appendix G). The results can be seen in Figure 4.12. As we can see a good match is obtained apart from a few results in the multi-ray code which are probably due to over-large ray-separation distances. This can be explained as follows: consider locating a caustic by evaluating the ray-tube area between two neighbouring rays, for example, the situation described in Figure 4.11. If the rays starting from $x$ and $x_2$ are neighbouring (ray starting from $x_1$ not present), then a caustic may be located at $z_2$. However if we halve the ray separation distance so that the rays starting from $x$ and $x_1$ are neighbouring, then a caustic may be located at $z_1$. The stray points in Figure 4.12 are not important and the match is good enough to be a successful benchmark test. However, it should be noted that the single-ray code is more robust.
4.5 Results

We now consider the statistics of caustic formation for rays propagating through a two-dimensional random wind. The random nature of the wind causes the rays to undergo many small deflections from their deterministic path and caustics can form. Figure 4.13 illustrates this by showing a typical run of the multi-ray code. The formation of caustics can clearly be seen at the dark areas where the rays overlap. Since the background wind is random, the altitude at which caustics are found is a random variable.

Figure 4.14 shows phase plots of the frequency \( \omega \) against \( k \cdot U \). Caustic events are shown as circles and critical level events as crosses. The separate clustering of caustics and critical levels (around the line \( \omega = k \cdot U \)) demonstrates that a ray hitting a critical level and a ray hitting a caustic are separate events and can be treated separately.

We now report results obtained using the single ray code and focus on caustic events. Figure 4.15 shows typical results of this code. Each point in a graph gives the altitude and relative wavenumber \( m/k \) of the first caustic encountered along a ray. It can be seen that the altitude at which caustics form is spread vertically. There is however, a dense cluster of points between about \( z = 50 \) and \( z = 100 \). This corresponds to dimensional altitudes of between 5 km and 10 km. Caustics are physically relevant to the stratosphere which is about 40 km deep.

Since the altitude \( z_c \) at which caustics occur is a random variable, we can obtain its PDF \( P(z_c) \), by a binning procedure (similar to that described in Section 3.5).
4.16 shows a typical $P(z_c)$, obtained for background wind parameters $\alpha = 1$ and $\sigma = 0.1$ (the curve has been smoothed using an arcsplines algorithm\(^7\)). There is a strong peak in the PDF at around $z_c = 10$, which is equivalent to an altitude of about 1 km.

The peak of $P(z_c)$ depends on the parameters used to generate the background wind ($\alpha$ and $\sigma$). This is clearly demonstrated by Figures 4.16 and 4.17 which show four PDFs for $\alpha = 1$ and different values of $\sigma$. It is natural to expect that, in the limit of small random perturbations $\sigma \to 0$, the PDF only depends on a simple scaling of $\sigma$. With a simple rescaling the PDFs for different background wind parameters can be made universal. The peak of the PDF will be in the same place irrespective of the statistics of the background wind. Specifically, we expect that if we define the new variable $\tilde{z}_c$ and the corresponding PDF $\tilde{P}$ by

$$\tilde{z}_c = \sigma^\nu z_c \quad \text{and} \quad \tilde{P}(\tilde{z}_c) = \sigma^{-\nu} P(z_c),$$

then $\tilde{P}(\tilde{z}_c)$ is independent of $\sigma$ in the limit $\sigma \to 0$ for a certain value of the scaling parameter $\nu$. We confirm that this is the case and test the hypothesis that $\nu = 1$. This value of $\nu$ is chosen based on the fact that the asymptotic analysis of White & Fornberg does not reveal any caustics for atmospheric gravity waves over the scale $\sigma^{-2/3}$; we postulate that caustics will occur over longer distances, i.e., $\sigma^{-1}$. We proceed as follows. For a fixed value of $\alpha$, we obtain $P(z_c)$ for several different values of $\sigma$. The maximum value of the PDF is obtained for each value of $\sigma$ and so we derive a relation between $z_{c(\text{max})}$ and $\sigma$.\(^8\) This relation is expected to be the power law $z_{c(\text{max})} \sim \sigma^{-\nu} \tilde{z}_{c(\text{max})}$, with $\tilde{z}_{c(\text{max})}$ independent of $\sigma$. This was proved by White & Fornberg (1998) with $\nu = 2/3$ for shallow water waves propagating through a random current.

\(^7\)In principle if the bin size was chosen correctly then smoothing is not necessary. However, this proved to be difficult in practice.

\(^8\)The universal PDF thus obtained will be highly biased towards the maximum value. It is possible to obtain a universal PDF by taking into account the functional form of $P(z_c)$ over the full range of $z_c$, but by doing this we would not obtain a simple power law.

![Figure 4.14](image)

Figure 4.14: Phase plots showing values of $k \cdot U$ against frequency $\omega$ for parameters:
(1) $\sigma = 0.3$; (2) $\sigma = 0.9$; $\alpha = 1$, $q = p = 1$ and $R = N = 5$ for 1000 rays. The crosses represent critical level events whereas the diamonds represent caustic events. The dotted line is that for $k \cdot U = \omega$.  

114
Figure 4.15: Phase plots showing the altitude at which caustics occur \( z_c \) against \( m/k \). Each point represents a ray hitting a caustic. 1000 rays were traced for two different values of the random background wind parameters \( \alpha \) and \( \sigma \): (1) \( \alpha = 1, \sigma = 0.01 \) and (2) \( \alpha = 2, \sigma = 0.01 \). Other parameters are \( q = p = 1 \) and \( \mathcal{R} = \mathcal{N} = 5 \).

For \( \alpha = 1 \), results were averaged over 1000 rays for 17 different values of \( \sigma \) in the range \((0.01, 1.0)\). \( 10^8 \) time steps are used for time \( t \) in the ranges \((0, 10000)\) for \( \sigma \) in the range \((0.01, 0.09)\) and \((0, 5000)\) for \( \sigma \) in the range \((0.1, 1.0)\). A background wind with random amplitude and five Fourier modes i.e., \( \mathcal{R} = \mathcal{N} = 5 \) is used. Figure 4.18 shows \( z_c(\text{max}) \) as a function of \( \sigma \) in log–log coordinates. The best fit lines given by the power laws

\[
\ln z_c = -\nu \ln \sigma - 0.2,
\]

where \( \nu = 1 \) and \( \nu = 2/3 \) are shown (dashed lines).

For \( \alpha = 2 \), results were averaged over 1000 rays for 17 different values of \( \sigma \) in the ranges \((0.07, 5)\) with \( t \) in the ranges \((0, 50000)\) for \( \sigma \) in the range \((0.07, 0.09)\) and \((0, 5000)\) for \( \sigma \) in the range \((0.1, 5.0)\). \( 10^8 \) time steps were used for time \( t \). The same background wind was used as for \( \alpha = 1 \). The relation between \( z_c \) and \( \sigma \) is shown in

Figure 4.16: PDF for \( z_c \), the height at which a caustic occurs. Background wind parameters are \( \alpha = 1 \) and \( \sigma = 0.1 \). The curve has been smoothed using an arcsplines algorithm.
Figure 4.17: PDFs for the altitude at which a caustic is found $z_c$ for background wind parameters: $\alpha = 1$ and (1) $\sigma = 0.03$; (2) $\sigma = 0.05$ and (3) $\sigma = 0.5$. The curves have been smoothed using an arcsplines algorithm.

Figure 4.18. In this case, the best fit lines given by the power laws

$$\ln z_c = -\nu \ln \sigma + 2.3,$$

where $\nu = 1$ and $\nu = 2/3$ are shown.

Thus, based only on the maximum $z_{c(\text{max})}$ of $P(z_c)$, it appears that both values $\alpha = 1$ and $\alpha = 2$ lead to the same scaling, namely $\nu = 1$. This result can be tested by plotting the scaled PDF $\tilde{P}(\tilde{z}_c)$ against $\tilde{z}_c$ for different values of $\sigma$ on the same graph. If the scaling holds, then all the curves should collapse on a single PDF. This is demonstrated in Figure 4.19 for $\alpha = 1$ and Figure 4.20 for $\alpha = 2$. The fit is not perfect but it is convincing nevertheless, given the relatively limited number of realisations used to approximate the PDF. Thus we can be relatively confident in our hypothesis that $\nu = 1$. Incidentally, it may also be noted that the universal PDF is not unlike a Poisson distribution.

The distance to the first caustic along a ray can be seen to scale like $z_c \sim \sigma^{-1}$. This scaling is different to that of the White & Fornberg (1998) theory. They found the scaling $z_c \sim \sigma^{-2/3}$. In other words, assuming the background media for the White & Fornberg theory and the theory in this thesis have the same characteristics, then caustics occur further down the ray for atmospheric gravity waves than they do for the general systems considered in White & Fornberg. This is consistent with the argument given at the beginning of this chapter. Note that the White & Fornberg theory also predicts that the PDF depends only on a few properties of the random velocity field.
Figure 4.18: Maximum $z_{c(\text{max})}$ of $P(z_c)$ as a function of $\sigma$ in log-log coordinates for: (1) $\alpha = 1$ and (2) $\alpha = 2$. Best of fit lines are shown for: (i) $\nu = 2/3$ and (ii) $\nu = 1.$
Figure 4.19: The universal PDF for $\tilde{P}(\tilde{z}_c)$ against $\tilde{z}_c$ for $\alpha = 1$. Five values of $\sigma$ are plotted: $\sigma = 0.04, 0.09, 0.2, 0.6$ and 0.9. The curves have been smoothed using an arcsplines algorithm.

Figure 4.20: Same as Figure 4.19 but with $\alpha = 2$ and $\sigma = 0.3, 0.5, 0.7, 0.9, 1.0, 2.0$ and 3.0.

and that this dependence is of a scaling form. Our numerical results do not allow us to test whether this is the also the case for gravity wave caustics.

4.6 Summary and discussion

Atmospheric gravity waves are modelled in a two-dimensional random background medium that is independent of time. When the correlation length of random fluc-
tations is much less than the length scale of the waves, the WKB approximation can be applied. This yields the ray-tracing equations. The solution of these equations along with the transport equation implies that there are amplitude singularities along the rays, known as caustics. In actual fact the singularities represent a breakdown in the modelling assumptions (the wavenumber and frequency are not slowly varying at a caustic) and the amplitudes are large but finite (if non-linear solutions are allowed) at the caustics. Since large wave amplitudes cause instability and hence wave breaking, the modelling of caustics is important.

White & Fornberg (1998) applied an asymptotic theory to the ray-equations when the standard deviation of the amplitude of the random fluctuations is assumed to be small. They showed that for generic dispersive systems, after propagating long $O(\sigma^{-2/3})$ distances, the rays develop caustics and the probability distribution of the distance along a ray to first caustic formation is given by a universal curve, i.e., the statistics of the random medium do not influence the shape of the caustic probability curve and only contribute scaling factors. This theory was described in Section 4.2 and it was shown that when applied to gravity waves it leads to the prediction that caustics do not appear over the generic $\sigma^{-2/3}$ distance. This suggests that the typical distances for the formation of gravity waves are much larger than $\sigma^{-2/3}$. The asymptotic theory required to capture this is likely to be extremely complicated, therefore a numerical approach was taken.

Caustics are located by using the fact that the ray-tube area is zero on caustic curves. A system of nine equations to model a ray-tube propagation were derived: the four ray equations; the four derived ray equations and the ray-tube area evolution equation, (see Appendix E). These were derived entirely geometrically, without appealing to a transport equation. The equations were then implemented in a numerical code.

Firstly, a code that traced multiple rays (using only the four ray-tracing equations) was developed which was used to visually spot the occurrence of caustics. This code was thoroughly tested against cases where rays can be traced analytically. Secondly, a code was developed to trace a single ray along with the associated ray-tube area. This was benchmark tested against the first code. This second code was used to trace a statistical ensemble of rays through a random background wind consisting of five Fourier modes.

PDFs were obtained for the altitude at which caustics occur in the atmosphere. It was observed that the maximum value of the PDF scales with $\sigma$ as $z_c \sim \sigma^{-1}$. This was found for both $\alpha = 1$ and $\alpha = 2$, where $\alpha$ is a parameter controlling the relative strength of consecutive Fourier modes of the background wind.

The scaling found is such that caustics occur later along a ray than in the theory of White & Fornberg who found the generic scaling $z_c \sim \sigma^{-2/3}$. In other words, assuming
the background media for the White & Fornberg theory and the theory in this thesis
have the same characteristics, caustics occur further down the ray for gravity waves
(and indeed for any conical waves) than they do for the generic systems considered in

Further improvements and extensions to the work described in this chapter could be
carried out:

1. The numerical solution of the derived ray equations requires highly intensive
computations, (see Section 4.4.3). With the given computer resources, it was
not possible to obtain results for large numbers of realisations, so the PDFs are
quite crude. More computer time would be necessary to produce more definitive
results. Also, to obtain PDFs for smaller values of $\sigma$, it is necessary to increase
the time interval and hence the number of time steps. We therefore had to limit
our calculations to moderately small values of $\sigma$. Calculations for smaller values,
for which the scaling should hold better, are desirable.

Given the highly intensive nature of the single-ray code, it is probably necessary
to develop a more specialised code if more realisations are to be calculated. The
stiffness in particular would need to be dealt with. It would be possible to calcu-
late the overlap of rays using the multi-ray code (as stiffness errors come from
the $J, k_x, k_z, m_x$ and $m_z$ equations), but then the problem of robustness must
be tackled (see Section 4.4.5).

2. The model may be extended to three dimensions. In principle, this is fairly easy
to do numerically as the codes written for the two-dimensional model can be very
easily generalised. However, this would require the solution of a coupled system
of 16 ordinary differential equations. More computer power would be required for
the solution of these.
Chapter 5

Conclusion

Internal gravity waves have an important effect on the large-scale circulation of the middle atmosphere. This thesis has contributed to furthering the knowledge of gravity waves in three main parts.

In Chapter 2 we derive transport equations for a general class of Hamiltonian systems in random inhomogeneous background media whose deterministic properties vary slowly compared with typical wavelengths. We consider the intermediate regime where the random perturbations of the media have spatial scale comparable to the wavelengths. These transport equations lead to the dispersion relation and ray equations. In this chapter we further generalisation the work by Ryzhik et al. (1996) and Guo & Wang (1999). We apply this theory to the system of atmospheric gravity waves propagating through a random background wind (varying in the vertical only) and obtain the transport equation

\[
\partial_t a_s(x, k, t) = \pm |k|^{-3} N k_2 (k_2 \partial_{x_1} - k_1 \partial_{x_2}) a_s(x, k, t) + U_0(z) \partial_{x_1} a_s(x, k, t) -
\begin{align*}
\frac{8\pi^2 k_1 (k_2^2 - k_1^2)^2}{k_2 |k| N} & R(2k_2) \sum_s \left( a_s(x, k^p, t) - a_s(x, k, t) \right),
\end{align*}
\]

where \( k^p = (k_1, -k_2) \) and \( a(x, k, t) \) is a scalar energy density. The two wave modes are represented by \( s = + \), corresponding to the positive sign in the dispersion relation and \( s = - \), corresponding to the negative sign.

In Chapter 3 we examine the Doppler spreading of gravity waves through a simple random background media modeled by an Ornstein-Uhlenbeck process. We consider a one-dimensional model by revisiting the work of Souprayen et al. (2001). We derive simple closed form analytical expressions in terms of altitude for the the energy spectrum induced by the gravity waves and the wave induced force of gravity waves on the background wind.

When the vertical wave number \( m \) is bounded by an upper value \( m_c \) (at which
breaking is assumed to occur) we obtain the energy spectrum when \( m_0 \gg k_0 \)

\[
\mathcal{E}(m; z) \simeq \frac{CNK\varepsilon_0m_0}{\sqrt{2\sigma m^3}} \exp \left[ -\frac{\omega^2}{2} \left( \frac{1}{m} - \frac{1}{m_0} \right)^2 \right] U \left( -\lambda - \frac{1}{2}, \sqrt{2\sigma} \left( \frac{1}{m} - \frac{1}{m_0} \right) \right) e^{-\lambda z},
\]

(5.1)

where \( C \) is a constant that depends on the density of wavepackets, \( N \) is the Brunt-Väisälä frequency, \( K \) and \( \lambda \) are constants derived from the first eigenfunction, \( \varepsilon_0 \) is the initial energy, \( m_0 \) is the initial vertical wavenumber, \( \sigma^2 \) is the variance of the Orstein-Uhlenbeck process that characterises the background wind and \( \omega = N/\sqrt{2\sigma} \).

It can be seen that in the large \( m \) limit the energy approaches an \( m^{-3} \) functional form, which is characteristic of observations. The stochastic model of this chapter can be viewed as one of the simplest models to date that predicts an \( m^{-3} \) spectral tail. The wave induced force as a function of altitude in this setting is given by

\[
\mathcal{F}(z) = \frac{A_0}{2} c_g(u_c)e^{-u_c^2/4} \sum_{i=1}^{P} \left\{ K_i \left[ -uU(-\lambda_i - 1/2, u_c) + \lambda_i U(-\lambda_i + 1/2, u_c) \right] e^{-\lambda_i z} \right\},
\]

(5.2)

where \( U \) is a parabolic cylinder function, \( A_0 \) is the initial action and \( p \) is the number of eigenfunctions used to control the desired level of accuracy.

The formation of caustics, due to the interaction of atmospheric gravity waves propagating through a random background wind, is another possible mechanism for gravity wave steepening and hence breaking, that is often overlooked.

In Chapter 4, a mainly numerical approach is taken to study atmospheric gravity waves in a two-dimensional time-independent random background wind. Probability density functions are provided for the altitude \( z_c \) at which a caustic forms as a function of the amplitude \( \sigma \) of the random part of the background wind. A universal scaling is obtained in the form \( z_c \sim \sigma^{-1} \). This scaling is different to that of the White & Fornberg (1998) theory. They found the scaling \( z_c \sim \sigma^{-2/3} \). In other words, assuming the background media for the White & Fornberg theory and the theory in this thesis have the same characteristics, then caustics occur further down the ray for atmospheric gravity waves than they do for the general systems considered in White & Fornberg. This is consistent with the argument given at the beginning of this chapter.
Appendix A

Calculations for Chapter 2

A.1 Details concerning the functions \( H \) and \( J \)

The function \( \hat{H}(x, \epsilon \partial_x) \) is given by the following relationship

\[
H(x, \epsilon \partial_x)f(x) = \int \hat{H}(k, \epsilon l) \tilde{f}(l)e^{i(k+l)x} \, dk \, dl.
\]

A similar expression exists for \( \tilde{J} \) but with a minus sign present in front of the integral.

Relationships between \( H \) and \( H^* \), \( J \) and \( J^* \) are derived as follows: Since \( \hat{H}(x, \epsilon \partial_x) \) is self-adjoint, then for two arbitrary vectors \( f(x) \) and \( g(x) \), we have the property

\[
\int g^*(x)H(x, \epsilon \partial_x)f(x) \, dx = \int f^*(x)H(x, \epsilon \partial_x)g(x) \, dx.
\]

Taking Fourier transforms of both sides we obtain

\[
\int \int \int \int \tilde{g}^*(k) \hat{H}(l, \epsilon m) \tilde{f}(m)e^{imx} \, dk \, dl \, dm = \int \int \int \tilde{f}^*(m) \hat{H}(l, \epsilon k) \hat{g}(k)e^{i(k+l)x} \, dk \, dl \, dm.
\]

Integration over \( x \) gives

\[
\int \int \int \tilde{g}^*(k) \hat{H}(l, \epsilon m) \tilde{f}(m) \delta(-k + l + m) \, dk \, dl \, dm = \int \int \int \tilde{f}^*(m) \hat{H}(l, \epsilon k) \hat{g}(k) \delta(k + l - m) \, dk \, dl \, dm.
\]

Integration over \( l \) gives

\[
\int \int \tilde{g}^*(k) \hat{H}(k - m, \epsilon m) \tilde{f}(m) \, dk \, dm = \int \int \tilde{f}^*(m) \hat{H}(m - k, \epsilon k) \hat{g}(k) \, dk \, dm
\]

\[
= \left( \int \int \tilde{f}^*(m) \hat{H}(m - k, \epsilon k) \hat{g}(k) \, dk \, dm \right)^*
\]

\[
= \int \int \tilde{g}^*(k) \hat{H}^*(m - k, \epsilon k) \tilde{f}(m) \, dk \, dm.
\]

This holds for all \( \hat{g} \) and \( \tilde{f} \) and hence

\[
\hat{H}^*(l, k) = \hat{H}(-l, k + \epsilon l).
\]
Similarly, \( \hat{J}^*(l, k) = -\hat{J}(-l, k + \epsilon l) \). \hfill (A.2)

Now to obtain expressions for \( \hat{H}^*_0, \hat{H}^*_{1/2} \) and \( \hat{H}^*_1 \) we substitute (2.8) into (A.1) and expand in powers of \( \epsilon \):

\[
\hat{H}^*_0(k, l) + \epsilon^{1/2} \hat{H}^*_{1/2}(k/\epsilon, l) + \epsilon \hat{H}^*_1(k, l) \\
= \hat{H}_0(-k, l + \epsilon k) + \epsilon^{1/2} \hat{H}_{1/2}(-k/\epsilon, l + k) + \epsilon \hat{H}_1(-k, l + \epsilon k) \\
= \hat{H}_0(-k, l) + \epsilon k \nabla_l \hat{H}_0(-k, l) + \epsilon^{1/2} \hat{H}_{1/2}(-k/\epsilon, l + k) + \epsilon \hat{H}_1(-k, l) + O(\epsilon^2).
\]

Equating powers of \( \epsilon \) we find

\[
\text{at } O(1) : \quad \hat{H}^*_0(k, l) = \hat{H}_0(-k, l), \\
\text{at } O(\epsilon^{1/2}) : \quad \hat{H}^*_{1/2}(k, l) = \hat{H}_{1/2}(-k, l + \epsilon k) \\
\text{and at } O(\epsilon) : \quad \hat{H}^*_1(k, l) = \hat{H}_1(-k, l) = k \nabla_l \hat{H}_0(-k, l).
\]

Multiplying this last equation by \( e^{-ikx} \) and integrating with respect to \( k \) we get

\[
\left( \int \hat{H}^*_1(k, l)e^{ikx} \, dk \right)^* - \int \hat{H}_1(-k, l)e^{-ikx} \, dk = \int k \nabla_l \hat{H}_0(-k, l)e^{-ikx} \, dk \\
\Rightarrow \quad \hat{H}^*_1(x, l) - H_1(x, l) = i \nabla_x \cdot \int \nabla_l \hat{H}_0(-k, l)e^{-ikx} \, dk = i \nabla_x \cdot \nabla_l \hat{H}_0(x, l) \\
\Rightarrow \quad H_1(x, l) - \hat{H}^*_1(x, l) = -i \nabla_x \cdot \nabla_l \hat{H}_0(x, l).
\]

A similar calculation can be made for \( J_1 \) exploiting the skew-symmetric property (A.2) to obtain

\[
J_1(x, l) + \hat{J}^*_1(x, l) = -i \nabla_x \cdot \nabla_l J_0(x, l).
\]

### A.2 Correlation functions

We give a derivation of the symmetries of the correlation functions given by (2.15).

Using the definition of the correlation function (2.11) and (2.14) we get

\[
J_{\alpha\gamma\delta\star}(l, k, n)\delta(l + m) = \langle \hat{J}_{\gamma\delta\star}^\dagger(m, n)\hat{J}_{\alpha\delta\star}(l, k + m) \rangle \\
= \langle \hat{J}_{1/2}^\dagger(-m, m + n)\hat{J}_{1/2}(l, k + m + l) \rangle \\
= \hat{J}_{\gamma\delta\alpha}^\dagger(-m, m + n + l, k + m + l)\delta(-l - m).
\]

Non-trivial solutions exist only when \( m = -l \). Hence, we obtain

\[
J_{\alpha\beta\gamma\delta\star}(l, k, n) = J_{\delta\gamma\alpha}(l, n, k).
\]

The proofs for \( H \) and \( K \) work in the same manner.
A.3 Energy and Wigner matrix

Introducing the Fourier expansion of $u(x, t)$ into (2.5) leads to

$$\mathcal{H} = \frac{1}{2} \int \int \int \dot{u}(l, t) H(x, \imath \epsilon k) \dot{u}(k, t) e^{i(k-l)x} \, dx \, dk \, dl$$

$$= \frac{1}{2} \operatorname{Tr} \int \int \int H(x, \imath \epsilon k) \dot{u}(k, t) \dot{u}^*(l, t) e^{i(k-l)x} \, dx \, dk \, dl.$$  

Changing the variables of integration and using (A.11) then gives

$$\mathcal{H} = \frac{1}{2(2\pi)^d} \operatorname{Tr} \iiint H(x, \epsilon k + \epsilon l/2) W_c(y, k, t) e^{i(x-y)} \, dx \, dy \, dk \, dl$$

$$= \frac{1}{2} \operatorname{Tr} \int \int H(x, \epsilon k + \epsilon \partial_x/2) W_c(x, k, t) \, dx \, dk.$$  

A.4 Dirac delta function

Informally consider the function

$$d_n(x) = \frac{n}{\pi(1 + n^2 x^2)}.$$  

This function has the following limits and integral properties

$$\lim_{n \to \infty} d_n(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases} \quad \text{and} \quad \lim_{n \to \infty} \int_a^b d_n(x) \, dx = 1 \quad \text{for} \quad a < 0 < b.$$  

In the limit $n \to \infty$, $d_n(x)$ is known as the Dirac delta function, (e.g. Riley & Hobson 1998, chapter 11), i.e.,

$$\delta(x) = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases}, \quad \int_{-\infty}^{\infty} \delta(x) \, dx = 1.$$  

The word "informally" was used earlier as this is not a proper mathematical definition due to $d_n$ not having a limit for all $x$. However the delta function usually appears inside an integral and if $\delta$ was replaced by $d_n$ and the integral calculated, then the limit $n \to \infty$ taken, it will generally be found that the integral has finite limits and is well behaved. The following properties hold for the delta function

$$\delta(-x) = \delta(x),$$

$$x \delta(x) = 0,$$  

$$x \frac{d \delta(x)}{dx} = -\delta(x),$$

$$\delta(ax) = \frac{1}{|a|} \delta(x)$$

and

$$\delta(x^2 - a^2) = \frac{1}{|a|} \left[ \delta(x-a) + \delta(x+a) \right].$$

\footnote{This is just one of the many representations that can be used to model the delta function.}
where \( a \) is a constant.

In \( n \) dimensions we define the delta function as

\[
\delta(x) = \delta(x_1)\delta(x_2) \cdots \delta(x_n) = \prod_{i=1}^{n} \delta(x_i).
\]

The following integral properties hold for \( \delta \) in \( n \) dimensions (e.g. Griffel 1993).

\[
\int_{-\infty}^{\infty} x \delta(x) \, dx = 0 \quad (A.8)
\]

and

\[
\int_{-\infty}^{\infty} F(x) \delta(x-y) \, dx = F(y). \quad (A.9)
\]

### A.5 Calculation of equation for \( W_\varepsilon \)

\[
W_\varepsilon(x, k, t) = \frac{1}{(2\pi)^d} \int u(x - \varepsilon y/2, t) u^*(x + \varepsilon y/2, t) e^{ik \cdot y} \, dy
\]

\[
= \frac{1}{(2\pi)^d} \iint \dot{u}(p, t) \dot{u}^*(q, t) e^{ip(x-\varepsilon y/2)} e^{-iq(x+\varepsilon y/2)} e^{ik \cdot y} \, dp \, dq.
\]

Now using the fact that

\[
\int e^{j f} \, dy = (2\pi)^d \delta(f),
\]

where \( f \) is independent of \( y \), then integrating over \( p \) (see appendix A.4) we get

\[
W_\varepsilon(x, k, t) = \iint \dot{u}(p, t) \dot{u}^*(q, t) e^{i\varepsilon p (x-\varepsilon y/2)} \delta \left( k - \frac{\varepsilon q}{2} - \frac{\varepsilon p}{2} \right) \, dp \, dq.
\]

Substituting \( q = k/\varepsilon - h \), we arrive at

\[
W_\varepsilon(x, k, t) = \left( \frac{2}{\varepsilon} \right)^d \int \dot{u}(k/\varepsilon + h, t) \dot{u}^*(k/\varepsilon - h, t) e^{2\varepsilon h \cdot x} \, dh.
\]

We now multiply by \( e^{-2\varepsilon y \cdot t} \) and integrate over \( y \) to obtain an expression for

\[
\dot{u}(k/\varepsilon + h, t) \dot{u}^*(k/\varepsilon - h, t), \text{ i.e.,}
\]

\[
\int W_\varepsilon(y, k, t) e^{-2\varepsilon y \cdot t} \, dy = \left( \frac{2}{\varepsilon} \right)^d \iint \dot{u}(k/\varepsilon + h, t) \dot{u}^*(k/\varepsilon - h, t) e^{2\varepsilon y \cdot (h-t)} \, dy \, dh
\]

\[
= \left( \frac{4\pi}{\varepsilon} \right)^d \int \dot{u}(k/\varepsilon + h, t) \dot{u}^*(k/\varepsilon - h, t) \delta(2(h-t)) \, dh.
\]

Integrating over \( h \) we get

\[
\dot{u}(k/\varepsilon + h, t) \dot{u}^*(k/\varepsilon - h, t) = \left( \frac{\varepsilon}{2\pi} \right)^d \int W_\varepsilon(y, k, t) e^{-2\varepsilon y \cdot h} \, dy. \quad (A.11)
\]
A.6 Span of null space of $Q_{00}$

The null space of $Q_{00}$ is all vectors $p$ such that $Q_{00}p = 0$, i.e., $L_0p + pL_0^* = 0$. This can be proved as follows: let $p = e_{(s)}^*e_{(s)}$, then

$$L_0e_{(s)}^*e_{(s)}^*L_0 = \omega_{(s)}e_{(s)}^*e_{(s)}^* + (\omega_{(s)}e_{(s)}^*e_{(s)}^*)^* = \omega_{(s)}(e_{(s)}^*e_{(s)}^* - e_{(s)}^*e_{(s)}^*) = 0.$$ 

Hence the null space of $Q_{00}$ is spanned by

$$e_{(s)}(x,k)e_{(s)}^*(x,k).$$

A.7 Calculation of $\mathcal{P}_{(s,t)}$

We introduce (2.44) into (2.43), and left- and right-multiply by $e_{(s)}^*(x,k+l)$ and $e_{(t)}^*(x,k-l)$. Using (2.34), (2.38) and (2.40), we obtain

$$\mathcal{P}_{(s,t)}(x,l,k,t) = a_{(s)}(x,k+l,t)e_{(s)}^*(x,k+l)V^*(x,-2l,k+l)e_{(t)}^*(x,k-l) + a_{(t)}(x,k-l,t)e_{(s)}^*(x,k+l)V(x,2l,k-l)e_{(t)}^*(x,k-l).$$

Using definition (2.29) of $V$, (2.36) and (2.37), this can be rewritten as

$$\mathcal{P}_{(s,t)}(x,l,k,t) = a_{(s)}(x,k+l,t)e_{(s)}^*(x,k+l)V^*(x,-2l,k+l)e_{(t)}^*(x,k-l) + a_{(t)}(x,k-l,t)e_{(s)}^*(x,k+l)V(x,2l,k-l)e_{(t)}^*(x,k-l).$$

This expression reduces to (2.45) on using (2.14).

A.8 Deterministic contributions to the transport equations

The deterministic terms in (2.47), namely $\tilde{e}_{(s)}^*(x,k)(Q_{01} + Q_{10})W_0(x,k,t)e_{(s)}(x,k)$, can in principle be simplified by introducing the form of $Q_{10}$ and $Q_{01}$, and the expansion (2.40) of $W_0$, and by making extensive use of the eigenvalue equation (2.34) and of its consequences. The relationships (2.9) between $J_1,H_1$ and $J_0,H_0$ are also crucial. Here, we circumvent most of the tedious computations that this entails by exploiting the conservation of energy. We first note that, in the absence of random terms, (2.47) takes the form

$$\partial_t a_{(s)}(x,k,t) + \xi_{(s)}(x,k) \cdot \nabla_x a_{(s)}(x,k,t) - g_{(s)}(x,k) \cdot \nabla_k a_{(s)}(x,k,t)$$

$$= a_{(s)}(x,k,t)h_{(s)}(x,k) + \sum_t h_{(s,t)}(x,k)a_{(t)}(x,k,t).$$

(A.12)
This holds for \( s = t \) when we make use of the fact that \( \delta_{st} \). Here, we have defined

\[
f_{(s)}(x, k) = \frac{1}{2} \left[ \hat{e}_{(s)}^*(x, k) \nabla_k L_0(x, k) e_{(s)}(x, k) - \hat{e}_{(s)}^*(x, k) \nabla_k L_0^*(x, k) \hat{e}_{(s)}(x, k) \right],
\]

\[
g_{(s)}(x, k) = \frac{1}{2} \left[ \hat{e}_{(s)}^*(x, k) \nabla_x L_0(x, k) e_{(s)}(x, k) - \hat{e}_{(s)}^*(x, k) \nabla_x L_0^*(x, k) \hat{e}_{(s)}(x, k) \right],
\]

\[
b_{(s, s)}(x, k) = \frac{1}{2} \hat{e}_{(s)}(x, k) \left[ \nabla_k L_0^*(x, k) + \nabla_x L_0^*(x, k) - \nabla_k \nabla_x L_0(x, k) \right.
\]
\[+ \left( J_1(x, k) H_0(x, k) + J_0(x, k) H_1(x, k) \right)^* \]
\[\left. \hat{e}_{(s)}^*(x, k) \right] - \text{c.c.}
\]

\[
h_{(s, t)}(x, k) = \frac{1}{2} \hat{e}_{(s)}(x, k) \left[ \nabla_x e_{(s)}(x, k) e_{(t)}(x, k) \nabla_k L_0^*(x, k),
\right.
\]
\[+ \nabla_k e_{(s)}(x, k) e_{(t)}(x, k) \nabla_x L_0^*(x, k) \right] \hat{e}_{(t)}(x, k) - \text{c.c.}
\]

and we have grouped all the terms involving undifferentiated amplitudes \( a(t)(x, k, t) \) and \( a(s)(x, k, t) \) on the right-hand side. The scalars \( h_{(s, t)} \) and \( b_{(s, s)} \) are written in terms of \( J_0, H_0, J_1, H_1, L_0, e_{(s)}, \hat{e}_{(s)}, e_{(t)}, \hat{e}_{(t)} \) and their derivatives with respect to \( x \) and \( k \). Now, differentiating (2.34) with respect to \( k \) and \( x \), respectively, leads, after left-multiplying by \( \hat{e}_{(s)}^*(x, k) \) and using (2.38), to

\[
\hat{e}_{(s)}^*(x, k) \nabla_k L_0(x, k) e_{(s)}(x, k) = -i \nabla_k \omega_{(s)}(x, k)
\]

and

\[
\hat{e}_{(s)}^*(x, k) \nabla_x L_0(x, k) e_{(s)}(x, k) = -i \nabla_x \omega_{(s)}(x, k).
\]

Therefore, \( f_{(s)}(x, k) = \nabla_k \omega_{(s)}(x, k) \) and \( g_{(s)}(x, k) = \nabla_x \omega_{(s)}(x, k) \).

We now argue that the scalars \( h_{(s, t)}(x, k) \equiv 0 \) for all \( s \) and \( t \). Differentiating (2.34) with respect to \( k \) or \( x \) yields

\[
\hat{e}_{(t)}^* \nabla_k L_0 e_{(s)} = i(\omega_{(t)} - \omega_{(s)}) \hat{e}_{(t)}^* \nabla_k e_{(s)} \quad \text{and} \quad \hat{e}_{(t)}^* \nabla_x L_0 e_{(s)} = i(\omega_{(t)} - \omega_{(s)}) \hat{e}_{(t)}^* \nabla_x e_{(s)}
\]

for \( s \neq t \). Taking this into account, we find that \( h_{(s, t)} = 0 \) for \( s \neq t \) and thus that (A.12) reduces to

\[
\partial_t a_{(s)}(x, k, t) + \nabla_k \omega_{(s)}(x, k) \cdot \nabla_x a_{(s)}(x, k, t) - \nabla_x \omega_{(s)}(x, k) \cdot \nabla_k a_{(s)}(x, k, t)
\]
\[= B_{(s, s)}(x, k) a_{(s)}(x, k, t),
\]

where \( B_{(s, s)} = b_{(s, s)} + h_{(s, s)} \). It is now easy to show that \( B_{(s, s)}(x, k) \equiv 0 \): in view of (2.41), the conservation of the energy \( \mathcal{H} \) implies at leading order in \( \epsilon \) that

\[
\frac{d}{dt} \sum_s \iint a_{(s)}(x, k, t) \, dx \, dk = \sum_s \iint B_{(s, s)}(x, k) a_{(s)}(x, k, t) \, dx \, dk = 0.
\]

Since this holds for arbitrary \( a_{(s)}(x, k, t) \), \( B_{(s, s)}(x, k) \equiv 0 \) and the Liouville form (2.48) of the transport equations follow. 

128
A.9 Derivation of $I_J$, $I_H$ and $I_K$

In this appendix, we omit the dependence of the various objects on $x$ and $t$. We introduce (2.44)-(2.45) into (2.49) and use (2.11)-(2.13) to obtain explicit expressions for the three terms $I_J$, $I_H$ and $I_K$. This gives

\[ \tilde{e}^{*}_{(s)}(k) (\tilde{Q}_{1,0} W_{1/2}(\xi, k)) \tilde{e}_{(s)}(k) \]

\[ = 2^d \tilde{e}^{\alpha*}_{(s)}(k) \int F_{\alpha\beta}(m, k + l) V^{*}_{\beta\gamma}(-l, k + l - m) \]

\[ + V_{\alpha\beta}(2l, k - l + m) F_{\beta\gamma}(m, k - l) e^{2i(m+l)\xi} \, dl \, dm \tilde{e}^{*}_{(s)}(k) \]

\[ = -2^d i \tilde{e}^{\alpha*}_{(s)}(k) \int \frac{j_{1/2}^{\alpha\beta}(2l, k - l + m) H_{0}^{\alpha\beta}(i(k - l + m))}{\omega(t)[k - l + m] - \omega(s)[k - l - m]} \times \]

\[ \times \left\{ -[a(t)(k - l + m) - a(s)(k - l - m)] \tilde{e}^{\alpha*}_{(t)}(k - l + m) \times \right\} \times \]

\[ \times \tilde{e}^{\alpha*}_{(t)}(k - l + m) \tilde{e}^{b*}_{(t)}(k - l - m) e^{2i(m+l)\xi} \, dl \, dm \tilde{e}^{*}_{(s)}(k) + \text{c.c.} \]

Now taking the contribution from $J_{1/2} J_{1/2}$ only and using the correlation function (2.11), we obtain

\[ I_J = 2^d \tilde{e}^{\alpha*}_{(s)}(k) \int H_{0}^{\alpha\beta}(i(k - 2l)) \frac{\alpha(t)(k - 2l) - \alpha(s)(k)}{\omega(t)(k - 2l) - \omega(s)(k)} j_{0}^{\alpha\beta}(2l, k, k) \times \]

\[ \times \tilde{e}^{\alpha*}_{(t)}(k - 2l) \tilde{e}^{b*}_{(t)}(k) e^{2i(l+m)\xi} \, dl + \text{c.c.} \]

A simple computation using (2.36) and (2.38) gives

\[ I_J = 2^d \sum_{l} \int \frac{\alpha(t)(k - 2l) - \alpha(s)(k)}{\omega(t)(k - 2l) - \omega(s)(k)} \Phi_{(s,t)}(2l, k) \, dl + \text{c.c.}, \]

with $\Phi_{(s,t)}$ defined in (2.51). This expression is of the form

\[ \int \frac{f(\omega)}{\Omega - \omega} d\omega, \]

which has a single pole at $\omega = \Omega$. To allow this integral to be well defined we must regularise the denominator so that the pole does not lie on the real line (cf. Ryzhik et al. 1996). Causality indicates that this is achieved by adding $-i\theta$ to the denominator. Taking the limit $\theta \downarrow 0$ we obtain

\[ \frac{1}{\Omega - \omega - i\theta} = \frac{\Omega - \omega}{(\Omega - \omega)^2 + \theta^2} + \frac{i\theta}{(\Omega - \omega)^2 + \theta^2} \rightarrow PV \frac{1}{\Omega - \omega} + i\pi \delta(\Omega - \omega), \]

where $PV$ denotes the Cauchy principal value and we have used

\[ \frac{\theta}{(\Omega - \omega)^2 + \theta^2} = \frac{n}{1 + n^2 (\Omega - \omega)^2} = \pi d_n(\Omega - \omega), \]

129
with \( n = 1/\theta \) and \( d_n \) defined in Section A.4. So, in other words, we make the substitution

\[
\frac{1}{\omega(t)(k-2l) - \omega(s)(k)} \rightarrow \text{PV} \frac{1}{\omega(t)(k-2l) - \omega(s)(k)} + \pi \delta[\omega(t)(k-2l) - \omega(s)(k)]. \tag{A.13}
\]

Using the property (2.15) of \( J \), it is easy to show that \( \Phi^J_{s,t}(2l,k) \) is real. Since the energy density and frequency are also real, then \( I_J \) is also real. The only contribution to \( I_J \) comes from the regularisation of the denominator, leading to

\[
I_J = -2^d \sum_t \int \Phi^J_{s,t}(2l,k) \delta[\omega(t)(k-2l) - \omega(s)(k)][\alpha(t)(k-2l) - \alpha(s)(k)] \, dl + \text{c.c.}
\]

Changing the integration variable yields (2.50). The treatment of \( I_H \) is similar: using (2.37) and (2.38), it can be written in the form

\[
I_H = -2^d \omega(s)(k) \sum_t \int \frac{\omega(s)(k)\alpha(t)(k-2l) - \omega(t)(k-2l)\alpha(s)(k)}{\omega(t)(k-2l) - \omega(s)(k)} \Phi^H_{s,t}(2l,k) \, dl + \text{c.c.},
\]

with \( \Phi^H_{s,t} \) defined in (2.53). Using (2.15) shows that \( \Phi^H_{s,t} \) is real, so that only the regularisation of the denominator contributes to \( I^H \). We use (A.13) and change the variable of integration. We also use the fact that the principle value gives a factor of \( \delta[\omega(t)(k) - \omega(s)(k-2l)] \) which is only non-zero when \( \omega(t)(k) = \omega(s)(k-2l) \). This leads to (2.52).

The third term, \( I^K \) is computed in a similar, though lengthier, manner. Using (2.36), (2.37) and (2.38), it can be written as

\[
I^K = -2^d \left[ \omega(s)(k) \sum_t \int \frac{\alpha(t)(k-2l) - \alpha(s)(k)}{\omega(t)(k-2l) - \omega(s)(k)} \Phi^H_{s,t}(2l,k) \, dl + \text{c.c.,}
\right. \]

\[
+ \left. \alpha(s)(k) \sum_t \int K^{\alpha\beta\gamma\delta}(2l,k,k)\hat{e}^{\alpha\gamma}(k)\hat{e}^{\beta\delta}(k-2l)\hat{e}^{\gamma\delta}(k-2l)\hat{e}^{\alpha\delta}(k) \, dl \right] + \text{c.c.,}
\]

where \( \Phi^H_{s,t} \) defined in (2.55). Taking (2.39) into account, the term on the second line can be simplified into

\[
-2^d \alpha(s)(k) \int K^{\alpha\beta\gamma\delta}(2l,k,k)\hat{e}^{\alpha\gamma}(k)\hat{e}^{\beta\delta}(k) \, dl + \text{c.c.,}
\]

since \( K^{\alpha\beta\gamma\delta}\hat{e}^{\beta\delta} = K^{\alpha\beta\gamma\delta}I^{\beta\gamma} = K^{\alpha\beta\gamma\delta} \).

Using (2.15), it can be shown that \( \Phi^K_{s,t} \) is real. Thus, when substituting (A.13) into the first line of \( I^K \), only the regularisation term remains, leading to (2.54) after changing the variable of integration.
Appendix B

Critique of work by C.O. Hines

In Hines (1992) gravity waves are modelled by Doppler spreading. The one-dimensional (in altitude) ray tracing system is used, where the rays are refracted and advected by a statistically random background wind given by the Gaussian probability function

\[
p(U) = \frac{1}{\sqrt{\pi \sigma}} e^{-U^2/2\sigma^2}, \tag{B.1}
\]

(equation preceded by an “H” refer to equation numbers in Hines (1992)) where each ray experiences a different realisation of the wind, (see Section 3.2). The resultant energy spectra obtained have a \(m^{-3}\) form for large \(m\). However, these energy spectra are highly dependent on the input spectra and so do not appear to be universal.

In his analysis Hines arrives at the equation

\[
\mathcal{E} = \frac{e^{-(m_0^{-1} - m^{-1})^2}}{\sqrt{\pi mm_0}}, \tag{B.2}
\]

for the energy spectra \(\mathcal{E}\), where \(m\) is a non-dimensional vertical wavenumber and \(m_0\) is an initial vertical wavenumber. This is Hines’ equation numbered [H2.20]. Hines uses slightly different notation: \(\Theta_2^2\) for the energy spectrum and \(m_0\) for the initial vertical wavenumber. The notation of this thesis is used here throughout.

Equation (B.2) clearly tends to a \(m^{-1}\) form in the large \(m\) limit. However Hines claims this shows an \(m^{-3}\) form! This is achieved by only showing a limited range in \(m\) so that the true large \(m\) limit cannot be seen. In Figure B.1, a reproduction of the plot used to show the \(m^{-3}\) tail can be seen for various values of \(m_0\), but here the horizontal axis has been extended further than that of the original paper. Hines claims that in the region immediately to the large \(m\) side of the peak, a slope of order \(m^{-3}\) can be observed. Hines himself realized (Hines 1996a), that his Doppler spread theory fails by predicting the \(m^{-3}\) tail only for intermediate \(m\) and not for large \(m\). This, he argues might be a consequence of the Doppler shift approximation being inadequate for large values of the vertical wavenumber, where neglected vertical components of the wind become important.\(^1\) He also mentions that to conform with observations it may not be

\(^1\)Background winds with vertical shears are modeled by Kuo & Lue (1994). They highlight the possibility of energy spectra different to that of \(m^{-3}\).
necessary to resolve the energy spectrum for large values of $m$. These claims are argued in detail in Hines (1992). Here however, we present a minor correction to the analysis of Hines (1992) which shows that his Doppler spreading theory does, in fact, lead to an $m^{-3}$ tail for large $m$. His calculation is now repeated with slight modifications.

For vertically propagating gravity waves in a vertically dependent horizontal background wind, the frequency $\omega$ is given by

$$\omega = \omega - kU,$$

where in the hydrostatic limit (used by Hines) $\omega = -Nk/m$. The ray equations (see Section 1.5.1) give

$$\frac{dm}{dz} = -\frac{m^2}{N} \frac{dU}{dz},$$

which when integrated gives

$$m^{-1} - m_0^{-1} = \frac{U(z)}{N}. \tag{B.3}$$

Now the product of action and ray-tube area is conserved along a ray (see Appendix E):

$$J(m)A(m) = J_0(m)A_0(m). \tag{B.4}$$

In the one dimensional system considered here, $A(m) = \mathcal{E}(m)\hat{\omega}(m)$ and $J(m) = dm/dm_0$. So for a single ray we can obtain an expression for the energy along a ray. Using (B.4) and the derivative of (B.3) with respect to $m_0$ we find [H2.11]:

$$\mathcal{E}(m) = \frac{\mathcal{E}_0m_0}{m}. \tag{B.5}$$

It is at this stage that Hines essentially makes his mistake. He takes an ensemble of rays, each one starting with a different value of $m_0$ and experiencing a different realisation.
of the wind, so the total energy $\mathcal{E}_t$ is given by [H2.13]:

$$\mathcal{E}_t(m) = \int \mathcal{E}(m)p(U)dU.$$  

After calculating this [H2.18], he then assumes $m_0$ to be given by a $\delta$ distribution which leads to (B.2), [H2.20]. This assumption would not in itself create problems, but then to make this calculation Hines assumes that there is a connection between the wind and the initial wavenumber $m_0$, [H2.12]:

$$dU = \left| \frac{dU}{dm_0} \right| dm_0.$$  

This is conceptually flawed: there is no such connection between $U$ and $m_0$; the connection is between $U$ and $m$. This can be corrected. At a given altitude the probability of finding a ray with wavenumber $m$, given the ray had initial wavenumber $m_0$ is

$$p(m|m_0) = p(U|m_0) \left| \frac{dU}{dm} \right| = p(N(m^{-1} - m_0^{-1})|m_0) \frac{N}{m^2}.$$  

(B.6)

Substituting into this (B.1) for the wind we obtain

$$p(m|m_0) = \frac{N}{\sqrt{\pi} \sigma m^2} e^{-N^2(m^{-1} - m_0^{-1})^2/2\sigma^2}.$$  

Substituting this into (B.5) we obtain

$$\mathcal{E}(m|m_0) = \frac{\mathcal{E}_0 m_0 N}{\sqrt{\pi} \sigma m^3} e^{-N^2(m^{-1} - m_0^{-1})^2/2\sigma^2}.$$  

(B.7)

This equation clearly shows we have a $m^{-3}$ spectral tail. Notice that the energy does not scale like $N^2$ but like $N$. This is equivalent to the Doppler theory of Section 3.6.

Non-dimensionalising (B.7) by letting $m = \hat{m}N/\sqrt{2\sigma}$ and dropping hats, we obtain

$$\mathcal{E}(m|m_0) \sim \frac{1}{m^3} e^{(m^{-1} - m_0^{-1})^2}.$$  

(B.8)

This result is replotted in figure B.2. It is quite clear that an $m^{-3}$ tail in the energy spectrum can be seen for large $m$.
Appendix C

Basic stochastics

This section will describe three basic stochastic processes: Brownian motion, the Wiener process and the Ornstein–Uhlenbeck process, with properties given. We also list some basic results in stochastic analysis: the correlation formula, Fokker–Plank equations and the normal distribution.

C.1 Stochastic processes

Brownian motion is that which describes the path of a particle that is frequently colliding with other particles. This path is naturally extremely rough. If a whole cloud of particles were to be observed then it would be seen that the cloud will spread out over time and the concentration of particles in the cloud will vary smoothly. This spreading out of the cloud is known as diffusion.

The Wiener process $W(t)$ is a mathematical function designed to model the path of a single particle undergoing Brownian motion over time $t$. It is described by the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$$

and satisfies the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \frac{\partial^2 f}{\partial x^2}.$$ 

The diffusion equation is equivalent to a Fokker–Planck equation (see Section C.3) with drift coefficient zero and diffusion coefficient 1.

Wiener processes are often used to model noisy systems. If the noise is due to many independent random increments then the net effect of the increments will (by the central limit theorem, Grimmett & Welsh e.g. 1991) be distributed normally. There are some properties of the Wiener process that will be useful in further chapters (see for example any standard text on stochastic analysis, such as Gardiner 1985, Breźniak & Zastawniak 2000) and will be listed here for convenience.

1Named after Robert Brown, the British botanist who observed the random movement of pollen particles in water in 1827.
Figure C.1: Schematic showing sample paths for the Wiener process \( W(t) \) and the O-U process \( U(t) \). The variance of the Wiener process increases with \( t \) whereas the variance for the O-U process remains fixed.

1. \( W(0) = 0 \).

2. \( \mathbb{E}[W(t)] = 0 \), and \( \text{Var}[W(t)] = t \).

3. For any \( 0 \leq t' < t \) the increment \( W(t) - W(t') \) has the normal distribution with \( \mathbb{E}[W(t) - W(t')] = 0 \) and \( \text{Var}[W(t) - W(t')] = t - t' \).

4. \( \mathbb{E}[(W(t) - W(t'))^2] = |t - t'| \).

Since the Wiener process has an increasing variance with time then it is not stationary. However if a drift term is added to the Wiener process as a restoring force, we have a new process: the Ornstein–Uhlenbeck process \( U(t) \), which is stationary after a long time. The difference in sample paths between the Wiener process and the O-U process is demonstrated in Figure C.1. The Ornstein–Uhlenbeck process is described by the probability density function \( f(x) \), which satisfies the Fokker–Planck equation

\[
\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}(kxf) + \frac{1}{2}D \frac{\partial^2 f}{\partial x^2}
\]

and has the following properties (e.g. Gardiner 1985, Breźniak & Zastawniak 2000)

1. \( U(0) = 0 \).

2. \( \mathbb{E}[U(t)] = 0 \) and \( \text{Var}[U(t)] = \frac{D}{2k}(1 - e^{-2kt}) \).

### C.2 The correlation formula

If \( g(t) \) and \( h(t) \) are arbitrary continuous non-anticipating (i.e., values of \( h(t') \) for \( t' > t \) do not depend on \( g(t) \) and vice versa.) functions then the correlation formula for Wiener processes (e.g. Gardiner 1985) states that

\[
\int_{t_0}^{t} g(t')dW(t') \int_{t_0}^{t} h(t')dW(t') = \int_{t_0}^{t} g(t')h(t')dt',
\]

where the over-bar denotes the statistical average.
C.3 Fokker–Planck equations

For a given random function a partial differential equation can be derived (e.g. Gardiner 1985) for the conditional probability \( f(x, z) = p(x, z|z_0, x_0) \), for any initial \( x_0, z_0 \), with the initial condition \( p(x, z_0|z_0, x_0) = \delta(x - x_0) \), where \( \delta \) is the Dirac delta function. This PDE is known as the Fokker–Plank equation (also known as the Smoluchowski equation, the Klein–Kramer equation and the Kolmogorov equation in some mathematical texts) and is given by

\[
f(x, z) = -[A(x, z)f(x, z)]_x + \frac{1}{2}B^2(x, z)f(x, z),
\]

where \( A \) and \( B \) are functions that depend on the random function that the equation represents. \( A \) is a drift term and \( B \) is a diffusion term. The Fokker–Plank equation has an equivalent stochastic differential equation given by

\[
dx(z) = A[x(z), z]dz + B[x(z), z]dW(z).
\]

C.4 The normal distribution

The probability density function \( f(x) \) of a normal distribution is given by (e.g. Grimmett & Welsh 1991)

\[
f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),
\]

where \( \mu \) is the mean and \( \sigma^2 \) the variance.
Appendix D

Eigenvalues for finite domain solution

To calculate the eigenvalues for various different initial conditions for the finite domain solution it was necessary to find the solution to equation (3.4.3). This was carried out using Maple's "fsolve" routine. There are many texts available on Maple, for example, Char et al. (1992) and Heck (2003).

The exact eigenvalues for various different initial conditions and upper boundaries are given in the tables below. If the values of \( u_l \) and \( u_c \) derived from \( m_0, k_0, m_l \) and \( m_c \) are too large, i.e., about order 10 then the parabolic cylinder routines used in Maple are unable to compute values. When this occurs we must rely on the asymptotic expression (3.31) for the eigenvalue \( \lambda \).

The following tables give the first and second eigenvalues for different values of \( m_0 \), \( k_0 \) and \( m_c \). These are values that are used within Chapter 3 for graphs and other results.

<table>
<thead>
<tr>
<th>( m_c/m_0 )</th>
<th>( k_0 = 0.0003, m_0 = 0.0003 )</th>
<th>( k_0 = 0.0007, m_0 = 0.05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \lambda_1 )</td>
<td>( \lambda_2 )</td>
</tr>
<tr>
<td>2</td>
<td>1.4915 \times 10^{-4}</td>
<td>1.000266448</td>
</tr>
<tr>
<td>4</td>
<td>1.2423 \times 10^{-5}</td>
<td>1.000266386</td>
</tr>
<tr>
<td>6</td>
<td>1.2423 \times 10^{-5}</td>
<td>1.000266386</td>
</tr>
<tr>
<td>8</td>
<td>1.2423 \times 10^{-5}</td>
<td>1.000266386</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( m_c/m_0 )</th>
<th>( k_0 = 0.0007, m_0 = 0.007 )</th>
<th>( k_0 = 0.0007, m_0 = 0.005 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \lambda_1 )</td>
<td>( \lambda_2 )</td>
</tr>
<tr>
<td>2</td>
<td>0.741633</td>
<td>2.601287997</td>
</tr>
<tr>
<td>4</td>
<td>0.629396</td>
<td>2.419327042</td>
</tr>
<tr>
<td>6</td>
<td>0.594497</td>
<td>2.361337242</td>
</tr>
<tr>
<td>8</td>
<td>0.577516</td>
<td>2.332847145</td>
</tr>
</tbody>
</table>

Table D.1: Values of first and second eigenvalues for various values of upper boundary \( m_c \) and initial values \( k_0 \) and \( m_0 \). With parameters \( \gamma = 1/400 \) m\(^{-1}\), \( N = 0.02 \) s\(^{-1}\) and \( \sigma = 4 \) ms\(^{-1}\)
<table>
<thead>
<tr>
<th>$m_c/m_0$</th>
<th>$k_0 = 0.0007, m_0 = 0.001$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_1$</td>
</tr>
<tr>
<td>2</td>
<td>0.167434</td>
</tr>
<tr>
<td>4</td>
<td>0.027527</td>
</tr>
<tr>
<td>6</td>
<td>0.016126</td>
</tr>
<tr>
<td>8</td>
<td>0.013429</td>
</tr>
</tbody>
</table>

Table D.2: Same as Table D.1

<table>
<thead>
<tr>
<th>$m_c/m_0$</th>
<th>$k_0 = 0.0007, m_0 = 0.0007$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_1$</td>
</tr>
<tr>
<td>2</td>
<td>0.258852</td>
</tr>
<tr>
<td>3</td>
<td>0.110989</td>
</tr>
<tr>
<td>4</td>
<td>0.089977</td>
</tr>
<tr>
<td>5</td>
<td>0.854683</td>
</tr>
<tr>
<td>6</td>
<td>0.081111</td>
</tr>
<tr>
<td>7</td>
<td>0.835868</td>
</tr>
<tr>
<td>8</td>
<td>0.083344</td>
</tr>
<tr>
<td>9</td>
<td>0.083171</td>
</tr>
<tr>
<td>10</td>
<td>0.083142</td>
</tr>
<tr>
<td>11</td>
<td>0.083096</td>
</tr>
</tbody>
</table>

Table D.3: Same as Table D.1 but including third eigenvalue. These are used for calculating wave induced force in Section 3.7.

<table>
<thead>
<tr>
<th>$m_c/m_0$</th>
<th>$k_0 = 0.0007, m_0 = 0.0007$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_1$</td>
</tr>
<tr>
<td>2</td>
<td>-0.806219</td>
</tr>
<tr>
<td>3</td>
<td>-0.667667</td>
</tr>
<tr>
<td>4</td>
<td>-0.637900</td>
</tr>
<tr>
<td>5</td>
<td>-0.625453</td>
</tr>
<tr>
<td>6</td>
<td>-0.628025</td>
</tr>
<tr>
<td>7</td>
<td>-0.627041</td>
</tr>
<tr>
<td>8</td>
<td>-0.626572</td>
</tr>
<tr>
<td>9</td>
<td>-0.626319</td>
</tr>
<tr>
<td>10</td>
<td>-0.626169</td>
</tr>
<tr>
<td>11</td>
<td>-0.626074</td>
</tr>
</tbody>
</table>

Table D.4: Same as Table D.3, but for values of the constants $K_i$ used in the eigenfunction expansion (3.22).
Appendix E

System of equations for ray tube code

The location of a caustic is found by determining when the ray-tube area tends to zero. This problem was tackled by Hayes (1970) who introduced the notion of “derived ray equations”. However, Hayes relies on the principle of conservation of wave action or wave (pseudo) energy which does not necessarily hold when the background flow is not a solution of the unforced fluid equations. Alternatively, we follow the derivation of White & Fornberg (1998) which is entirely geometrical.

Referring to Figure E.1, the derivation of the derived ray-equations is as follows. Consider all rays to be initialized on a curve $\mathbf{x}_0(\alpha)$ parametrized by arclength $\alpha$. In the case of two-dimensional gravity waves, $\mathbf{x}_0$ could be along the $x$-axis, i.e., $\mathbf{x}_0 = (x, 0)$. We let the rays be parametrized by $\mathbf{x}(t, \alpha)$ and $k(t, \alpha)$, where $t$ is an along ray coordinate and $\alpha$ denotes the starting point of the ray along $\mathbf{x}_0$. The ray equations in these coordinates are given by

$$\frac{\partial}{\partial t} \mathbf{x}(t, \alpha) = \frac{\partial}{\partial k} \omega(\mathbf{x}(t, \alpha), k(t, \alpha))$$

(E.1)

and

$$\frac{\partial}{\partial t} k(t, \alpha) = - \frac{\partial}{\partial x} \omega(\mathbf{x}(t, \alpha), k(t, \alpha)).$$

(E.2)

The unit tangent, $t$ to a ray is given by

$$t(t, \alpha) = \frac{\omega_k(x, k)}{|\omega_k(x, k)|}$$

(E.3)

and we denote the unit normal to $t$ by $n(t, \alpha)$. The Serret–Frennet formulae (e.g. Roe 1995) now give relationships between the normal and tangent to a ray:

$$\frac{\partial t}{\partial t} = |\omega_k| \kappa n \quad \text{and} \quad \frac{\partial n}{\partial t} = -|\omega_k| \kappa t,$$

(E.4)

where $\kappa$ is the ray curvature. An expression for the curvature is found by substituting (E.3) into the left equation in (E.4) and we find

$$\kappa = \frac{1}{|\omega_k|} \left( n_i \frac{\partial^2 \omega}{\partial k_i \partial x_j} \xi_j - n_i \frac{\partial^2 \omega}{\partial k_i \partial k_j} \frac{\partial \omega}{\partial x_j} \frac{1}{|\omega_k|} \right).$$
Figure E.1: Schematic showing two rays and the various vectors involved in the derivation of the evolution equations.

For simplicity, let us now restrict our attention to a two-dimensional system. We let $\gamma = (t, \alpha)^T$ so that

$$x_\gamma = \begin{pmatrix} x_t & x_\alpha \\ z_t & z_\alpha \end{pmatrix}$$

is the matrix representing the transformation from ray coordinates $\gamma$ to physical coordinates $(x, z)$. By using (E.1), (E.2) and (E.4) we find

$$x \frac{\partial}{\partial t} (w_k x - I \int w_{k1} \cdot z)$$

and so the Jacobian $J$ of the transformation can be written

$$J = \det(x_\gamma) = |w_k| (-t_2, t_1) \cdot (x_\alpha, z_\alpha)^T = |w_k| n \cdot x_\alpha.$$

Here $A = n \cdot x_\alpha$ represents the distance between two infinitesimally close rays in the direction of the normal, (see Figure E.1); in view of its meaning in dimensions higher than 2, we term it the ray-tube area.

Evolution equations for $x_\gamma$ and $k_\gamma$ along a ray can be obtained by differentiating (E.1) and (E.2) with respect to $\gamma$. Using the chain rule, we obtain

$$\frac{\partial}{\partial t} \left( \frac{\partial x_i}{\partial \gamma_j} \right) = \frac{\partial^2 \omega}{\partial k_i x_i \partial \gamma_j} + \frac{\partial^2 \omega}{\partial k_i k_i \partial \gamma_j} \quad \text{and} \quad \frac{\partial}{\partial t} \left( \frac{\partial k_i}{\partial \gamma_j} \right) = -\frac{\partial^2 \omega}{\partial x_i x_i \partial \gamma_j} - \frac{\partial^2 \omega}{\partial x_i k_i \partial \gamma_j}.$$  

(E.5)

An evolution equation for $J$ can also be obtained by differentiating with respect to $t$ and using (E.5):

$$\frac{\partial J}{\partial t} = \epsilon_{ij} \frac{\partial}{\partial t} \left( \frac{\partial x_i}{\partial \gamma_j} \right) = \epsilon_{ij} \left( \frac{\partial^2 \omega}{\partial k_i x_i \partial \gamma_j} + \frac{\partial^2 \omega}{\partial k_i k_i \partial \gamma_j} \right).$$  

(E.6)
Now, by using the chain rule,
\[ \frac{\partial k_i}{\partial x_j} = \frac{\partial k_i}{\partial \gamma_i} \frac{\partial \gamma_i}{\partial x_j} \implies \frac{\partial k_i}{\partial x_j} = \frac{\partial k_i}{\partial x_j} + \frac{\partial k_i}{\partial x_j} \implies \epsilon_{ij} \frac{\partial k_i}{\partial \gamma_i} = \epsilon_{ij} \frac{\partial k_i}{\partial \gamma_i} \frac{\partial J}{\partial x_j}. \] (E.7)
Combining (E.7) with (E.6) and writing \( J = \epsilon_{ij} \frac{\partial x_i}{\partial \gamma_j} \) we obtain
\[ \frac{\partial}{\partial t}(\ln J) = \epsilon_{ij} \left( \frac{\partial^2 \omega}{\partial k_i \partial x_j} + \frac{\partial^2 \omega}{\partial k_i \partial k_i \partial x_j} \right). \] (E.8)
We can also obtain an evolution equation for \( k_x \). Differentiating (E.7) with respect to \( t \) we obtain
\[ \frac{\partial k_i}{\partial x_i} \frac{\partial (\partial x_j)}{\partial \gamma_j} + \frac{\partial}{\partial t} \left( \frac{\partial k_i}{\partial \gamma_j} \right) \partial x_i = -\frac{\partial^2 \omega}{\partial x_i x_i} \partial x_j \partial \gamma_j = -\frac{\partial^2 \omega}{\partial x_i x_i} \partial x_j \partial \gamma_j. \]
Using (E.7) again and (E.5) we find
\[ \frac{\partial}{\partial t} \left( \frac{\partial k_i}{\partial x_j} \right) = -\frac{\partial^2 \omega}{\partial x_i x_i} - \frac{\partial^2 \omega}{\partial x_i x_i} \frac{\partial k_i}{\partial x_i} - \frac{\partial^2 \omega}{\partial x_i x_i} \frac{\partial k_i}{\partial x_i} - \frac{\partial k_i}{\partial x_i} \frac{\partial^2 \omega}{\partial x_i x_i} \frac{\partial k_i}{\partial x_i} \frac{\partial k_i}{\partial x_i}. \] (E.9)
To find the location of a caustic we solve together the ray equations (4.6), derived ray equations, (E.9), and the evolution equation for \( J \), (E.8), i.e., a system of nine ordinary differential equations. These equations have been derived entirely geometrically without appealing to the conservation of wave action. To find an expression for the amplitude \( a \) in terms of the ray-tube area we now use the conservation of action, (see (2.60))
\[ \frac{\partial}{\partial t} \left( \frac{a^2}{\omega} \right) + \nabla \cdot \left( \frac{a^2}{\omega} \omega_k \right) = 0. \]
In a steady system, where we make use of (E.1) we find
\[ \nabla \cdot \left( \frac{a^2}{\omega} \omega_k \right) = \left( \frac{a^2}{\omega} \right) \nabla \cdot \omega_k + \omega_k \nabla \cdot \left( \frac{a^2}{\omega} \right) = \epsilon_{ij} \frac{a^2}{\omega} \left( \frac{\partial^2 \omega}{\partial k_i \partial x_j} + \frac{\partial^2 \omega}{\partial k_i \partial k_i \partial x_j} \right) + \nabla \cdot \left( \frac{a^2}{\omega} \right) \frac{\partial x}{\partial t}. \] (E.10)
Using (E.8) and the definition of ray-tube area we obtain
\[ \frac{\partial}{\partial t} \left( \frac{a^2}{\omega} \right) = -\left( \frac{a^2}{\omega} \right) \frac{\partial}{\partial t} \ln(|\omega_k| A). \]
Integrating we find
\[ a(t, \alpha) = a_0 \left( \frac{|\omega_k| A_0 \omega_k}{|\omega_k| A \omega_k} \right)^{1/2} \]
and so it is now evident that as the ray-tube area \( A \) tends to zero so the amplitude on the ray becomes large.

We now write (E.9) and (E.8) in component form and non-dimensionalise them. The derivatives of \( \omega \) are
\[ \omega_x = kU_x + mW_z, \quad \omega_z = kU_z + mW_z, \]
\[ \omega_k = -m^2 N(k^2 + m^2)^{-3/2} + U, \]
\[ \omega_m = k m N(k^2 + m^2)^{-3/2} + W, \]
\[ \omega_{xz} = U_x, \quad \omega_{xm} = W_x, \quad \omega_{x} = U_x, \quad \omega_{mz} = W_z, \]
\[ \omega_{zz} = k U_{zz} + m W_{zz}, \quad \omega_{zz} = k U_{zz} + m W_{zz}, \]
\[ \omega_{kk} = 3 k m^2 N(k^2 + m^2)^{-5/2}, \]
\[ \omega_{km} = m N(m^2 - 2k^2)(k^2 + m^2)^{-5/2} \]
and \[ \omega_{mm} = k N(k^2 - 2m^2)(k^2 + m^2)^{-5/2}. \]

Expanding (E.9) and (E.8) using the derivatives of \( \omega \) given above, we obtain the following equations

\[ k_{xt} = -2k_x U_x - m_x W_x - k_x W_x - k U_{xx} - m W_{xx}, \]
\[ -N(k^2 + m^2)^{-5/2} \left[ (3k_x k m^2 + k_x m(m^2 - 2k^2))k_x \right. \]
\[ + (k_x m(m^2 - 2k^2) + k_x k(k^2 - 2m^2))m_x \left. \right] \]
\[ k_{zt} = -k_z U_x - m_z W_x - k_z W_x - k U_{zz} - m W_{zz}, \]
\[ -N(k^2 + m^2)^{-5/2} \left[ (3k_z k m^2 + k_z m(m^2 - 2k^2))k_z \right. \]
\[ + (k_z m(m^2 - 2k^2) + k_z k(k^2 - 2m^2))m_z \left. \right] \]
\[ m_{xt} = -k_x U_z - m_x W_z - m_x U_z - m_x W_z - k U_{xz} - m W_{xz}, \]
\[ -N(k^2 + m^2)^{-5/2} \left[ (3m_x k m^2 + m_x m(m^2 - 2k^2))k_x \right. \]
\[ + (m_x m(m^2 - 2k^2) + m_x k(k^2 - 2m^2))m_x \left. \right] \]
\[ m_{zt} = -k_z U_z - 2m_z W_z - m_z U_z - k U_{zz} - m W_{zz}, \]
\[ -N(k^2 + m^2)^{-5/2} \left[ (3m_z k m^2 + m_z m(m^2 - 2k^2))k_z \right. \]
\[ + (m_z m(m^2 - 2k^2) + m_z k(k^2 - 2m^2))m_z \left. \right] \]
and \[ \frac{1}{J} \frac{dJ}{dt} = N(k^2 + m^2)^{-5/2} \left[ 3k_x k m^2 + m(m^2 - 2k^2)(k_x + m_x) \right. \]
\[ + m_x k(k^2 - 2m^2) \left. \right] + U_x + W_z. \]

Notice here that the divergence of the wind \( \nabla \cdot U \) appears in the evolution equation for \( J \). This implies that for non-divergent winds ray convergence can occur independently of the random nature of the wind. We can now non-dimensionalise these equations by using the following scales (described in Section 4.3)

\[ x = l \hat{x}, \quad z = l \hat{z}, \quad k = k_0 \hat{k}, \quad m = \frac{N}{U_0} \hat{m}, \quad t = \frac{l}{U_0} \hat{t}, \quad U = U_0 \hat{U}, \]
\[ \omega = U_0 k_0 \hat{\omega}, \quad W = \delta U_0 \hat{W} \quad \text{and} \quad \delta = \frac{k_0}{m_0}, \]

142
Applying these and dropping hats we obtain the following system of non-dimensional equations

\[
\begin{align*}
k_{xt} &= -2k_x U_x - m_x W_x - k_x W_x - kU_{xx} - mW_{xx}, \\
&\quad - (\delta^2 k^2 + m^2)^{-5/2} \left[(3\delta^2 k_x km^2 + k_x m(m^2 - 2\delta^2 k^2))k_x \\
&\quad + (k_x m(m^2 - 2\delta^2 k^2) + k_x k(\delta^2 k^2 - 2m^2))m_x \right]
\end{align*}
\]

\[
\begin{align*}
k_{zt} &= -k_z U_z - m_z W_z - k_z U_z - k_2 W_z - kU_{zz} - mW_{zz}, \\
&\quad - (\delta^2 k^2 + m^2)^{-5/2} \left[(3\delta^2 k_x km^2 + k_x m(m^2 - 2\delta^2 k^2))k_z \\
&\quad + (k_x m(m^2 - 2\delta^2 k^2) + k_x k(\delta^2 k^2 - 2m^2))m_z \right]
\end{align*}
\]

\[
\begin{align*}
m_{xt} &= -k_x U_x - k_x W_x - m_x U_x - m_x W_x - kU_{xx}, \\
&\quad -mW_{xx} - (\delta^2 k^2 + m^2)^{-5/2} \left[(3\delta^2 m_x km^2 + m_x m(m^2 - 2\delta^2 k^2))k_x \\
&\quad + (m_x m(m^2 - 2\delta^2 k^2) + m_x k(\delta^2 k^2 - 2m^2))m_x \right]
\end{align*}
\]

\[
\begin{align*}
m_{zt} &= -k_z U_z - 2m_z W_z - m_z U_z - kU_{zz} - mW_{zz}, \\
&\quad - (\delta^2 k^2 + m^2)^{-5/2} \left[(3\delta^2 m_x km^2 + m_x m(m^2 - 2\delta^2 k^2))k_z \\
&\quad + (m_x m(m^2 - 2\delta^2 k^2) + m_x k(\delta^2 k^2 - 2m^2))m_z \right]
\end{align*}
\]

\[
\begin{align*}
1 \frac{dJ}{dt} &= (\delta^2 k^2 + m^2)^{-5/2} \left[3\delta^2 k_x km^2 + m(m^2 - 2\delta^2 k^2)(k_x + m_x) \\
&\quad + m_x k(\delta^2 k^2 - 2m^2) \right] + U_x + W_z.
\end{align*}
\]

For \(\delta \ll 1\) these equations become

\[
\begin{align*}
k_{xt} &= -2k_x U_x - m_x W_x - k_x W_x - kU_{xx} - mW_{xx}, \quad (E.11) \\
&\quad - m^{-5}\left[k_x m^3 k_x + (k_x m^3 - 2k_x km^2)m_x \right], \quad (E.12) \\
k_{zt} &= -k_z U_z - m_z W_z - k_z U_z - k_z W_z - kU_{zz} - mW_{zz}, \quad (E.13) \\
&\quad - m^{-5}\left[k_x^2 m^3 + (k_x m^3 - 2k_x km^2)k_z \right], \quad (E.14) \\
m_{xt} &= -k_x U_x - k_x W_x - m_x U_x - m_x W_x - kU_{xx} - mW_{xx}, \quad (E.15) \\
&\quad - m^{-5}\left[m_x m^3 k_x + (m_x m^3 - 2m_x km^2)m_x \right], \quad (E.16)
\end{align*}
\]

\[
\begin{align*}
m_{zt} &= -k_z U_z - 2m_z W_z - m_z U_z - kU_{zz} - mW_{zz}, \quad (E.17) \\
&\quad - m^{-5}\left[m_x m^3 k_z + (m_x m^3 - 2m_x km^2)m_z \right]
\end{align*}
\]

\[
\begin{align*}
\text{and } 1 \frac{dJ}{dt} &= m^{-5}(m^3(k_x + m_x) - 2m_x km^2) + U_x + W_z. \quad (E.18)
\end{align*}
\]
Appendix F

Parabolic caustic

Consider a two-dimensional parametric curve $C(x)$ given by

$$C(x) := \{(x, f(x)) \in \mathbb{R} \}.$$ 

The tangent $T(x)$ and normal $N(x)$ to this curve are given by (e.g. Kreyszig 1993)

$$T(x) = (1, f'(x))(1 + [f'(x)]^2)^{-1/2} \quad \text{and} \quad N(x) = (-f'(x), 1)(1 + [f'(x)]^2)^{-1/2}.$$ 

Now consider a parabola given by $C(x) = (x, x^2)$ as a wavefront. By Huygen’s principle, each point on the wavefront evolves in equal distances in a direction normal to the wavefront. So the evolved wavefront $C_t(x)$ is given by

$$C_t(x) = C(x) + tN(x),$$

which for the parabola gives

$$C_t(x) = (\alpha_1(x), \alpha_2(x)) = \left( x - 2xt(1 + 4x^2)^{-1/2}, x^2 + t(1 + 4x^2)^{-1/2} \right). \quad (F.1)$$

We can find the equation for the caustic by considering the tangent to $C_t(x)$. This is given by

$$C'_t(x) = (\alpha'_1(x), \alpha'_2(x)) = \left( 1 - 2t(1 + 4x^2)^{-3/2}, 2x(1 - 2t(1 + 4x^2)^{-3/2}) \right).$$

The caustic curve is defined as that linking up all the points where the tangent is singular (i.e., not well defined), this can be seen by refering to Figure 4.2. The tangent is singular when $C'_t(x) = (0, 0)$, i.e., when

$$t = \frac{1}{2}(1 + 4x^2)^{3/2}. \quad (F.2)$$

This only has a solution for $x$ when $t > 1/2$, i.e., the first singularity forms at $t = 1/2$. Substituting (F.2) back into (F.1) gives a relationship between $\alpha_1$ and $\alpha_2$

$$\frac{1}{27} \left( \alpha_2 - \frac{1}{2} \right)^3 = \frac{\alpha_1^2}{16}. \quad (F.3)$$
The figures in Chapter 4 show the curve $C_t(x)$ for various values of $t$. Figure 4.1 shows the curves for $t = 0.1$ and $t = 0.4$. Figure 4.2 shows the curves for $t = 0.1, 0.3, 0.5, 0.7, 0.9$ and $t = 1.5$. The caustic curve shown by the dashed line in Figure 4.2 is given by (F.3).

A nice demonstration of this wavefront propagating in time can be seen by running the following code in Maple.

```maple
with(plots);
animate([x-2*0.1*t*x/sqrt(1+4*x^2),x^2+0.1*t/sqrt(1+4*x^2),x=-1..1],
       t=0..10,frames=100);
```
Appendix G

Runge–Kutta numerical methods for systems of ODEs

Runge–Kutta methods are used in Chapter 4 and a brief description of them is given here.

There are various different methods for solving a system of coupled ordinary differential equations with specified initial values in the form

\[
y' = f(x, y) \quad \text{with} \quad y(x_0) = y_0. \tag{G.1}
\]

Euler’s method – which can be found in most texts on ODEs or numerical analysis (e.g. Kreyszig 1993) is one of the most simple examples, but requires a very small step size for reasonable accuracy compared to other methods. This method is generally of little use despite its simplicity as it is too computationally intensive.

Another method is Taylor’s algorithm (e.g. Conte & de Boor 1983) which will be described here as it is useful in the derivation of Runge–Kutta methods. We can expand the solution to (G.1) using Taylor series as

\[
y(x) = y_0 + (x - x_0)y'(x_0) + \frac{(x - x_0)^2}{2!}y''(x_0) + \ldots,
\]

where the derivatives of \(y\) with respect to \(x\) are given by the chain rule as

\[y' = f, \quad y'' = f' = f_x + f_y f,\]

and

\[y''' = f_{xx} + 2f_{xy} f + f_{yy} f^2 + f_x f_y + f_y^2 f, \tag{G.2}\]

with

\[f_y = \sum_i \frac{\partial f_i}{\partial y_i}.
\]

To solve (G.1) over the interval \(x \in [x_a, x_b]\) we choose a step size \(h = (x_a - x_b)/N\), where \(N\) is the number of steps that divide the interval into discrete points, given by

\[x_n = a + nh \quad \text{for} \quad n = 0, 1, \ldots, N.\]
Then $y$ is calculated on the discrete points $n = 0, 1, \ldots, N - 1$ using the following recursion relation

$$y_{n+1} = y_n + hT_k(x_n, y_n)$$

and

$$T_k(x, y) = f(x, y) + \frac{h}{2!} f'(x, y) + \ldots$$

It is quite clear that unless $f$ is a simple function then higher order derivatives can be very complicated and hence numerically intensive to compute. Runge–Kutta methods improve on this and attempt to obtain a high degree of accuracy without using high order derivatives (e.g. Conte & de Boor 1983, Burden & Faires 1993).

**G.1 Runge-Kutta method of order 2**

We first derive a two-step Runge-Kutta method of order 2: define the recursive relation for $y$ as

$$y_{n+1} = y_n + a k_1 + b k_2,$$  \hspace{1cm} (G.4)

$$k_1 = h f(x_n, y_n) \quad \text{and} \quad k_2 = h f(x_n + ah, y_n + \beta k_1),$$

where $a$, $b$, $\alpha$ and $\beta$ are constants such that (G.4) agrees with the Taylor algorithm (G.3) to as high an order as possible. The Taylor expansion in two variables of the function $k_2$ is given by

$$k_2 = hf + \alpha h^2 (x + f_y f) + \beta k_1 f_{xy} + \frac{\alpha^2 h^3}{2} f_{xx} + \alpha h^2 \beta k_1 f_{xy} + \frac{\beta^2 h k_1^2}{2} f_{yy} + O(h^5).$$

Substituting this and $k_1 = hf$ into (G.4) we get

$$y_{n+1} = y_n + (a+b)hf + bh (\alpha f_x + \beta f f_y) + bh^3 \left( \frac{\alpha^2}{2} f_{xx} + \alpha \beta f f_{xy} + \frac{\beta^2}{2} f_{yy} \right) + O(h^4).$$

(G.5)

Substituting the values for the derivatives of $f$, equation (G.2), into (G.3) we get

$$y_{n+1} = y_n + h f + \frac{h^2}{2} (x + f y y) + \frac{h^3}{6} (f_{xx} + 2 f f_{xy} + f_{yy} f^2 + f_x f_y + f_y f) + O(h^4).$$

Comparing this with (G.5) and equating coefficients of $a$ and $b$ in the $h$ and $h^2$ terms, we find

$$a + b = 1 \quad \text{and} \quad b \alpha = b \beta = \frac{1}{2}.$$

Unfortunately this is the best we can do as it can clearly be seen that it is impossible to match terms in the $h^3$ powers. Many solutions to the above equations exist (four unknowns, three equations), so we can solve in terms of one of the parameters, $b$ say and obtain a family of two stage, second order Runge–Kutta methods. Popular values for $b$ are $b = 1/2, 3/4$ and 1. When $b = 0$ the system reduces to the first order Euler method.

In Chapter 4 it was mentioned that the ODEs being solved can be stiff. The second order two-step R–K method just derived is not the best to deal with stiffness (see Nag 1995). At a time cost it is better to use a fourth order R–K method.
G.2 Runge-Kutta method of order 4

The above calculation can be generalised to order 4 by repeating with

\[ y_{n+1} = y_n + \alpha k_1 + \beta k_2 + \gamma k_3 + \delta k_4. \]

Without doing the calculation (which is rather long and tedious), we get the following method:

\[ y_{n+1} = y_n + \frac{1}{6} (k_1 + 2k_2 + 2k_3 + k_4), \]

\[ k_1 = hf(x_n, y_n), \quad k_2 = hf(x_n + h/2, y_n + k_1/2), \]

\[ k_3 = hf(x_n + h/2, y_n + k_2/2) \quad \text{and} \quad k_4 = hf(x_n + h, y_n + k_3). \]

This is the numerical method used in Chapter 4 to solve the system of nine ordinary differential equations.
Bibliography


Burden, R. L. & Faires, J. D. 1993, Numerical analysis, 5 edn, PWS-KENT.


Conte, S. D. & de Boor, C. 1983, Elementary numerical analysis, 3 edn, McGraw-Hill.

Dewar, E. M. & Good, R. E. 1986, Saturation and the “universal” spectrum for vertical profiles of horizontal scalar winds in the atmosphere, J. Geophys. Res. 91(D2), 2742-2748.


Fritts, D. C. & Rastogi, P. K. 1985, Convective and dynamical instabilities due to gravity wave motions in the lower and middle atmosphere: theory and observations, Radio Sci. 20, 1247-1277.


